



# CSCL

## 2017

**Making a Difference:**

**Prioritizing Equity and Access in CSCL**

**12<sup>th</sup> International Conference on Computer Supported Collaborative Learning**

CONFERENCE PROCEEDINGS  
VOLUME 1

EDITED BY

BRIAN K SMITH  
MARCELA BORGE  
EMMA MERCIER  
KYU YON LIM

**18-22 June**

PHILADELPHIA, PA USA



DREXEL UNIVERSITY  
School of  
Education



International Society of  
the Learning Sciences



**Making a Difference: Prioritizing Equity and  
Access in CSCL**

12<sup>th</sup> International Conference on Computer  
Supported Collaborative Learning (CSCL) 2017

Volume 1

12<sup>th</sup> International Conference on  
Computer Supported Collaborative Learning

June 18-22, 2017, Philadelphia, PA  
Drexel University School of Education  
University of Pennsylvania Graduate School of  
Education

Editors:

Brian K. Smith, Marcela Borge, Emma Mercier, and Kyu Yon Lim

*Making a Difference: Prioritizing Equity and Access in CSCL, 12th International Conference on Computer Supported Collaborative Learning (CSCL) 2017*

June 18-22, 2017, Philadelphia, PA

Drexel University School of Education and University of Pennsylvania Graduate School of Education.

© 2017 International Society of the Learning Sciences, Inc. [ISLS]. Rights reserved.



[www.isls.org](http://www.isls.org)

ISBN: 978-0-9903550-0-7 (Volume 1, PDF Version)

ISSN: 1573-4552

Cite as: Smith, B. K., Borge, M., Mercier, E., and Lim, K. Y. (Eds.). (2017). *Making a Difference: Prioritizing Equity and Access in CSCL, 12th International Conference on Computer Supported Collaborative Learning (CSCL) 2017, Volume 1*. Philadelphia, PA: International Society of the Learning Sciences.

All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear the copyright notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior written permission of the International Society of the Learning Sciences. The International Society of the Learning Sciences is not responsible for the use which might be made of the information contained in this book.

Conference logo and cover design by Sally Im and Healthy Moeung.

# Conference Organizers and Committees

## Local Organizing Committee

Aroutis Foster, *Drexel University*  
Yasmin Kafai, *University of Pennsylvania*  
Brian K Smith, *Drexel University*  
Susan Yoon, *University of Pennsylvania*  
Nancy Butler Songer, *Drexel University*

## Program Co-Chairs

Brian K Smith, *Drexel University, USA*  
Marcela Borge, *The Pennsylvania State University, USA*  
Emma Mercier, *University of Illinois at Urbana-Champaign, USA*  
Kyu Yon Lim, *Ewha Womans University, Korea*

## Logistics Co-Chairs

Aroutis Foster, *Drexel University, USA*  
Nur Akkuş, *Hacettepe University, Turkey*  
Evrin Baram, *Middle East Technical University, Turkey*

## Pre-Conference Workshop Co-Chairs

Aaron Kessler, *Concordia University Chicago, USA*  
Mamta Shah, *Drexel University, USA*

## Doctoral Consortium Co-Chairs

Jun Oshima, *Shizuoka University, Japan*  
Kylie Pepler, *Indiana University, USA*

## Early Career Workshop Co-Chairs

Susan Yoon, *University of Pennsylvania, USA*  
Manu Kapur, *ETH Zurich, Switzerland*

## Mid-Career Workshop Chair

Kristine Lund, *University of Lyon, France*

## Interactive Events Chair

Heather Toomey Zimmerman, *The Pennsylvania State University, USA*

## Communications and Publicity Chair

Gabriela Richard, *The Pennsylvania State University, USA*

## Student Volunteer Coordinators

Amanda Barany, *Drexel University, USA*  
Noora Noushad, *University of Pennsylvania, USA*  
Jooeun Shim, *University of Pennsylvania, USA*

## Advisory Committee

Oskar Lindwall, *University of Gothenburg, Sweden*  
Sadhana Puntambekar, *University of Wisconsin at Madison, USA*

## Senior Reviewers

Dor Abrahamson  
June Ahn  
Nur Akkuş  
Richard Alterman  
Hans Christian Arnseth  
Michael Baker  
Marcela Borge  
Murat Cakir  
Britte Cheng  
Ming Ming Chiu  
Cynthia D'Angelo  
Joshua Danish  
Stavros Demetriadis  
Yannis Dimitriadis  
Gijsbert Erkens  
Bernhard Ertl  
Aroutis Foster  
Xun Ge  
Yotam Hod  
Feng Feng Ke  
Valerie Klein  
Timothy Koschmann  
Kristiina Kumpulainen  
Susan Land  
Victor Lee  
Kyu Yon Lim  
Oskar Lindwall  
Chee-Kit Looi  
Rose Luckin

Sten Ludvigsen  
Kristine Lund  
Emma Mercier  
Anders Morch  
Hanni Muukkonen  
Mitchell Nathan  
William R. Penuel  
Joseph Polman  
Luis P. Prieto  
Ingvill Rasmussen  
Carolyn Rosé  
Joshua Rosenberg  
Regina Ruane  
Baruch Schwarz  
Gerard Sensevy  
David Williamson Shaffer  
Mamta Shah  
Hajime Shirouzu  
Jason Silverman  
Brian K. Smith  
Erica Snow  
Gerry Stahl  
Seng Chee Tan  
Pierre Tchounikine  
Michael Twidale  
Armin Weinberger  
Alyssa Wise  
Heather Toomey Zimmerman

## Reviewers

Anthony Aakre  
Bunmi Adewoyin  
Rama Adithya  
Tugce Aldemir  
Katelyn Bright Aldever  
Elizabeth Marie Allen  
Isabel Alvarez  
Alia Ammar  
Ricardo Amorim  
Alejandro Andrade  
Golnaz Arastoopour  
Robert Ashley  
Tutaleni I. Asino  
Christa Asterhan  
Na'ama Av-Shalom  
Afaf Baabdullah  
Nilufar Baghaei  
Amanda Barany  
Brian Belland  
Yifat Ben-David Kolikant  
Dani Ben-Zvi  
Matthew Berland  
Heather J. Birch  
Daniel Bodemer  
Ivica Boticki  
Tharrenos Bratitsis  
Karen Brennan  
Paul Brna  
Juergen Buder  
Erin Buehler  
Huiying Cai  
Teresa Cerratto Pargman  
Margaret Chan  
Elizabeth Charles  
Clement Chau  
Bodong Chen  
Ching-Huei Chen  
Min Chen  
Wenli Chen  
John Cherniavsky  
Ellina Chernobilsky  
Gi Woong Choi  
Irene-Angelica Chounta  
Tamara Clegg  
Charlie Cox  
Ulrike Cress  
Mutlu Cukurova  
Mihai Dascalu  
Frank de Jong  
Bram De Wever  
Carrie Demmans Epp  
Ben DeVane  
Pierre Dillenbourg  
Ning Ding  
Betsy DiSalvo  
Jan Arild Dolonen  
Leona Donaldson  
Jonan Donaldson  
Dermot Donnelly  
Matt Duvall  
Gregory Dyke  
Julia Eberle  
Noel Enyedy  
Kiran Eranki  
Howard Everson  
Georgios Fesakis  
Deborah Fields  
Eilis Flanagan  
Paul Flynn  
Dominique Forest  
Nobuko Fujita  
Judi Fusco  
Tamara Galoyan  
Iolanda Garcia Gonzalez  
Maria Teresa Gastardo  
Sébastien George  
Aristotelis Gkiolmas  
Michael Glass  
Alex Sandro Gomes  
Kim Gomez  
Jeffrey Greene  
Gabriela Groza  
Frode Guribye  
Joerg Haake  
Alan Hackbarth  
Tony Hall  
Erica Halverson  
Insook Han  
Jiangang Hao  
Qiang Hao  
Lisa Hardy  
Brenna Hassinger-Das  
Katie Headrick Taylor  
Sven Heimbuch  
Cindy Hmelo-Silver  
Ty Hollett  
Kian Sam Hong  
H. Ulrich Hoppe  
Mark Horney  
Anesa Hosein  
Iris Howley  
Pi-Sui Hsu  
Helen Hu  
Joey Huang  
Roland Hubscher  
So Hyo-Jeong  
Shinya Ikubo  
Isa Jahnke  
Ellen Jameson  
Sanna Jarvela  
Allan Jeong  
Heisawn Jeong  
Jo Inge Johansen Frøytlog  
Emily Johnson  
Zywica Jolene  
Andrew Joyce-Gibbons  
Yong Ju Jung  
Yael Kali  
Anastasios Karakostas  
Ilias Karasavvidis  
Petter Karlstrom  
Susan Kelly  
Md. Saifuddin Khalid  
Anastasia Khawaja  
Khusro Kidwai  
Andrea Kienle  
ChanMin Kim  
Dongsik Kim  
Soo Hyeon Kim  
Mi Song Kim  
Joachim Kimmerle  
Simon Knight  
Ingo Kollar  
Alexa Kottmeyer  
Brian Krisler  
Samuel Kwon  
Eleni Kyza  
Yacine Lafifi  
Minna Lakkala  
Niki Lambropoulos  
Elise Lavoué  
Ard Lazonder  
Heather Leary  
Alwyn Vwen Yen Lee  
Chien-Sing Lee  
Sy-ying Lee  
Wincy Wing Sze Lee  
Teemu Leinonen  
Dalit Levy  
Wenjuan Li  
Ken W. Li  
Xiao Li  
Jian Liao  
Rasheda Likely  
Lijia Lin  
Feng Lin  
Andreas Lingnau  
Max Louwerse  
Rosemary Luckin

Michelle Lui  
Johan Lundin  
Leanne Ma  
Jasmine Ma  
Moseli Alexander Mafa  
Rachel Magee  
Alecia Marie Magnifico  
Stefania Manca  
Crystle Martin  
Roberto Martinez Maldonado  
Alejandra Martínez-Monés  
Anthony Matranga  
Steven McGee  
Zachary McKinley  
Richard Medina  
Veronique Mertl  
Toshio Mochizuki  
Gaelle Molinari  
Savitha Moorthy  
Chrystalla Mouza  
Magdalene Moy  
Hani Murad  
Denise Nacu  
Lijun NI  
Iolie Nicolaidou  
Malzahn Nils  
Nicolae Nistor  
Jalal Nouri  
Jessica Nowak  
Claire O'Malley  
Amanda Ochsner  
Francisco Kelsen de Oliveira  
Jennifer Olsen  
Yann Shiou Ong  
Jun Oshima  
Hiroki Oura  
Xueying Ouyang  
Annemarie Palincsar  
Pantelis M. Papadopoulos  
Frederick Peck  
Anthony Pellicone  
Koraly Perez-Edgar  
Elaine Perignat  
Elizabeth Pier  
Johanna Poysa-Tarhonen

Michael Prilla  
David Quigley  
Tanushree Rawat  
Traian Rebedea  
Mimi Recker  
Christophe Reffay  
Martin Rehm  
Peter Reimann  
Christoph Richter  
Antti Rissanen  
Giuseppe Ritella  
Jessica Roberts  
Michael Rook  
Ricarose Roque  
Jeremy Roschelle  
Ellen Rusman  
Donna Russell  
Stephen Rutherford  
Ornit Sagy  
Fariah Hayat Salman  
Daouda Sawadogo  
Giovanna Scalone  
Bertrand Schneider  
Kristin Searle  
Colleen Seifert  
Paul Seitlinger  
David Shaenfield  
R. Benjamin Shapiro  
Kshitij Sharma  
Priya Sharma  
Saadeddine Shehab  
Joshua Sheldon  
Patrick Shih  
Amanda Siebert-Evenstone  
Jim Slotta  
Ole Smørdal  
Amber Solomon  
Sergey Sosnovsky  
Hans Spada  
Daniel Spikol  
Karsten Stegmann  
Slavi Stoyanov  
Jan-Willem Strijbos  
Florence Sullivan  
Patrick Sunnen

Zachari Swiecki  
Elisabeth Sylvan  
Antti Syvanen  
Aiko Takazawa  
Hamideh Talafian  
Dan Tao  
Jakob Tholander  
Christopher Thorn  
Mike Tissenbaum  
Dhvani Toprani  
Stefan Trausan-Matu  
Michael Tscholl  
Thrasylvoulos Tsiatsos  
Dimitra Tsovaltzi  
Selen Turkey  
Hakan Tuzun  
Judith Uchidiuno  
Joshua Underwood  
Ralph Vacca  
Phil Vahey  
Jan van Aalst  
Charles Vanover  
Yen Verhoeven  
Himanshu Verma  
Sylvi Vigmo  
Freydis Vogel  
Iro Voulgari  
Yianna Vovides  
Chieu Vu Minh  
Hanna Vuojarvi  
Jo Wake  
Xu Wang  
Joshua Weese  
Jennifer L. Weible  
Joanna Weidler-Lewis  
Tobin White  
Caro Williams-Pierce  
Iwan Wopereis  
Shulong Yan  
Wei-Yu Yang  
Jason Yip  
Junxiu Yu  
Alan Zemel  
Gaoxia Zhu  
Katerina Zourou



## Student Volunteers

Alia Ammar  
Amanda Barany  
Katelyn Bright  
Jessica Cellitti  
Matt Duvall  
Tamara Galoyan  
Jamie Gorson  
Sally Im  
Rasheda Likely  
Debra Lui  
Anthony Matranga  
Healthy Moeung

Magdalene Moy  
Noora Noushad  
Miyoung Park  
Elaine Perignat  
Jooeun Shim  
Matthew Shuster  
Hamideh Talafian  
Shawna L. Thomas-EL  
Justice Toshiba Walker  
Aileen Tschiderer  
Olivia Yutong-Wang  
Shizhu Zhang



# Preface

Drexel University and The University of Pennsylvania are co-hosting the 12th International Conference on Computer Supported Collaborative Learning from June 18 to June 22, 2017. The CSCL conference has an explicit focus on how and why collaboration can enhance learning processes and outcomes. CSCL emerged in the late 1980's and early 1990's to bring together researchers from cognitive science, educational research, psychology, computer science, artificial intelligence, information sciences, anthropology, sociology, neurosciences, and other fields to study learning in a wide variety of formal and informal contexts (see <http://www.isls.org> for more details).

Before the establishment of the biannual CSCL conferences, there was a NATO-sponsored workshop in Maratea, Italy in 1989 and another workshop sponsored by Xerox PARC in 1991 at Southern Illinois University. The first international conference was held in 1995 at Indiana University, followed by meetings in Toronto, ON, Canada (1997); Maastricht, Netherlands (2001); Boulder, CO, USA, (2002); Bergen, Norway (2003), Taipei, Taiwan (2005); New Brunswick, NJ, USA (2007); Rhodes, Greece (2009); Hong Kong, China (2011); Madison, WI, USA (2013); Gothenberg, Sweden (2015). There is also a scholarly journal, the *International Journal of Computer-Supported Collaborative Learning*, and a book series published by Springer.

Submissions for CSCL 2017 were received in November 2016 and sent out for peer review. 386 paper and poster submissions were received from 28 countries, and the overall acceptance rate for submissions was 45%. We accepted 60% of symposium submissions, 35% of full papers, 31% of short papers, and 48% of posters. 295 experts completed 1287 reviews, and an additional 61 senior reviewers assigned papers to reviewers and provided summary reflections on each submission to guide the development of the program.

## Making a Difference—Prioritizing Equity and Access in CSCL

CSCL 2017's theme, Making a Difference—Prioritizing Equity and Access in CSCL, revisits the concepts of equity and access to learning opportunities that have always been central to collaborative learning pedagogies and research. Work in the 1960s sought to address issues of classroom authority structures with group activities. Work in the 1980's and 1990's attempted to provide young people with access to safe, collaborative, after-school learning environments. Research on learning communities also empowered students to have agency over their learning processes and to see themselves as creators rather than merely consumers of knowledge. More recent work has sought to provide opportunities for a wider range of students through resident and online university courses, new collaborative learning technologies, and Massive Open Online Courses. Throughout this work, there have been two common themes that focus on equity and access: equity at a small, community scale and equity at a larger, societal level.

The most common theme in CSCL is the promotion of equity within the classroom community. Many researchers have emphasized the need to provide students with more agency over their own learning processes. Others have focused on breaking down social hierarchies that can interfere with important social learning processes. For example, work on communities of learners and learning forums has examined how students take on increasingly active roles in deciding what is learned and how. Some questions that emerge as part of this work include:

- How much and what kind of participation is equitable?
- How important is equitable participation for learning?
- How do we measure participation?
- How do emerging technologies and methods allow us to address and understand participation?
- How do we teach students to participate and encourage others to participate in a manner that allows equal opportunity and access to content learning and skill development for all learners?
- How do we distribute responsibility over learning across teachers and students such that all have opportunities to develop the ability to monitor, regulate, and make decisions about collaborative practices and learning outcomes?

Another common theme within CSCL is the promotion of educational equity and access on a broader scale. Namely, how collaborative learning can attract, support, and engage underrepresented groups while ensuring that all students have access to high-quality and productive cognitive and social learning contexts. Common questions that emerge as part of this work include:

- How do we design activities and tools that meet the needs of different populations?
- How do we balance required content learning with the development of necessary skills?
- How can we develop important collective thinking and discourse processes in ways that engage all learners?
- How do we narrow gaps in learning and educational access?
- How do we build partnerships with schools and communities to ensure that our designs are informed by multiple voices and sustainable beyond the span of a research grant or program?

The CSCL community has additional questions to ask since collaboration, in and of itself, can be a barrier to many students. This is particularly the case for students with physical or learning disabilities and socio-emotional problems. The special education community is underrepresented in the learning sciences. Addressing this absence would increase the richness and diversity of our community. Experts in special education could help us address design issues for students with a range of abilities and developmental needs and make CSCL more accessible to a larger population.

We should also evaluate our designs in the context of cultural, social, and technological change, identifying potential unintended consequences of technology use and ways that we can improve our work to develop the types of skills learners will need in the future. This means not only examining how our designs impact a particular learning outcome for a current population but to carefully consider their effects on related learning and socio-emotional processes and future populations.

Finally, an important consideration is how we can scale CSCL in ways that maintain essential principles of pedagogy and equity. As technology allows for more forms of interaction, we need to ensure that we go beyond providing access to collaborative activities and towards supporting the development of important learning processes within these environments. For example, the need to maintain social relationships between students and teachers is an important concern at a time when technology use, automation, and social isolation is rapidly growing.

Addressing these larger questions will ensure that the core principles and practices that are central to CSCL do not get lost as technologies and educational practices evolve and proliferate. Focusing on these questions

can help us inform policy and provide access to higher quality, meaningful, collaborative learning environments for a broader population of students.

Our three keynote speakers are at the forefront of examining these broader questions. Dr. Laura Czerniewicz highlights the inequalities that exist in higher education and how we can redesign learning environments to mitigate inequalities. Dr. D. Fox Harrell examines the use of growing technologies and their impacts at the intersection of technology use, personal identity, and societal identity. Dr. Teo Chew Lee focuses on larger implementations in ways that maintain core CSCL principles and attend to important social relationships between teachers and students.

Many classic and returning research themes remain stable within these proceedings. Classic research themes include the examination of knowledge building practices and communities, using technology to disrupt traditional teaching practices, and examining discourse, feedback, and argumentation. Returning themes include an emphasis on regulation and awareness at the level of the group and many technologically supported methodological approaches to evaluate learning and social interaction. One of the fastest growing returning themes is learning analytics. This strand gained prominence in the CSCL community in 2015 and had an even stronger representation this year.

Additionally, this year's submissions showcase significant shifts in education and the growing influence of CSCL in some new domains. We noticed four growing trends in CSCL this year:

1. A continued increase in studies of CSCL in informal learning contexts.
2. A growing focus on supporting scientific modeling.
3. A larger representation of CSCL in higher education, especially in the information and computer sciences.
4. An increasing emphasis on scaling CSCL through the creation of massive online courses and large-scale assessments, as well as through community-level participatory and technology design.

Given these growing trends, it was not surprising to see many submissions that were taking the time to step back and assess the state of the field to examine important methodological and practical issues.

As we consider this year's submissions in light of the conference theme, the challenge is to continue holding the principles of equity and access at the forefront of our activities as we grow and expand as a field. Even with a call for papers that addressed the theme, representation for research examining equity and accessibility was relatively small. While there is much to address and embrace regarding the potential of new methods and technologies to advance our field, the values that drive our research should remain the same. We cannot risk losing sight of the reasons why we want to promote discourse as access to new technologies make discussion and collaboration more accessible and easy to evaluate. Otherwise, we run the risk of expanding the computer supported aspect of CSCL without supporting collaborative learning for all.

In these volumes, you will find a collection of thoughtful papers that examine collaborative learning at different levels of scale, question our current practices and assumptions about learning and assessment, and take innovative approaches to support learning both in and out of school. Many of the papers focus on these by addressing issues of equity and accessibility within the classroom community and a few take on the challenge of addressing our theme at a broader scale.

We end by acknowledging the contributions of the many members of our community that made this conference possible: The organizing committee, the mentors that volunteered their time to help young students, mid/early career scholars, and doctoral students, our leading and supporting reviewers, the staff at both host institutions, the session chairs and discussants, and all the presenters and participants. We especially thank our copy editor, Allison Hall, who worked countless hours over many months to prepare the proceedings. We also thank our student volunteers who put in personal time and effort to put together the poster sessions, help organize submissions, and assist the program and organizing committee. We extend special thanks to the following students: Amanda Barany, Kaitlyn Bright, Heather Tanner from Drexel University; Noora Noushad and Joe Eun Shim from the University of Pennsylvania; Shulong Yan and Dhvani Toprani from Penn State University. Finally, many thanks to Aroutis Foster for his leadership and coordination of the conference logistics.

Brian K Smith, *Drexel University, USA*

Marcela Borge, *The Pennsylvania State University, USA*

Emma Mercier, *University of Illinois at Urbana-Champaign, USA*

Kyu Yon Lim, *Ewha Womans University, Korea*



# Table of Contents

## Volume 1

### Keynotes

Unbundling and Inequality in Higher Education .....	3
<i>Laura Czerniewicz</i>	
Virtual Selves and Learning .....	4
<i>D. Fox Harrell</i>	
Symmetrical Advancement: Teachers and Students Sustaining Idea-Centered Collaborative Practices.....	5
<i>Teo Chew Lee</i>	

### Full Papers

Collaborative Learning on Multi-Touch Interfaces: Scaffolding Elementary School Students.....	9
<i>Lara Johanna Schmitt, Armin Weinberger</i>	
Behavioral and Relationship Patterns in an Online Collaborative Reading Activity.....	17
<i>Hai-Peng Wan, Qi Wang, Sheng-Quan Yu</i>	
Contrasting Explicit and Implicit Support for Transactive Exchange in Team Oriented Project Based Learning.....	25
<i>Xu Wang, Miaomiao Wen, Carolyn Rosé</i>	
Making Engagement Visible: The Use of Mondrian Transcripts in a Museum .....	33
<i>Ben Rydal Shapiro, Rogers Hall</i>	
Examining the Flow of Ideas During Critique Activities in a Design Project.....	41
<i>Elizabeth McBride, Jonathan Vitale, Lauren Applebaum, Marcia Linn</i>	
Secondary School Peer-to-Peer Knowledge Sharing Through Social Network Technologies.....	49
<i>Christa Asterhan, Edith Bouton</i>	
Collective Knowledge Advancement and Conceptual Understanding of Complex Scientific Concepts in the Jigsaw Instruction.....	57
<i>Jun Oshima, Ayano Ohsaki, Yuki Yamada, Ritsuko Oshima</i>	
Technology-Mediated Teacher Noticing: A Goal for Classroom Practice, Tool Design, and Professional Development .....	65
<i>Janet Walkoe, Michelle Wilkerson, Andrew Elby</i>	
Girls' Interest in Computing: Types and Persistence .....	71
<i>Michelle Friend</i>	
Educational Technology Support for Collaborative Learning With Multiple Visual Representations in Chemistry .....	79
<i>Martina Rau, Sally Wu</i>	
Whose Culture Is It? Modeling the Designs of Authentic Learning Environments and the Cultures They Mediate.....	87
<i>Yotam Hod, Ornit Sagy</i>	
Learning Alone or Together? A Combination Can Be Best!.....	95
<i>Jennifer K. Olsen, Nikol Rummel, Vincent Alevan</i>	
Predicting Success in Massive Open Online Courses (MOOCs) Using Cohesion Network Analysis .....	103
<i>Scott Crossley, Mihai Dascalu, Danielle McNamara, Ryan Baker, Stefan Trausan-Matu</i>	

Anchored Annotation to Support Collaborative Knowledge Construction .....	111
<i>Justin Plevinski, Jennifer Weible, Michael DeSchryver</i>	
Collaborative Game Design: A Bounded Case Study of Undergraduate Students in a Capstone Course .....	119
<i>Helen Fake, Jennifer Whiteman, Brenda Bannan</i>	
Which Visualization Guides Learners Best? Impact of Available Partner- and Content-Related Information on Collaborative Learning .....	127
<i>Melanie Erkens, Daniel Bodemer</i>	
Articulating Uncertainty Attribution as Part of Critical Epistemic Practice of Scientific Argumentation .....	135
<i>Hee-Sun Lee, Amy Pallant, Sarah Pryputniewicz, Trudi Lord</i>	
Creating Socially Relevant Mobile Apps: Infusing Computing into Middle School Curricula in Two School Districts .....	143
<i>Lijun Ni, Fred Martin</i>	
Exploring a Text-Mining Approach as Rapid Prototyping Tool for Formative Assessments in Inquiry-Based Online Learning .....	151
<i>Alejandro Andrade, Chris Georgen, Michael Stucker</i>	
Beyond Demographic Boxes: Relationships Between Students' Cultural Orientations and Collaborative Communication .....	159
<i>Nishan Perera, Alyssa Friend Wise</i>	
Collaborative Scientizing in Pokémon GO Online Communities.....	167
<i>Jason Yip, Travis Windleharth, Jin Ha Lee</i>	
Making the Invisible Visible: A New Method for Capturing Student Development in Makerspaces .....	175
<i>Richard Lee Davis, Bertrand Schneider, Paulo Blikstein</i>	
Dual Gaze as a Proxy for Collaboration in Informal Learning.....	183
<i>Kshitij Sharma, Ioannis Leftheriotis, Jama Noor, Michail Giannakos</i>	
A Mixed-Methods Approach for Studying Collaborative Learning Processes at Individual and Group Levels .....	191
<i>Catherine Dornfeld, Naxin Zhao, Sadhana Puntambekar</i>	
Learning Biology Coherently Through Complex Systems, Scientific Practices, and Agent-Based Simulations.....	199
<i>Miyoun Park, Emma Anderson, Susan Yoon</i>	
Expressing and Addressing Uncertainty: A Study of Collaborative Problem-Solving Dialogues.....	207
<i>Fernando J. Rodríguez, Kimberly Michelle Price, Kristy Elizabeth Boyer</i>	
Collaborative and Individual Scientific Reasoning of Pre-Service Teachers: New Insights Through Epistemic Network Analysis (ENA).....	215
<i>Andras Csanadi, Brendan Eagan, David Shaffer, Ingo Kollar, Frank Fischer</i>	
Time and Semantic Similarity – What is the Best Alternative to Capture Implicit Links in CSCL Conversations? .....	223
<i>Gabriel Gutu, Mihai Dascalu, Traian Rebedea, Stefan Trausan-Matu</i>	
The Dangers of Assuming Before Analysis: Three Case Studies of Argumentation and Cognition.....	231
<i>Kristine Lund, Matthieu Quignard</i>	
Teaching Accessibility in a Technology Design Course .....	239
<i>Kristen Shinohara, Cynthia L. Bennett, Jacob O. Wobbrock, Wanda Pratt</i>	
Group and Individual Level Effects of Supporting Socio-Cognitive Conflict Awareness and Its Resolution in Large SNS Discussion Groups: A Social Network Analysis.....	247
<i>Dimitra Tsovaltzi, Nikita Dutta, Thomas Puhl, Armin Weinberger</i>	



High School Students' Collaboration and Engagement With Scaffolding and Information as Predictors of Argument Quality During Problem-Based Learning.....	255
<i>Brian R. Belland, Nam Ju Kim, David M. Weiss, Jacob Piland</i>	
Using Multimodal Learning Analytics to Identify Aspects of Collaboration in Project-Based Learning .....	263
<i>Daniel Spikol, Emanuele Ruffaldi, Mutlu Cukurova</i>	
Student Re-Design of Deprived Neighbourhoods in Minecraft: Community-Driven Urban Development .....	271
<i>Rikke Magnussen, Anna Lindenhoff Elming</i>	
How Technology and Collaboration Promote Formative Feedback: A Role for CSCL Research in Active Learning Interventions .....	279
<i>Sally P. W. Wu, Martina Rau</i>	
On the Adoption of Social Network Analysis Methods in CSCL Research – A Network Analysis .....	287
<i>Marielle Dado, Tobias Hecking, Daniel Bodemer, H. Ulrich Hoppe</i>	
Think First: Fostering Substantive Contributions in Collaborative Problem-Solving Dialogues .....	295
<i>Mehmet Celepkolu, Joseph B. Wiggins, Kristy Elizabeth Boyer, Kyla McMullen</i>	
Finding the Community in Online Education: It's in the Instructors' Eyes.....	303
<i>Na Sun, Mary Beth Rosson</i>	
Revealing Interaction Patterns Among Youth in an Online Social Learning Network Using Markov Chain Principles.....	311
<i>Sarah Bishara, Jennifer Baltes, Taha Hamid, Taihua Li, Denise C. Nacu, Caitlin K. Martin, Jonathan Gemmell, Chris MacArthur, Daniela Raicu, Nichole Pinkard</i>	
Analyzing Students' Collaborative Regulation Behaviors in a Classroom-Integrated Open Ended Learning Environment .....	319
<i>Mona Emar, Michael Tscholl, Yi Dong, Gautam Biswas</i>	
Integrating Physical and Virtual Models in Biology: A Study of Students' Reasoning While Solving a Design Challenge.....	327
<i>Nicole Martin, Dana Gnesdilow, Sadhana Puntambekar</i>	
High Accuracy Detection of Collaboration From Log Data and Superficial Speech Features.....	335
<i>Sree Aurovindh Viswanathan, Kurt Vanlehn</i>	
Scripting and Orchestrating Learning Communities: A Role for Learning Analytics.....	343
<i>James D. Slotta, Alisa Acosta</i>	
Individual Versus Shared Design Goals in a Graph Construction Activity .....	351
<i>Jonathan Vitale, Lauren Applebaum, Marcia Linn</i>	
Who Signs Up and Who Stays? Attraction and Retention in an After-School Computer-Supported Program.....	359
<i>Maggie Renken, Jonathan Cohen, Tugba Ayer, Brendan Calandra, Aeslya Fuqua</i>	
Creating Parentopia: Design-Based Research to Develop an Interface for Parent Learning Communities and Networks .....	367
<i>Susan Walker</i>	
How Middle School Students Construct and Critique Graphs to Explain Cancer Treatment.....	375
<i>Camillia Matuk, Jiayuan Zhang, Marcia C. Linn</i>	
The Impact of Peer Tutors' Use of Indirect Feedback and Instructions .....	383
<i>Michael Madaio, Justine Cassell, Amy Ogan</i>	
Mobile City Science: Technology-Supported Collaborative Learning at Community Scale .....	391
<i>Katie Headrick Taylor, Deborah Silvis</i>	

Scientific Discourse of Citizen Scientists: A Collaborative Modeling as a Boundary Object.....	399
<i>Joey Huang, Cindy Hmelo-Silver, Rebecca Jordan, Troy Frensley, Steven Gray, Greg Newman</i>	
Cross-Community Interaction for Knowledge Building in Two Grade 5/6 Classrooms .....	407
<i>Jianwei Zhang, Maria Bogouslavsky, Guangji Yuan</i>	
Inclusive Collaborative Learning With Multi-Interface Design: Implications for Diverse and Equitable Makerspace Education.....	415
<i>Gabriela Richard, Sagun Giri</i>	
Investigating Immersion in Relation to Students' Learning During a Collaborative Location-Based Augmented Reality Activity .....	423
<i>Yiannis Georgiou, Eleni A. Kyza</i>	
A Self-Organizing Network of Schools That Transform Teacher and Student Learning Through Socio-Technical Co-Evolution.....	431
<i>Nancy Law, Leming Liang, Kenneth Cheng</i>	
A Meta-Synthesis of CSCL Literature in STEM Education.....	439
<i>Jessica McKeown, Cindy E. Hmelo-Silver, Heisawn Jeong, Kylie Hartley, Roosevelt Faulkner, Navo Emmanuel</i>	
Exploring Computational Modeling Environments as Tools to Structure Classroom-Level Knowledge Building .....	447
<i>Michelle Wilkerson, Becca Shareff, Brian Gravel, Yara Shaban, Vasiliki Laina</i>	
Using Rotating Leadership to Visualize Students' Epistemic Agency and Collective Responsibility for Knowledge Advancement .....	455
<i>Leanne Ma, Samuel Tan, Chew Lee Teo, Muhamad Ansar B. Kamsan</i>	
GroupWork: Learning During Collaborative Assessment Activities .....	463
<i>William T. Tarimo, Timothy J. Hickey</i>	
Explanation-Giving in a Collaborative Tangible Tabletop Game: Initiation, Positionality, Valence, and Action-Oriented .....	471
<i>Alyssa Wise, Alissa Nicole Antle, Jillian Warren</i>	
Mediating Access: How Visually Impaired Users Leverage Collaborative Learning to Keep Up With Mobile Phone Innovations.....	479
<i>Joyjeet Pal, Anandhi Viswanathan, Aditya Johri</i>	
Scoring Qualitative Informal Learning Dialogue: The SQuILD Method for Measuring Museum Learning Talk .....	487
<i>Jessica Roberts, Leilah Lyons</i>	
Learning to Model Ecosystems With Interaction Food Webs in Middle School Classrooms .....	495
<i>Michelle Lui, Tom Moher</i>	
The Role of Visual Representations Within the Scientific Practice of Explanation .....	503
<i>Rebecca Quintana, Tom Moher, James Slotta</i>	
Through the (Thin-Slice) Looking Glass: An Initial Look at Rapport and Co-Construction Within Peer Collaboration .....	511
<i>Jennifer K. Olsen, Samantha Finkelstein</i>	
Finding Collaboration Partners in a Scientific Community: The Role of Cognitive Group Awareness, Career Level, and Disciplinary Background.....	519
<i>Julia Eberle, Karsten Stegmann, Frank Fischer, Alain Barrat, Kristine Lund</i>	

## Volume 2

### Short Papers

Can We Rely on IRR? Testing the Assumptions of Inter-Rater Reliability .....	529
<i>Brendan R. Eagan, Bradley Rogers, Ronald Serlin, Andrew R. Ruis, Golnaz Arastoopour Irgens, David Williamson Shaffer</i>	
To What Extent Students' Epistemic Beliefs Influence Their Engagement in Argumentative Discourse and Attitudinal Change .....	533
<i>Omid Noroozi, Javad Hatami, Martin Mulder, Harm Biemans</i>	
Children's Emergent Leadership and Relational Thinking in Collaborative Learning.....	537
<i>Jingjing Sun, Julia Jackson, Mary Burns, Richard C. Anderson</i>	
Exploring Student Engagement in an Augmented Reality Game .....	541
<i>Nicolaas VanMeerten, Keisha Varma</i>	
Maker Portfolios as Learning and Community-Building Tools Inside and Outside Makerspaces .....	545
<i>Anna Keune, Kylie Pepler</i>	
Videoconferencing in Peer Review: Exploring Differences in Efficiency and Outcomes .....	549
<i>Elizabeth L. Pier, Joshua Raclaw, Cecilia E. Ford, Anna Kaatz, Molly Carnes, Mitchell J. Nathan</i>	
Self-Organizing Collaborations as Blueprints for CSCL Design .....	553
<i>Uzi Zevik Brami, Iris Tabak</i>	
CSCL and Vocational Education: A Bond Worthy of Investigation? .....	557
<i>Beat A. Schwendimann, Bram De Wever, Raija Hämäläinen, Alberto A. P. Cattaneo</i>	
Newcomer Integration Strategies in Blogger Online Knowledge Building Communities: A Dialog Analysis .....	561
<i>Nicolae Nistor, Yvonne Serafin</i>	
Designing Spaces for Collaboration in Practice-Based Learning .....	565
<i>Donal Healion, Sam Russell, Mutlu Cukurova, Daniel Spikol</i>	
Reflections on Pair E-Crafting: High School Students' Approaches to Collaboration in Electronic Textiles Projects .....	569
<i>Breanne K. Litts, Debora A. Lui, Sari A. Widman, Justice T. Walker, Yasmin B. Kafai</i>	
Embracing Learners With Visual Impairments in CSCL .....	573
<i>JooYoung Seo, Mona AlQahtani, Xueying Ouyang, Marcela Borge</i>	
Appropriating a Climate Science Discourse About Uncertainty in Online Lessons.....	577
<i>Kenneth Wright, Amy Pallant, Hee-Sun Lee</i>	
Engaging Everyday Science Knowledge to Help Make Sense of Data .....	581
<i>Susan B. Kelly, LuEttaMae Lawrence, Emma Mercier</i>	
Productive Knowledge Building Discourse Through Student-Generated Questions .....	585
<i>Ahmad Khanlari, Monica Resendes, Gaoxia Zhu, Marlene Scardamalia</i>	
Building Arguments Together or Alone? Using Learning Analytics to Study the Collaborative Construction of Argument Diagrams.....	589
<i>Irene-Angelica Chounta, Bruce M. McLaren, Maralee Harrell</i>	
Effects of Perspective-Taking Through Tangible Puppetry in Microteaching Role-Play .....	593
<i>Toshio Mochizuki, Takehiro Wakimoto, Hiroshi Sasaki, Ryoya Hirayama, Hideo Funaoi, Yoshihiko Kubota, Hideyuki Suzuki, Hiroshi Kato</i>	

Integrating Social Problem Solving with Programming to Enhance Science Agency Through Creation of Mobile Apps in Middle School .....	597
<i>Noora F. Noushad, Jooeun Shim, Susan A. Yoon</i>	
Symbiotic Learning Partnerships in Youth Action Sports .....	601
<i>Ty Hollett</i>	
Showing <i>and</i> Telling: Response Dynamics in an Online Community of Makers .....	605
<i>Omaima Almatrafi, Aditya Johri</i>	
Instant Sharing Makes Task More Engaging In Computer Aided Classroom .....	609
<i>Rafikh Shaikh, Harshit Agrawal, Nagarjuna G, Mrunal Nachankar</i>	
The Digital Use Divide and Knowledge Building .....	613
<i>Thérèse Laferrière, Alain Breuleux</i>	
Preparing Pre-Service Early Childhood Teachers to Teach Mathematics With Robots .....	617
<i>ChanMin Kim, Jiangmei Yuan, Cory Gleasman, Minyoung Shin, Roger B. Hill</i>	
Collaborating With Stakeholders in STEM Studios .....	621
<i>Kate Thompson, Les Dawes, Tanya Doyle, Harry Kanasa, Katherine Nickels, David Nutchey</i>	
Challenges in Implementing Small Group Collaboration in Large Online Courses .....	625
<i>Julia Erdmann, Nikol Rummel, Nina Christmann, Malte Elson, Tobias Hecking, Thomas Herrmann, H. Ulrich Hoppe, Nicole C. Krämer, Elias Kyewski, Astrid Wichmann</i>	
Collaborative Intelligent Tutoring Systems: Comparing Learner Outcomes Across Varying Collaboration Feedback Strategies .....	629
<i>Rachel Harsley, Barbara Di Eugenio, Nick Green, Davide Fossati</i>	
Examining Positive and Negative Interdependence in an Elementary School CSCL Setting .....	633
<i>Christian Hartmann, Jennifer K. Olsen, Charleen Brand, Vincent Aleven, Nikol Rummel</i>	
Transgressing Ideologies of Collaborative Learning and Working Spaces .....	637
<i>Jarek Sierschynski, Scott Spaulding</i>	
A Preliminary Study of University Students' Collaborative Learning Behavior Patterns in the Context of Online Argumentation Learning Activities: The Role of Idea-Centered Collaborative Argumentation Instruction .....	640
<i>Ying-Tien Wu, Li-Jen Wang, Teng-Yao Cheng</i>	
Reflective Structuration of Knowledge Building Practices in Grade 5 Science: A Two-Year Design-Based Research .....	644
<i>Dan Tao, Jianwei Zhang, Dandan Gao</i>	
Integrating Eye-Tracking Activities Into a Learning Environment to Promote Collaborative Meta-Semiotic Reflection and Discourse .....	648
<i>Stephen Sommer, Leighanna Hinojosa, Hilary Traut, Joseph Polman, Joanna Weidler-Lewis</i>	
Children's Participation in Rulemaking to Mitigate Process Problems in CSCL .....	652
<i>Yong Ju Jung, Dhvani Toprani, Shulong Yan, Marcela Borge</i>	
Anchor Code: Modularity as Evidence for Conceptual Learning and Computational Practices of Students Using a Code-First Environment .....	656
<i>Aditi Wagh, Sharona Levy, Michael Horn, Yu Guo, Corey Brady, Uri Wilensky</i>	
How Did a Grade 5 Science Community Co-Construct Collective Structures of Inquiry? .....	660
<i>Dan Tao, Jianwei Zhang</i>	
Learning About Climate Change Through Cooperation .....	664
<i>Lauren R. Applebaum, Kyle W. Fricke, Jonathan M. Vitale, Marcia C. Linn</i>	

Evaluating the Distribution of Students' Contributions in Theorizing: Idea Evenness in Knowledge Building Communities .....	668
<i>Gaoxia Zhu, Marlene Scardamalia, Ahmad Khanlari, Haipeng Wan</i>	
Developing Professional Competency in a CSCL Environment for Teamwork: Two TPACK Case Studies of Teachers as Co-Designers.....	672
<i>Elizabeth Koh, Helen Hong</i>	
Collaborative Argumentation During a Making and Tinkering Afterschool Program With Squishy Circuits.....	676
<i>Soo Hyeon Kim, Heather Toomey Zimmerman</i>	
Participatory Design With Students for Technology Integration: Shifting Power and Organizational Practices in an Urban School.....	680
<i>Ung-Sang Lee, Kimberly Gomez</i>	
Assessing Student Generated Infographics for Scaffolding Learning With Multiple Representations .....	684
<i>Engida Gebre</i>	
Role of Socio-Emotional Interactions on Mutual Trust and Shared Mental Models in a Case Study of Programming Teams.....	688
<i>Maedeh A. Kazemitabar, Susanne P. Lajoie</i>	
Fostering a Knowledge Building Community in a Primary Social Studies Class to Develop Humanistic View on Real World Problem .....	692
<i>Andy Ng Ding Xuan, Teo Chew Lee, Zahira Mohd Sedik, Lee Yu Ling</i>	
Exercising the Heart of History Education: Negotiating the Past Through a Principle-Based, Technological Driven Knowledge Building Culture .....	696
<i>Melvin Chan, Teo Chew Lee</i>	
<b>Symposia</b>	
Toward a Multi-Level Knowledge Building Innovation Network .....	703
<i>Marlene Scardamalia, Carl Bereiter, Thérèse Laferrière, Katerine Bielaczyc, Shaoming Chai, Carol K.K. Chan, Bodong Chen, Mei-Hwa Chen, Frank de Jong, Fernando Diaz del Castillo, Kai Hakkarainen, Yoshiaki Matsuzawa, Alexander McAuley, Mireia Montané, Cesar Nunes, Richard Reeve, Pirita Seitamaa-Hakkarainen, Jun Oshima, Hajime Shirouzu, Seng Chee Tan, Chew Lee Teo, Jan van Aalst, Telma Vinha, Jianwei Zhang</i>	
Making a Difference: Analytics for Quality Knowledge-Building Conversations.....	711
<i>Frank de Jong, Joan van den Ende, Hennie van Heijst, Yoshiaki Matsuzawa, Paul Kirschner, Jianwei Zhang, Mei-Hwa Chen, Feng Chen, Carolyn Rosé, Erick Velazquez Godinez, Sylvie Ratté, Bodong Chen, Carol Chan, Jan van Aalst, Christine Yang, Jun Oshima, Cindy Hmelo-Silver, Alyssa Wise</i>	
Technology and Applications for Collaborative Learning in Virtual Reality.....	719
<i>Scott W. Greenwald, Alexander Kulik, André Kunert, Stephan Beck, Bernd Fröhlich, Sue Cobb, Sarah Parsons, Nigel Newbutt, Christine Gouveia, Claire Cook, Anne Snyder, Scott Payne, Jennifer Holland, Shawn Buessing, Gabriel Fields, Wiley Corning, Victoria Lee, Lei Xia, Pattie Maes</i>	
CSCL and Eye-Tracking: Experiences, Opportunities and Challenges.....	727
<i>Kshitij Sharma, Patrick Jermann, Pierre Dillenbourg, Luis P. Prieto, Sarah D'Angelo, Darren Gergle, Bertrand Schneider, Martina Rau, Zach Pardos, Nikol Rummel</i>	
Libraries as Emerging Spaces for Computer-Supported Collaborative Learning in Schools and Communities .....	735
<i>Victor R. Lee, Carrie Tzou, Megan Bang, Philip Bell, Shelley Stromholt, Nancy Price, Meixi Ng, Yasmin Kafai, Orkan Telhan, Richard Davis, K-Fai Steele, Barrie Adleberg, Jennifer Kahn, Rogers Hall, Abigail Phillips, Jennifer Hansen, Mimi Recker, Brigid Barron</i>	

Collaborative Problem Solving: Innovating Standardized Assessment .....	743
<i>Lei Liu, Jiangang Hao, Jessica J. Andrews, Mengxiao Zhu, Robert J. Mislevy, Patrick Kyllonen, Alina A. von Davier, Deirdre Kerr, Thales Ricarte, Art Graesser</i>	
Scripted and Unscripted Aspects of Creative Work With Knowledge .....	751
<i>Carl Bereiter, Ulrike Cress, Frank Fischer, Kai Hakkarainen, Marlene Scardamalia, Freydis Vogel</i>	
<b>Posters</b>	
Does Collaboratively Constructing Contrasting-Case Animations Facilitate Learning?.....	761
<i>David Shaenfield</i>	
Socio-Semantic Network Analysis of ijCSCL Articles: Development of CSCL Ideas in ISLS.....	763
<i>Jun Oshima, Takashi Tsunakawa</i>	
Students' Engagement in a Science Classroom: Does Cognitive Diversity Matter? .....	765
<i>Lijia Lin, Jiangshan Sun, Xudong Zheng, Jia Yin, Jian Zhao</i>	
Technology Affordances for CSCL: A Preliminary Review .....	767
<i>Navo Emmanuel, Cindy Hmelo-Silver, Heisawn Jeong, Kylie Hartley, Jessica McKeown</i>	
Epistemic Game Design for Democratic and Media Education .....	769
<i>Jeremy Stoddard, Kimberly Rodriguez, Mason Rayner, Zachari Swiecki, David Williamson Shaffer</i>	
Text Chatting in Collaborative Writing: Its Role in Coordinating Activities .....	771
<i>Martina Bientzle, Wiebke Steffen, Heisawn Jeong, Ulrike Cress, Joachim Kimmerle</i>	
Designing for Collaborative Literary Inquiry .....	773
<i>Allison H. Hall, Renato Carvalho</i>	
Exploring the Road to Place-Based Collaborative Learning via Telepresence Robots .....	775
<i>Jian Liao, Jaclyn Dudek</i>	
The Effect of Peer Interaction on Task Efficiency and Learning Engagement in Digital Game-Based Learning .....	777
<i>Jewoong Moon, Fengfeng Ke, Xinhao Xu, Yanjun Pan, Zhaihuan Dai</i>	
Breaking the SEAL: A CSCL History Teaching Methodology to Support Transition Into Undergraduate Education .....	779
<i>Paul Flynn, Mary Fleming, Barry Houlihan, Niall McSweeney</i>	
The Effect of Varied Gender Groupings on Science Knowledge and Argumentation Skills Among Middle Level Students.....	781
<i>Pi-Sui Hsu, Margot Van Dyke, Eric Monsu Lee, Thomas J. Smith</i>	
Democratic Engagement: A Progressive Approach to CSCL.....	783
<i>Bob Coulter</i>	
“You switch, and I press”: Comparing Children’s Collaborative Behavior in a Tangible and Graphical Interface Game .....	785
<i>David Kim</i>	
Transformational Change in Humanistic Learning Communities: A Case Study of Person- and Idea-Centered Integration .....	787
<i>Yotam Hod, Dani Ben-Zvi</i>	
An Emotion Awareness Tool for the Sharing of Emotions: What Impact on Computer-Supported Collaborative Processes? .....	789
<i>Sunny Avry, Gaëlle Molinari, Guillaume Chanel, Thierry Pun, Mireille Bétrancourt</i>	
In Search of Helpful Group Awareness Metrics in Closed-Type Formative Assessment Tools .....	791
<i>Pantelis Papadopoulos, Antonis Natsis, Nikolaus Obwegeser</i>	

Adding Time to Social Networks: A New Perspective on Using Learning Analytics for Learning Environment Design.....	793
<i>Yang Xu</i>	
Promoting Equity and Access in Public Libraries' Computer-Supported Youth Programming.....	795
<i>Ligaya Scaff, Saba Kawas, Katie Davis, Mega Subramaniam, Kelly H. Hoffman</i>	
Exploring Ways of Contributing to Math Talk in a Knowledge Building Community.....	797
<i>Stacy A. Costa, Marlene Scardamalia</i>	
Brokering Collaboration Among Children for Equity .....	799
<i>Yanghee Kim, Sherry Marx, Tung Nguyen</i>	
Capturing and Visualizing: Classroom Analytics for Physical and Digital Collaborative Learning Processes .....	801
<i>Sarah K. Howard, Kate Thompson, Jie Yang, Jun Ma, Abelardo Pardo, Harry Kanasa</i>	
Orchestration Challenges Raised by Transporting a Traditional Writing Activity Into a Web-Based Computer Supported Collaborative Language Learning Activity .....	803
<i>Eirini Dellatola, Thanasis Daradoumis, Yannis Dimitriadis</i>	
Teacher Regulation of Collaborative Learning: Research Directions for Learning Analytics Dashboards.....	805
<i>Anouschka van Leeuwen, Nikol Rummel</i>	
How Do K-12 In-Service Teachers Plan for Collaboration in Game-Based Lessons? .....	807
<i>Kathryn Wozniak, Aaron Kessler</i>	
Collaborative Scientific Modeling in the Classroom.....	809
<i>David Quigley, Tamara Sumner</i>	
Go GRASP: A Mobile Application to Facilitate Orchestration in Active Learning Classrooms .....	811
<i>Nathaniel Lasry, Michael Dugdale, Elizabeth S. Charles, Chris Whittaker, Kevin Lenton</i>	
Cued Gestures: Their Role in Collaborative Discourse on Seasons .....	813
<i>Robert C. Wallon, Robb Lindgren</i>	
Habits of Civic Collaboration in a Digital Carnival: Fostering Other-Oriented Collaboration in a High School Game Making Workshop.....	815
<i>Gideon Dishon, Yasmin Kafai</i>	
Co-Regulation Competences: Can They Be Measured?.....	817
<i>Christopher A. Williams, Tina Seufert, Armin Weinberger</i>	
SynergyNet Into Schools: Facilitating Remote Inter-Group Collaborative Learning Using Multi-Touch Tables.....	819
<i>Andrew Joyce-Gibbons, James McNaughton, Elaine Tan, Nick Young, Gary Beauchamp, Tom Crick</i>	
Affordances and Constraints of Immersive Virtual Environments for Identity Change.....	821
<i>Tamara Galoyan, Mamta Shah, Aroutis Foster</i>	
The Effect of the Screen Size of Multi-Touch Tables on Collaborative Problem Solving Interactions .....	823
<i>Saadeddine Shehab, Emma Mercier</i>	
Designing Engineering Tasks for Collaborative Problem Solving .....	825
<i>Saadeddine Shehab, Emma Mercier</i>	
Designing Simulations for Evaluating Collaborative Problem Solving in Electronics.....	827
<i>Jessica J. Andrews, Paul Horwitz, John Chamberlain, Al Koon, Cynthia McIntyre, Alina A. von Davier</i>	
Teachers' Cultural Competency: Media Interactive Case Studies as a Low-Stake Practice Space.....	829
<i>Yoon Jeon Kim, Kevin Robinson, Kesiena Owho-Ovuakporie, Justin Reich</i>	

Laboratory of Co-Inquiry, Co-Design, Co-Teaching, and Co-Regulation (Co4-Lab).....	831
<i>Pirita Seitamaa-Hakkarainen, Kati Sormunen, Tiina Korhonen, Anniina Koskinen, Jari Lavonen, Kai Hakkarainen</i>	
Cultivating a Culture of Learning to Foster Socioscientific Reasoning.....	833
<i>Hava Ben-Horin, Carmit Pion, Yael Kali</i>	
Increasing Access and Engagement Through Iterative Design.....	835
<i>Kimberly Rodriguez, Mason Rayner, Jeremy Stoddard, Zachari Swiecki, David Williamson Shaffer</i>	
Tablets in the CSCL Classroom: A Lens on Teachers’ Instrumental Geneses .....	837
<i>Teresa Cerratto Pargman, Jalal Nouri</i>	
Context and Collaborative Problem Solving (CPS): The Development of Observable Signifiers to Inform the Design of CPS Learning Analytics .....	839
<i>Rosemary Luckin, Mutlu Cukurova, Manolis Mavrikis, Eva Millan</i>	
Girls, Robotics Learning, and Internalized Stereotypes: Is There a Relationship?.....	841
<i>Florence R. Sullivan, P. Kevin Keith, Ricardo Poza</i>	
Framing the Design Space for Mobile Facilitation Tools in Exhibit Settings .....	843
<i>Priscilla Jimenez Pazmino, Leilah Lyons, Brian Slattery</i>	
Distributed Teaching and Learning in Pokémon Go .....	845
<i>Kelly M. Tran</i>	
Sequencing and Fading Worked Examples and Collaboration Scripts to Foster Mathematical Argumentation – Working Memory Capacity Matters for Fading .....	847
<i>Matthias Schwaighofer, Freydis Vogel, Ingo Kollar, Anselm Strohmaier, Sarah Ottinger, Ilka Terwedow, Stefan Ufer, Kristina Reiss, Frank Fischer</i>	
Embodied Activities As Entry Points for Science Data Literacy .....	849
<i>Stephen R. Sommer, Joseph L. Polman</i>	
Touch   Don’t Touch: Exploring the Role of Interactive Displays in Natural History Museums to Help Visitors Appreciate Objects Behind Glass.....	851
<i>Michael Horn, Jessica Roberts, Amartya Banerjee, Steven McGee, Matt Matcuk</i>	
The Impact of Play, Gesture, and Teacher Prompts on Student Explanations About the Particulate Nature of Matter .....	853
<i>Bria Davis, Xintian Tu, Joshua A. Danish, Noel Enyedy</i>	
The Effects of Explicit Collaborative Argumentation Instruction in Collaborative Argumentation-Based Learning Activities in High School Context.....	855
<i>Ying-Tien Wu, Li-Jen Wang, Teng-Yao Cheng</i>	
Visualizations to Support Facilitation: The Instructors’ View .....	857
<i>Yuxin Chen, Gurpreet Birk, Cindy. E. Hmelo-Silver, Maedeh Kazemitabar, Stephen Bodnar, Susanne. P. Lajoie</i>	
Bilingual Learning Spaces: Lessons From Using WhatsApp Videos in a Ghanaian Rural Context .....	859
<i>Mama Adobea Nii Owoo</i>	
Research on an International Network of STEM Media Making and Student-Led Participatory Teaching .....	861
<i>Eric Hamilton, Nicholas Nardi, Joyce Ndegemo, Danielle Espino</i>	
Collaborative Sense Making in a Tablet-Mediated Informal, Place-Based Learning Environment .....	863
<i>Susan M. Land, Heather Toomey Zimmerman, Chrystal Maggiore, Soo Hyeon Kim, Jessica Briskin</i>	



The Dragon Swooping Cough: Mass Community Participation in a Virtual Epidemic Within a Tween Online World .....	865
<i>Deborah A. Fields, Yasmin Kafai, Michael T. Giang, Nina Fefferman, Jacqueline Wong</i>	
Bridging Students' Practical and Formal Epistemology of Science Through Epistemic Reflection Embedded in a Computer-Supported Knowledge-Building Environment .....	867
<i>Feng Lin, Carol K.K. Chan</i>	
Individual Role-Based Profiles for Successful Team Engagement in Knowledge Building Environments.....	869
<i>Ahmad Khanlari, Monica Resendes, Marlene Scardamalia, Gaoxia Zhu</i>	
Misconceptions and Their Evolution in Knowledge Building Communities .....	871
<i>Ahmad Khanlari, Carl Bereiter, Marlene Scardamalia</i>	
Networks in Small-Group Structure in Knowledge Building Discourse .....	873
<i>Xueqi Feng, Jan van Aalst, Carol K.K. Chan, Yuqin Yang</i>	
Asking Semantically Similar Questions in Knowledge Building Communities: Patterns and Effects.....	875
<i>Gaoxia Zhu, Monica Resendes, Ahmad Khanlari, Marlene Scardamalia, Ying-Tien Wu</i>	
Evaluation of an Online-Environment to Prevent Frustration and Procrastination in Literature-Based Inquiry Learning .....	877
<i>Julia Eberle, Tim Schönfeld, Selma Arukovic, Nikol Rummel</i>	
Multi-User Framework for Collaboration and Co-Creation in Virtual Reality.....	879
<i>Scott W. Greenwald, Wiley Corning, Pattie Maes</i>	
Examining Regulation of Idea Improvement and Knowledge Advances in a Principle-Based Knowledge Building Environment .....	881
<i>Yuyao Tong, Carol K.K. Chan, Jan van Aalst, Kun Liu</i>	
Epistemic Understanding of Discourse and Collective Responsibility in a Knowledge Building Community .....	883
<i>Yuyao Tong, Carol K.K. Chan, Jan van Aalst</i>	
<b>Interactive Events</b>	
Braincandy: A Cloud-Based Platform Providing Students Authentic, Engaging, and Safe Spaces to Articulate and Refine Oral Argumentation.....	887
<i>Kevin Close, J. Bryan Henderson</i>	
Investigating Computer Supported Collaborative Learning in Collegiate E-sports.....	891
<i>Gabriela T. Richard, R. William Ashley, Zachary McKinley</i>	
<b>Workshops</b>	
Synthesizing CSCL Perspectives on the Theory, Methods, Design, and Implementation of Future Learning Spaces.....	897
<i>Yotam Hod, Julia Eberle, Maya Benichou, Elizabeth Charles, Ulrike Cress, Frank Fischer, Peter Goodyear, Yael Kali, Ingo Kollar, Jim Slotta, Kate Thompson, Phil Tjietjen, Pippa Yeoman</i>	
Publishing in the Learning Sciences: A Journal Writers' Workshop .....	901
<i>Mitchell Nathan, Erica Halverson, Jeremy Roschelle, Carol Chan, Susan Yoon, Sten Ludvigsen, Jan van Aalst</i>	
Emergent Practices and Material Conditions in Tablet-mediated Collaborative Learning and Teaching .....	905
<i>Teresa Cearrto-Pargman, Isa Jahnke, Crina Damsa, Miguel Nussbaum, Roger Säljö</i>	
Enabling and Understanding Embodied STEM Learning .....	909
<i>Caro Williams-Pierce, Candace Walkington, David Landy, Robb Lindgren, Sharona Levy, Mitchell J. Nathan, Dor Abrahamson</i>	

EPCAL: Computer-Supported Collaboration at Scale.....	913
<i>Jiangang Hao, Lei Liu, Jessica Andrews, Diego Zapata, Alina von Davier, Art Graesser</i>	
Establishing a Foundation for Collaborative Process Evaluation and Adaptive Support in CSCL .....	916
<i>Cynthia D'Angelo, Cindy Hmelo-Silver, Marcela Borge, Alyssa Wise, Bodong Chen</i>	
Mobile Computing in CSCL: A Hands-On Tutorial on the ARIS Game Design Platform .....	920
<i>Breanne K. Litts, Stephanie Benson, Whitney Lewis, Chase Mortensen</i>	
Reflections and Discussions about NAPLeS Learning Resources for the Learning Sciences.....	922
<i>Freydis Vogel, Frank Fischer, Yotam Hod, Kris Lund, Daniel Sommerhoff</i>	
Digitally-Mediated Design Thinking in CSCL Environments.....	924
<i>Jonan Phillip Donaldson, Amanda Barany, Brian K. Smith</i>	
<b>Early Career Workshop</b>	
CSCL 2017 Early Career Workshop .....	931
<i>Susan A. Yoon, Manu Kapur, Armin Weinberger</i>	
Subgoal Learning in Online STEM Instruction .....	932
<i>Lauren Margulieux</i>	
Analysing Collaborative Problem-Solving From Students' Physical Interactions.....	934
<i>Muthu Cukurova</i>	
Using Computer Models and Collaboration to Explore Energy Concepts .....	936
<i>Lauren Applebaum</i>	
Intercultural Computer Supported Collaborative Learning .....	938
<i>Vitaliy Popov</i>	
Fostering Epistemic Growth in CSCL Environment .....	940
<i>Feng Lin</i>	
Technology-Enhanced Collaborative Learning for Improved Interactivity, Collaboration, and Flexibility in Higher Education and Corporate Training.....	942
<i>Annelies Raes</i>	
ECW Contribution: Anouschka van Leeuwen.....	944
<i>Anouschka van Leeuwen</i>	
How Tangible User Interfaces Can Contribute to Collaborative Language Learning .....	946
<i>Yun Wen</i>	
Designing Collective Learning in Mixed Reality Environments.....	948
<i>Michelle Lui</i>	
Enhancing Collaboration and Assessment: A Learning Analytics Approach.....	950
<i>Wanli Xing</i>	
An Exploration in Learning Through Art, Science, and Making.....	952
<i>Emma Anderson</i>	
Peer Assessment: Students Helping Peers to Learn.....	955
<i>Melissa Patchan</i>	
The Hidden Curriculum of Online Learning: Discourses of Whiteness, Social Absence, and Inequity.....	957
<i>Murat Oztok</i>	

Understanding Learning Through, With, and About Data.....	959
<i>Jessica Roberts</i>	
Growing Teamwork Competency in 21st Century Learners .....	962
<i>Elizabeth Koh</i>	
Seeking and Designing for Educational Equity Within the Maker Movement.....	964
<i>Debora Lui</i>	
Exploring the Embodied Aspects of Imaginative and Creative Processes.....	966
<i>Rolf Steier</i>	
<b>Mid Career Workshop</b>	
CSCL 2017 Mid Career Workshop .....	971
<i>Kristine Lund, Frank Fischer</i>	
Exploring Social, Cognitive, and Representational Issues in Learning Through Playful Co-Design.....	972
<i>Camillia Matuk</i>	
Creativity in Post-Secondary STEM Teaching and Learning.....	974
<i>Jennifer D. Adams</i>	
Looking at Technology in CSCL.....	976
<i>Teresa Cerratto Pargman</i>	
<b>Doctoral Consortium</b>	
The CSCL 2017 Doctoral Consortium Workshop.....	981
<i>Jun Oshima, Kylie Peppler, Pirita Seitamaa-Hakkarainen, Kai Hakkarainen, Yasmin B. Kafai</i>	
Connecting Science and Engineering Practices: Using Collaborative Annotation to Improve Student Design Justifications.....	983
<i>Elizabeth McBride</i>	
Idea Identification and Analysis (I2A) for Sustained Idea Improvement in Knowledge Building Discourse .....	985
<i>Alwyn Vwen Yen Lee, Seng Chee Tan</i>	
Public Peer Review for Collaborative Learning in MOOCs.....	987
<i>Xu Wang</i>	
Embodied Learning With Gesture Augmented Computer Simulations in Middle School Science Classrooms .....	989
<i>Robert Wallon</i>	
Toward Adaptive Collaborative Support for Elementary Students Learning Computer Science.....	991
<i>Jennifer Tsan, Collin F. Lynch, Kristy Elizabeth Boyer</i>	
Knowledge Building Discourse in a Large Community.....	993
<i>Xueqi Feng, Jan van Aalst, Carol K.K. Chan</i>	
Fostering Sustained Knowledge Building Practices in Grade 5 Science: A Reflective Structuration Approach.....	995
<i>Dan Tao, Jianwei Zhang</i>	
Visualizing Networked Relations to Support Computer-Supported Collaborative Learning .....	997
<i>Marielle Dado, Daniel Bodemer</i>	
Promoting Productive Failure in Collaborative Design Contexts: A Collaborative Failure-Management Learning Model .....	999
<i>Shulong Yan</i>	

The Effect of Playing Portal 2 on Collaborative Problem Solving .....	1001
<i>Dima Kassab, Caro Williams-Pierce</i>	
The Role of Context in Virtual Environments: Investigating Student Reasoning With Online Places .....	1003
<i>Amanda Siebert-Evenstone, David Williamson Shaffer</i>	
Conceptualizing Scaffolding for Science Learning in Classrooms and Museums Using Mixed-Methods Approaches .....	1005
<i>Catherine Dornfeld, Sadhana Puntambekar</i>	
Evolution of Knowledge Building Teacher Professional Development Communities.....	1007
<i>Derya Kici, Marlene Scardamalia</i>	
Teacher Leadership in Information and Communications Technology (ICT) Reform .....	1009
<i>Helen Hong</i>	
<b>Indexes</b>	
Author Index.....	A1-A7
Keyword Index .....	K1-K12

# **Keynote Presentations**



# Unbundling and Inequality in Higher Education

Laura Czerniewicz, University of Cape Town, South Africa, [laura.czerniewicz@uct.ac.za](mailto:laura.czerniewicz@uct.ac.za)

**Abstract:** The unbundling and rebundling of higher education refer to the ways that the components of the traditional university experience—including resources, provision, support, assessment, accreditation, and research—become disaggregated or disintermediated, and reorganised and redefined through new restructured relationships, reassembled and available in new ways. These profound changes may play out in quite different ways at different levels of the system: Within the university, across the system nationally, or across the entire sector globally. This talk explores what these emergent relationships might look like with a particular concern for the implications regarding inequality and the quality of the educational experience. Depending on the interests served and the models developed, reintermediation offers opportunities to either exacerbate or ameliorate educational inequality, with concomitant profound implications for teaching and learning itself.

*Laura Czerniewicz* (@Czernie), the Director of the Centre for Innovation in Learning and Teaching (CILT) at the University of Cape Town in South Africa, is an associate professor in the Centre for Higher Education Development. She is committed to equity of access and success in higher education. Her research interests include the technologically-mediated practices of students and academics, the nature of the changing higher education environment and the geopolitics of knowledge, underpinned by a commitment to surfacing the expressions of inequality within and across contexts. Laura is involved with policy work, is a contributor to national and global conversations in varied formats and serves on the advisory boards of a variety of international higher education educational and technology publications. She blogs intermittently at <http://lauraczerniewicz.uct.ac.za/> and can be followed as @czernie on Twitter.

## Virtual Selves and Learning

D. Fox Harrell, Massachusetts Institute of Technology, fox@csail.mit.edu

**Abstract:** Educational technologies such as adaptive learning systems, educational games, and Massive Open Online Courses (MOOCs) have proliferated. Almost all students these days play videogames. Given the widespread and growing use of such technologies, which invariably involve virtual identities such as user profiles and avatars, it is important to better understand their impacts and to establish innovative and best practices for learning. In this talk, Harrell explores how our social identities are complicated by their intersection with computing and learning technologies including videogames, virtual worlds, social media, and related digital media forms. With an emphasis on equity, Harrell will explore how virtual identities both implement and transform persistent issues of class, gender, sex, race, ethnicity, and the dynamic construction social categories more generally.

*D. Fox Harrell* is Professor (as of July 1, 2017) of Digital Media in both the Comparative Media Studies Program Computer Science and Artificial Intelligence Laboratory at MIT. His research focuses on the relationship between imaginative cognition and computation. He founded and directs the MIT Imagination, Computation, and Expression Laboratory (ICE Lab) to develop new forms of computational narrative, gaming, social media, and related digital media based in computer science, cognitive science, and digital media arts. He is the author of the book *Phantasmal Media: An Approach to Imagination, Computation, and Expression* (MIT Press, 2013). In 2010, Professor Harrell received a NSF CAREER Award and, in 2014-2015, he was awarded a Fellowship at the Center for Advanced Study in the Behavioral Sciences (CASBS) at Stanford University and was the recipient of the Lenore Annenberg and Wallis Annenberg Fellowship in Communication.



# Symmetrical Advancement: Teachers and Students Sustaining Idea-Centered Collaborative Practices

Teo Chew Lee, Ministry of Education, Singapore, teo\_chew\_lee@moe.gov.sg

**Abstract:** With insights established by learning sciences research, educating in the 21st century requires changing not just the procedures of the classroom practice but the underlying principles that guide the practice. There are many strategic approaches to scale such principle-based innovative practice, many involved coordinated efforts across school organization and school administrators. Regardless of the scaling approach, the most important likely remained to be the concerted effort to shift teachers' conception of their students, the trust on their students and the imagination required to see possibilities of deep learning. Similarly, we seek to shift these conceptions in middle managers and finally the idea of 21st-century school perceived by school leaders.

The content of this talk is taken from an 8-year old Knowledge Building Project in Singapore. In this talk, we attempt to trace the growth of four visible dimensions of the project and the challenges embed within each dimension. Knowledge building practice requires a significant shift from knowledge deepening to knowledge creation paradigm. This particular KB project has been focusing on working with teachers to design and enact idea-centered and collaborative classrooms while tackling all curriculum and assessment demands along with physical and time constraints prevalent in every school. It warrants a detailed study of the areas of growth to ensure symmetry in advancement in all stakeholders in schools and considering all dimensions of schools and teaching and learning processes. This is needed so that intensive innovations, such as KB, have a chance to take root in practice. In fact, creating symmetrical knowledge advancement in all our collaborators and collaborating schools has always been a core principle of design in the research.

Four visible areas of growth include (i) growth in number and connectedness of teachers in practice; (ii) growth in the dimensions of teaching and learning involved in the innovation; (iii) growth in ownership of practice; (iv) growth in research considerations; (v) growth in the role of the researcher.

*Teo Chew Lee* is the Lead Specialist in Learning Partnership in Educational Technology at the Educational Technology Division in the Ministry of Singapore. She began exploring Knowledge Building (KB) technologies in her classroom at the beginning of her career as a science educator about two decades ago. She completed her Ph.D. at OISE/UT, Canada and joined the ministry in 2009 to lead a research group on translating KB theories technologies into Singapore classrooms. Chew Lee uses a design-based research approach to study ways to facilitate Singapore teachers in designing knowledge building environments and has worked at various level and subjects from primary school to junior colleges. She focuses her work on understanding teachers' problem spaces in their discourse and their work to design idea-centered learning environments. From 2013, she extended the impact of the work to create a KB network learning community in Singapore that builds new understanding of the practice. At the ministry, Chew Lee also does extensive work on Educational Technology in curriculum design & development at the policy level. She currently heads a group of specialists and teacher-researchers in exploring educational technology for active learning with technology in English Language, Chinese Language, Sciences, and the Humanities.



# **Full Papers**



# Collaborative Learning on Multi-Touch Interfaces: Scaffolding Elementary School Students

Lara Johanna Schmitt and Armin Weinberger  
l.schmitt@edutech.uni-saarland.de, a.weinberger@edutech.uni-saarland.de  
EduTech, Saarland University

**Abstract:** Multi-touch interfaces allow for direct and simultaneous input by several co-present learners. Additional scaffolding may or may not be needed to ease or problematize tasks that involve intuitive bodily experiences. In this study, a tablet app (“Proportion”) is supposed to enable two novices (about 10 years old) to collaboratively construct an understanding of proportional relations. In a  $2 \times 2$  factorial design ( $n = 162$ ), effects of facilitating strategy prompts (with / without) and problematizing verbalization prompts (with / without) regarding the variables task focus, emotions, quality of dialogue and learning gains have been investigated. While the strategy prompts did not have any significant influence, the verbalization prompts had versatile effects: On one hand, quality of talk was improved, on the other hand, task focus and emotions were negatively affected. Learning gains were limited to near transfer task types and comparable over conditions.

**Keywords:** collaborative learning, embodiment, proportional reasoning, scaffolding, tablets

## Multi-touch interfaces for collaborative embodied learning

Multi-touch interfaces allow for co-present collaborative learning and, specifically, for equal, simultaneous, and direct manipulation of a learning environment (Roschelle et al., 2010). Using a multi-touch interface together can support beneficial forms of interactions, like whole-group discussions, fluid interactions (Alvarez, Brown, & Nussbaum, 2011), equal participation, and joint time on task, while process losses are reduced (Mercier, Higgins, & da Costa, 2014). The direct manipulation is supposed to reduce the “cognitive distance between intent and execution” (Rick, 2012; p. 316) and also enables forms of embodied learning experiences (Schneps et al., 2014), using the body to interact with a tool in order to construct knowledge. Despite the possible benefits of embodied approaches in learning mathematics (Abrahamson & Lindgren, 2014), embodied hands-on learning activities may provoke off-task behavior and distract from the actual learning goals (e.g., Danish, Enyedy, Saleh, Lee, & Andrade, 2015), thus come at the cost of verbalization, abstraction, and reflection of knowledge. Hence, additional scaffolding for reflection may constructively complement, but may also disrupt the embodied learning experiences. To address this issue, the present study examines to what extent scaffolding can support learning in an embodied learning environment or not.

## Scaffolding CSCL with collaboration scripts

Supporting learners’ (inter-)actions for knowledge construction can take various forms. One widely applied form is scripting of collaborative processes, which focuses on fostering reflection (Kobbe et al., 2007). Building and sharing arguments together may productively coalesce with embodied learning experiences, helping to translate between embodied experience and abstract conceptualization. A large body of research shows that argumentative practices can be facilitated in CSCL scenarios through scripting (e.g., Gijlers, Weinberger, van Dijk, Bollen, & van Joolingen, 2013).

One of the core design questions of scripting to scaffold collaborative learning processes is whether scripts should make the task easier or harder (Reiser, 2004): On one hand, learners often need structural or strategic support to proceed in task solution processes; providing this support, by e.g. reducing the complexity of a task or learning environment, is making the task easier. For example, scaffolding can guide learners to better understand what they have to do, what is a sensible sequence of activities, or where should they focus their attention (Jackson, Krajcik, & Soloway, 1998). Those strategic types of scaffolds are prospective, i.e. directed to future behavior. One goal of scripting is internalization of effective scripts by the learner (Fischer, Kollar, Stegmann, & Wecker, 2013). On the other hand, learners’ understanding can be enhanced by problematizing aspects of the tasks; to this end, prompting reflection and verbalization is feasible; contrary to the first way of scaffolding, it is making the task harder (Reiser, 2004). Stopping learners from superficial and fast problem solving and prompting them to engage in verbalizations, aims at fostering deep elaboration of the learning material (King, 1990). The quality of learners’ interactive talk has been found to positively relate to learning (Paus, Werner, & Jucks, 2012; Teasley,

1995; Webb, 1989). Scaffolding learners to reflect typically is a retrospective, thus reflective activity, directed into the past.

So potentially, there is a need for enriching embodied learning experiences on multi-touch interfaces with scripting for deliberate activities, reflective collaborative discourse, or both. In this paper, we consider whether prompting learners to apply heuristics that make the task easier and / or prompting learners to verbalize underlying mathematical concepts, abstract from the immediate experience and hence making the task more difficult, can facilitate processes and outcomes of co-located CSCL using multi-touch devices. While positive effects can be assumed, the opposite might happen: Do well-meant prompts actually interfere with intuitive embodied learning environments? While prompts theoretically would improve cognition and problem solving strategies, their sudden appearance during gameplay can also seriously ruin the experience (Wouters et al., 2015). Scaffolding could be both: counterproductive, thus ruining and disrupting the bodily, hands-on learning experience, or complementary, thus improving and supplementing it with processes of encoding that are necessary for memorization and learning.

## Proportional reasoning

Proportional reasoning is defined as “reasoning with ratios, rates, and percentages” (Jitendra, Star, Rodriguez, Lindell, & Someki, 2011, p. 731) and is a central topic in mathematics (Boyer, Levine, & Huttenlocher, 2008). Children typically have difficulties with it; particularly handling fractions is a problem (Mix, Levine, & Huttenlocher, 1999). Reinholz, Trninic, Howison, and Abrahamson (2010) blame a lack of senso-motorical, embodied experiences for causing difficulties in handling proportions. Researchers identified typical misconceptions that children face in proportional reasoning tasks: First, application of counting strategies to proportions, especially when concrete units are presented (Boyer et al., 2008). Second, application of addition rules to proportions (Mix et al., 1999). Third, failure to form correct proportional representations from discrete units, that is not building a relative relationship between numerator and denominator of a fraction (Boyer et al., 2008; Mix et al., 1999). Fostering proportional reasoning requires effective learning set-ups. Research showed that collaboration in combination with hypothesis testing has the potential to live up to that (Ellis, Klahr, & Siegler, 1993). Recent developments also aim at including embodied learning experiences. Similar to the Mathematical Image Trainer (Reinholz et al., 2010), we have developed the "Proportion" iPad app that aims at improving children's proportional reasoning by letting them directly manipulate proportional relations.

## Research question and hypotheses

This study clarifies the following research question utilizing a 2×2 design: To what extent can collaborative learning with tablets be supported by different types of prompts (facilitating strategy prompts / “STRAT” vs. problematizing verbalization prompts / “VERB”), regarding learning processes and outcomes?

- **Hypothesis 1:** STRAT prompts and VERB prompts and their combination will result in higher task focus. Facilitating the task with the STRAT prompts should help learners to make progress and stay on track. Problematizing the task with the VERB prompts also should direct learners' attention to relevant task features.
- **Hypothesis 2:** The STRAT prompts will induce more positive emotions; the VERB prompts will induce more negative emotions. Making the task easier (STRAT) is supposed to trigger positive emotions like enjoyment, because task progression is facilitated. Making the task harder (VERB) is supposed to result in negative emotions like frustration or anger, because task progression is slowed down by difficult verbal tasks.
- **Hypothesis 3:** The VERB prompts will enhance the quality of dialogic interactions. Explicitly requesting learners to externalize their knowledge and engage in discussions should result in higher transactivity and higher epistemic quality.
- **Hypothesis 4:** STRAT prompts and VERB prompts and their combination will result in higher learning gains, regarding both near and far transfer task types. The internalization of task solving strategies suggested by the STRAT prompts should have a positive impact on learning. Higher-order verbalizations as scaffolded by the VERB prompts impact learning positively by promoting deeper elaboration and multiple perspectives.

## Methods

### Sample

Participants (fourth graders; mean age: 10.34 years (SD=.55); 50% male) were acquired from seven primary schools in Germany. All participants had a consent form signed by their legal representatives, informing on data collection and analyses. In total,  $n=162$  participants took part in the experiments, being tested in four different experimental conditions (control, STRAT, VERB, and VERB-STRAT).  $\chi^2$ -tests and ANOVAs did not reveal statistically significant differences between conditions regarding the control variables gender, handedness, experience with or owning of a multi-touch device, pre-test or –questionnaire.

### Material

The learning environment “Proportion” is an iPad application (Rick, 2012). The interface is designed to afford incorporation of hand / arm movements, i.e. aiming at actively experiencing and embodying proportional relations. The app consists of a fixed sequence of 21 levels with 5 to 23 tasks each. In Proportion, learners control two bars (orange and blue) that are positioned vertically next to each other. To solve the tasks, learners need to resize the bars, so that they are in the right relation to each other as indicated by the associated numbers. See figure 1a for an example: In this case, the bars need to be resized so that the left bar’s height would be  $\frac{2}{3}$  compared to the right one. Once a task is solved, the numbers of the next task appear. An owl acts as a pedagogical agent and provides feedback, e.g. announces “correct” once a task is solved. The owl also voices the varied prompts; see figure 1b and 1c. The prompt versions (A, B, or C; see table 1) alternate in the same fashion for all dyads in the prompted conditions: After the first task of each level, one STRAT (in the STRAT condition), one VERB (in the VERB condition), or both prompts (in the VERB-STRAT condition) appear on the screen.

Table 1: Overview on prompts

Version	Level	STRAT	VERB
A	1, 4, 7, 10, 13, 16, 19	“Tip for all tasks: What is higher, orange or blue? First say it out loud, then you do!”	“Explain to your learning partner: What did one need to do in order to solve the task?”
B	2, 5, 8, 11, 14, 17, 20	“Tip for all tasks: First think and provide an estimate, then set the bars' correct height!”	“Describe to your learning partner: What could one learn in this task?”
C	3, 6, 9, 12, 15, 18, 21	“Tip for all tasks: If the task is hard and you're stuck, what might help is to discuss and talk!”	“Explain to your learning partner: What do all of these tasks have in common?”

Pre- and post- questionnaires and math tests were applied. The pre-questionnaire collected socio-demographic data, previous experiences with multi-touch devices, and attitudes towards math, school, and collaborative learning. The post-questionnaire measured participants’ acceptance of the app, subjective learning gain, and aspects of the collaboration. The math test consisted of tasks related to fractions and proportions; the tasks were classified as requiring lower vs. higher levels of transfer (near transfer tasks vs. far transfer tasks). The near transfer tasks were designed to capture the strategies that were used to progress within Proportion. The far transfer tasks aimed at capturing knowledge on proportions and fractions more broadly. At maximum, one could reach 21 points in the math test: 13 for the far transfer tasks and 8 for the near transfer tasks.



Figure 1. The interface of Proportion (a), example of displaying one STRAT (b) and one VERB (c) prompt, two children collaboratively using Proportion (d).

Seven iPads of the second generation were used for the experiments. Video cameras and microphones recorded participants’ interactions with each other and Proportion.

## Experimental procedure

The experimental procedure followed a pre-test – intervention – post-test design and has been carried out by one of several trained experimenters. Experiments took place inside the respective schools. Participants were randomly assigned to conditions by lot. After a general welcoming and introduction to the learning session, the participants individually filled in the pre-questionnaire and the pre- math test (10 minutes). Next, the students worked collaboratively with the Proportion app for 40 minutes, see figure 1d; this phase was video-taped. After a 5 minutes break, the participants individually filled in the post-questionnaire and post- math test (10 minutes). Altogether, one experiment cycle covered about two regular school lessons (90 minutes).

## Variables

The dependent variables have been aggregated (i.e. averaged) on the dyad level as we cannot assume statistical independence of the dyadic learners. Regarding the analysis of learners' non-verbal behavior and dialogues, we chose to sample every second problem of every second of the up to 21 levels learners reached in the given 40 minutes. This allowed us to focus on continuous interactions (reactions to success, reactions to the prompts for the prompted conditions, and the problem solving process) on typical tasks throughout the learning experience rather than special cases of initial coordination or final conclusions. The samples started with the appearance of "Correct" of the first problem of the level, covered the prompt, extended over the second problem and ended when this problem has been solved. The video samples have been transcribed and video coded using coding schemes that have been developed based on previous work (e.g., Weinberger & Fischer, 2006) and the data at hand. Interrater reliability was measured using Krippendorff's  $\alpha$ . Dependent variables analyzed for this contribution comprise task focus (measured as off-task behavior), negative and positive emotions (gestures), transactivity and epistemic quality, and learning gains in near and far transfer tasks (math-tests). Instances of off-task behavior, positive and negative emotions have each been summed up and their average occurrence per coded segment has been calculated, see table 2 for the coding criteria.

Table 2: Coding criteria for nonverbal behavior (video analysis)

Variable	Krippendorff's $\alpha$	Indicators
Off-task behavior	.95	looking around the classroom, looking into the camera, interactions with participants outside the own group
Positive emotions	.81	clapping into one's own or the learning partner's hands, throwing hands up in the air, clenching the fist, showing thumbs-up
Negative emotions	.77	threatening the iPad, facing the palms upwards, dismissive hand gesture, face-palming

Regarding the transcripts of dialogues, the utterance / turn was the unit of analysis and one category has been assigned to each. Transactivity refers to the extent that participants react to and base their verbal contributions on their partners' contributions. Epistemic quality refers to the content of participants' utterances: Are the utterances off- or on-topic and are they a pure regulation of their interaction or (different levels) of actual task-related explanations?

Table 3: Coding criteria for transactivity and epistemic quality (transcripts of videos)

Variable	Krippendorff's $\alpha$	Categories	Relative quality
Transactivity	.78	Externalization	0
		Externalization as reaction	0
		Acceptance	1
		Refusal	1
		Elicitation	2
		Integration	4
		Conflict-oriented consensus building	8
Epistemic quality	.90	Off-topic utterance	0
		On-topic: regulation of the interaction	0
		On-topic: concrete task-related regulation	1
		On-topic: abstract content-related regulation	4
		On-topic: Strategies / procedural knowledge	8



To illustrate the coding scheme, here is one example from the data: Student A: “This needs to go higher”, student B: “You have to go lower. It needs to be three times the size”. Regarding transactivity, we coded externalization for student A, and conflict-oriented consensus building for student B, as student B refutes the partner’s suggestion, but also provides an alternative suggestion and justifies it. Regarding epistemic quality, we coded concrete task-related regulation for student A, because the remark is closely tied to what can be seen on the screen, and we coded abstract content-related regulation for student B, because the remark refers to more abstract knowledge (“three times the size”), going beyond what can be seen on the screen. Taking their relative quality into account (see e.g., Teasley, 1997; Weinberger & Fischer, 2006), the raw categories of transactivity and epistemic quality, see table 3, have each been aggregated to global scores.

## Results

Hypothesis 1 stated a main effect for both STRAT and VERB as well as an interaction between the prompts on task focus. A two-factorial ANOVA revealed no statistically significant main effect of STRAT ( $F(1,76)=.159, p=.691$ ), but a highly significant main effect of VERB:  $F(1,76)=18.190, p=.000, \eta^2=.19$ . However, this highly significant effect is contrary to our hypothesis, as off-task behavior is actually reinforced, and not reduced, with the presence of the VERB prompt, see figure 2. The interaction of the prompts was not statistically significant ( $F(1,76)=2.451, p=.122$ ).

Hypothesis 2 stated more positive emotions for STRAT and more negative emotions for VERB. A two-factorial ANOVA revealed no statistically significant effect of STRAT on positive emotions:  $F(1,76)=1.919, p=.170$ . The effect of VERB on negative emotions is statistically significant with  $F(1,76)=7.019, p=.010, \eta^2=.09$ , see figure 2.

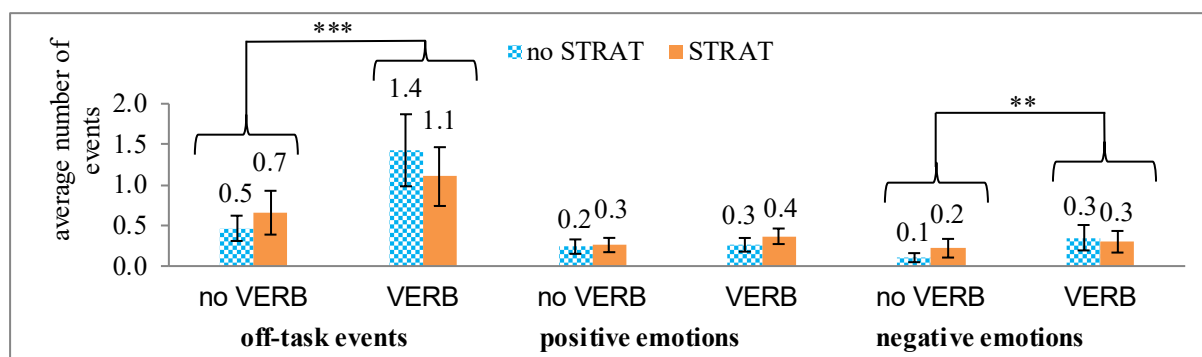


Figure 2. Average number of off-task events, and positive and negative emotions per coded segment; error bars represent 95% CIs; \*  $p \leq .05$ ; \*\*  $p \leq .01$ ; \*\*\*  $p \leq .001$ .

Hypothesis 3 stated a main effect of VERB on the quality of dialogic interactions, regarding transactivity as well as epistemic quality. Two-factorial ANOVAs were conducted. Regarding transactivity, there was a significant main effect of VERB ( $F(1,70)=7.241, p=.009, \eta^2=.094$ ), indicating higher transactivity in the presence of the VERB prompts than without it, see figure 3a. Also regarding the epistemic quality score, there was a significant main effect of VERB ( $F(1,70)=9.437, p=.003, \eta^2=.119$ ), indicating a higher epistemic quality in the presence of the VERB prompts than without it, see figure 3b.

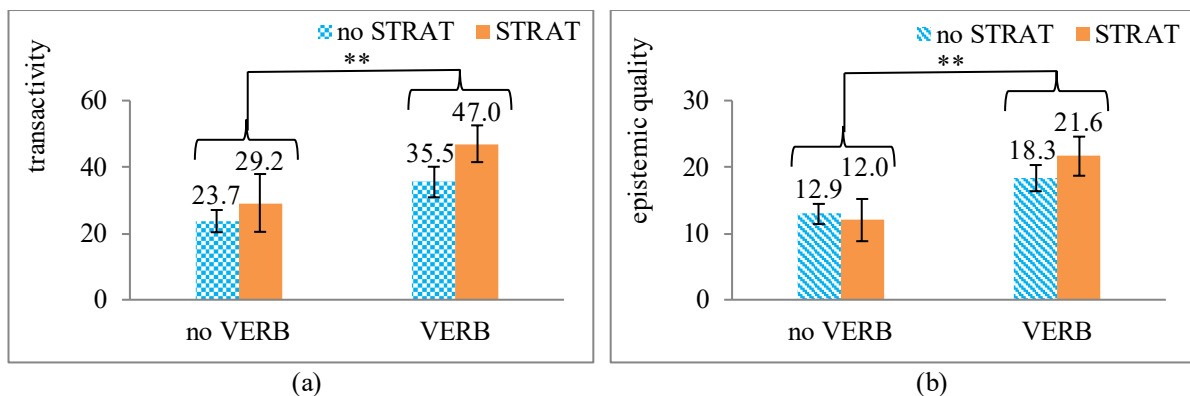


Figure 3. Scores in transactivity (a) and epistemic quality (b); error bars represent 95% CIs; \*  $p \leq .05$ ; \*\*  $p \leq .01$ ; \*\*\*  $p \leq .001$ .

Hypothesis 4 stated that STRAT and VERB and their combination will increase learning gains, regarding both near and far transfer task types. Two-factorial repeated measures ANOVAs were conducted. Regarding near transfer tasks, there was no statistically significant interaction of STRAT  $\times$  point in time ( $F(1,77)=.585, p=.447$ ), VERB  $\times$  point in time ( $F(1,77)=.061, p=.805$ ) or STRAT  $\times$  VERB  $\times$  point in time ( $F(1,77)=1.502, p=.224$ ). Only an improvement independent from conditions could be found (main effect of point in time):  $F(1,77)=11.179, p=.001, \eta^2=.13$ . Regarding far transfer tasks, there was no statistically significant interaction of STRAT  $\times$  point in time ( $F(1,77)=3.231, p=.076$ ), VERB  $\times$  point in time ( $F(1,77)=.112, p=.739$ ) or STRAT  $\times$  VERB  $\times$  point in time ( $F(1,77)=.003, p=.953$ ). Contrary to near transfer tasks, no generic improvement from pre- to post-test (main effect of point in time) in the far transfer tasks could be found:  $F(1,77)=.174, p=.677$ .

## Discussion

Collaborative learning with a shared multi-touch device might enable young learners to bodily experience mathematical properties. The learning experiences may be enhanced by different ways of scaffolding, which is being investigated in this study.

Hypothesis 1 claimed a main effect on task focus for the strategy prompts, the verbalization prompts as well as their combination. While the strategy prompts did not have a statistically significant effect on task focus, the verbalization prompts' influence was strong but reverse to our hypothesis: The verbalization prompts actually increased off-task behavior. An interaction effect of the prompts on task focused behavior could not be found. Hypothesis 1 needs to be rejected. On one hand, the strategy prompts could not measurably help learners to stay on track, on the other hand, while we intended to direct learners' attention to relevant aspects of the task with the verbalization prompts, we actually achieved the opposite: Learners got distracted more easily.

Those results get solidified when looking at hypothesis 2 which predicted more positive emotions as a consequence of the strategy prompts and more negative emotions as a consequence of the verbalization prompts. Again, presence of the strategy prompts did not make a difference regarding positive emotions; possibly, any prompt, even if it is there to help learners, might be unwelcome in an embodied learning environment, as it disrupts the ongoing immersive activities. As hypothesized, the verbalization prompts induced more negative emotions, which may be linked to an increase in perceived task difficulty. Hypothesis 2 is being rejected concerning positive emotions affected by the strategy prompts, but confirmed regarding negative emotions affected by the verbalization prompts. Taken together with the results of hypothesis 1 (verbalization prompts increase off-task behavior) those results are disconcerting. Participants being confronted with the verbalization prompts showed about twice as much negative emotions and off-task behavior as participants without it. There seems to be a thin line between enriching a game-like learning environment in a way that facilitates learning, while keeping students' engagement high (Deater-Deckard, El Mallah, Chang, Evans, & Norton, 2014). However helpful prompts like "Explain to your learning partner..." have proven to be in the past, they may still come at the cost of raising the difficulty too much, interrupting the flow, and disengaging learners in specific immersive, embodied CSCL experiences. Hence, these results merit further investigation of when, how, and why prompting may produce these problematic side effects.

Hypothesis 3 predicted higher levels of transactivity and higher epistemic quality caused by the verbalization prompts. Indeed, this could be confirmed. Verbalization prompts had positive medium-sized effects on students' dialogues on both, transactivity as well as epistemic quality. Similar findings have already been established with adult learners, e.g. showing positive effects of scripting on argumentation (Weinberger,

Stegmann, & Fischer, 2010). Here we can show that prompts can raise discussion quality even with young learners (see also van Dijk, Gijlers, & Weinberger, 2014).

Hypothesis 4 claimed that strategy prompts and verbalization prompts and their combination will result in higher learning gains. Only for near transfer tasks, a medium-to-large statistically significant improvement from pre- to post- math test over all conditions could be found. There was no interaction with any of the prompts, there was also no significant improvement regarding the far transfer tasks.

The strategy prompts were widely ineffectual. We could not find support for the assumption that they would actually make the task easier. Motivational as well as cognitive factors could have played a role. We could not force participants to thoroughly read the prompts and in the light of the engaging and fun gaming activity, the prompts might have simply been ignored and the gaming activity continued instead. Also, limited cognitive capacities in processing, understanding and applying the suggested strategies might account for the limited effects of the strategy prompts.

The verbalization prompts impacted many variables, in beneficial and detrimental ways. They showed to enhance quality of talk, not to affect learning gain, and to have adverse effects on task focus and negative emotions, presumably having raised the difficulty (too much). Since the verbalization prompts did augment quality of learners' talk (see results of hypotheses 3), a positive effect on learning could have been expected. Different alternative explanations may apply why that was not the case. Verbalization of difficult to verbalize cognitive representations can corrupt performance in non-verbal tasks (verbal overshadowing effect; Schooler, 2002). Also, social preference might have played a role: Van Dijk et al. (2014) found that social aspects are crucial in young learners' collaborative learning, i.e. students who got along well with each other produced better results. In our study, the dyad formation was random, so maybe that is why participants did not always collaborate effectively. Regarding the topical embodiment perspective, however, our study does not find much support for embodiment to be a huge factor in learning. Likewise, limited effects of game-based learning environments have been attributed to the sole acquisition of intuitive knowledge that is never verbalized and translated to out-of-game contexts like formal tests (Wouters et al., 2015). More studies comparing Proportion to a comparable learning environment that does not feature embodied approaches could firstly, shed more light on the actual role of embodiment in learning and secondly, inspire the development of new learning environments that sensibly bring together embodied, hands-on learning experiences with phases of activities that serve abstraction and reflection.

In this study, we investigated how prompting for strategic behavior and/or reflective dialogue effects collaborative learning of proportional reasoning in an intuitive, embodied learning environment. Challenges of incorporating phases of reflection into bodily learning environments while keeping flow and engagement high have been highlighted. Future research could focus on how to further address this issue.

## References

- Abrahamson, D., & Lindgren, R. (2014). Embodiment and embodied design. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 358-376). Cambridge, UK: Cambridge University Press.
- Alvarez, C., Brown, C., & Nussbaum, M. (2011). Comparative study of netbooks and tablet PCs for fostering face-to-face collaborative learning. *Computers in Human Behavior*, 27(2), 834-844.
- Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: where young children go wrong. *Developmental Psychology*, 44(5), 1478-1490.
- Danish, J. A., Enyedy, N., Saleh, A., Lee, C., & Andrade, A. (2015). Science through technology enhanced play: Designing to support reflection through play and embodiment. In O. Lindwall, P. Häkkinen, T. Koschman, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015: Vol. 1* (pp. 332-339). Gothenburg, Sweden: The International Society of the Learning Sciences.
- Deater-Deckard, K., El Mallah, S., Chang, M., Evans, M. A., & Norton, A. (2014). Student behavioral engagement during mathematics educational video game instruction with 11-14 year olds. *International Journal of Child-Computer Interaction*, 2(3), 101-108.
- Ellis, S., Klahr, D., & Siegler, R. S. (1993). *Effects of feedback and collaboration on changes in children's use of mathematical rules*. Paper presented at the Meetings of the Society for Research in Child Development, New Orleans.
- Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, 48(1), 56-66.
- Gijlers, H., Weinberger, A., van Dijk, A. M., Bollen, L., & van Joolingen, W. (2013). Collaborative drawing on a shared digital canvas in elementary science education: The effects of script and task awareness support. *International Journal of Computer-Supported Collaborative Learning*, 8(4), 427-453.

- Jackson, S. L., Krajcik, J., & Soloway, E. (1998). The design of guided learner-adaptable scaffolding in interactive learning environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI* (pp. 187–194). Los Angeles, CA: ACM Press.
- Jitendra, A. K., Star, J. R., Rodriguez, M., Lindell, M., & Someki, F. (2011). Improving students' proportional thinking using schema-based instruction. *Learning and Instruction, 21*(6), 731–745.
- King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal, 27*(4), 664–687.
- Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*(2), 211–224.
- Mercier, E. M., Higgins, S. E., & da Costa, L. (2014). Different leaders: Emergent organizational and intellectual leadership in children's collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning, 9*(4), 397–432.
- Mix, K. S., Levine, S. C., & Huttenlocher, J. (1999). Early fraction calculation ability. *Developmental Psychology, 35*(5), 164–174.
- Paus, E., Werner, C. S., & Jucks, R. (2012). Learning through online peer discourse: Structural equation modeling points to the role of discourse activities in individual understanding. *Computers and Education, 58*(4), 1127–1137.
- Reinholz, D. L., Trninic, D., Howison, M., & Abrahamson, D. (2010). It's not easy being green: Embodied artifacts and the guided emergence of mathematical meaning. In P. Brosnan, D. Erchick, & L. Flevaris (Eds.), *Proceedings of the 32nd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education: Vol. 6* (pp. 1488–1496). Columbus, OH: PME-NA.
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13*(3), 273–304.
- Rick, J. (2012). Proportion: A tablet app for collaborative learning. In *Proceedings of the 11th Annual Interaction Design and Children Conference (IDC 2012)* (pp. 316–319). New York, NY: ACM Press.
- Roschelle, J., Rafanan, K., Bhanot, R., Estrella, G., Penuel, B., Nussbaum, M., & Claro, S. (2010). Scaffolding group explanation and feedback with handheld technology: Impact on students' mathematics learning. *Educational Technology Research and Development, 58*(4), 399–419.
- Schneps, M. H., Ruel, J., Sonnert, G., Dussault, M., Griffin, M., & Sadler, P. M. (2014). Conceptualizing astronomical scale: Virtual simulations on handheld tablet computers reverse misconceptions. *Computers & Education, 70*, 269–280.
- Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology, 16*(8), 989–997.
- Teasley, S. (1995). The Role of Talk in Children's Peer Collaborations. *Developmental Psychology, 31*(2), 207–220.
- Teasley, S. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. B. Resnick, R. Säljö, C. Pontecorvo, & B. Burge (Eds.), *Discourse, tools and reasoning: Essays on situated cognition* (pp. 361–384). Berlin, Germany: Springer.
- Van Dijk, A. M., Gijlers, H., & Weinberger, A. (2014). Scripted collaborative drawing in elementary science education. *Instructional Science, 42*(3), 353–372.
- Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13*(1), 21–39.
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71–95.
- Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior, 26*(4), 506–515.
- Wouters, P., van Oostendorp, H., ter Vrugte, J., Vandercruyssen, S., de Jong, T., & Elen, J. (2015). The role of curiosity-triggering events in game-based learning for mathematics. In J. Torbeyns, E. Lehtinen, & J. Elen (Eds.), *Describing and studying domain-specific serious games. Advances in game-based learning* (pp. 191–207). Springer.

## Acknowledgements

The authors would like to thank Jochen Rick for designing and programming the Proportion app and the various student assistants supporting the research project.

# Behavioral and Relationship Patterns in an Online Collaborative Reading Activity

Hai-Peng Wan, Faculty of Education, Beijing Normal University, haipengwan@gmail.com  
Qi Wang, Faculty of Education, Beijing Normal University, wangqi.20080906@163.com  
Sheng-Quan Yu, Advanced Innovation Center for Future Education, Beijing Normal University, yusq@bnu.edu.cn

**Abstract:** Online collaborative reading has been widely implemented as an instructional activity in various context, with many studies demonstrating effective learning outcomes. Based on knowledge construction theory, we put forward an online collaborative reading approach to learning from an academic handbook in a graduate-level course. Through examination of the behavioral patterns and relationship patterns of different phases in the course, we found that student contributions to peers' micro-courses were not symmetric; some students would submit irrelevant comments in different collaborative phases, and almost all students kept in touch with each other directly. Our study also indicated that students' task load and consistency were two important factors to affect their collaborative performance. Our findings would help course teachers design and conduct collaborative reading activities at the postsecondary level in future.

## Introduction

Knowledge construction has been widely used and discussed, which emphasizes that students construct new knowledge through social interactions (Huang, 2002; Kanuka & Anderson, 2007). Advancement in internet technology has led to an increase in instructional activities with computer support, such as English language reading instruction (Chen, Chen, & Sun, 2010). Based on knowledge construction theory, previous studies of online collaborative reading mainly focus on students' reading attitude, reading comprehension, reading strategy, motivation, and learning effectiveness, and reveal that students in collaborative learning environments demonstrate stronger cognitive development, more positive learning attitude, and higher learning motivation than control groups (Chen & Chen, 2014; Ding, 2009; Lin, Chen, Yang, Xie, & Lin, 2014).

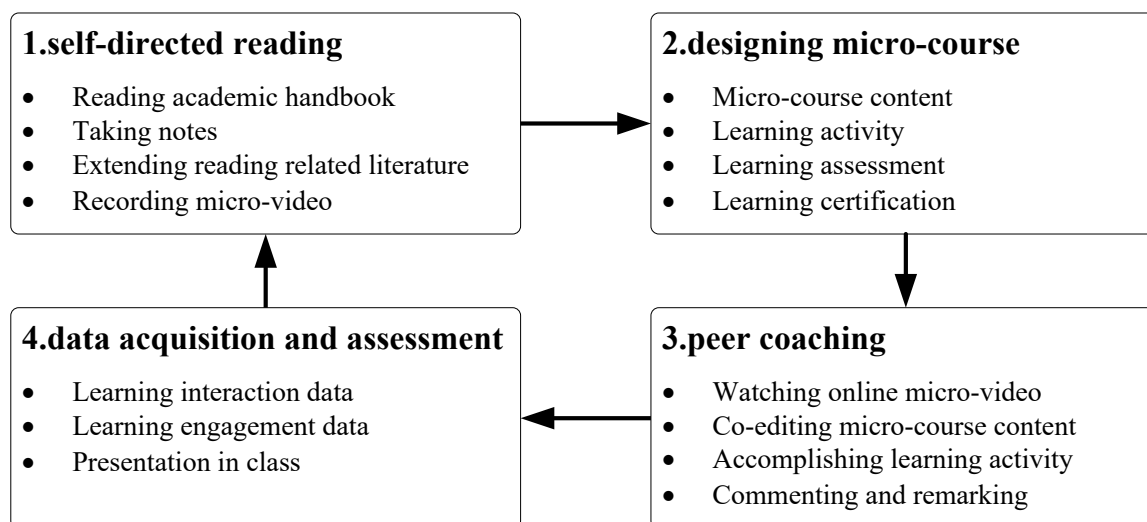


Figure 1. The online collaborative reading procedure. Adopted from Wan et al. (2015)

Peer tutoring is vital to collaborative learning (Robinson & Hullinger, 2008). However, students in conventional collaborative learning environments tend to share and compare the available information rather than to construct new knowledge (Ma, 2009; Schellens, Van Keer, De Wever, & Valcke, 2008). They are usually only required to finish tasks according to reading materials rather than generate new knowledge for peer to study (Chen & Chen, 2014; Lin et al, 2014). In this study, we put forward an innovative collaborative reading approach with four stages: self-directed reading, designing micro-course, peer coaching, and data acquisition and assessment (as shown in Figure 1; for detail referring to Wan, Yu, Cui & Chiang, 2015). Apart from sharing information, students not only need to generate new information through their own reading, but they also need to finish learning the

information generated by their peers in this innovative collaborative reading activity. Identification of students' collaboration pattern is of value to pedagogical and technical design (Lin et. al, 2014). For example, the sequential analysis technique could demonstrate the sequences of students' action and has been widely used to analyze online collaborative discussion (Hou & Wu, 2011; Shukor, Tasir, Van der Meijden, & Harun, 2014). Therefore, this study attempted to investigate the students' behavior and relationship patterns by lag sequential analysis and social network analysis to provide reference for course teachers to design and conduct collaborative reading activities in higher education.

## Method

### Participants

The participants were twelve graduate students and one visiting scholar in a graduate course, New Development of Educational Technology, at a university in China. The course contained lectures implemented by professors and the reading activity of an English academic handbook which made up the students' course assignment. Apart from simply reading the English academic handbook, the course required students to make a micro-course of each article they read according to their own understanding. Students were also required to learn and contribute to peers' micro-courses with the Learning Cell System (an online collaborative learning system described below). All the students received prior training and were capable of using this learning system with ease.

### Procedures

At the beginning, the course teacher selected the Handbook of Research on Educational Communications and Technology (4th edition) published by Springer as reading materials. This handbook was written in English and included nine sections with seventy-four articles, covering foundations, methods, assessment and evaluation, general instructional strategies, domain-specific strategies and models, design, planning, and implementation, emerging technologies, technology integration, and look forward. The goal of reading the handbook was to support the students in developing a systematic understanding of educational technology research and its development.

Afterwards, each student randomly chose five or six articles. The course teacher divided the whole semester into three phases and each phase lasted six weeks. During each phase, the students completed four tasks, (i.e., reading two articles, making two micro-courses, learning twenty-four micro-courses of peers with learning system, and making one presentation in an offline class). Those micro-courses required students to create a complete teaching structure, including a micro digital resource (e.g., micro-video), a learning activity and a learning assessment.

Finally, all of the interaction data generated in the process of the collaborative reading activity were exported to one Excel file for further lag sequential analysis and social network analysis.

## Instruments

### Learning Cell System

An online collaborative learning system entitled Learning Cell System (LCS, <http://lcell.bnu.edu.cn>) (Yu, Yang, Cheng, & Wang, 2015) was used to observe the behavioral and social network patterns by supporting the whole process of the collaborative reading activity. The heart of LCS is an open, networked, communal knowledge community. Its main functions are learning cell, knowledge group, knowledge cloud, learning tool, personal space, and learning community. A learning cell serves as a micro-course, which usually includes learning content, learning activity and learning assessment. Students could share their ideas and information, and contribute to peers' ideas through authoring or coauthoring a learning cell.

### Coding scheme

To understand the learners' process of social knowledge construction, the items in Gunawardena, Lowe and Anderson's (1997) coding scheme were adopted as the scheme has been widely used in many studies of online collaborative learning patterns (Choo, Kaur, Fook, & Yong, 2014; Hou & Wu, 2011; Yang, Li, Guo, & Li, 2015). Gunawardena et. al (1997) divided the knowledge construction process into five dimensions: 1) sharing and comparing information, 2) discovery and exploration of dissonance or inconsistency, 3) negotiation of meaning and co-construction of knowledge, 4) testing and modification of the proposed synthesis and co-construction, and 5) agreement statement(s) and applications of newly constructed meanings (see B1 to B5 in Table 1). In addition, we added a new dimension B6 to express irrelevant information to this collaborative reading task. Thus, the coding scheme for content analysis in online collaborative reading behaviors of English academic handbook is shown in

Table 1, which also provides behavior type and content example for each item to more clearly to clarify different behaviors.

Each log or comment message was treated as a unit and coded, and the codes were then arranged in chronological order. 9343 log messages and 851 comment messages were yielded during the 120-day observation. These log messages were coded according to their categories (e.g., creating learning cell, browsing, cooperative editing learning cell, remark, reflection) defined in LCS. These comment messages were coded by two coders with the same expertise according to the scheme and the kappa value was 0.73.

Table 1: Coding scheme for knowledge collaborative construction behaviors

Code	Dimension	Behavior types and examples
B1	Sharing/comparing of information	Creating learning cell, adding learning activities, uploading learning material, and releasing reading work and concept map.
B2	Discovery and exploration of dissonance or inconsistency among participants	Browsing, collecting, and giving feedback on learning cell created by peer; coming up with confusion during learning. Can “qualitative research” be translated into “质性研究” or “定性研究”?
B3	Negotiation of meaning/co-construction of knowledge	Cooperative editing learning cell, modifying video and content, adjusting content structure, comment. Discussion with peer on topics and give suggestion on problems.
B4	Testing and modification of proposed synthesis or co-construction	Remark, comment, annotation, pointing out problem. I cannot hear clearly of the back of video. I think “educational design research” translated into “教学设计研究” will lead to misunderstanding. The micro-course does not include learning activity.
B5	Agreement statement(s)/application of newly constructed meaning	Reflection, comment, annotation. Writing reflective journal entries. I think teacher cannot be replaced by pedagogical agent. I agree that both the internal validity and external validity are important for a study.
B6	Other interactions with no relations with the reading task	Irrelevant information. Very good. You have done a good job. You are an idol for me. I have got a lot from it.

## Results and discussion

At the end of semester, we found that the students did not strictly follow the pre-class requirement made by course teacher, (i.e., reading two articles and making two micro-courses in each phase). Four micro-courses were submitted after the end of course and one micro-course was incomplete. Hence, the coded 10194 messages were about those sixty-nine micro-courses. The sum frequency of B1 was 312, of B2 was 7490, of B3 was 963, of B4 was 729, of B5 was 401 and of B6 was 299. The distribution of those coded messages in each phase is shown in Figure 2. As shown in Figure 2, the behavior frequency in phase 3 is more than phase 1 and phase 2. And the behavior frequency difference was very big because students only made 18 micro-courses during the first phase, 12 micro-courses during the second phase and 41 micro-courses during the last phase. In each phase, the behavior frequency of B2 was always larger than other behaviors, even the sum of other behaviors.

GSEQ 5.1 (Bakeman & Quera, 2011) was used to conduct lag sequential analysis by analyzing the behavioral patterns of knowledge construction in collaborative reading process. Table 2 shows the frequency of each behavioral category immediately following another behavioral category in different phases (Phase 1, Phase 2, and Phase 3). The columns represent the starting behaviors, whereas the rows represent the behaviors that occurred after the starting behaviors finished. The numbers represent the total number of times a column behavior occurred immediately after a row behavior ended (e.g., in row 2 column 3, the number 216 meant that “B3 occurred immediately after B1,” which happened 216 times in Phase 1).

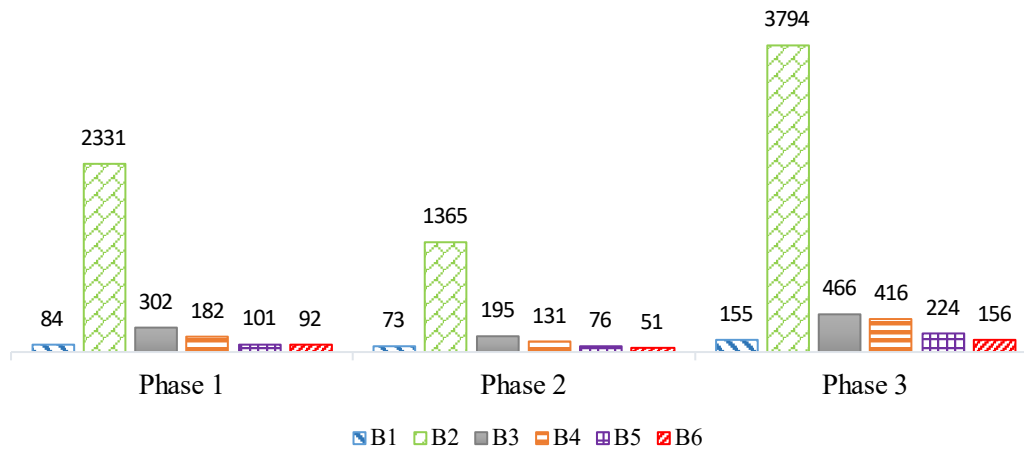


Figure 2. Frequencies of knowledge collaborative construction behavior (Phase 1 to Phase 3).

Table 2: Frequency transition table (Phase 1 to Phase 3)

Frequency		B1	B2	B3	B4	B5	B6	Total
Phase 1	B1	42	26	12	1	0	2	83
	B2	16	1889	216	136	26	32	2315
	B3	7	217	42	26	6	4	302
	B4	0	43	13	8	68	50	182
	B5	0	84	10	6	1	0	101
	B6	1	72	9	5	0	4	91
	Total	66	2331	302	182	101	92	3074
Phase 2	B1	24	34	12	1	2	0	73
	B2	24	1077	118	95	26	15	1355
	B3	10	115	46	19	3	2	195
	B4	1	35	10	10	42	32	130
	B5	1	62	5	5	2	0	75
	B6	2	41	4	1	1	2	51
	Total	62	1364	195	131	76	51	1879
Phase 3	B1	68	57	26	1	1	2	155
	B2	37	2944	71	289	71	44	3756
	B3	7	356	33	57	5	6	464
	B4	0	97	24	139	139	101	416
	B5	0	199	6	7	7	1	223
	B6	2	141	6	1	1	2	156
	Total	114	3794	466	416	224	156	5170

Table 3 shows the results of adjusted residuals. The Z-score of a sequence greater than 1.96 means that the connectivity of this sequence reached statistical significance ( $p < 0.05$ ) (Bakeman & Gottman, 1997). According to those 22 statistically significant sequences with Z-score greater than 1.96 in Table 3, we formed the behavioral transition diagrams (see Figure 3.) The node represents one of the six behavioral categories, the numerical value represents the Z-value for the sequence, the arrowheads represent the transitional direction, and the connecting line thickness represents the level of significance.

Table 3: Adjusted residuals table (Z-scores) (Phase 1 to Phase 3)

Z-score		B1	B2	B3	B4	B5	B6
Phase 1	B1	<b>30.88*</b>	-9.60	1.44	-1.85	-1.70	-0.32
	B2	-9.73	<b>13.05*</b>	-1.61	-0.19	-11.75	-9.15
	B3	0.22	-1.70	<b>2.51*</b>	<b>2.08*</b>	-1.33	-1.79



	<b>B4</b>	-2.06	-16.96	-1.25	-0.90	<b>26.59*</b>	<b>19.98*</b>
	<b>B5</b>	-1.51	1.75	0.03	0.01	-1.32	-1.79
	<b>B6</b>	-0.70	0.74	0.02	-0.17	-1.78	0.80
<b>Phase 2</b>	<b>B1</b>	<b>14.43*</b>	-5.08	1.73	-1.92	-0.58	-1.46
	<b>B2</b>	-5.96	<b>10.77*</b>	-3.82	0.11	-7.52	-6.89
	<b>B3</b>	1.51	-4.50	<b>6.39*</b>	1.61	-1.88	-1.53
	<b>B4</b>	-1.67	-12.10	-1.04	0.33	<b>16.95*</b>	<b>15.93*</b>
	<b>B5</b>	-0.97	<b>2.00*</b>	-1.08	-0.11	-0.62	-1.48
	<b>B6</b>	0.25	1.27	-0.60	-1.42	-0.77	0.54
<b>Phase 3</b>	<b>B1</b>	<b>35.87*</b>	-10.47	<b>3.43*</b>	-3.44	-2.29	-1.28
	<b>B2</b>	-9.74	<b>13.25*</b>	<b>3.54*</b>	-1.52	-14.06	-12.65
	<b>B3</b>	-1.07	1.71	-1.50	<b>3.52*</b>	-3.61	-2.28
	<b>B4</b>	-3.19	-24.10	-2.41	<b>4.05*</b>	<b>30.38*</b>	<b>26.44*</b>
	<b>B5</b>	-2.29	<b>5.48*</b>	-3.37	-2.00	-0.90	-2.29
	<b>B6</b>	-0.80	<b>4.88*</b>	-2.29	-2.56	-2.30	-1.29

Figure 3 shows that there were remarkably different behavior sequences in different phases. In phase 1, the significant behavioral sequences included: B1→B1, B2→B2, B3→B3, B3→B4, B4→B5, and B4→B6. Meanwhile, phase 2 also had six significant behavioral sequences, just B5→B2 substituting B3→B4. In phase 3, the significant behavioral sequences included: B1→B1, B1→B3, B2→B2, B2→B3, B3→B4, B4→B4, B4→B5, B4→B6, B5→B2 and B6→B2. In addition, B1→B3, B2→B3, B6→B2 were three new emerging behavioral paths. These sequences demonstrated the whole behavioral patterns in online collaborative reading activity.

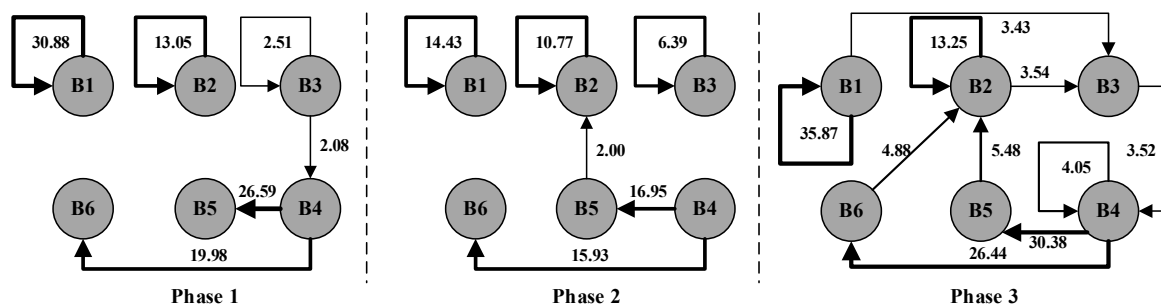


Figure 3. Behavioral transition diagrams in different phases.

First, let us turn to those uniform behavioral sequences in the three phases. The behavioral path B1→B1 in the three phase indicates that students tended to preserve their behavioral transition when they shared or compared information. This is because students usually added learning activities or uploaded resources after creating learning cells. And the Z-score of behavioral path B1→B1 in three phase seems to be positively correlated to the number of micro-courses created by students in each phase. The behavioral path B2→B2 in the three phases indicates that students tended to maintain their behavioral transition when they discovered and explored the dissonance or inconsistency. In order to understand peers' micro-course, students needed to watch the micro-videos again and again, and participated in learning activities. The Z-score of behavioral path B4→B5 in each phase is relatively large, which demonstrates B4 and B5 have significant correlations with each other in the collaborative reading process. But this result is inconsistent with the previous research findings (Hou, Chang, & Sung, 2007; Hou & Wu, 2011; Yang et. al, 2015), which held the view that B4 and B5 rarely occurred in the overall cooperation process. During this online collaborative reading activity, however, students were prompted to revise their micro-courses according to peers' comments before reaching an agreed upon understanding of the core idea of the article related to the micro-course. The behavioral path B4→B6 in each phase reveals that students discussed some irrelevant topics with the current collaborative reading task after pointing out the problem or rating. In addition, the Z-score of this behavioral path is very high in each phase, which indicates that the teacher needed to give some guides to help students solve the problem rather than just let the students explore freely.

Next, we explain the disparate behavioral sequences in each phase as shown in Figure 2. In phase 1, students always maintained their behavioral transition when they collaboratively edited the learning cell, adjusted content or learning activity, and discussed with peer about article idea (B3→B3, Z-score=2.53). Meanwhile, the behavioral path B3→B4 (Z-score=2.08) suggests that students would often give a rating after they had completed

micro-course learning or proposed questions. In phase 2, the behavioral path B3→B3 shows that students maintained their collaborative editing of the learning cell, adjusting content or learning activity and discussing ideas in the article with peers. Those behaviors could facilitate the advancement of the micro-course, and that may explain why the quality of the micro-course in the first two phases was better than the last phase. In addition, students did not make reflections or state their point of view all the time, rather, they put forward new questions or expressed confusion during their agreement statements (B5→B2, Z-score=2.00). Being a coauthor of peers' micro-course means that the student would have the same authority as the micro-course creator, such as editing learning content without checking, and cooperatively designing the learning activity and learning assessment. In phase 3, students coauthored peers' micro-courses and added learning activity and uploaded the resources by themselves (B1→B3, Z-score=3.43). Moreover, students coauthored their peer's micro-courses and provided some solutions for problems when they learned in the micro-courses (B2→B3, Z-score=3.54). In addition, a helpful behavioral path B6→B2 indicates that students did not do irrelevant things repeatedly, but returned to learn in the micro-courses or declare their confusion. Moreover, students sustained their behavioral path B3→B4 appearing in phase 1, and B5→B2 appearing in phase 2.

Next, Ucinet 6 (Borgatti, Everett, & Freeman, 2002) was used to conduct social network analysis by analyzing the patterns of relationship among members in collaborative reading process. Figure 4 illustrates that the social network of collaborative reading activity is a connected graph. The node represents the student, the line represents the relationship between students, and the arrow direction represents the information flow. Cohesion means that a network of individuals contains many ties and yields a tighter structure, which is usually identified by density, reciprocity, and actor distance (Hu & Racherla, 2008). The density of this network is 0.92, which implies that it is high-density network. Students almost kept in touch with every other student. The hybrid reciprocity of the network is 0.83, which implies lots of reciprocal interactions generated among students. The average distance of the network is 1.01, which implies that each student could almost directly contact with other students. In short, the whole social relationship network was symmetric, and all the students maintained a relatively frequent contact with each other, except the visiting scholar who only designed her own micro-courses without learning from other students' micro-courses. The reason for it may be she did not hold any pressure to obtain the course credit. Hence, it required course teachers to take the consistency of participants into consideration before implementation.

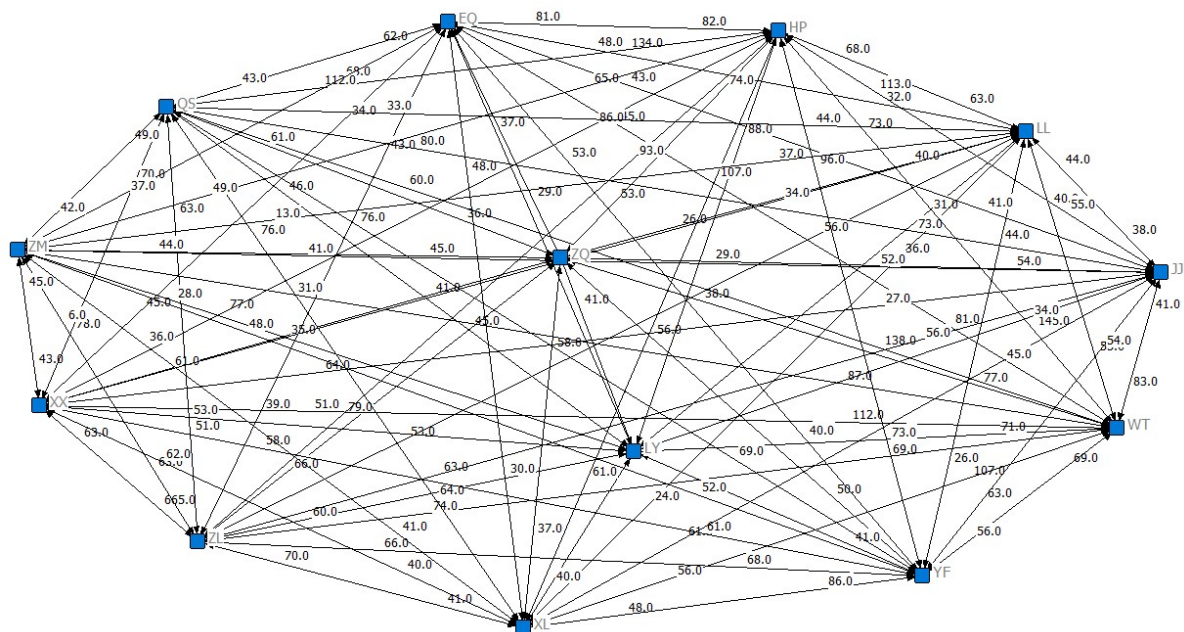


Figure 4. Social relationships network in online collaborative reading activity.

## Conclusions and suggestions

In this study, we coded the logs and comment message contents, and conducted a sequential analysis of behaviors and a social network analysis in an online collaborative reading activity. We found that 1) the behavioral sequences of students' knowledge construction presented different characteristics in three phases, though some

behavioral paths, such as B1→B1, B2→B2, B4→B5, and B4→B6, appeared all the time; 2) the behavioral path became more and more abundant with further deepened collaboration, such as the path B6→B2 emerging in the third phase, which might be caused by students' increasing interest and adeptness in this innovative collaborative reading approach; 3) students maintained relatively frequent contact with each other, which might be due to peer coaching instruction strategy. In addition, we also discovered that 1) contributions that students made to peers' micro-courses were not symmetric, such as someone contributing a lot to peers' micro-courses but receiving little contribution from peers on his or her own micro-courses; 2) students would submit some irrelevant comments in order to increase their course score. One reason for the irrelevant information may be that LCS could not make semantic analysis of students' comment content automatically at this moment which resulted in assessment according to the quantity rather quality. A possible solution to these problems is that the course teacher designs a better assessment scheme including artificial assessment and word segmentation. Moreover, contributions to peers' micro-courses and the quality of comments should be covered in artificial assessment.

In summary, this study explored interactive behavioral patterns and relationship patterns in an online collaborative reading activity through an innovative approach. This innovative approach is very different from previous studies: we used adult participants while previous studies used primary and secondary school students (Chen & Chen, 2014; Goh, Chai, & Tsai, 2013; Lin et. al, 2014); each of our participants used different reading material, rather than having participants use the same material (Chen et. al, 2010; Looi, Lin, & Liu, 2008); and we employed a new learning platform(LCS) for knowledge building, rather than using wiki (Chang, 2009; Kimmerle, Moskaliuk, & Cress, 2011) or knowledge forum (Hong, 2014; Hong, Chang, & Chai, 2014). Furthermore, our findings are helpful to further study collaborative reading among EFL students in higher education. For course teachers, they need to provide an effective incentive mechanism and assessment scheme, take the students' load of reading task and participants' consistency into considerations, and allocate the materials of the same theme to one person. Nevertheless, there exist some limitations to this study. Firstly, only thirteen students participated in this study which led to some analysis outcomes that are not statistically significant. Second, the study lasted a long time and generated many behaviors, with the result that some behaviors were inappropriately coded according to the categories defined in LCS. In the future, we will increase the number of participants and set up a control group to investigate the actual effect of this innovative collaborative reading approach on learning performance.

## References

- Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: an introduction to sequential analysis* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge, UK: Cambridge University Press.
- Borgatti, S.P., Everett, M.G., & Freeman, L.C. (2002). *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.
- Chang, C. K. (2009). Augmenting Wiki system for collaborative EFL reading by digital pen annotations. In *Ubiquitous Virtual Reality, 2009. ISUVR'09. International Symposium on* (pp. 43-46). IEEE.
- Chen, C. M., & Chen, F. Y. (2014). Enhancing digital reading performance with a collaborative reading annotation system. *Computers & Education, 77*, 67-81.
- Chen, J. M., Chen, M. C., & Sun, Y. S. (2010). A novel approach for enhancing student reading comprehension and assisting teacher assessment of literacy. *Computers & Education, 55*(3), 1367-1382.
- Choo, L. P., Kaur, G., Fook, C. Y., & Yong, T. C. (2014). Patterns of Interaction among ESL Students During Online Collaboration. *Procedia-Social and Behavioral Sciences, 123*, 307-314.
- Ding, N. (2009). Visualizing the sequential process of knowledge elaboration in computer-supported collaborative problem solving. *Computers & Education, 52*, 509-519
- Goh, A., Chai, C. S., & Tsai, C. C. (2013). Facilitating students' development of their views on nature of science: A knowledge building approach. *The Asia-Pacific Education Researcher, 22*(4), 521-530.
- Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of educational computing research, 17*(4), 397-431.
- Hong, H. Y. (2014). Exploring college students' perceptions of learning and online performance in a knowledge building environment. *The Asia-Pacific Education Researcher, 23*(3), 511-522.
- Hong, H. Y., Chang, Y. H., & Chai, C. S. (2014). Fostering a collaborative and creative climate in a college class through idea-centered knowledge-building. *Instructional Science, 42*(3), 389-407.
- Hou, H. T., Chang, K. E., & Sung, Y. T. (2007). An analysis of peer assessment online discussions within a course that uses project-based learning. *Interactive learning environments, 15*(3), 237-251.

- Hou, H. T., & Wu, S. Y. (2011). Analyzing the social knowledge construction behavioral patterns of an online synchronous collaborative discussion instructional activity using an instant messaging tool: A case study. *Computers & Education*, 57(2), 1459-1468.
- Hu, C., & Racherla, P. (2008). Visual representation of knowledge networks: A social network analysis of hospitality research domain. *International Journal of Hospitality Management*, 27(2), 302-312.
- Huang, H. M. (2002). Toward constructivism for adult learners in online learning environments. *British Journal of Educational Technology*, 33(1), 27-37.
- Kanuka, H., & Anderson, T. (2007). Online social interchange, discord, and knowledge construction. *International Journal of E-Learning & Distance Education*, 13(1), 57-74.
- Kimmerle, J., Moskaliuk, J., & Cress, U. (2011). Using Wikis for Learning and Knowledge Building: Results of an Experimental Study. *Educational Technology & Society*, 14(4), 138-148.
- Lin, C. P., Chen, W., Yang, S. J., Xie, W., & Lin, C. C. (2014). Exploring students' learning effectiveness and attitude in Group Scribbles-supported collaborative reading activities: a study in the primary classroom. *Journal of Computer Assisted Learning*, 30(1), 68-81.
- Looi, C. K., Lin, C. P., & Liu, K. P. (2008). Group Scribbles to support knowledge building in jigsaw method. *IEEE Transactions on Learning Technologies*, 1(3), 157-164.
- Ma, W. W. A. (2009). Computer-supported collaborative learning and higher order thinking skills: A case study of textile studies. *Interdisciplinary Journal of E-learning and Learning Objects*, 5, 145-167.
- Goh, A., Chai, C. S., & Tsai, C. C. (2013). Facilitating students' development of their views on nature of science: A knowledge building approach. *The Asia-Pacific Education Researcher*, 22(4), 521-530.
- Robinson, C. C., & Hullinger, H. (2008). New benchmarks in higher education: Student engagement in online learning. *Journal of Education for Business*, 84(2), 101-109.
- Schellens, T., Van Keer, H., De Wever, B., & Valcke, M. (2008, June). Student elaborations and knowledge construction in asynchronous discussion groups in secondary education. In *Proceedings of the 8th international conference on International conference for the learning Sciences-Volume 3* (pp. 413-415). International Society of the Learning Sciences.
- Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). Exploring Students' Knowledge Construction Strategies in Computer-Supported Collaborative Learning Discussions Using Sequential Analysis. *Educational Technology & Society*, 17(4), 216-228.
- Wan, H. P., Yu, S. Q., Cui, J. J., & Chiang, F. K. (2015). An Innovative Model of English Academic Works Collaborative Reading on Chinese EFL Graduate Students. Ogata, H. et al. (Eds.), *Proceedings of the 23rd International Conference on Computers in Education* (pp.160-165). Hangzhou, China: Asia-Pacific Society for Computers in Education.
- Yang, X. M., Li, J. H., Guo, X. S., & Li, X. J. (2015). Group interactive network and behavioral patterns in online English-to-Chinese cooperative translation activity. *The Internet and Higher Education*, 25, 28-36.
- Yu, S. Q., Yang, X. M., Cheng, G., & Wang, M. J. (2015). From Learning Object to Learning Cell: A Resource Organization Model for Ubiquitous Learning. *Journal of Educational Technology & Society*, 18(2), 206-224.

## Acknowledgments

This work is supported by Advanced Innovation Center for Future Education, Beijing Normal University and China Scholarship Council.

# Contrasting Explicit and Implicit Scaffolding for Transactive Exchange in Team Oriented Project Based Learning

Xu Wang, Miaomiao Wen, and Carolyn Rosé  
xuwang@cs.cmu.edu, mwen@cs.cmu.edu, cprose@cs.cmu.edu  
Carnegie Mellon University

**Abstract:** Script based support for collaborative learning may employ either explicit or implicit scaffolding. If the effect of script based support is mainly by means of its effect on collaborative processes, it would be reasonable to expect that if two forms of support manipulate the same process variable, they would provide redundant rather than synergistic effects when offered together. However, explicit forms of scaffolding may provide additional benefits from the reification of the processes that they provide, which could be received as instruction. This paper contributes to research on dynamic support for collaborative learning by proposing a novel form of micro-level prompt designed to elevate the amount of Transactive exchange. In a  $2 \times 2$  factorial design, we measure the impact of this explicit form of scaffolding and an implicit form of support for Transactive exchange developed in prior work, alone and in combination in terms of impact on two key outcomes, namely collaborative product quality and acquisition of multi-perspective knowledge. In this study, the pair of manipulations provide synergistic support for group product quality, but only the explicit support contributes to acquisition of multi-perspective knowledge.

## Introduction

Work related to both static and dynamic script based support for collaborative learning has leveraged the theoretical construct of Transactivity (Teasley et al., 2008; Azmitia & Montgomery, 1993; De Lisi & Golbeck, 1999), often demonstrating a mediating or moderating effect on learning, especially conceptual learning of difficult content (Azmitia & Montgomery, 1993; Schellens et al., 2007) and acquisition of argumentative knowledge (Weinberger et al., 2010). In notable recent work, Wen (2016) provides evidence that Transactivity in collaborative discussion can be increased through careful assignment of students to teams based on observation of a past history of Transactive exchange earlier in the course. Building on Wen's prior work, in this paper we address the question of whether explicit scaffolding of collaborative processes provided during collaboration offers an additional beneficial impact on outcomes above that achieved through implicit support. The findings highlight the importance of explicitly reifying desired collaborative processes over elevating those processes through an implicit mechanism such as a team formation protocol.

More specifically, this paper presents a study of dynamic support for collaborative learning at scale as a contribution to the line of research aiming to import research from the CSCL community to the context of Massive Open Online Courses (MOOCs). Interest in collaborative learning in MOOCs has been present since the earliest MOOCs, especially cMOOCs (Yeager et al., 2013). However, the reality remains that participation in typical MOOCs is a solitary experience. The bulk of prior work related to analysis of collaborative or discussion based learning in MOOCs has largely focused on information exchange in discussion forums (Wang et al., 2015; Wang et al., 2016) or designed collaborative learning experiences that are very short activities, such as collaborative reflection activities (Rosé & Ferschke, 2016). While there has been a desire to incorporate more ambitious project-based learning in MOOCs (Ronaghi, 2014), limited success has been reported. Wen (2016) offers hope that this problem can be overcome, by presenting evidence from a combination of controlled experiments and MOOC deployments that demonstrate positive impact of a team formation approach that offers project teams in MOOCs a more productive starting condition. Despite a successful MOOC deployment in which all formed teams turned in a final project, inspection of group processes suggest that additional improvements could be achieved through support of collaborative processes after team formation.

This paper contributes new insights related to dynamic support of at scale collaborative learning, with the goal of providing an empirical foundation for a later high external validity MOOC deployment. The most similar prior work to our own combines high internal validity controlled experimentation in a crowdsourcing environment such as Amazon's Mechanical Turk (Coetzee et al., 2015; Wen et al., 2016) with subsequent high external validity deployments in at scale online learning environments such as online courses or team based MOOCs. In this paper, we present a controlled study that provides design recommendations for a subsequent MOOC deployment that will offer dynamic support of collaborative processes during teamwork in that context.

In the remainder of this paper we first offer an overview of related work leading into the specific hypotheses underlying our study as well as its proposed contributions. Next we offer the details of the design and preparation for our study. Next we offer an empirical analysis of the results of our study. Finally, we discuss what we can conclude from the results and how they motivate the design of a subsequent MOOC deployment.

### **Novel dynamic support prompts in relation to prior work**

Collaborative work can be structured either at the macro level or the micro level. A frequent method for structuring collaborative work at the macro level is the use of what is known as the Jigsaw paradigm (Aronson, 1978), to increase the interdependence between team members. In this paradigm, students are provided with specialized expertise, and the task makes each piece of specialized knowledge that defines the Jigsaw to be necessary in order to achieve a satisfactory collaborative product. The Jigsaw used in Wen et al. (2016) consisted of four bodies of relevant knowledge for a task of constructing an integrated energy plan; each of the four bodies of knowledge focused on the pros and cons associated with one form of energy. Another way to introduce complementarity and interdependence is to assign students to roles that define their intended contribution (Strijbos & Weinberger, 2010; Schellens et al., 2007), such as assignment of different task responsibilities. In our work, we use the same task and Jigsaw paradigm used in Wen's study.

Micro-level structuring focuses more on the collaborative processes themselves. Sometimes this involves reification of desirable forms of contribution to the discussion, such as with scaffolds in a message authoring buffer (Weinberger et al., 2005) or buttons on a structured graphical user interface (Baker & Lund, 1997). Both static and dynamic forms of script based support for collaboration have been evaluated in the prior CSCL literature. A specific form of dynamic support that has been successful at increasing the intensity of substantive exchange of and improvement of reasoning contributions was originally designed as an automated form of what is referred to as Accountable Talk Classroom Facilitation (Michaels et al., 2002). A series of earlier studies of conversational agents employing Accountable Talk Facilitation have been successful at elevating collaborative processes and learning (Adamson et al., 2014). Results from this prior work demonstrate that Accountable Talk based support for collaborative learning significantly increases conceptual learning.

Typically macro and micro level support are treated separately. However, in our work we employ Accountable Talk prompts as micro-level support, but we tailor them to the Jigsaw role of each student. In particular, when we direct a student to evaluate and respond to another student's contribution, we ask them to do so from the perspective of their assigned jigsaw role. Thus, in addition to supporting Transactive exchange, the goal is for the Accountable Talk prompts to intensify the interdependence between students by emphasizing their unique knowledge.

### **Theoretical foundation and hypotheses**

A key theoretical construct that underlies our work is that of Transactivity, where our operationalization of Transactivity is defined as the process of building on an idea expressed earlier in a conversation using a reasoning statement. Research has shown that such knowledge integration processes provide opportunities for cognitive conflict to be triggered within group interactions, which may eventually result in cognitive restructuring and learning (De Lisi & Golbeck, 1999). While the value of this general class of processes in the learning sciences has largely been argued from a cognitive perspective, these processes undoubtedly have a social component. From the cognitive perspective, Transactivity has been shown to positively correlate with students' increased learning, since transactive discussion provides opportunities for cognitive conflict to be triggered (Azmitia & Montgomery, 1993; De Lisi & Golbeck, 1999). It has also been shown to result in collaborative knowledge integration (Gweon, 2012), since optimal learning between students occurs when students both respect their own ideas and those of others' (De Lisi & Golbeck, 1999). From the social perspective, Transactivity demonstrates good social dynamics in a group (Teasley et al., 2008).

Wen (2016) showed that by using Transactivity in one context to index collaboration potential in another context, we are able to significantly improve collaborative product quality. This raises the question of whether we can also increase learning with the same implicit support used in that study. However, in the learning sciences, there has often been a tension observed between emphasizing performance and emphasizing learning. In project courses, for example, students tend to take up roles where they can use the knowledge they already have in order to achieve a high quality product, which undercuts the learning that could take place. Often, learning requires focus on skills that are just beyond a person's ability level. Thus, engagement that leads to learning may frequently appear less successful in terms of performance. We cannot assume that a manipulation that supports a high quality product will necessarily support higher learning. In the case of the manipulation of group composition used in Wen (2016), Transactivity played a central role, and as already mentioned much prior work associates Transactivity with learning (Teasley et al., 2008; Azmitia &

Montgomery, 1993;De Lisi & Golbeck, 1999). Thus, **(1) we hypothesize that the composition manipulation that implicitly supports Transactivity will increase learning.** Conversely, prior work has demonstrated that dynamic support that intensifies collaborative processes leads to increased learning. Thus, **(2) we also hypothesize that the dynamic support manipulation of accountable talk that uses an explicit means to increase knowledge integration processes would increase quality of a collaborative product where quality is related to knowledge integration.**

As discussed above, in this study we introduce a novel form of adaptive prompting behavior that aligns the prompts with the students' roles in a Jigsaw task, aiming to elicit more discussion and reasoning related to the unique information the student has. Because such adaptive prompts reinforce the perspectives held by students, **(3) we hypothesize that the dynamic support we provide that explicitly scaffolds collaborative processes will lead to increased acquisition of multi-perspective knowledge.**

Wen (2016) shows that group composition has a positive effect on group product through moderating Transactivity in collaboration. Because dynamic support using Accountable Talk prompts also aims at increasing Transactivity in collaboration, there's a question of whether the implicit group composition manipulation and the explicit dynamic support manipulation during the collaboration process would interact synergistically or whether they would prove to be redundant forms of support. Prior work suggests that the support for learning offered through scaffolding a structured reasoning process is synergistic rather than redundant with the support received through intensive human interaction when conceptual knowledge is the target of learning (Kumar et al., 2007). Furthermore, explicit scaffolding may be received differently than implicit scaffolding, and thus lead to a different learning effect. **(4) Thus, we hypothesize a synergistic effect on group product quality if Accountable Talk prompts designed to explicitly scaffold a structured reasoning process are combined with a manipulation designed to intensify collaboration in a more organic implicit way, such as the composition manipulation.**

## Method

In order to test hypotheses 1-4, we conduct a  $2 \times 2$  factorial design where we independently manipulate the presence of the *implicit* group composition manipulation and the *explicit* Accountable Talk manipulation. This enables us to test the impact of the implicit support manipulation and the explicit support manipulation on group product quality as well as testing whether the two manipulations have a synergistic or redundant effect. As a key part of the paradigm for conducting this experiment, we use the same knowledge integration task used in prior research related to the implicit support composition manipulation (Wen, 2016). In order to examine the effects of both factors on individual learning and team product quality, we designed two outcome measures. 1) Individual pre-test and post-test to measure individual learning gains, including both measures of conceptual knowledge and multi-perspective knowledge. 2) A group proposal to measure team knowledge integration quality. The overview of the theoretical model is shown in Fig.1. In the rest of section, we will describe the task we used, the logic of the prompting behaviors in our agent, measurement of learning, participant recruitment and a manipulation check.

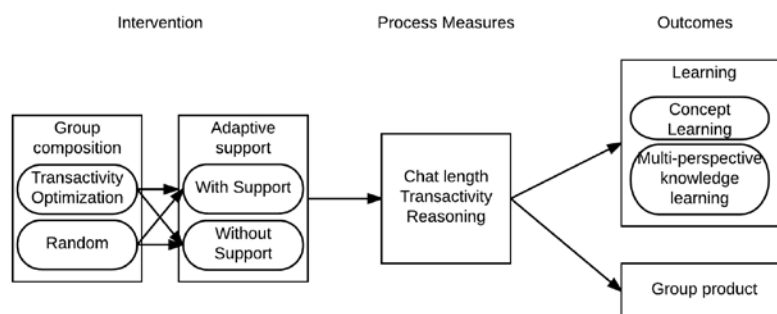


Figure 1. Overview of theoretical model for the study.

## Task description

In order for collaborative learning to be successful, the preconditions for interdependence and substantive exchange and integration should be established. We followed the Jigsaw paradigm and designed this Jigsaw knowledge integration task. Because each student represents a different perspective and receives unique information, it becomes critical for the students to transactively talk to each other and integrate their information to complete the task, which was found to be a successful knowledge integration task (Wen, 2016).

In the experiment, the student works through 6 steps, which take approximately 45 minutes. Step 1 is a pre-test where students are asked to write an individual proposal on an open-ended task involving proposing an energy plan for a city. In Step 2, students read an article. We designed the task to be a Jigsaw task, so that different students read materials focusing on their one assigned energy type, including coal, wind, nuclear or hydroelectric energy as types. In Step 3, students write one individual proposal based on what they have read from their own assigned perspective corresponding to the energy type, including “economical”, “environmental friendliness and low startup cost”, “carbon neutrality and economy in the long run”, “environmental friendliness and reliability”. The proposal will be posted to a discussion forum. In Step 4, students comment on each other’s proposals. For students in the Transactivity grouping condition that provides implicit support for group processes, they are then assigned to teams of four based on the group formation paradigm developed in Wen et al. (2016) to maximize the observed pairwise Transactivity across all teams, which means team members are assigned based on whom they’ve had the most Transactive exchanges with in the past, while also enforcing the Jigsaw paradigm. For students in the random grouping condition, they’re randomly assigned to teams of four based only on the Jigsaw constraint. In Step 5, students work in teams on a collaborative task using an interface where they write their group proposals on the left, and they will at the same time chat in the chat window on the right. The right window is where we provide the explicit scaffolded adaptive support to them as micro-scripting. Both conditions receive macro-level task structuring in the chat. We will in detail discuss the prompts we provide in the explicit support manipulation in the next section. In Step 6, students do a post-test, which is an isomorphic task to the pre-test with minimal rewording.

### Scaffolded adaptive scripting

In this section, we give a brief introduction to how the role-aligned adaptive scripting works. We inherited the conversation agent architecture Bazaar from Kumar (2011) and use it to introduce new conversational agent behaviors in this work. The system needs to understand two features from each contribution a student has typed in. The first is whether the student is in support of a plan. We designed this to be key-word based, we keep a dictionary of all possibilities we’ve found a student has used to refer to a certain plan, such as plan 1, plan A, option A, etc. We also keep a dictionary of all possibilities we’ve found students using to show being against a plan, such as, “totally against”, “don’t agree with”, etc. We use both rules to decide whether a student is in support of a plan. The second is whether the student shows reasoning. We labeled contributions from our pilot studies as reasoning or not and trained a machine learning model using a machine learning toolbench to assign this label to contributions in our study. At the same time, the system also keeps track of which student has talked least, and which plans have been brought up or fully discussed (i.e., discussed with reasoning).

When the student doesn’t show reasoning towards a certain plan, there are two possible prompting behaviors from the agent:

- 1) Ask the student to elaborate on the plan from his perspective, where there are three templates for this. One example is “Hey xx, can you elaborate on the reason you chose plan # from your perspective of most economical?”

- 2) Ask the student to compare the plan with another plan that hasn’t been fully discussed. For example, “Hey xx, you have proposed plan #, and xx has proposed plan #. What do you think are the pros and cons of the two plans from your perspective of #?”

When the person does show reasoning towards a certain plan, there are two possible prompting behaviors from the agent as well:

- 1) Ask someone else who has proposed a different plan that hasn’t been fully discussed to compare the two plans. “Hey xx, you have proposed plan #, and xx has proposed plan #. Can you compare the two plans from your perspective of #?”

- 2) Ask someone else to evaluate this plan, for example: “Hey xx, can you evaluate xx’s plan from your perspective of #?”

When there’s nobody talking for 3 minutes, the agent will pick the person who has talked the least and prompt: “Hey xx, which plan do you prefer from your perspective of xx?”

We went through an iterative design process to keep the amount of support at a reasonable level, and avoid over-scripting. We added additional rules, including 1) One person will not be prompted twice. 2) If a plan has already been fully discussed, it will not be prompted again. 3) The elaboration prompts will wait after having been triggered before being inserted into the conversation. In particular, the agent will only prompt if the student doesn’t talk in the next 10 seconds. After adding the constraints, there will be no more than 3 prompts in the 15-minute collaborative task. For most groups, there are 1 or 2 prompts.

In addition to the micro-level adaptive prompts that are only used in the explicit support condition, we also provide a starter prompt and a finishing-up prompt as macro-level support in all conditions.



## Measurement of learning and group product

To measure individual learning gains, we administered a pair of identical pre-test and post-test activities. In both cases, it is an open-ended question that asks the student to write an 80-120-word energy proposal for a city based on some basic information we provided about the city. We developed a coding manual to grade the open-ended proposal. Students can learn about energy from all steps in the task, including reading the article, and commenting in the discussion forum. There are two ways students can improve in their ability to articulate a plan from their experience in the collaborative task. In particular, they may learn new pros and cons about an individual energy source, which we consider conceptual knowledge. Or they may acquire new tradeoffs between energy sources, each related to a specific perspective, for example, coal is economical, but hydroelectric support is less so, and nuclear is even less so. We thus introduce two constructs to measure individual learning. 1) concept learning and 2) multi-perspective knowledge learning respectively.

Concept learning refers to learning of correct concept points, such as “coal is very cheap and economical” or “wind is a renewable energy source”. We counted incorrect knowledge points and removed these from the total. Multi-perspective knowledge learning refers to learning of comparisons between different energy types and tradeoffs of one energy type from different perspectives. For example, “nuclear is very reliable for hospitals, whereas wind is very unreliable.” or “Although wind is very environmentally friendly, it can be harmful for bird habitats.” We also took off incorrect comparison/tradeoff points from the total.

To measure group product quality, we graded the team proposals using the same rubric. The inter-rater reliability between two independent coders on a sample of the dataset is Kappa = 0.74. The two coders split up and then coded the remaining pre/post tests and group proposals without knowing which condition they came from. In addition to these outcome measures, we also introduced process measures to measure the quality of the collaboration process. More specifically, we assessed each team’s chat logs, using metrics including 1) number of contributions, 2) number of words, 3) number of reasoning contributions, 4) number of transactive contributions, 5) number of reasoning contributions that are aligned with each member’s assigned role. As mentioned in the scaffolded adaptive scripting section, we trained a machine learning model to automatically predict whether a contribution contains reasoning. We used the same model to compute the number of reasoning contributions in each team’s chat log. Among these labeled logs, we labeled contributions as transactive or not manually. In addition to general reasoning and transactivity, we also looked into how effective our intervention is in eliciting role-aligned reasoning. For example, a student represents coal energy, and he is supposed to argue from the perspective of which energy types are economical; if the student’s contribution either reasons about coal energy, or reasons about other energy types from an economical perspective in a contribution, it is considered as role-aligned reasoning.

## Participant assignment and manipulation check

We ran the study on Amazon Mechanical Turk from June to August in 2016. We ran the experiment in batches, with each batch associated with one or the other condition of the implicit group composition manipulation. Thus, each batch is assigned either to the Transactivity maximization condition or the control condition, so that all students within a batch are assigned to a team using the same process. In the rest of the paper, we refer to this manipulation as the Transactivity factor. Within each batch, teams are randomly assigned to an explicit support condition, i.e., either having adaptive scaffolding from Bazaar as micro-scripting in the experimental condition, or no explicit micro-level support in the control condition. In the rest of the paper, we refer to this manipulation as the Bazaar factor. We ran 14 batches in total. Because 3 batches did not end up including at least 16 students, we removed them from our dataset. This is to guarantee that at least two teams are generated in each four of the conditions in each batch. Among the 11 batches, we have 4 batches of random grouping and 7 batches of transactivity grouping. They generated 63 teams total, the distribution among conditions is displayed in Table 1.

Table 1. Number of teams in each condition

	Group composition: Random	Group composition: Transactivity
Adaptive support: Without Bazaar	14	18
Adaptive support: With Bazaar	13	18

We first did a manipulation check to make sure our grouping assignment manipulation successfully assigned students to groups such that they had a higher history of prior transactive exchange than expected by chance based on the distribution of transactive exchanges in the whole batch. We used one-way ANOVA to compare the average transactivity score during deliberation discussion within groups between those in control condition and those in Transactivity maximization condition and found a significant difference ( $F(1, 61) = 8.19$ ,  $p = 0.006$ ), with random assigned groups having a lower transactivity score ( $M = 8.56$ ,  $SD = 5.87$ ) compared to

Transactivity maximization groups ( $M = 13.28$ ,  $SD = 6.91$ ). The effect size value computed by Cohen's  $D$  is  $0.73$ , suggesting a moderate to high practical significance. We then checked to make sure the random assignment of students across the four conditions was successful in terms of ability to contribute to the task. As a proxy, we evaluated the length of the individual proposal each student wrote prior to the deliberation phase and did not find any systematic difference between conditions. ( $F(3, 59) = 1.25$ ,  $p = 0.3$ )

## Results

The results in correspondence to the hypotheses are summarized in Table 2. As a summary of our results, our finding is that, consistent with prior results of accountable talk prompts on conceptual learning, the *explicit* support intervention increases acquisition of multi-perspective knowledge (as measured by trade-offs). It does not impact collaborative product quality. Conversely, consistent with earlier studies of the *implicit* group composition manipulation, we observe here an impact on collaborative product quality, but not learning. However, an observed interaction effect reveals that the biggest impact on the collaborative product is achieved when both explicit and implicit interventions are combined.

Table 2. Hypotheses testing and results

Hypothesis	Support or not?
(1) The composition manipulation will increase learning, and a key process variable will be transactivity	Not supported
(2) The dynamic support manipulation of accountable talk will increase collaborative product quality.	Partly supported
(3) The dynamic support we provide that scaffolds collaborative processes aligned with students' perspectives will lead to increased acquisition of multi-perspective knowledge	Supported
(4) There is a synergistic effect on group product if accountable talk designed to scaffold a structured reasoning process is combined with a manipulation designed to intensify collaboration in a more organic way, such as the composition manipulation.	Supported

In response to Hypothesis (3), we built an ANCOVA model, using comparison/tradeoff points in post-test as the dependent variable, Transactivity grouping and Bazaar as main effects, and comparison/tradeoff points in the pre-test as a covariate, while also including group ID as a random intercept, to control for the fact that students in the same group may be correlated. We found Bazaar has a significant effect on the learning of multi-perspective knowledge, as represented by comparison/tradeoff points ( $F(1, 230) = 5.135$ ,  $p = 0.024$ ), which is consistent with our Hypothesis (3). The effect size value computed by Cohen's  $d$  is  $0.30$ , suggesting a moderate practical significance. In response to Hypotheses (1), we didn't find an effect of Transactivity grouping on either concept learning or multi-perspective knowledge learning. Thus hypothesis (1) is not supported. And there was not an interaction effect between Transactivity and Bazaar on individual learning.

In response to Hypotheses (2) and (4), we evaluated the two factors on the score of group proposals. We first built an ANCOVA model using group proposal score as the dependent variable, Transactivity grouping, Bazaar, and the interaction term as independent variables, as well as the total length of individual proposal of each group as a covariate. We found an interaction effect between the two factors ( $F(1, 62) = 5.240$ ,  $p = 0.026$ ). We then recoded the two main factors Transactivity and Bazaar into one variable indicating 4 conditions. As a planned contrast analysis we compared the group proposal scores across the four conditions. Based on Dunnett t-test, the average score of the condition where both explicit and implicit support were present is significantly higher than the other three conditions (with  $p = 0.015$ ,  $0.043$  and  $0.001$  respectively). The other three conditions were not different from one another. This partially supports our Hypothesis (2) that Bazaar is helpful for group product when transactivity is also present. And this confirms our Hypothesis (4) that there is a synergistic effect between the scaffolded adaptive support and the team's composition on group product quality.

In terms of impact on process variables, our finding is that both manipulations influence collaborative processes, but in different ways. In chat logs, Bazaar groups have a marginally higher percentage of reasoning compared to control condition groups. ( $F(1, 61) = 3.213$ ,  $p = 0.078$ ) In addition to general reasoning, we also looked at whether our intervention leads to more role-aligned reasoning, which is a direct result of the intervention. We found for Bazaar groups, students displayed marginally significantly more reasoning behavior consistent with students' assigned Jigsaw roles. ( $F(1, 61) = 0.09$ ,  $p = 0.098$ ) The effect is more salient for students in the random grouping condition. ( $F(1, 25) = 4.49$ ,  $p = 0.044$ ) We don't see a difference on other process measures between the experimental and control conditions. We see that Bazaar increases the concentration in chat, and also increases students' explanation aligned with their perspectives and roles. We also tested whether these process measures had a mediating or moderating effect on outcome measures. Among the

process measures, we only see the percentage of role-aligned reasoning has a moderating effect on both group product and multi-perspective knowledge learning.

## Discussion

From the above analysis, we found that the explicit adaptive support we provide in the chat is helpful for students' multi-perspective knowledge learning. But the group composition manipulation, which offers implicit support, increases transactivity in the chat, but doesn't show an effect on individual learning. On the other hand, the group composition manipulation has a significant positive effect on group product quality. Thus, in connection with group product quality, we see both manipulations having a synergistic effect. Based on these results, the recommended intervention would depend upon whether acquisition of multi-perspective knowledge or collaborative product quality is the primary target. If multi-perspective knowledge learning is the target, explicit support such as scaffolded accountable talk prompts should be emphasized, which reifies the value of including an integration of perspectives in the discussion. Both manipulations should be used together as they have been observed to work well in tandem for achieving impact on the collaborative product quality.

From the analysis of process measures, we found that the explicit support for transactive exchange encourages students to focus their most sophisticated articulation of domain reasoning in the chat rather than in the collaborative product. The responses to the Bazaar prompts increased the effort required on average per contribution (in particular when students were responding directly to the prompts), which dampened the tendency of the group composition manipulation to increase amount of discussion and integration in the collaborative product. Thus, while the Bazaar prompts improved discussion processes and learning, we do not see this value reflected in the collaborative product quality. On the other hand, it is important to consider that in this task, where effort cannot be simultaneously expended towards the discussion process and the product producing process, a manipulation that intensifies the discussion process may draw attention away from the product producing process. This is true when collaborative discussion and work on collaborative products occurs simultaneously. That is a difference between our setup and many phased collaboration setups in earlier studies. Nevertheless, it reflects the reality of many collaborative setups both in learning contexts and in the workplace. It is possible that if we required a more strict phasing structure, so that discussion occurred strictly before the collaborative work, we could employ both manipulations without any interference or dampening on the learning effect. We leave this for future studies.

## Conclusion

In this paper we reported on an experiment to contrast the effect of implicit support through manipulation of group composition and explicit support through providing scaffolded accountable talk prompts during collaboration. This high internal validity investigation motivates subsequent work where we will implement both interventions in future team-based MOOCs, as done with the same study paradigm in earlier work (Wen, 2016). In addition to practical implication for team-based learning in MOOCs, the theoretical contributions of this study are five fold: First, we investigated a novel form of accountable talk prompt that focuses on transactivity from a specific knowledge-based specialization, which was found to be helpful for multi-perspective learning and shows promise to be provided in future project-based courses to reinforce the roles and perspectives of team members. Second, we investigated explicit dynamic collaboration support in the form of accountable talk prompts in connection with a new form of learning (i.e., including multi-perspective knowledge as a learning measure rather than conceptual knowledge alone, which was the target in earlier studies of Accountable Talk prompting). Third, we investigated the generality of the impact of dynamic support in the form of accountable talk prompts to collaborative product quality and found that although the dynamic support is helpful for learning, it wasn't observed to improve collaborative product quality when provided alone. Fourth, we investigated whether the effect of the group composition manipulation demonstrated to be effective for improving collaborative product generalizes to learning. We found although group composition was effective in encouraging team members to chat transactively, it wasn't helpful for either of our learning measures. Finally, we investigated the extent to which the respective effects of the group composition and explicit micro-level scaffolding manipulations, which are assumed to be similar in that they are both grounded in the concept of transactivity, are synergistic or redundant. Consistent with our hypothesis, we found both them to be synergistic.

## References

- Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, 24(1), 92-124.
- Elliot Aronson. (1978). *The jigsaw classroom*. Sage.

- Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social development*, 2(3), 202-221.
- Baker, M., & Lund, K. (1997). Promoting reflective interactions in a CSCL environment. *Journal of computer assisted learning*, 13(3), 175-193.
- Berkowitz, M. W., & Gibbs, J. C. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly* (1982-), 399-410.
- Coetzee, D., Lim, S., Fox, A., Hartmann, B., & Hearst, M. A. (2015). Structuring interactions for large-scale synchronous peer learning. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1139-1152). ACM.
- De Lisi, R., & Golbeck, S. L. (1999). Implications of Piagetian theory for peer learning.
- Gahgene Gweon. Assessment and support of the idea co-construction process that influences collaboration. PhD thesis, Carnegie Mellon University, 2012
- Kumar, R., & Rose, C. P. (2011). Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies*, 4(1), 21-34.
- Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., & Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. *Frontiers in artificial intelligence and applications*, 158, 383.
- Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick, L. (2002). *Accountable talk: classroom conversation that works*. Pittsburgh: University of Pittsburgh.
- Schellens, T., Van Keer, H., De Wever, B., & Valcke, M. (2007). Scripting by assigning roles: Does it improve knowledge construction in asynchronous discussion groups?. *International Journal of Computer-Supported Collaborative Learning*, 2(2-3), 225-246.
- Strijbos, J. W., & Weinberger, A. (2010). Emerging and scripted roles in computer-supported collaborative learning. *Computers in Human Behavior*, 26(4), 491-494.
- Ronaghi, F., Saberi, A., & Trumbore, A. (2014). *NovoEd, a Social Learning Environment. Massive Open Online Courses: The MOOC Revolution*, 96.
- Rosé, C. P., & Ferschke, O. (2016). Technology Support for Discussion Based Learning: From Computer Supported Collaborative Learning to the Future of Massive Open Online Courses. *International Journal of Artificial Intelligence in Education*, 26(2), 660-678.
- Teasley, S. D., Fischer, F., Weinberger, A., Stegmann, K., Dillenbourg, P., Kapur, M., & Chi, M. (2008). Cognitive convergence in collaborative learning. In *Proceedings of the 8th international conference on International conference for the learning sciences-Volume 3* (pp. 360-367).
- Weinberger, A., Fischer, F., & Mandl, H. (2002). Fostering computer supported collaborative learning with cooperation scripts and scaffolds. *Proceedings of the conference on computer support for collaborative learning* (pp. 573-574). Boulder, CO.
- Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science*, 33(1), 1-30.
- Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction*, 17(4), 416-426.
- Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human behavior*, 26(4), 506-515.
- Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating How Student's Cognitive Behavior in MOOC Discussion Forums Affect Learning Gains. In *proceedings of the 8<sup>th</sup> International Conference on Educational Data Mining*.
- Wang, X., Wen, M., & Rosé, C. P. (2016). Towards triggering higher-order thinking behaviors in MOOCs. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*
- Wen, M. (2016). *Investigating Virtual Teams in Massive Open Online Courses: Deliberation-based Virtual Team Formation, Discussion Mining and Support* (Doctoral dissertation, Carnegie Mellon University).
- Wen, M., Maki, K., Wang, X., Dow, S. P., Herbsleb, J., & Rose, C. (2016) Transactivity as a Predictor of Future Collaborative Knowledge Integration in Team-Based Learning in Online Courses. In *Proceedings of the 9<sup>th</sup> International Conference on Educational Data Mining*.
- Yeager, C., Hurley-Dasgupta, B., & Bliss, C. A. (2013). cMOOCs and Global Learning: An Authentic Alternative. *Journal of Asynchronous Learning Networks*, 17(2), 133-147.

## Acknowledgements

This work was funded in part by NSF grant ACI-1443068 and funding from Google.

# Making Engagement Visible: The Use of Mondrian Transcripts in a Museum

Ben Rydal Shapiro, Vanderbilt University, benrydal@gmail.com  
Rogers Hall, Vanderbilt University, r.hall@vanderbilt.edu

**Abstract:** This paper describes how we have used a new transcription system we call Mondrian Transcripts to study visitor engagement and expand professional vision (Goodwin, 1994) in a museum. Methods, concepts and findings from this paper contribute to research concerning interest driven learning (Azevedo, 2013; Ito et al., 2009, Crowley & Jacobs, 2002), how people “make places” for learning while moving through different types of physical or information environments (Taylor & Hall, 2013; Ma & Munter, 2014; Marin, 2013; Lave et al., 1984) and the design of learning environments that can advance professional design practice. Empirical data include 1) 22 case studies of complete museum visits that captured continuous, multi-perspective video and audio records of visitor mobility and interaction and 2) audio, video and survey based data from a professional development and design session with museum educators, exhibit designers, curators and archivists.

## Introduction and organization of paper

The setting and empirical basis of this research is a two year ethnographic study to understand how visitors cultivate interests in and learn about the diverse historical and cultural heritage of American Roots and Country music while visiting a nationally renowned museum located in the mid-South region of the United States. As part of this work, we collected a purposive sample of complete museum visits across 22 visitor group cases including 11 family groups (2-5 visitors per group) over a period of six weeks (twenty-four days of data collection). These 22 case studies captured continuous, multi-perspective video and audio records of visitor group mobility and interaction through small GoPro cameras worn by visitors with no researchers present. In addition, following their visit we conducted 1-2 hour post visit interviews with all visitor groups and we connected with visitors on various social media platforms in order to follow online the content that visitors collected and shared from their visit. Table 1 summarizes this work.

Table 1: Overview

### Overview of 22 Visitor Groups & Followed Social Media Posts

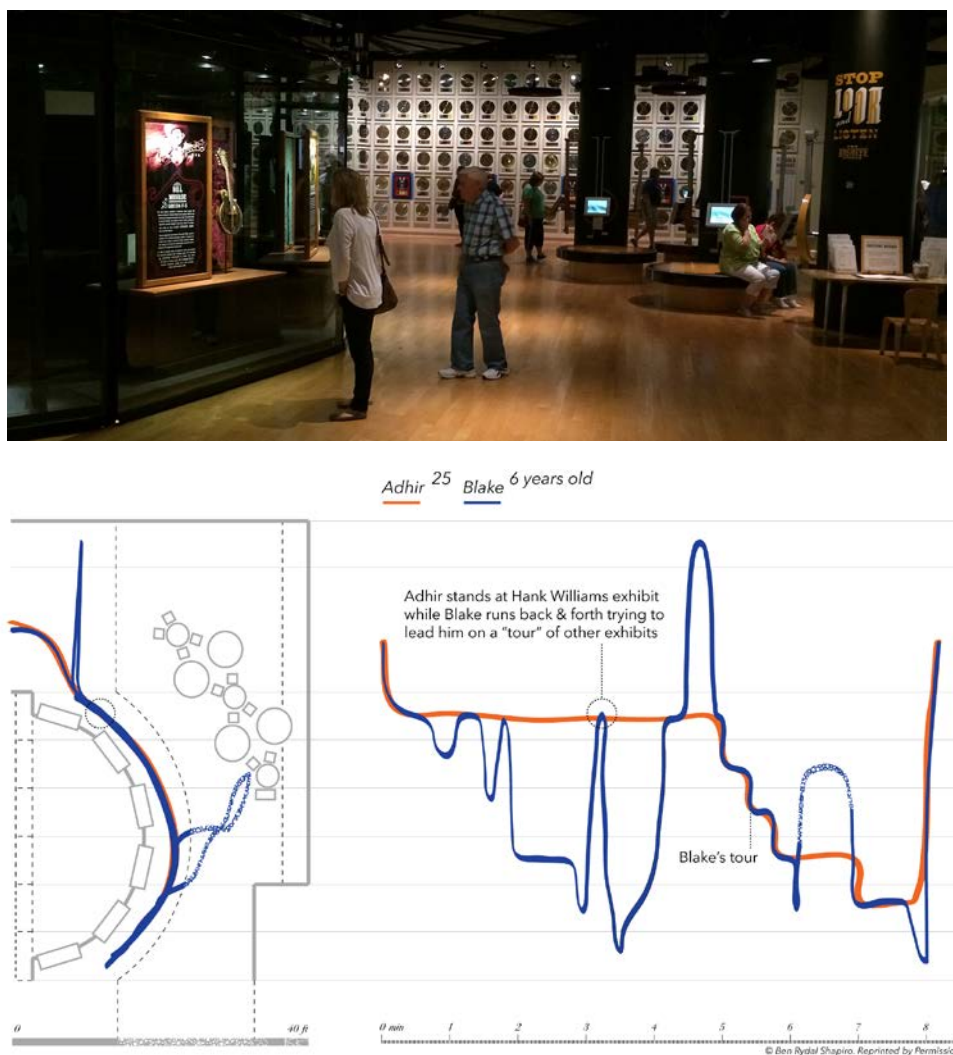
The table reads from left to right with each row corresponding to one of 22 groups of museum visitors. For example, Group 1 from Pittsburgh, PA completed their visit together in 1:40. Of the 3 people in the group, 2 shared single posts to the followed social media platforms of instagram and facebook. Together, these posts received 16 likes & comments.

Hometown	Visit Length	People	People w/ Post	Type of Post	Likes & Comments
<i>* family group</i>	<i>(hr: min)</i>	<i>* musician</i>	○ 40-60 yrs old ● 20-30 yrs old ● 10-20 yrs old	📷📷 single post/platform x <sup>x</sup> # of photographs 📷 album/collage { } multiple visitors	
1 Pittsburgh, PA	1:40	●●●	●●	{📷} {📷}	16
2 Fresno, CA	1:09	●●	●	{📷} {📷}	101
3 Staten Island, NY	47 1:04 1:08	●● ● ●			
4 South Korea	52 56	● ●	●	{📷} 10	29
5 TX *	3:43	●●●●	●	{📷}	7
6 Nova Scotia & MI	1:37	●●	●	{📷}{📷}{📷}{📷}	169
7 Cordele, GA	1:34	●●			
8 Holland, MI	1:09	●●●●	●	{📷} 25 {📷} 10	5
9 Staten Island, NY	44	●●	●●	{📷} {📷}	58
10 Iowa City, IA *	1:46	●●●			
11 St. Mary's, PA *	1:09	●●●●	●	{📷}	4
12 Owings, MD *	1:16	●●●●	●●●	{📷}{📷}{📷}{📷}{📷}{📷}{📷}{📷}{📷}{📷}	103
13 London, UK *	1:07	●●●●			
14 Atlanta, GA	1:39 2:12	● ●			
15 GA & England	1:05	●●●●			
16 Hazlet, NJ *	1:25	●●●●			
17 Washington, D.C.	58	●●	●●	{📷} {📷}	58
18 Big Sur, CA *	49	●●●●●	●	{📷} {📷}{📷}{📷}{📷}	55
19 Sunrise, FL *	1:56	●●●			
20 Milwaukee, WI *	1:08	●●●●	○	{📷} 12	5
21 Port Charlotte, FL *	1:04	●●●●	●○	{📷} 21 {📷} 12	98
22 Chicago, IL *	38 1:33	● ●●	○	{📷} 9	8
Range:	38 to 3:43	2 to 5	1 to 3	0 to 13 Online Postings	4 to 169

The first part of this paper introduces Mondrian Transcripts as a transcription system to: 1) visually transcribe museum visitor's physical movement and conversation over space and through time and 2) study how visitors engage in museum gallery spaces. This work draws from and extends Interaction Analysis (Jordan & Henderson, 1995) and Time Geography (Hagerstrand, 1970) and reflects our recent efforts to describe and use what we call "Interaction Geography" in studies of learning (Shapiro & Hall, 2017; Shapiro, 2017). The second part of this paper describes how we have used a dynamic visualization environment that allows for multi-modal analysis of Mondrian Transcripts, which we call the *Interaction Geography Slicer (IGS)* (Shapiro & Hall, 2017; Shapiro, 2017), to support a professional development and design session with museum professionals at this museum. This session aimed to demonstrate new ways to conceptualize and design for visitor engagement and learning. This work is part of a larger design study (Cobb, Confrey, diSessa, Lehrer & Schauble, 2003) that aims to advance museum professional's learning about how design practice can create opportunities for interest-driven learning in and beyond their gallery spaces (Azevedo, 2013; Ito et al., 2009, Crowley & Jacobs, 2002).

### Mondrian Transcripts and visitor engagement

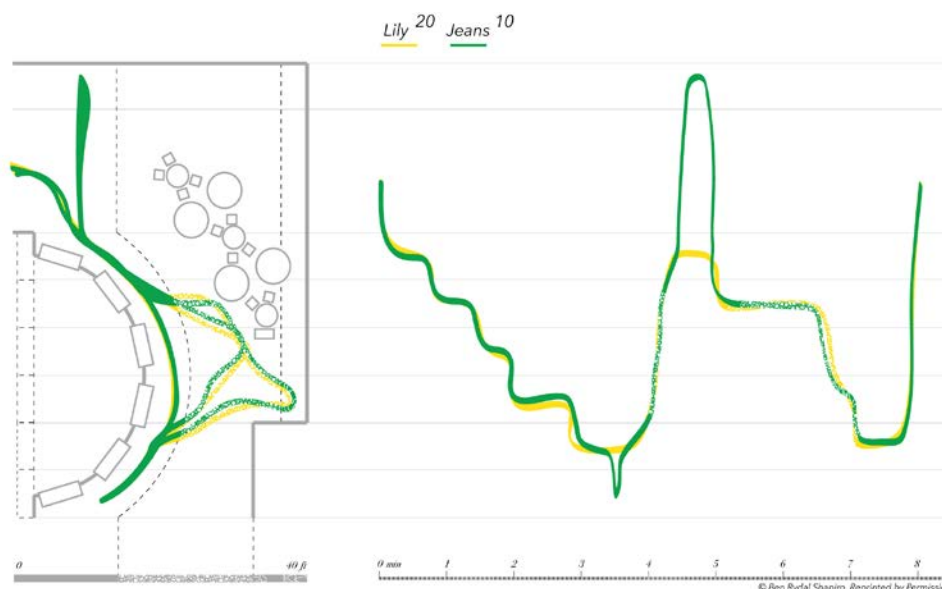
Figure 1 below displays a museum gallery space and maps the physical movement within that space of two (of five) members of a family we call the "Bluegrass Family" (not pictured in the image). On the left in "floor plan display", movement is shown over a floor plan of the gallery space (e.g. as if you were looking "down" onto the space). On the right in "timeline display", movement unfolds continuously over a timeline. On the timeline display, vertical position corresponds to vertical position on the floor plan. In addition, line pattern corresponds to horizontal position on the floor plan. Adhir (*orange*) is 25 years old and Blake (*blue*) is a 6-year old boy.



**Figure 1.** A museum gallery space & Adhir and Blake's mobility over space and space-time.  
 Source: Copyright © by Ben Rydal Shapiro. Reprinted by permission.

The plan display shows where Blake and Adhir go within the gallery space, while the timeline display shows how they interact with exhibits and one another in space and over time. For example, using the timeline display, one can see that during the first five minutes Blake is moving quickly (apparently running) back and forth across the gallery space in what appears to be multiple, frantic attempts to draw Adhir away from an exhibit dedicated to Hank Williams where he remains standing (straight orange path in space-time from roughly 0-5 minutes). After four failed attempts, Blake finally appears to succeed in leading Adhir on what we describe as a “tour” of other exhibits in the gallery, indicated by their intertwined paths from approximately minutes 5-6.

Figure 2 below maps the physical movement of two other members of the Bluegrass Family, Blake’s brother Jeans (green) and their sister Lily (yellow). The timeline display illustrates how Jeans and Lily move through the gallery space nearly always together (apart only during minutes 4-5).

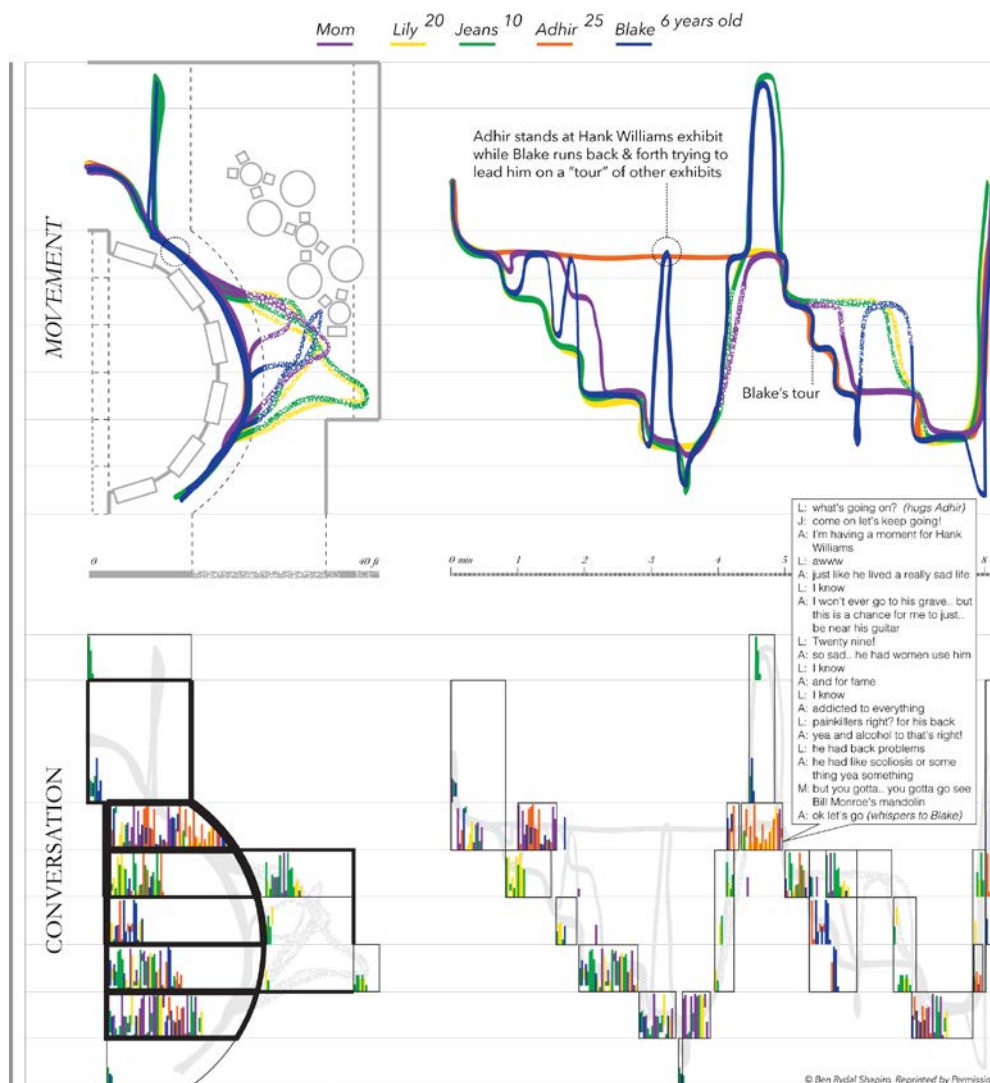


**Figure 2.** Jean’s & Lily’s mobility over space and space-time.  
*Source: Copyright © by Ben Rydal Shapiro. Reprinted by permission.*

Together, Figures 1 and 2 show how pairs within the Bluegrass Family move and engage with exhibits and one another in starkly different ways. While Blake produces a “recruitment” mobility pattern in response to Adhir’s persistent pattern of “reverence” (e.g., he later explains his attachment to the troubles in Hank’s life), Jeans and Lily produce an intertwined path that is similar to the “tour” mobility pattern later achieved between Blake and Adhir. On closer examination, all four young people and their paths are entangled in ways that allow: 1) Lily to soothe the emotions of Adhir (her fiancé), 2) Jeans to lead Blake away from the Hank Williams exhibit to give them privacy, and 3) Lily and Jeans (with help from Mom) to help Blake to succeed in “recruiting” Adhir to follow a tour. Blake’s dramatic (blue) path is produced in relation to other members of the family, who eventually help him take Adhir on a tour of other musicians.

Figures 1 and 2 also demonstrate how the visible qualities (e.g., pace, duration, shape, distance) and relationships (e.g., intersections, weaving, splitting, proximity) among movement paths in Mondrian Transcripts support and deepen different analytical framings of engagement. For example, they provide a means to study how people engage by managing personal and social distance between family members (Hall, 1966). However, Mondrian Transcripts also demonstrate that these distances are not only interactional phenomena as traditionally conceived but are also related to the spatial layout of the gallery space. For example, Blake and Adhir’s respective mobility patterns of recruitment and reverence are partly a response to the spatial location of the Hank Williams exhibit in relation to other artists in the gallery and how this sequence can be experienced as a path over time.

Figure 3 below extends the design of Mondrian Transcripts in a variety of ways. First, it maps the simultaneous physical movement of all five members of the Bluegrass Family (now including the mother in purple). Second, it similarly maps the Bluegrass Family’s conversation. To do this, talk is transcribed using standard conventions of conversation analysis, colored by speaker, and organized in relation to physical movement through space and over time. Conversation “boxes” group topically related talk. Thicker boxes on the floor plan show repeated conversations in the same area of space.



**Figure 3.** Mondrian Transcript of the Bluegrass Family's Interaction Geography.  
*Source: Copyright © by Ben Rydal Shapiro. Reprinted by permission.*

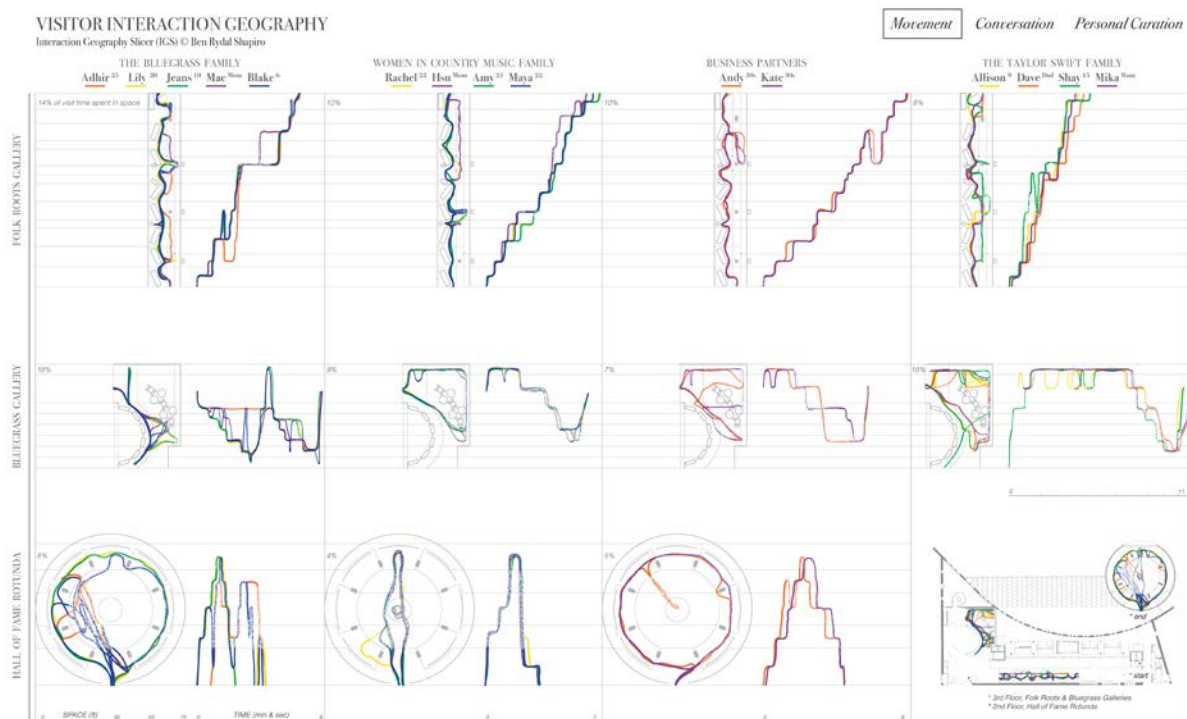
Figure 3 shows how the design of Mondrian Transcripts supports and deepens an understanding of engagement as both a response to the built environment and produced in interaction and mobility (Cleveland & Fisher, 2014). The Bluegrass Family's mobility shows how the family manages personal and social distance and likewise produces patterns of mobility that can be studied as a "meshwork" or as a form of "learning on the move" (Ingold, 2007; Taylor, 2013; Taylor & Hall, 2013). For example, the mother's mobility (named Mae) appears to "lag" behind other members of the family in a manner that could indicate she is overseeing and managing her family. Lagging patterns in space-time are common and in this case, indicate the need for closer analysis of Mae's mobility. In this example, closer analysis reveals how Mae often joins her family to make connections across exhibits for her children, thus helping to manage their engagement and learning. Additionally, the family's mobility and conversation illustrates how this family is intimately connected to one another as well as a particular semi-circular set of exhibits within the gallery space. Put another way, the family's mobility and conversation patterns show how the family selectively creates a "personally edited" (Lave et al., 1984; Ma & Munter, 2014) version of the gallery space that extends or elaborates the meaning of exhibits in ways relevant to the personal and social history of the Bluegrass Family.

Figure 3 also illustrates the embodied work of museum visitor groups that we call "engagement contours" (ECs). Engagement contours are comprised of topically bounded sequences of movement and talk that often repeat and accumulate over space and through time in relation to the physical environment and other



people. Engagement contours have a spatial and temporal footprint and are connected in ways that further the concept of “personal editing” as previously described. In Mondrian Transcripts, each visual box in the timeline display that bounds a set of intersecting movement paths and conversation indicates an engagement contour. Within engagement contours members of the Bluegrass Family arrange themselves in different types of interactional formations, similar to what (Kendon, 1990) calls “facing formations” and what (Marin, 2013) calls “ambulatory sequences”. One result of this, as shown in Figure 3, is that repeated engagement contours in the same area of space accumulate to produce dense conversation “boxes” or “textures” on the floor plan. Another result is that things like “Blake’s tour” become visible as forms of “place making” or “inhabitation” that are spatially and sequentially experienced in relation to the physical environment and other people. Thus, the concept of engagement contours furthers Marin’s innovative efforts to begin to put Kendon’s concept of a “facing formation” (e.g. how people spatially organize their bodies in interaction) into motion in studies of learning on the move (Taylor, 2013; Taylor & Hall, 2013). Moreover, it specifically theorizes engagement as both 1) a response to the built environment at different grain sizes (e.g. ECs can be studied at one exhibit or at the scale of a gallery space) and 2) produced through people’s interaction and mobility. Research rarely considers both of these aspects of engagement simultaneously primarily due to methodological limitations.

Figure 4 below is a set of “small multiples” from the *Interaction Geography Slicer (IGS)* and further illustrates the kinds of comparative analysis that can be explored using Mondrian Transcripts. The figure compares the movement of four different families/groups in three different types of gallery spaces within the museum. Columns distinguish different families while rows distinguish different museum gallery spaces. All displayed information is set to the same spatial and temporal scales. The Taylor Swift Family did not visit the Hall of Fame Rotunda Gallery thus we substituted an image of the entire museum and superimposed the movement of all families in each gallery space.



**Figure 4.** Small multiples of family mobility from Interaction Geography Slicer (IGS).  
*Source: Copyright © by Ben Rydal Shapiro. Reprinted by permission.*

Figure 4 illustrates broader use of Mondrian Transcripts in a variety of ways. First, it highlights similarities and differences in how engagement can be a response to the built environment. For example, for all groups the “Folk Roots Gallery” (a narrow, linear space) conditions particular linear ways of moving indicated by similarly sloped lines in space-time that rarely lead to repeated engagements. In contrast, the “Bluegrass Gallery” and “Hall of Fame Rotunda Gallery” (both open plan spaces with different supports for sequential engagement) promote a wide variety of movement patterns (in both space and space-time) and a great number of repeated engagements in some visitor groups.

Second, Figure 4 helps to further show how people and groups “personally edit” gallery spaces (Lave et al., 1984; Ma & Munter, 2014). For example, for those who know these gallery spaces (e.g., museum curators), it is immediately apparent that the Bluegrass Family primarily visits and spends time at exhibits that reflect content from Bluegrass and early Country Music, whereas the Women in Music Family engages with exhibits featuring female artists. Mondrian Transcripts provide a new means to conceptualize and visualize personal editing as processes of selecting and “making places” within the museum for group exchange and in ways that are driven by the personal and social history of individuals and groups.

Third, Figure 4 further illustrates how Mondrian Transcripts aim to meet provocations to develop a “graphic anthropology” to study meshworks of mobility (Ingold, 2007). In particular, Figure 4 shows how Mondrian Transcripts make processes by which visitor groups come together and split apart (e.g. meshworking) visible as a form of space-time mobility and how these can vary across visitor groups and gallery spaces. Mondrian Transcripts cannot tell us what goes on inside these meshworks, but they do draw our attention to moments and places of potential importance where, in this work, multi-party engagement with museum exhibits and their content rises and falls over space and through time.

Lastly, Figure 4 begins to demonstrate a developing space-time vocabulary of what we call “interaction geography” (Shapiro & Hall, 2017) that draws from both established “constraints” paradigms and emerging “new mobilities” paradigms in human geography (Hagerstrand, 1970; Sheller & Urry, 2006; Cresswell, 2010). For example, the Bluegrass and Women in Country Music Families illustrate significant variation in family “path braiding” (e.g., the pace and spatial density of engagement contours). In comparison, the Business Partner’s movement within the Hall of Fame Rotunda Gallery shows a lagged pattern of “path following”, in which Cindy follows Andy’s movement closely, but almost always trails about one minute behind. Likewise, across visitor groups, sharp “cuts” in space-time typically indicate young children who are running between family formations or trying to draw adults to other parts of the museum.

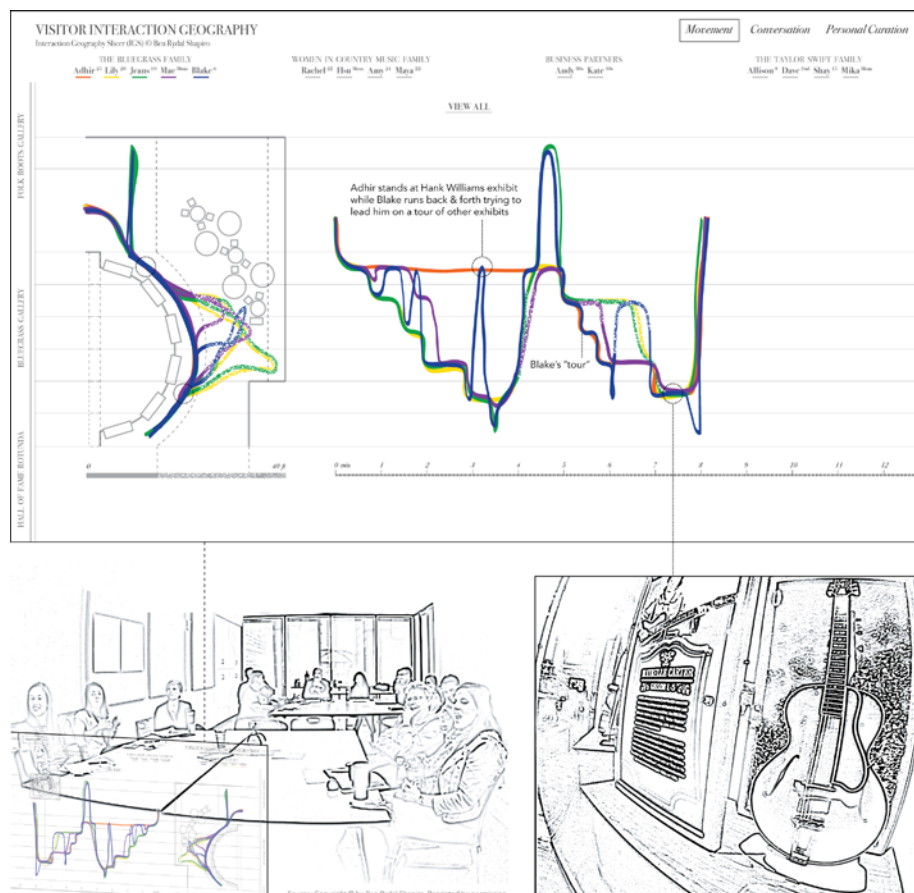
## Using Mondrian Transcripts to extend professional vision

We now shift our focus to describing how we used a dynamic visualization environment that allows for multi-modal analysis of Mondrian Transcripts, which we call the *Interaction Geography Slicer (IGS)* (Shapiro & Hall, 2017), during a professional development and design session with museum professionals at this museum.

Two *starting points* informed our design. First, visitor learning is not the primary focus of this museum’s design departments (e.g., they primarily design exhibits, marketing campaigns, and social media presence). For example, museum educators see their mission as “fitting” learning programs to museum content and exhibits after these have been designed/built. Second, as in any museum organization (Schauble et al., 1997), departments within the museum possess what we describe as an idealized view or model of their visitors. For example, there is a prevalent view across all departments that museum exhibits are a fixed curriculum that visitors succeed or fail to understand as opposed to a view of visitor engagement and interaction as an “enacted curriculum”, where learning is in the hands of visitors (Crowley & Jacobs, 2002; Louw & Crowley, 2013).

Our *design goals* were thus: 1) to provide methods and concepts to bring a more expansive view of learning to museum professionals and 2) to challenge with empirical cases the “idealized” view of gallery spaces and museum visitors described above. Our long-term goal, shared with our partners at the museum, is to identify opportunities for designing more equitable, expansive and productive learning in museum gallery spaces. Bringing learning sciences and museum design together is a promising but challenging design space.

In a half day workshop with 15 museum professionals, we used Mondrian Transcripts within the Interaction Geography Slicer (IGS) to help create an environment for joint exploration and discussion of what 4 different visitor families/groups were doing in 3 different gallery spaces (e.g., the visitor groups shown in Figure 4). The IGS allowed museum professionals to use Mondrian Transcripts as a means to study in highly interactive ways visitor’s movement, conversation and what we call “personal curation” (e.g. people’s use of personal information devices to capture, edit and share exhibit content during their visit). Moreover, the IGS also synced transcript and audio and video data from multiple-perspectives (e.g. from cameras worn by each visitor within a group) to each visitor’s movement, conversation and personal curation as visualized in Mondrian Transcripts. Thus, museum professionals could study a visitor group’s mobility, switch seamlessly to study their conversation and personal curation, isolate particular members of that visitor group, isolate particular regions of space and sequences of time during their visit, read transcript (e.g. what people were saying) and listen/watch audio and video from the perspective of each family member. Figure 5 below provides a snapshot from the session of museum professionals using the IGS to study the Bluegrass Family as previously described. The bottom right image in the figure displays video (from a camera worn by Lily) selected by a museum curator at a point in space and time when the family is gathered together at an exhibit dedicated to Maybelle Carter.



**Figure 5.** Museum Professionals using the Interaction Geography Slicer (IGS) to study the Bluegrass Family.  
*Source: Copyright © by Ben Rydal Shapiro. Reprinted by permission.*

One of the more powerful sequences during the workshop was one that triggered a dramatic shift in professional vision (Goodwin, 1994) among lead designers and educators regarding the movement and experiences of Blake from the Bluegrass Family. When museum professionals first saw Blake’s highly mobile paths, as previously described, few believed that he could possibly be learning. Some expressed concern that his seemingly erratic mobility might even be undermining the intended design of exhibits by distracting other members of his family during their museum visit. However, the workshop provided numerous opportunities for museum professionals to unpack Blake’s (and other young children’s) mobility patterns as a form of learning on the move. For example, in addition to confronting Blake’s “tour”, museum professionals studied how in one gallery space Blake, after failing to get an adequate answer to a question he asked from Adhir, ran to another gallery space across the museum to find and get an adequate answer from his brother Jeans only to run back across the museum once again to deliver his “found” answer to Adhir in the original gallery space. By the end of the workshop, museum professionals were studying and asking questions about Blake’s (and other young children’s) mobility that demonstrated an understanding that young children’s seemingly erratic patterns of mobility could be quite intentional efforts to engage and learn and were also opportunities for learning design. There were even jokes about hiring Blake as a museum ambassador for bluegrass music.

While there is not space to present further empirical material in this paper, we believe that exploring Mondrian Transcripts with the Interaction Geography Slicer (IGS) allowed museum professionals to see their visitors in new ways and to challenge idealized models of museum visitors as relatively passive consumers of intended design in ways that were previously impossible. Likewise, this work supported conversations between museum professionals that rarely occur. As one museum educator described in the post-survey, “I recall the productive cross-department conversation about visitor behavior, engagement, learning. We seldom (never?) have the opportunity to discuss visitor experience in the gallery—with our content—across departments. I also enjoyed and benefited from the visitor conversations in relation to specific space and artifacts—good to “see” the exhibit through their eyes and mind rather than assume their view, takeaways, paths, etc.”.

## Conclusion and next steps

Current and future work continues to develop and address inherent limitations in this early work in order to a) advance Mondrian Transcripts (theoretically and computationally) for use by others working in a variety of settings and at different scales (e.g., schools, neighborhoods, cities) and b) use Mondrian Transcripts to support professional development and new design practices. While illustrated with data from a museum, these methods and concepts are quite general purpose and may provide new possibilities for research and learning designs that consider space, learning and mobility.

## References

- Azevedo, Flávio S. (2013). The Tailored Practice of Hobbies and Its Implication for the Design of Interest Driven Learning Environments. *The Journal of the Learning Sciences*, 22(3), 462-510.
- Cleveland, B., & Fisher, K. (2014). The evaluation of physical learning environments: A critical review of the literature. *Learning Environments Research*, 17(1), 1-28.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9-13.
- Cresswell, T. (2010). Towards a politics of mobility. *Environment and planning D: society and space*, 28(1), 17-31.
- Crowley, K. & Jacobs, M. (2002). Building islands of expertise in everyday family activity. In G. Leinhardt, K. Crowley, & K. Knutson (Eds.) *Learning conversations in museums* (pp. 333-356). Mahwah, NJ: Lawrence Erlbaum Associates.
- Goodwin, C. (1994). *Professional vision*. *American Anthropologist*, 96(3): 606-633.
- Hagerstrand, T. (1970). *What about people in regional science?* *Papers in Regional Science*, 24(1), 6-21.
- Hall, E. T. (1966) *The Hidden Dimension*. Garden City, N.Y., Doubleday.
- Ingold, Tim (2007). *Lines: A Brief History*. London, Routledge.
- Ito, M., et al. (2009). *Hanging Out, Messing Around, and Geeking Out: Kids Living and Learning with New Media*. Cambridge, MA: MIT Press.
- Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences*, 4(1), 39-103.
- Kendon, A. (1990). Spatial organization in social encounters: The F-formation system. *Conducting interaction: Patterns of behavior in focused encounters*, pp. 209-238.
- Lave, J., Murtaugh, M., & de la Rocha, O. (1984). The dialectics of arithmetic in grocery shopping. In B. Rogoff and J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 67-94). Cambridge, UK: Cambridge University Press.
- Louw, M., & Crowley, K. (2013). New ways of looking and learning in natural history museums: The use of gigapixel imaging to bring science and publics together. *Curator: The Museum Journal*, 56(1), 87-104.
- Ma, J. Y., & Munter, C. (2014). The Spatial Production of Learning Opportunities in Skateboard Parks. *Mind, Culture, and Activity*, 21(3), 238-258.
- Marin, Ananda M. (2013) *Learning to Attend and Observe: Parent-child Meaning Making in the Natural World*. Ph.D. Dissertation. Northwestern University.
- Schauble, L., Leinhardt, G., & Martin, L. (1997). A framework for organizing a cumulative research agenda in informal learning contexts. *Journal of Museum Education* 22: 3-7.
- Shapiro, B.R. (2017). Using Space Time Visualization in Learning Environment Design. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, Denver, CO, USA.
- Shapiro, B.R., & Hall, R. (2017). Interaction Geography in a Museum. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, Denver, CO, USA.
- Sheller, M., & Urry, J. (2006). The new mobilities paradigm. *Environment and planning A*, 38(2), 207-226.
- Taylor, K. H. (2013). Counter-mapping the neighborhood: A social design experiment for spatial justice. Dissertation, Vanderbilt University.
- Taylor, K. H., & Hall, R. (2013). Counter-mapping the neighborhood on bicycles: Mobilizing youth to reimagine the city. *Technology, Knowledge and Learning*, 18, 1-2: 65-93.

## Acknowledgments

This work is made possible by the National Science Foundation, wonderful and ongoing collaborations with our museum partners, and Vanderbilt University's Space, Learning & Mobility Lab.

# Examining the Flow of Ideas During Critique Activities in a Design Project

Elizabeth A. McBride, Jonathan M. Vitale, Lauren Applebaum, and Marcia C. Linn  
bethmcbride@berkeley.edu, jonvitale@berkeley.edu, lauren.applebaum@berkeley.edu, mclinn@berkeley.edu  
University of California, Berkeley

**Abstract:** Peer critique activities in design projects give students the opportunity to share ideas, receive feedback, and revise their work. Critique can increase student feelings of ownership of science ideas and help students to distinguish between different ideas they may have about how things work. In this paper, we examine how students use their own ideas and ideas from a partner group to revise and improve a physical solar oven they have built using guidance from an online curriculum. We find that students fall into two groups: distinguishing ideas and adding new ideas. Within distinguishing ideas, students can further be separated by whether or not they kept only their own ideas or also added the ideas from their partner group. We look at case studies to determine how these groups changed their ideas before, during, and after the critique activity.

**Keywords:** science, engineering, peer critique, technology, knowledge integration

## Introduction

Design projects allow students to use science concepts to solve meaningful problems in topics such as energy efficiency. In addition to improving understanding of disciplinary concepts, middle school design activities engage students in the NGSS practices of engineering design (NGSS Lead States, 2013). We study a design project on solar ovens supported by an online curriculum. The project includes visualizations and interactive simulations to help students develop meaningful plans. Students draw on their scientific ideas and interpret the data collected using their physical artifacts to design and refine their solar ovens. We study how collaborative critique of student designs contributes to effective science learning.

Critique is common as a way to improve engineering designs (Krajcik, Blumenfeld, Marx & Soloway, 1994). Critique activities can guide students to justify their designs and identify flaws in their plans (Chang & Linn, 2013). Peer critique often succeeds when students discuss ideas together and justify their claims. This can help students clarify their ideas and reveal weaknesses in their understanding (Blumenfeld, Kempler & Krajcik, 2006). We examine how middle school students used their own ideas, others' ideas, and new ideas during critique of solar ovens. We then develop categories that are common types of student interactions during the peer critique activity in our curriculum.

By situating thinking in a social context, peer critique may increase student motivation (Wentzel, 1997). If students feel responsible for the success of their group, they may be more likely to engage and offer ideas. On the other hand, peer critique activities must be structured to ensure that all participants feel comfortable giving and receiving criticism (Scardamalia & Bereiter, 1994; Sato, 2015). Research on peer critique has shown the value of having students evaluate written work by their peers as a way to help students examine their own writing with a more critical eye (Black, Harrison, Lee & Marshall, 2003).

In addition to improving their designs, peer critique can help students develop a coherent understanding of underlying scientific concepts. According to the knowledge integration framework (Linn & Eylon, 2011), learning is achieved by first eliciting student ideas, then giving students opportunities to add new ideas and distinguish between these ideas. Peer critique provides an opportunity for students to express their ideas and hear new ideas from their peers. Students bring different prior experiences with them to the project, and can offer unique ideas (Matuk, Linn & Eylon, 2015). In a successful critique activity students will then distinguish between these ideas according to agreed-upon criteria. Students may be called upon to distinguish between their own ideas when developing explanations, but distinguishing the ideas of a group based on agreed-upon criteria may require further discussion to develop criteria.

Since students often receive feedback only from their teachers or other authority figures they may benefit from peer feedback that is worded more like their own thinking (Cole, 1991; Linn & Songer, 1991). In an environment where students feel comfortable sharing ideas and providing reasonable criticism, in addition to receiving feedback, students can develop general criteria for evaluating designs, which they may then apply to later activities (Clark et al., 2012). By building an awareness of what makes a good design, students develop greater agency and a sense of ownership over their designs and ideas.

However, it is often the case that students have difficulty establishing and applying appropriate, mutual criteria for evaluating ideas. If students have many ideas to distinguish, developing criteria for distinguishing could be quite a challenging task. This could be due in part to the fact that students may not practice critique often in the classroom setting. Numerous studies document the challenges faced by students when they are new to peer critique activities (e.g., Tsivitanidou, Zacharia, & Hovardas, 2011; Gan & Hattie, 2014; van Zundert, Könings, Sluijsmans, & van Merriënboer, 2012). In cases where students have not agreed upon criteria based upon underlying scientific ideas, they may focus on superficial features of designs. Furthermore, without criteria grounded in scientific concepts, students may feel reluctant to provide criticism, as it could be misinterpreted as a personal offense.

In previous work (McBride, Vitale, Applebaum & Linn, 2016), we saw mainly positive critiques (“Add more tinfoil”) rather than negative (“use black paper instead of tinfoil”). This effect may be exacerbated if students are expected to engage in face-to-face, verbal critique. This could be because of the structure of the curriculum; students are given a new budget during the revision to add to their oven, so during the critique activity they may be trying to decide how to spend their new budgets. However, students may also find it easier from a social standpoint to give positive critiques rather than negative (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001).

In this study, students often seemed to provide feedback without scientific justification, providing a peer group with a new idea, but no rationale for that new idea. This could be because students were reluctant to provide explanations, they did not feel they had to provide an explanation, or if they did provide an explanation during conversation, they simply did not write it down. In practice, engineers are often called upon to provide rationales for design changes. For professional engineers, providing rationales and being reflective is an important practice in improving design skills (Adams, Turns, & Atman, 2003; Schön, 1983). Rationales may be centered in the artifact and may also provide a way to understand and talk about dependencies in a complex project (Gruber & Russell, 1996). We introduce students to dependencies and tradeoffs by having them use a budget during the curriculum.

Types of student interactions during the peer critique activity could be categorized into groups using the knowledge integration framework. Two possible groups are “idea distinguishers”, and “idea adders”. In this curriculum, we aim to help students integrate their design choices with science concepts, so we consider the “integration” component of the knowledge integration framework to be specifically science and design integration. Ideally, peer critique activities will support students across groups in improving their practices of scientific idea integration, while also giving students the opportunity to add and distinguish ideas.

In this study, we examine how students’ original ideas change and grow from their first ideas about revisions they will make to their final ideas. We will evaluate the extent to which groups maintain their original ideas, add new ideas, distinguish between ideas, and integrate ideas with scientific justifications. Based on our findings from this analysis, we can also inform future design of critique activities to encourage certain types of student activity, like providing justifications and rationale.

## Methods

### Participants and procedures

One 6<sup>th</sup> grade teacher and her 150 students participated in this study. Following individual pretest, the teacher assigned students to a total of 55 dyad or triad workgroups; students in this class often work together on group activities. Following curricular activities, students engaged in an individual posttest.

### Curricular materials

This study was implemented in a curriculum module entitled *Solar Ovens and Solar Radiation* (referred to as *Solar Ovens* in this paper). The goal of the unit is to familiarize students with the way energy transforms from solar radiation to heat through a hands-on project and interactive models, covering the modeling aspect of the Science and Engineering Practices of the NGSS, as well as the standards associated with energy, specifically standards related to the transformation of thermal energy (NGSS Lead States, 2013). Students engage with the curriculum online through WISE (Web-based Inquiry Science Environment), utilizing a variety of instructional and assessment tools (Linn & Eylon, 2011).

Students follow a design, build, test approach with two iterations. For added support for distinguishing between and reflecting upon ideas, we include explanation and critique activities between iterations. Prior to building, students engage in a series of design activities intended to make science concepts central to students’ design plans. In a budgeting activity students are prompted to choose and justify materials they plan to use. Figure 1 shows the material and budget list. For their initial design, students are allowed to spend \$20 on their

materials (excluding the box). In later design revisions students receive an additional \$13 budget. By limiting their access to materials, students are forced to consider the most important elements of their design. Following budgeting, students engage with an interactive virtual model to investigate various design options (e.g., materials), and familiarize themselves with underlying mechanisms (Wilensky, 1999). In addition to selecting materials and testing them in the virtual model, students are also prompted to draw pictures of their ovens and explain how energy transfer will occur in their oven. After building, students test the ovens using digital temperature probes that collect data and generate a graph in real-time. The physical ovens are tested under lamps with a common set of requirements so that results are comparable between trials and groups. After revising, building, and testing a second time, students in this class cooked marshmallows in their ovens. Table 1 displays the general layout and features of the *Solar Ovens* curriculum unit.

We specifically investigate the use of the critique activity that occurs between iterations of designing, building, and testing. In this activity, students were first asked to describe the changes they would make to their own ovens during the next iteration. Then, students were instructed to work with the group next to them to exchange ideas about the ovens. This activity required some facilitation from the teacher. Students were asked to give at least one idea to the group they were working with, and to take at least one idea from the other group. Students wrote these ideas in WISE during the activity. In the next activity, redesigning the oven, we asked students to describe the changes they would actually make to their ovens.

Table 1: *Solar Ovens* Curriculum Outline

Activity	Description & Items of Interest
Introduction to Solar Ovens	Elicit initial student ideas about energy transformation
Solar Radiation and the atmosphere	Energy comes as radiation from the sun; energy can be absorbed or reflected. Students use a simulation to investigate energy.
Solar Radiation and Greenhouse Gases (GHGs)	Describes how energy interacts with greenhouse gases. Students use a model to investigate how addition of GHGs impacts energy.
Model Activity	Students use an interactive model to investigate how radiation works in a solar oven
Design, Build, Test 1	Design oven under budgetary constraints using a draw tool, build, test under a heat lamp using a temperature probe to collect data
Reflect & Critique	Students think about changes they will make to their oven and engage in critique activity with other student group
Design, Build, Test 2	Students reflect on what was learned from the first iteration; use new budget constraints to repeat process
Reflect	Students describe how their solar ovens work using energy from the sun; make connections between solar ovens and the atmosphere






Materials & Costs		
1 sheet of construction paper (8.5 inch x 11 inch)		\$2
12 inch x 12 inch Sheet of tin foil		\$7
12 inch x 12 inch Sheet of plastic wrap		\$6
You can RENT: 12 inch x 12 inch sheet of Plexiglas (thick plastic)		\$10
3 feet of tape (Duct, masking, clear)		\$3

Figure 1. Materials and cost list for Solar Ovens curriculum

### Analysis materials

We examined four items for each group. These items come from the reflection stage, the critique activity, and the very first activity during the redesign process. These items are:

1. *Reflect*: Students describe the changes they wish to make during the redesign of their oven

2. *Take*: Students write the idea(s) they received from the group they worked with
3. *Give*: Students write the idea(s) they gave to the group they worked with
4. *Redesign*: Students describe the changes they wish to make during the redesign of their oven

## Analysis approach

Our analysis approach for these items included developing a list of ideas students mentioned in any of their four responses. This list of ideas was then grouped into several categories that encompassed the majority of student ideas. We used the shortened list of ideas to code students' responses for the presence of ideas in order to track where ideas occurred within these four items. The shortened list is made up of the ten student ideas in Table 2. It would be useful to also have student reasoning for their design decisions, but many students did not provide reasoning, though the question specifically asked for it. Many students may have provided reasoning during the conversation that took place during the peer critique activity, but that reasoning was not recorded.

Table 2: Student idea categories and counts during each activity

Student Ideas	Reflect	Take	Give	Redesign	Total
Add aluminum foil	22	12	8	25	67
Add tape	5	0	3	10	18
Add more paper (black or white)	12	9	13	17	51
Add plastic wrap	9	4	6	10	29
Add or adjust reflective flap	16	13	9	11	49
Add Plexiglas	6	4	1	5	16
Tighten the plastic wrap over the top of the oven	3	3	3	3	12
Patch holes anywhere in the oven	4	4	6	1	15
Adjust or add a flap to insert food or a temperature probe	3	5	7	1	16
Adjust box construction (size, shape, structure, etc.)	2	2	0	2	6

## Results

Overall, we found that students wrote more ideas about their own ovens than peers' ovens. Across the 55 groups, students wrote an average of 1.55 ideas in the *reflect* item and 1.60 ideas in the *redesign* item, while only writing an average of 1.06 ideas in each of the *give* and *take* items. The greater number of ideas written for *reflect* and *redesign* are not surprising since students are more familiar with their own designs and ovens than the designs of other groups.

We next break down the flow of ideas from each item to the next in Table 3. In this table, we look at the interaction between each possible pair of items.

Table 3: Flow of ideas between items

	Items	# Ideas Carried Over
<i>Groups use ideas from their peer group</i>	<i>Reflect to Take</i> (1 to 2)	24
	<i>Take to Give</i> (2 to 3)	4
	<i>Take to Redesign</i> (2 to 4)	17
<i>Groups use their own ideas</i>	<i>Reflect to Redesign</i> (1 to 4)	35
	<i>Reflect to Give</i> (1 to 3)	13
	<i>Give to Redesign</i> (3 to 4)	23

Taking a closer look at whether students took their own ideas (*reflect*) or the ideas they received from the other group (*take*) with them to the redesign stage, of the 53 groups, 9 groups (16%) kept ideas in the *redesign* item from both the *reflect* and *take* items, 14 groups (25%) only kept ideas from *reflect* item in their *redesign* response (no ideas carried over from *take*), 2 groups (4%) only took ideas from the *take* item in their *redesign* response (no ideas carried over from *reflect*), and 22 groups (40%) did not use ideas from their *reflect* response or the *take* item in their *redesign* response. There were 8 groups (15%) that did not mention any specific ideas in their *redesign* response (e.g. "Well, we will use materials that will effect the box the most, so we only use 10 dollars.").



This interaction between ideas students generated themselves and those generated by another group shows that some students seem to keep only their own ideas, while others seem to engage in idea generation during the critique activity. Generating new ideas during the critique activity may help these students to generate more new ideas for themselves later in the design process.

Examining some of the other interactions between items, we see that students often carried ideas over from those that they gave to another group (*give*) to their own redesign. This happened 24 times (22 groups carried over one idea, 1 group carried over two ideas). This signals that having students generate ideas for another group is a useful activity for helping students to add more ideas to their repertoire for their own solar oven. Students generally did not carry over ideas from the *take* item to the *give* item. There were 4 ideas carried over, but in this case 3 of the ideas were carried over by the same group, with only one other group using the same idea for both *give* and *take* items.

During the analysis of this data, we also noticed that students did not often give negative critiques to other groups. However, this may be because of the structure of the unit. During the critique activity, students are thinking about what they can now add or change about their oven, so these are the types of critiques they get and give. This may also reflect a difficulty students have in giving their peers negative feedback, possibly because they do not yet consider themselves experts on the topics covered in the unit.

Students seem to fall into one of two categories: idea adders or idea distinguishers. Idea distinguishers can then be further broken down into students take ideas from others or those who keep only their own ideas. Students fall into the group of idea adders if they did not use any of their own previous written ideas or the ideas given by their peer group during *Redesign*. Students fall into the group of idea distinguishers (*keep*) if they kept only their own written ideas at *Redesign*. These students may not have liked the idea given to them by their peer group, the given idea may have been incorrect, or the budget may have been prohibitive. In any of these cases, however, the students distinguish between ideas and choose to carry forward with their own ideas. Students in the group of idea distinguishers (*take*) use ideas given to them by their peer group in *Redesign*. They can do this in conjunction with keeping their own written ideas from *Revise*. We discuss each of these four categories further through case studies.

Table 4 shows the breakdown of how many groups fell into each category from our data. In the knowledge integration framework, students should also integrate their ideas together. In this context, integration is considered integrating a design idea with reasoning that comes from science concepts. Only 14 groups integrated their design ideas with science reasoning in *Redesign*, even though the question specifically asked for reasoning. These groups were also spread across our three categories.

Table 4: Number of groups in each category and groups who integrate design ideas with science concepts

Category	# Groups / (Total)	# With Integration
<i>Idea Distinguishers (Take)</i>	17 / (55)	6
<i>Idea Distinguishers (Keep)</i>	14 / (55)	4
<i>Idea Adders</i>	23 / (55)	4

## Case studies

To further illustrate the results presented above, we use three case studies. Each of these cases offers a different view of how students use ideas presented during the critique activity. We will compare where ideas present in the last activity, *Redesign*, initially emerge. We specifically examine each of the following scenarios: groups who kept their own ideas and took ideas from their peer group, groups who kept only their own ideas, and groups who did not use any of their original ideas or the ideas from their peer group. Important parts of the responses are underlined.

### Case 1: Idea Distinguisher (Take)

In this case, the pair of students both kept their original ideas (*reflect*) and took the ideas given to them by their peer group during the critique activity (*take*). This type of case happened in 17 out of 55 groups. Each item answered by the pair is shown below:

- Reflect: “We could improve our solar oven by making the tin foil flaps bigger.”
- Take: “One idea that we got was to put black paper all around the inside of the box.”
- Give: “One idea I gave the group is that they should put a tin foil flap.”

- Redesign: *“we will buy black paper and more tin foil. The black paper is to absorb the heat inside the oven. The tin foil is to make a larger flap to direct the rays from the sun to the oven.”*

While students certainly write their own ideas for revising their oven in the *reflect* item, the nature of the project is such that students can watch other students in the classroom test their ovens and gain new ideas simply from looking around the classroom. Many of the *reflect* ideas likely come from watching other students, in addition to a group’s own ideas. This group starts out with an idea about improving the reflector flaps on their oven. This group got the idea of putting black paper on the inside of their box during the critique activity. When the group was asked how they would redesign their oven, they said they would use both ideas to improve their oven. Since each group has a limited budget, this was an interesting case in which the group was able to add materials to their budget to fulfill an idea given to them by another group. Another common occurrence seen in the *give* item was groups giving their own ideas to the other group. This happened in this case as well, with the group suggesting to their peer group that they *“should put a tin foil flap”*. This group also exhibited integration, integrating their design idea in *Redesign* with science reasoning.

### Case 2: Idea Distinguisher (Keep)

In this case, the pair of students kept their original ideas (*reflect*), but did not take any new ideas (*take*) into the *redesign* activity. This type of case was fairly common, happening in 14 out of 55 groups. Each item answered by the pair is shown below:

- Reflect: *“We can add plexiglass and still have \$3 dollars left if we get the additional \$10 button.”*
- Take: *“One idea that they gave us is that they put black paper on every side, including the bottom of their flap.”*
- Give: *“They could use plastic wrap or plexiglass on top and ad a hole that can put the smore in it easier.”*
- Redesign: *“We will add plexiglass and tape to keep more heat in.”*

In this case, the group began the critique activity with the idea that they would revise their oven by adding Plexiglas (*reflect*). This also included a discussion of their budget. The Plexiglas cost the students \$10, their whole budget. However, it seems that the students did not utilize their entire budget during the first round of building. It was common in this classroom for the teacher to allow the students to carry over any additional budget to the second iteration of building. The group mentions that they have \$3 left in their budget. During the critique activity, the students are given the idea of using black construction paper on all surfaces of the oven; construction paper costs \$2/sheet, so it is within their remaining budget to add some construction paper. However, the group decides to use their new budget to add Plexiglas and tape, sticking with their original idea from the *reflect* item. Again, this group offers their own idea to their peer group in the *give* item: to use Plexiglas. While students are giving ideas to other groups that are relevant to improving the function of the oven, we would like to improve students’ critical thinking in order to provide more relevant critiques to other groups.

### Case 3: Idea Adders

In this case, the pair of students did not keep their original ideas (*reflect*) or take the ideas from their peer group in the critique activity (*take*). This type of case was most common in our data, occurring in 23 out of the 55 groups. Each item answered by the pair is shown below:

- Reflect: *“we could have had the alummin foil flap better postioned. Also Maybe the plastic wrap could have been tighter.”*
- Take: *“we got the idea to use plexiglas insted of plastic wrap.”*
- Give: *“was to use tape to cover the holes.”*
- Redesign: *“We will add more black paper and touch up on some things that looked bad.”*

This type of case was the most common in our data. In the *reflect* item, the group wrote about changing the position of their reflector flap and tightening their plastic wrap. Their peer group gave them the idea of using Plexiglas instead of plastic wrap as a cover on their oven (*take*). However, in the *redesign* item, the group wrote about something completely different, adding black paper. In addition, the group did not give this idea about black paper to their peer group during the critique activity (*give*). While it is difficult for us to say exactly where this idea came from, this example illustrates how the critique activity can help students think about new and different ideas. This group both had and was exposed to many ideas during the course of these four items, and

was able to then think of even more new ideas for their redesign. While the group may not have provided scientific justification for their ideas yet, this still fits with the underlying knowledge integration framework; students need to add ideas to their repertoire in order to later sort those ideas.

## Conclusions and implications

This work offers a view into how students are using critique activities during their work in hands-on science projects, and offers a way to categorize how students use the peer critique activity to add and distinguish between new ideas.

We provide support for peer critique activities in our curriculum through face-to-face interaction with peer groups. During this direct interaction, students were able to engage in further conversation and often had to provide scientific justifications for their critiques to their peer groups. This resulted in the vast majority of the critiques during this project being about scientific and design choices, rather than superficial choices (e.g., decorative features).

The results of our data analysis help to show the benefits of using critique activities during design projects. While some students will still utilize only their own, preexisting ideas, many other students add ideas during the critique activity. The students who add ideas may combine their ideas with ideas from other groups, come up with completely new ideas after the critique activity, or give up on their ideas in favor of the ideas from their peers. From a creativity perspective, as well as a knowledge integration perspective, it is beneficial for students to be exposed to many different types of ideas. Students may learn more about the scientific implications of their design choices by having to sort through multiple ideas for revising or creating their designs.

Understanding how students use ideas from the peer critique activity to help develop new ideas or criteria for distinguishing between ideas helps to inform how we can design curriculum to encourage better practices for students.

One shortcoming of this curriculum and study was the lack of scientific reasoning given by students in their explanations for their design choices. In future uses of the Solar Ovens curriculum, students will be prompted for their design choice and their scientific reasoning separately (instead of in the same question prompt). This data provided us with useful information showing that students do not often want to provide reasoning or what they may see as extraneous explanation, but in the future we would like to also be able to better understand students' reasoning and help them to develop their explanation and argumentation skills.

## References

- Adams, R. S., Turns, J., & Atman, C. J. (2003). Educating effective engineering designers: The role of reflective practice. *Design studies*, 24(3), 275-294.
- Barron, B. (2003). When smart groups fail. *The journal of the learning sciences*, 12(3), 307-359.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology*, 5(4), 323.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). The nature of value of formative assessment for learning. *Improving schools*, 6, 7-22.
- Blumenfeld, P. C., Kempler, T. M., & Krajcik, J. S. (2006). *Motivation and cognitive engagement in learning environments*.
- Clark, D. B., Sampson, V., Chang, H. Y., Zhang, H., Tate, E. D., & Schwendimann, B. (2012). Research on critique and argumentation from the technology enhanced learning in science center. In *Perspectives on Scientific Argumentation* (pp. 157-199). Springer Netherlands.
- Chang, H. Y., & Linn, M. C. (2013). Scaffolding learning from molecular visualizations. *Journal of Research in Science Teaching*, 50(7), 858-886.
- Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of educational research*, 64(1), 1-35.
- Cole, D. A. (1991). Change in self-perceived competence as a function of peer and teacher evaluation. *Developmental Psychology*, 27(4), 682.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences*, 13(1), 15-42.
- Gan, M. J., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, 42(6), 861-878.
- Gruber, T. R., & Russell, D. M. (1996). Generative design rationale: Beyond the record and replay paradigm. *Design rationale: concepts, techniques, and use*, 323-349.

- Krajcik, J. S., & Blumenfeld, P. (2006). Project based learning. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences*. New York: Cambridge University Press.
- Krajcik, J. S., Blumenfeld, P. C., Marx, R. W., & Soloway, E. (1994). A collaborative model for helping middle grade science teachers learn project-based instruction. *The elementary school journal*, 483-497.
- Linn, M. C., & Eylon, B. S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration*. Routledge.
- Linn, M. C., & Songer, N. B. (1991). Teaching thermodynamics to middle school students: What are appropriate cognitive demands?. *Journal of research in Science teaching*, 28(10), 885-918.
- Matuk, C. F., Linn, M. C., & Eylon, B. S. (2015). Technology to support teachers using evidence from student work to customize technology-enhanced inquiry units. *Instructional Science*, 43(2), 229-257.
- McBride, E.A., Vitale, J.M., Applebaum, L.R., Linn, M.C. (2016) Use of Interactive Computer Models to Promote Integration of Science Concepts Through the Engineering Design Process. In *Proceedings of the 12<sup>th</sup> International Conference of the Learning Sciences*.
- NGSS Lead States. (2013). Next Generation Science Standards: For States, By States.
- Sato, M. E. (2015). Designing Critique for Knowledge Integration (Doctoral Dissertation). Retrieved from ProQuest Database. (Accession No: 3733349)
- Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, 3(3), 265-283.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action* (Vol. 5126). Basic books.
- Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction*, 21(4), 506-519.
- Van Zundert, M. J., Könings, K. D., Sluijsmans, D. M. A., & Van Merriënboer, J. J. G. (2012). Teaching domain-specific skills before peer assessment skills is superior to teaching them simultaneously. *Educational Studies*, 38(5), 541-557.
- Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of educational psychology*, 89(3), 411.
- Williams, L., Wiebe, E., Yang, K., Ferzli, M., & Miller, C. (2002). In support of pair programming in the introductory computer science course. *Computer Science Education*, 12(3), 197-212.
- Wilensky, U. (1999). {NetLogo}.

# Secondary School Peer-to-Peer Knowledge Sharing Through Social Network Technologies

Christa S.C. Asterhan, Hebrew University of Jerusalem, asterhan@huji.ac.il  
Edith Bouton, Hebrew University of Jerusalem, Edith.Bouton@mail.huji.ac.il

**Abstract:** The promise of social network technology for learning purposes has been heavily debated, with proponents highlighting its transformative qualities and opponents its distracting potential. However, very little is known about the actual, everyday use of ubiquitous social network technologies for learning and study purposes in secondary schools. In the present work, we present findings from two survey studies on representative samples of Israeli, Hebrew-speaking teenagers ( $N = 206$  and  $N = 515$ ) which explored the scope, characteristics and reasons behind such activities. Findings show that such practices can be described best as *online knowledge sharing*, that is: the up- and downloading of knowledge and knowledge sources to social network peer groups. This teenage, school-related knowledge sharing is common and widespread, entails different types of knowledge, and is mainly motivated by prosocial motives, as well as expectations for future reciprocation. Sharing is predicted by individual differences, such as gender, collectivist values, mastery goal orientations and academic self-efficacy. Relations between competitive-individualist values and sharing are more complex, and are, among others, moderated by expectations for future benefits. Implications for educational practices and for collaborative learning theories are discussed.

## Introduction

The prominence of social network sites (SNSs) in everyday life has ignited musings and debates about the implications of these developments for formal learning and education. Skeptics of SNSs for learning purposes pitch online social networking and formal learning as two mutually exclusive activities, emphasizing that SNSs have been designed and are mainly used for pastime socializing (Kirschner, 2015). This pastime socializing comes at the expense of and distracts from academic pursuits (e.g., Hollis & Was, 2016; Junco 2012, Kirschner & Karpinski, 2010). Research shows that teenagers and young adults indeed use SNS technology for various social purposes (e.g., Back et al., 2010; Hew, 2011). However, it does not rule out the possibility that students use SNSs for formal learning purposes as well.

At the other end of the debate, proponents of SNSs for learning envision that social media technologies will have positive and even transformative effects on how students learn, collaborate, share and create knowledge. These envisioned changes are often described in terms of a move away from traditional, hierarchical teaching structures organized in formal institutions, toward self-organized communities of interest, in which peers collaborate and discuss content, tutor newcomers, and create new knowledge, without the interventions of adult, certified teachers or other formal authority figures (e.g., Bingham & Connor, 2015; Collins & Halverson, 2009; Greenhow & Robelia 2009; Harasim, 2000; Wegerif, 2013). Recent educational design research initiatives have shown that some aspects of that vision may be achieved, with the help of extensive technical and professional support and with specifically developed add-ons to existing SNSs (Greenhow et al, 2015; Tsovaltzi et al., 2014). However, even though this shows the *possibility* of SNS-based learning activities in special circumstances and with tailored support, it does not provide further insights about the *everyday* and *spontaneous* use of ubiquitous SNS technology for learning and study in secondary school settings.

Recent work has sought to address this gap by exploring how teachers and students interact in SNSs (Asterhan & Rosenberg, 2015; Ophir, Rosenberg, Asterhan & Schwarz, 2016; Rosenberg & Asterhan, in press; HersHKovitz & Forkosh-Baruch, 2013; Schwarz & Caduri, in press). The combined findings from those studies show that teachers use SNS-based communication with their pupils for social-relational and psycho-pedagogical purposes, but also to support academic-instructional activities. In the present work, we extend this work by focusing on teenage, peer-to-peer, self-organized use of ubiquitous SNS technologies (i.e., Facebook, WhatsApp) for learning purposes in formal school settings.

We first explored this space with a pilot study that consisted of semi-structured teenager interviews (Bouton & Asterhan, 2014). Based on those first findings, we concluded that teenagers' self-organized learning- and study-related SNS activities are best approached under the umbrella of *online knowledge sharing*. Knowledge sharing is a well-known construct in communication, business management and information science. We provide a short overview of these bodies of research and their main findings.

## Online knowledge sharing

Communication scholar Nicholas John (2012) has argued that "sharing" has become the constitutive activity of social media, and of social network sites (SNSs) in particular. *Knowledge sharing* refers to activities in which individuals share their own internally stored knowledge or external knowledge sources they have at their disposal by making it accessible to others. There are countless examples of online knowledge sharing, such as contributing to Wikipedia, posting a response to a question on a thematic Q&A forum, uploading a tutorial video to YouTube, or posting college lecture summaries on a blog. Access to this knowledge may or may not require membership. In the vast majority of cases, there is no direct monetary reward involved for making one's knowledge available. Moreover, knowledge sharing is not a zero-sum game, as when one shares a candy bar with a friend or when sharing a dormitory room with another student. It involves letting someone else have something that you have, without entailing any kind of material sacrifice on the part of the sharer (John 2012). In other words, upon sharing one's knowledge one does not become "less knowledgeable". Quite to the contrary, when a sufficient number of participants contribute, knowledge sharing leaves one with more.

However, it does require time and effort to assemble and share knowledge online, and this is done without receiving direct material benefits, without the assurance of reciprocation, and often without knowing who benefits from this knowledge. Moreover, if indeed "knowledge is power", then in a competitive environment the sharer may lose his/her relative advantage over others. In many ways, knowledge sharing has then similar features to a public good dilemma (Connolly & Thorn, 1990): In this case, the public good is knowledge, from which every member of a group may benefit, regardless of whether or not they personally contribute to its provision (Olson, 1965), but whose availability does not diminish with use. According to game theory, defecting (not contributing) is technically considered to be the strategy which from an individual member's standpoint maximizes utility, independent of whether others contribute to the resource or not (Cabrera & Cabrera, 2002; Dawes, 1980). This raises the question: What motivates individuals to voluntarily share knowledge?

One model that has been used as a framework for explaining willingness to share is the gift economy model (Mauss, 1967): In pre-monetary societies, people exchanged goods with their acquaintances in an intricate weave of reciprocal acts. As there were no formal financial systems to protect the wealthy against future misfortunes, 'gifts' donated to society served as a social guarantee that the family that was kind enough to share their good fortune today, will be protected and taken care of, should the need occur in the future. This *quid pro quo* motive for sharing has been used to explain how seemingly altruistic online sharing may be based on expectation for future economic rewards (e.g., Restivo & van de Rijt, 2014; Roberts et al., 2006).

Knowledge sharing, its motivations and rewards has interested scholars from various fields of research, but predominantly so in business management and organizational science, where employee knowledge sharing has been related to a range of success criteria, such as decreased production costs, innovation, revenues and team performance (see Wang & Noe, 2010, for a review). Factors that predict individual willingness to share knowledge with members in an organization are, among others, expectations of reciprocity, expectations of personal benefit (and reduced costs of sharing), interpersonal trust, collectivist values, self-efficacy and lack of evaluation apprehension (Wang & Noe, 2010). Knowledge sharing has also been studied in the context of online professional learning communities (e.g., Belous, 2014; Lin et al., 2008; Tseng & Kuo, 2014), where sharing is mainly motivated by intrinsic rewards, altruism and self-efficacy.

## Knowledge sharing in educational contexts

Against this background, it is surprising that, to date, knowledge sharing in educational contexts has received so little scholarly attention. Knowledge sharing in school contexts is different from the aforementioned settings in several ways: First of all, students from a secondary school know and interact extensively with one another offline as well as online. Secondly, individual excellence in secondary school settings does not translate into monetary or other materialistic rewards (such as, salary bonuses, promotions), and the potential of future rewards (such as college scholarships) may be less salient to secondary school students. On the other hand, competition for academic excellence (medals, awards, honors) are by definition based on relative standing in a group. Such considerations may therefore stymie students' motivation to share school-relevant knowledge. Finally, whereas knowledge sharing is actively promoted by managers and leaders in organizations, attitudes towards sharing in the educational realm tend to be more ambiguous: On the one hand, peer assistance and help-giving are valued and encouraged in schools. Collaborative group work, peer tutoring and discussion are cornerstones of progressive pedagogies. In essence, even teaching is a form of knowledge sharing. However, assessment and evaluation is predominantly based on individual performance. Thus, peer knowledge sharing in the sense of giving, receiving and using solved solutions, completed homework assignments, and answers to test items are considered unethical (cheating) and, therefore, prohibited. Moreover, sharing in the sense of

exchanging learning derivatives is often discouraged: By relying on adapted materials that are produced by others, instead of processing the materials by one's self, important learning gains may be forfeited.

## The present research

The overall aim of the present research is to map teenagers' school-related knowledge sharing practices in SNSs. Two panel data studies were conducted on representative samples of Israeli, Hebrew-speaking adolescents. The first panel study (Study 1) was exploratory, and designed to gain first insights into the extent and specifics of teenagers' school-related SNS knowledge sharing practices. Based on a representative sample of 206 teenagers, the findings showed that most of them are members of student-organized SNS study groups and that the vast majority (90%) participates in some form of school-related knowledge sharing in those groups, mainly through WhatsApp (and to a lesser extent on Facebook). Moreover, teenagers almost unanimously regard knowledge sharing as beneficial for their learning.

Because of space considerations we do not provide a full report of this study here, as the nature of Study 1 was mainly exploratory and formed the basis for formulating hypotheses to be tested in a larger sample, Study 2. The research questions and hypotheses of Study 2 are organized according to four different aspects of teenage school-related sharing in SNSs (whether, what, why and who):

(1) *Whether*: How common and widespread is the phenomenon? How often do they share, how often do they use shared materials? How many SNS study groups do they participate in, on average? Do they appreciate it or is it considered a nuisance? Based on the findings from the pilot study interviews and the Study 1 survey it is expected that the majority of high school students participate in knowledge sharing through SNS, are members of at least one SNS study groups, and regard sharing positively.

(2) *What*: What types of knowledge sources are shared most often by high school students? Based on findings from Study 1, we expect that materials that require little personal effort to produce (e.g., snapshot, technical messages) will be shared more frequently, compared to learning materials that require substantive individual effort to produce (e.g., reading material summaries).

(3) *Why*: What motivates high school students to share learning materials in SNS study groups, and why in their opinion, do others choose to share? Is this mainly motivated by pro-social motives (i.e., the wish to help others) or by more egocentric motives (i.e., self-enhancement, impression management)? Also, do they feel social pressure to comply with sharing norms and do they experience regret afterwards?

(4) *Who*: Is there a profile for frequent sharers, or central knowledge brokers, and can they be distinguished from less frequent sharers? Based on findings from the aforementioned literatures, we expect that sharing is positively predicted by collectivist, but negatively predicted by individualistic value orientations. We also expect that sharing occurs more frequently among students with higher academic self-efficacy. Based on achievement goal theory, it is expected that mastery orientations are positively associated with sharing, but negatively related with performance goal orientations. Finally, based on existing research on gender differences in peer help-seeking, we hypothesize that sharing is more frequent among female than male students.

## Method

### Participants

515 Hebrew-speaking Israeli adolescents were recruited from the largest national panel sample (with over 10,000 active adolescent members), which is subjected to state privacy law and ethical regulations. In the registration process, users provide basic biographical data (e.g., age, gender, residence, mother tongue, religious affiliation). This biographical information is used for selection procedures (e.g., mother tongue, ethnicity, religiosity, SES) as well as to build representative samples for surveys. Registration to the panel requires that adolescents as well as their parents read and sign consent forms. Invitations to participate in the current study were sent via e-mail to all registered adolescent members (age 15-17) from the majority population in Israel (mainstream, ethnically Jewish population). Because of substantive differences in school systems, cultural norms, internet availability and/or teacher-student relationships, we did not recruit participants from the ultra-orthodox Jewish and the Arab-speaking population. The invitation did not reveal the research topic. Participation was on a first-come, first-served basis and was closed once the goal of 500 adolescent participants with active SNS accounts was reached, while safeguarding a representative breakdown of gender, age, and the different religious sectors that is representative of mainstream Jewish adolescent population (53% secular, 18% traditional, 18% religious).

Relying on the results of study 1, we assumed that most of the educational uses of SNS were organized via specifically generated SNS learning groups, created mainly by students, in various social network sites such as Facebook or WhatsApp. However, in the current sample only less than two-third of the total sample ( $N =$

291) admitted they were members of at least one SNS study group. The remaining 224 respondents then did not complete survey items relating to SNS study group activities. Data analyses of research questions revolving around such group activities is then restricted to 291 participants, instead of 515. Mean age of the 291 remaining participants (127 male) was 16.73 ( $SD = 1.04$ ). All the questionnaires were completed online, during the first week of April, 2015. Net time to complete the survey was estimated at 15 - 20 min.

## Tools

The full survey included 93 closed-end items. It included items that provided information about demographic background, overall SNS use, features of SNS study groups (e.g., how many groups, who initiates them, number of members in a typical group), as well as the following measures:

*Sharing intensity* was assessed with five items, each referring to one of five different types of content categories (Cronbach's,  $\alpha = .778$ ): Administrative messages ("such as homework instructions/reminders, tests notifications and so on"), snapshots ("of the teacher's notes on the class board or handouts given in class"), content summaries ("of class notes, reading material summaries, and so on"), solved homework and other individual assignments, and peer learning ("such as, helping friends and explaining online points they didn't understand"). Frequency of sharing was rated on a 5 point Likert scale, separately for (1) content shared; and (2) using shared content posted by others.

*Students' feelings about sharing.* Five Likert scale, self-report items were included in which respondents indicated the extent to which they have experienced regret after sharing, feel (themselves / others) pressure to share, are dependent on sharing, their sharing helps others.

*Explicit motives for sharing.* Based on findings from Study 1, six common motives for (own / other) sharing were phrased: Improving academic achievements, Helping classmates succeed, Positive self-concept, Quid pro quo, Gain social stature and Lack of effort. Likert scales ranged from 1 ('not true') to 5 ('very true').

*Individualism and collectivism.* Students' individual orientation towards collectivist and individualist values was measured by using two scales from the Singelis et al. (1995) survey, which was translated to Hebrew and validated by Adar (2005). Examples of items are: "Competition is a law of nature", "It annoys me when others perform better than I do", "The well-being of my co-students is important to me", and "I feel good when I cooperate with others". Internal reliability was  $\alpha = 0.78$  for the collectivist and  $\alpha = 0.75$  for the individualist value orientation scale ( $N = 515$ ).

*Academic Self efficacy.* Efficacy was measured with 8 items from an adapted version of the English version of the General Self Efficacy scale (Schwarzer & Jerusalem, 1995), namely the NGSE (translated and validated by Chen et al., 2001), which was adapted to target academic SE. Internal reliability was high,  $\alpha = 0.93$ .

*Achievement goals.* Individual achievement goal orientations were assessed with 18 items extracted from the Elliott & Church (1997) scales. Internal reliability was  $\alpha = 0.76$  for the mastery scale,  $\alpha = 0.89$  for the performance approach scale and  $\alpha = 0.73$  for the ability-avoidance scale.

## Results

### Whether students share in SNS study groups

On average, teenage participants ( $N = 515$ ) reported receiving hundreds of notifications daily from the two SNS accounts together ( $M = 428.33$ ,  $SD = 554.89$ , range from 0 - 2,000). This questionnaire did not include separate questions on memberships in WhatsApp or Facebook, but rather asked generally about number of notifications, as stated above. Only two participants (< 1%) said they do not receive notifications at all.

As in Study 1, participants were asked to indicate the number of SNS study groups they are members of. However, the format was slightly different from that in Study 1: Participants could either tick the "none" option or write the number of SNS study groups in an open-ended format (rather than choose from a close set of predefined intervals, as in Study 1). Unlike study 1, where the vast majority of students admitted being members of SNS groups, in the current sample, only 57% of respondents reported participation in at least one SNS study group, which is significantly less than in the Study 1 sample. Because of the different test item format, it is not possible to infer whether this reflects a genuine difference between the two samples or is an artifact of the different test format (i.e., clicking a predefined answer requires less effort, which increased participants' tendency to choose the "none" option more often). Participants who choose the "none" option, did not receive any further items on sharing in SNS study groups. The remainder of the data analyses on sharing behavior in study groups is therefore limited to  $N = 291$ .

Overall sharing intensity was calculated based on the mean frequency of the five separate types of learning materials. Mean sharing intensity was  $M = 2.81$  ( $SD = 1.05$ ), which is similar to findings from Study 1. Forty-four percent considered themselves prominent sharers in their groups. The majority of respondents (89%)



are members of more than one study group, more than 4 groups on average ( $M = 4.33, SD = 2.97$ ). SNS study groups are typically initiated by students (56%), rather than teachers (10%).

### What is shared in SNS study groups?

The mean sharing frequency score for each of the five different types of content was calculated separately (see Figure 1). Paired sample *t*-test comparisons were conducted with Bonferroni alpha corrections for multiple comparisons (0.05/10) within each sharing activity (shared / used). For own sharing the pattern was as follows: administrative messages = snapshots = peer assistance > content summaries > copying. For using shared materials: administrative messages > snapshots > peer assistance > content summaries > copying.

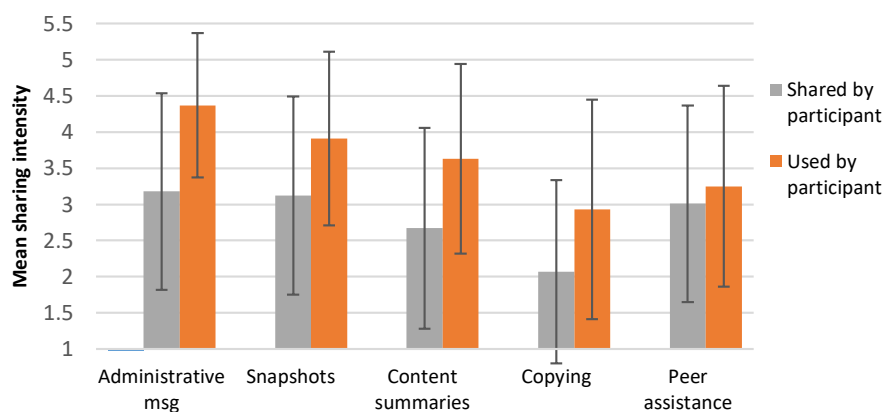


Figure 1. Mean (and SD) sharing intensity in SNS study groups, as shared and used by respondents ( $N = 291$ ).

### Why share?

The majority of participants (77%) strongly disagreed with the statements of experiencing regret after sharing ( $M = 1.42, SD = 0.88$ ), experiencing peer pressure to share (66%,  $M = 1.69, SD = 1.15$ ), or that others are pressured (51%,  $M = 1.97, SD = 1.23$ ). On the other hand, participants expressed overall agreement with positive statements endorsing sharing benefits: They feel that sharing their own learning materials helps their fellow classmates to improve their academic performance ( $M = 3.79, SD = 1.21$ ). Moreover, 21% of respondents strongly agreed with the statement that they are dependent on sharing to succeed ( $M = 3.19, SD = 1.26$ ).

Participants' responses to six predefined sharing motives were measured with separate items for one's own sharing and the sharing by other group members. To examine whether the mean differences between types of motives proved to be significant, paired sample *t*-test comparisons were conducted with Bonferroni adjustments for multiple comparisons (critical  $p = 0.05/15$ ) within each sharing category (i.e., participants' own motives for sharing and their perceptions on others' motives to share). The pattern that was obtained for own and others' sharing was identical: help classmates succeed ( $M = 4.43, SD = 0.84$ ) > improve self-concept = quid pro quo = improve own learning > gain social status ( $M = 2.05, SD = 1.36$ ).

### Who shares?

Gender differences were tested using a one-tailed independent sample *t* test. Female students were found to share more overall ( $M = 2.90, SD = 1.06$ ) than male students ( $M = 2.69, SD = 1.03$ ),  $t(289) = 1.70, p = .045$ . When separately tested for each type of sharing, Mann Whitney test showed that female students only shared content summaries more often ( $Mdn = 156.63$ ) than male students ( $Mdn = 132.27$ ),  $U = 12,158.0, p = .012$ .

Pearson correlations were calculated between the six individual characteristics and overall sharing intensity, as well as Spearman correlations with each of the five different sharing content categories (see Table 1). Inter-correlations between most scales were either non-existent or low. However, a strong, positive correlation was found between the individualist value orientation scale and the performance-approach goal scale,  $r(291) = .61, p < .001$ . Performance-approach goal scales were then omitted and we refer to the remaining scale as "competitive individualism" from here on onward.

Table 1. Bivariate correlations between individual characteristics and different sharing measures ( $N = 291$ )

	Collectivism values	Competitive individualism	Self-efficacy	Mastery goal	Performance-avoidance goal
Overall sharing intensity	.374**	.043	.267**	.327**	.092
Administrative msg	.340**	.018	.246**	.255**	.089
Snapshots /handouts	.316**	-.012	.204**	.225**	.087
Content summaries	.309**	.049	.186**	.252**	.081
Copying	.202**	.142*	.209**	.248**	.048
Peer learning	.291**	-.003	.303**	.288**	.060

As expected, positive correlations of moderate strength were found between collectivist value orientations, academic self-efficacy, and mastery goal orientation with sharing. Endorsement of performance-avoidance achievement goals was not related with any of the sharing categories. A multiple linear regression analysis showed that these predicted sharing intensity,  $F(3,287) = 24.78, p < .001$ , accounting for approximately 20% of the variance of overall sharing intensity,  $R^2 = .206$ , Adjusted  $R^2 = .197$ . Each of the three individual characteristics contributed separately to the prediction of overall sharing intensity (collectivism:  $\beta = .291, p < .001$ , mastery:  $\beta = .191, p = .003$ , self-efficacy:  $\beta = .149, p = .014$ ).

No correlation was found between overall actual sharing intensity and competitive individualism,  $r(291) = .04, p = .466$ . However, competitive individualism positively correlated with regret after sharing ( $r_s = .233, p < .001$ ), perceived peer pressure to share ( $r_s = .302, p < .001$ ), and the belief that others experience peer pressure ( $r_s = .316, p < .001$ ). Finally, competitive individualism was positively associated with on sharing category, namely the sharing (and using,  $r_s = .13, p = .024$ ) of solved homework tasks (cheating).

We examined the possibility that the relation between competitive individualism and sharing would be moderated by belief in quid pro quo benefits. Respondents were characterized as either endorsing or not endorsing gift economy views, based on the quid pro quo item. Competitive individualism was positively related with overall sharing when students expect future gains from it (48%),  $r = .24, p = .032$ , whereas among non-believers this correlation was negative,  $r = -.28, p = .011$ . This pattern was consistent across the five different types of sharing behavior as well.

## Discussion and significance

The combined findings presented here provide a first, descriptive account of teenage knowledge sharing via ubiquitous SNSs in secondary school settings. School-related knowledge sharing refers to the up- and downloading (posts, files) of knowledge and knowledge sources that pertain to the learning and studying of curricular topics to/from a SNS group. It includes sharing of logistical and organizational information, sharing of teacher-created materials, providing online peer assistance, and to a lesser extent the sharing of student-created content summaries and even completed individual assignments (cheating).

The findings show that knowledge sharing through SNSs is a widespread phenomenon that has become an integral part of routine study practices among secondary school students. Students have been known to borrow and copy content from each others' notebooks prior the introduction of Web 2.0 tools. Therefore, student peer-to-peer knowledge sharing is not a novel phenomenon in essence, nor is it created by SNS technology. What has changed, however, is the ease, and efficiency, and therefore the scale, with which information and knowledge can be duplicated and shared with the help of modern communication technologies.

We discuss our main findings, their contributions and the directions for future research from two separate angles: a knowledge sharing perspective and a learning theory perspective.

### Knowledge sharing in school settings

The present work extends the literature on online knowledge sharing as it is the first to address the phenomenon in formal, secondary education. Similar to findings from adult knowledge sharing in online communities, self-reported motivations for teenage sharing were predominantly pro-social in nature. In secondary school settings, interpersonal competition for material rewards and thus the personal costs of sharing is perhaps less salient than in professional settings. Not all sharing was purely motivated by altruistic motives, however, since quid pro quo motivations were found to play a role as well. Participation in SNS-based sharing is more likely when a teenager is female, endorses collectivist values, is guided by mastery goals, and has high academic self-efficacy.

Finally, in contrast to expectations, competitive individualism was not associated with less overall knowledge sharing or more overall use of shared materials, as may have been expected based on a straightforward utility maximization strategy. The results presented here showed that belief in quid pro quo, i.e., the gift economy rationale for sharing, serves as a moderating factor of the association between peer sharing and

competitive individualism orientation: Among teenagers who expect quid pro quo benefits from sharing, endorsement of competitive individualism was associated with increased sharing (and vice versa). Finally, even though competitively oriented teenagers did participate in sharing activities, they also expressed more regret and felt they were under more social pressure to share content. Moreover, and in alignment with findings on performance goals, competitive individualism was associated with more frequent sharing of the cheating type, that is: sharing and using copied homework and other assignments.

This first empirical study should be followed up by research that explores the phenomenon with additional methodologies (e.g., direct observations and qualitative research tools) and in additional cultures and countries. Given the popularity of SNSs among teenagers in many other countries, it is reasonable to expect that knowledge sharing is a widespread and common phenomenon there as well. However, different norms and practices may evolve locally and are likely to be influenced by local school cultures. Future research should include additional educational settings. In higher education, for example, competition for individual monetary rewards is more salient (scholarships, job offers, placement in graduate schools) and social cohesion not as strong, compared to secondary schools. This may affect the frequency of sharing as well as motivations behind it. Finally, future research should further explore the social structure of school-related knowledge sharing. Recent findings in the Netherlands reveal that, in spite of euphoric prediction about the democratizing effects of so-called consumer-directed sharing economies, sharing of under-utilized physical goods (such as cars, tools, and apartments) is in fact highly stratified within social class. Moreover, the supply and the demand of shared goods is dominated by middle-class participants, with considerably less participation from the upper and lower classes (de Beer & de Gier, 2015). The quid pro quo expectation may in fact deter the 'have-nots' from using shared goods, as they will have difficulty to return the favor in the future. Similar questions can and should be raised regarding knowledge sharing: Who gains most from knowledge sharing, who loses out on potential benefits and who is (purposefully) left out?

### Knowledge sharing, peer collaboration and learning

From an organizational point of view, knowledge sharing is a means to reach the organization's end goals more efficiently (Wang & Noe, 2010), but is not an end goal in itself. From an educational point of view, however, the desirability of online knowledge sharing between students is less clear cut. Whereas values of collaboration, sharing and pro-social behavior are encouraged and nurtured by society, parents and in schools, assessment and evaluation is predominantly based on individual performance. Individual mastery of knowledge is (one of) the end goal(s) of formal education. The most obvious case of undesirable sharing is that of solved homework tasks and other assignments. Even though it proved to be the least frequent type of sharing in the current study, still more than a quarter of the participants in both studies admitted to using it very frequently. Copying assignments and handing them in as one's own is considered unethical ('cheating'), since it provides an inaccurate picture of whether the end goal has been reached. Aside from the ethical aspects, however, the overarching question is whether SNS-based knowledge sharing is conducive to individual learning, or not?

Our findings showed that, overall, teenage students regard online peer-to-peer knowledge sharing positively and beneficial to their individual learning. However, these subjective perceptions may not necessarily reflect actual learning benefits. There are, in fact, several reasons that dampen such positive expectations.

First of all, a vast body of empirical research has shown that peer-based learning may indeed produce individual learning gains, provided that peers engage in particular rich forms of egalitarian, reasoned, transactive dialogue in which they co-construct knowledge (for reviews see Resnick, Asterhan & Clarke, 2015). Learners improve their individual knowledge and understanding through negotiating, externalizing and challenging (the reasons for) each other's knowledge structures. This collaborative knowledge construction shares some surface features with online knowledge sharing as it is described in the present work: It is a collaborative, peer-based effort in a formal learning context. However, it lacks the pivotal attributes of knowledge co-construction and can therefore not be assumed to improve individual learning in a similar vein. Quite to the contrary, by overly relying on learning derivatives that are produced by others, instead of self-made, students may forfeit important individual learning activities that produce knowledge gains as well as develop important competencies (e.g., summarizing, highlighting and integrating information).

A second reason to be cautious about expected learning benefits from online knowledge sharing stems from recent research on transactive memory systems and the increasing role of the Internet as the ultimate transactive memory partner (Sparrow, Liu, & Wegner, 2011; Ward, 2013a; Wegner, 1987). The Internet contains infinitely more expertise than a singular human partner, is accessible to all and is ever available. Recent research shows that people systematically overestimate their own internally stored knowledge, as they conflate it with the vast amounts of knowledge that are available through the Internet (Ward, 2013b). For example, Fisher, Goddu & Keil (2015) showed that the mere act of searching the Internet for knowledge creates an

illusion whereby people mistake potential access to Internet-stored information for their own personal understanding of the information even when the transactive memory partner is unavailable. Extrapolating from this research to the current settings, it is possible that the information gathered through and stored in online SNS study groups may cause a similar illusion of knowledge: The mere act of storing shared learning materials and derivatives in one's cell phone or cloud, combined with the knowledge that one can access this information at any time, may cause learners to overestimate their own internally stored knowledge and underestimate the need for extra study time. This could then paradoxically lead to less actual learning.

Finally, the findings reported here show that students share and gather shared materials on a regular basis. They do not provide further insight about whether and how they actually keep track, store, utilize and integrate these different knowledge resources, however. Are these shared resources mainly used as additional materials, or do they replace learning from the teacher-assigned, canonical materials? How do students select and decide what is relevant, important or helpful, especially when they have several knowledge resources at their disposal (e.g., shared summaries, lesson notes, whiteboard pictures, textbooks) from potentially different individuals?

The present work is a first step toward a better understanding of a novel, widespread phenomenon that was hitherto underexposed and could potentially have many implications for learning and study performances in formal education. More research is needed to broaden and deepen this understanding, not only for scientific purposes, but also to enable informed decision-making when addressing the practical, ethical and social questions that come along with it.

## Selected references

- Asterhan, C. S. C. & Rosenberg, H. (2015). The promise, reality and dilemmas of secondary school teacher-student interactions in Facebook: The teacher perspective. *Computers & Education*, 85, 134–148.
- Bingham, T. & Connor, M. (2015). *The New Social Learning: Connect. Collaborate. Work*. Alexandria, Virginia: ATD Press.
- Cabrera, A., Collins, W. C. & Salgado, J. F. (2007). Determinants of individual engagement in knowledge sharing. *The International Journal of Human Resource Management*, 17, 245-264.
- Chen, C. & Hung, S. (2010). To give or to receive? Factors influencing members' knowledge sharing and community promotion in professional virtual communities. *Information & Management*, 47, 226-236.
- Fisher, M., Goddu, M.K., & Keil, F. C. (2015). Searching for explanations: How the Internet inflates estimates of internal knowledge. *Journal of Experimental Psychology: General*, 144, 674-687.
- Greenhow, C., Gibbins, T., & Menzer, M. M. (2015). Re-thinking scientific literacy out-of-school: Arguing science issues in a niche Facebook application. *Computers in Human Behavior*, 53, 593-604.
- Hew, K. F. (2011). Students' and teachers' use of Facebook. *Computers in Human Behavior*, 27, 662-676.
- Hollis, R. J., & Was, C. A. (2016). Mind wandering, control failures, and social media distractions in online learning. *Learning & Instruction*, 42, 104–112.
- John, N. A. (2012). Sharing and Web 2.0: The emergence of a keyword. *New Media & Society*, 15, 167-182.
- Junco, R. (2012). Too much face and not enough books: the relationship between multiple indices of Facebook use and academic performance. *Computers in Human Behavior*. 28 (1), 187-198.
- Kirschner, P. A. & Karpinski, A. C. (2010). Facebook and academic performance. *Computers in Human Behavior*. 26, 1237–1245.
- Ophir, Y., Rosenberg, H., Asterhan, C. S. C., & Schwarz, B.B. (2016). In times of war, teachers do not fall silent: Teacher-student social network communication in wartimes. *Journal of Adolescence*, 46, 98-106.
- Resnick, L. B., Asterhan, C. S. C., & Clarke, S. N. (Eds.) (2015), *Socializing Intelligence through academic talk and dialogue*. Washington, DC: AERA.
- Rosenberg, H., & Asterhan, C. S. C. (in press). WhatsApp, sir? Teachers and students in Whatsapp groups (in Hebrew). In B. B. Schwarz, H. Rosenberg, & C. S. C. Asterhan (Eds.), *Breaking down barriers in education? Teachers, students and social network sites*. MOFET books.
- Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333, 776–778.
- Tseng, F., & Kuo, F (2014). A study of social participation and knowledge sharing in teachers' online professional community of practice. *Computer & Education*, 72, 37-47.
- Wang, S. & Noe, R. A. (2010). Knowledge sharing: a review and directions for future research. *Human Resource Management Review*. 20, 115-131.

# Collective Knowledge Advancement and Conceptual Understanding of Complex Scientific Concepts in the Jigsaw Instruction

Jun Oshima, Shizuoka University, joshima@inf.shizuoka.ac.jp  
Ayano Ohsaki, AIIT, ohsaki-ayano@aiit.ac.jp  
Yuki Yamada, Shizuoka University, y119y.b2.226@gmail.com  
Ritsuko Oshima, Shizuoka University, roshima@inf.shizuoka.ac.jp

**Abstract:** We examined jigsaw instruction related to the human immune system at a high school from the perspective of knowledge creation. Collective knowledge advancement was analyzed using socio-semantic network analysis (SSNA) and in-depth discourse analysis. Their conceptual understanding was collected through pre- and post-testing and evaluated by using the structure–behavior–function framework. SSNA revealed that higher conceptual understanding was related to improved collective knowledge advancement. Further in-depth discourse analysis clarified that students who acquired higher conceptual understanding engaged in shared epistemic agency through productive regulatory processes to promote their collective knowledge advancement. Based on the results, we discuss and propose scaffoldings for more students to engage in productive collaboration in jigsaw instruction.

**Keywords:** collective knowledge advancement, shared epistemic agency, regulation of collaboration, structure–behavior–function framework, jigsaw instruction

## Theoretical background

### Inquiry into collective knowledge advancement

In the knowledge-creation metaphor of learning (Paavola & Hakkarainen, 2005), students are expected to be involved in knowledge creation practice through collaborative construction of knowledge objects (Bereiter, 2002). Regarding creating knowledge in the classroom, Scardamalia (2002) discussed collective cognitive responsibility for contributing ideas toward collective knowledge advancement. She defines intentional engagement in collective knowledge advancement as the epistemic agency and proposes this agency as a new goal for instruction in the knowledge age (Scardamalia, Bransford, Kozma, & Quellmalz, 2012). Damşa, Kirschner, Andriessen, Erkens, and Sins (2010) further propose shared epistemic agency focusing more on the group-level agency. Through in-depth discourse analysis, Damşa et al. found that students in collaborative groups engage in the wholly joint epistemic actions of (1) being aware of their lack of knowledge, (2) alleviating this lack of knowledge, (3) creating shared understanding, and (4) generative collaboration. To regulate their joint epistemic actions, students were also found to engage in (1) projection by setting goals and creating joint plans, (2) regulation by monitoring and reflecting on their advancement, and (3) relation by transcending conflicts, redirecting critical feedback, and creating space for others' contributions.

### A new computational approach to collective knowledge advancement

In CSCL research, there have been discussions on the advantages of using social network analysis (SNA) to investigate collective knowledge advancement and learner engagement (e.g., Martinez et al. 2003; Reuven et al. 2003). De Laat et al. (2007) outlined an approach to synthesizing and extending comprehension of CSCL teaching and learning processes to balance SNA, content analysis, and critical event recall. In this complementary approach, SNA was used to study interaction patterns within a networked learning community, and to study how learners share and construct knowledge. They concluded that including SNA in any multi-method approach is advantageous, because doing so provides researchers and learners with tools for illustrating comprehension and cohesion of group activities, and because it provides researchers a method for selecting appropriate groups to study. Several studies have used SNA, especially in as a knowledge-creation metaphor. Over three years, Zhang et al. (2009) implemented a complementary approach that used SNA to visualize and compare classroom collaboration among fourth-grade elementary school students in a CSCL environment designed to support knowledge building. An analysis of online participatory patterns and knowledge advancement indicated that this learning process effectively facilitated knowledge advancement through critical changes in organizations within

the classroom, from fixed small groups in the first year of the study to appropriate collaboration through the dynamic formation of small teams based on emergent goals.

We extended the potential of SNA by describing a different type of network, socio-semantic network analysis (SSNA). Ordinary SNA illustrates the social patterns of learners, namely, their social network. As de Laat et al. (2007) suggested, this approach is thus informative when examining developments or changes in the participatory structure of a community. However, several studies argued that existing social network models are unable to examine how collective knowledge advances through learner collaboration (Oshima et al., 2007; Schaffer et al., 2009). Instead, we used a procedure similar to ordinary SNA but proposed a different type of social network, one based on the words learners use in their discourse in a CSCL environment. We compared this socio-semantic network—in which words were selected as nodes representing learners' knowledge or ideas during a discourse on a study topic—with a network of words from the discourse of a group of experts on the same topic. The results showed remarkable differences in the collective knowledge of elementary school students and experts regarding the words centered on the networks. We concluded that SSNA could provide a new representation of community knowledge building, enabling researchers to adopt a new complementary assessment technique for investigating models of knowledge-building communities. In recent years, this SSNA approach has been adopted in CSCL studies to analyze student roles in collaboration and to detect productive interaction patterns (e.g., Ma et al., 2016; Oshima, Oshima, & Matsuzawa, 2012).

### Conceptual understanding of complex scientific concepts

In knowledge-creation practices, learners take on complex tasks and comprehension of phenomena. Complex systems are defined as multiple levels of organizations locally interacting with one another such as financial economies and weather systems (Wilensky & Jacobson, 2013). Many students have difficulty mastering such complex subjects, despite their importance. One reason for this is that these concepts conflict with prior experience. Students usually have a “centralized” mindset and tend to provide explanations that assume central control and simple causality. In an interview study, Jacobson (2001) found that undergraduate students are more likely to generate simple causality, central control, and predictability in comparison with experts, who exhibit decentralized thinking of multiple causes as stochastic and equilibration processes.

Hmelo-Silver and Pfeffer (2004) proposed the structure–behavior–function (SBF) framework for assessing student understanding of complex systems. While the SBF framework has been used to examine the design of physical devices, they applied it to explaining student understanding of multiple interrelations and the dynamic nature of complex systems. To assess student understanding of an aquarium as a complex system, for instance, Hmelo-Silver and Pfeffer (2004) used the SBF framework as follows: Structures are elements of a system, and in an aquarium, there are fish, plants, and a filter. Behaviors represent how system structures achieve their purpose, such as filters that remove waste by trapping large particles, absorbing chemicals, and converting ammonia into harmless chemicals. Finally, functions express why an element exists within a given system, that is, the purpose of a system element. For example, the filter removes aquarium byproducts. They studied verbal responses and pictorial representations by middle school students, preservice teachers, and experts, and found that novices focused on perceptually available, static system components. Experts, on the other hand, focused more on interrelation among structures, functions, and behaviors. The results suggested that the SBF framework could be a useful formalism for understanding complex systems.

### Research purpose

This study examined how high school students engage in collective knowledge advancement through collaboration in jigsaw instruction (Brown & Campione, 1996; Miyake & Kirschner, 2013) and how their collective knowledge advancement is related to their learning outcome. Although studies have demonstrated that jigsaw instruction is effective for facilitating conceptual understanding (e.g., Miyake & Kirschner, 2013), few studies have shown how learners engage in collective knowledge advancement during collaboration. Through our design-based research on an immune system lesson (three class hours), we approached students' collective knowledge advancement by applying a multivocality approach (Suthers et al. (Eds.), 2013). First, we conducted a socio-semantic network analysis (SSNA) for numerically and visually representing collective knowledge advancement and comparing group performance based on their learning outcome, as evaluated by the SBF framework (in pre- and post-testing). We also conducted in-depth discourse analysis from the perspective of shared epistemic agency and the regulation of collaboration to examine how students interact with one another in collective knowledge advancement.

## Methods

### Student sample

Thirty-nine tenth grade students at a high school in Japan participated in this study as part of their regular curriculum. The high school is well-known and highly ranked in its district as a college prep school. Most graduating students go on to university. A science teacher with a Ph.D. in biology and more than ten years of teaching experience taught the students.

### Lesson unit design

#### Activity structure of collaborative learning: Constructive Jigsaw instruction (Figure 1)

We applied constructive jigsaw instruction (Miyake & Kirschner, 2013) in this study. In the jigsaw instruction, three students in a group were given a challenge such as “Can you explain how vaccinations protect us from infections?” then provided three study documents, each of which was necessary for solving the challenge. In the first phase of collaborative learning, one student from each group gathered to form an *expert* group and worked on their allocated materials over 1.5 lesson periods (each lesson period was 50 min.). After the expert group activity, students returned to their original group (the *jigsaw* group), where students had different pieces of knowledge. They were encouraged to share and integrate their individual knowledge to solve the challenge problem in the jigsaw group activity. This jigsaw activity took another 1.5 lesson periods. Group composition in the both group activities was designed by the teacher.

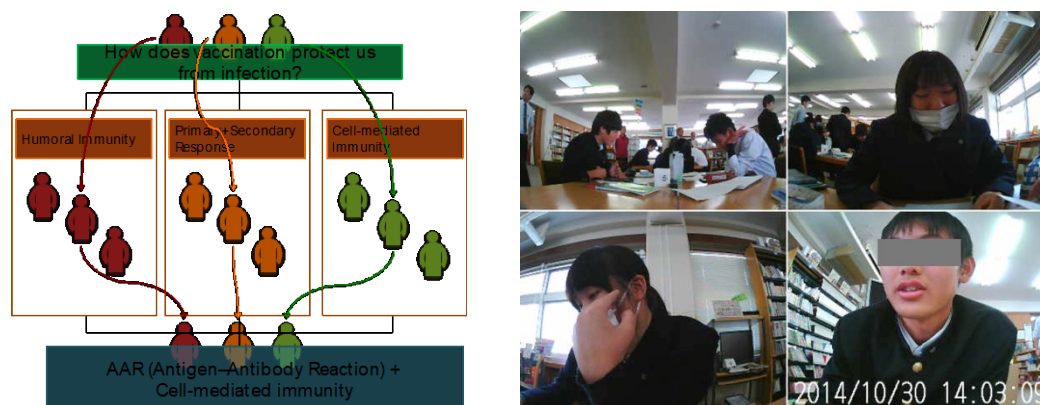


Figure 1. Jigsaw Activity Structure in the Immune System Lesson (Left) and a Snapshot of Recorded Data (Right).

#### Study documents

We first identified what knowledge and principles were covered in the school textbook. Figure 2 shows the SBF framework for the immune system as described in the textbook. We then discussed with the collaborating teacher how to separate the content into pieces of knowledge given to students in expert activities based on three key functions (the three areas separated by dashed lines in Figure 2): *humoral immunity*, *primary and secondary response*, and *cell-mediated immunity*.

### Study design

#### Pre- and post-tests

Before and after the lesson unit we applied pre- and post-tests to evaluate learning outcomes. Students were individually asked about their understanding of how the human immune system responds to vaccination. They were given a worksheet with a printed question, on which they could write or draw their ideas. The pre-test was conducted during the class period right before the lesson started. The post-test was conducted right after the lesson finished. Each test took one lesson period. Thirty-five students completed both tests and were further analyzed.

#### Process data collection

Student conversations during expert and jigsaw group activities were video recorded by a device with four cameras and an omnidirectional microphone placed at the center of the table (right half in Figure 1). Student conversations were transcribed and used for SSNA. We used transcriptions of the jigsaw group activity to examine how students

exerted agency in advancing their collective knowledge with three pieces of knowledge acquired in their expert group activity.

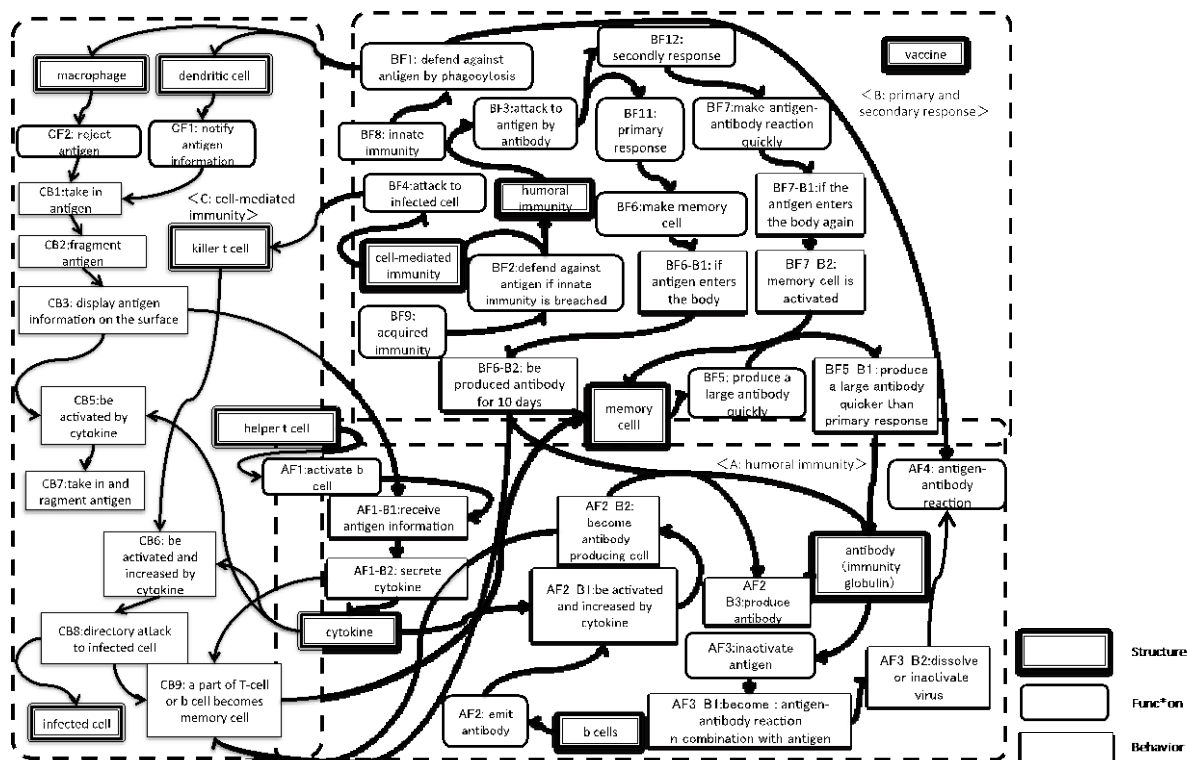


Figure 2. Structure-Function-Behavior Framework of the Human Immune System.

### Analysis procedure

We first investigated how students developed their biological understanding of the immune system as a complex system. To do so, we analyzed their writing and drawings on the pre- and post-test. Student SBF frameworks were categorized into the following types: When student explanations of how the immune system works considered the relationships among structures, their behaviors, and functions as described in a specific document (A, B, or C), we evaluated that they successfully constructed their understanding of the document. We did not count fragmented descriptions of structures, behaviors, or functions where connections were not identified. Based on this criterion, students were categorized as having (1) no understanding, (2) single-document understanding, (3) partially integrated understanding between two documents, or (4) fully integrated understanding across three documents. Referring to the SBF framework of the immune system, the first and second authors independently evaluated ten randomly selected students' SBF frameworks based on their explanatory discourse and pictures in each of the pre- and post-tests. Cohen's Kappa coefficient for the agreement between the two raters was 0.92. Disagreements were resolved through discussion. The first author evaluated the remaining data. Because no students demonstrated knowledge of the SBF framework of the immune system in the pre-test, we focused on their SBF frameworks in the post-test for analysis.

To analyze students' collective knowledge advancement, we next focused on discourse in the jigsaw group. Students on average engaged in discourse exchange 358.5 times in jigsaw groups (SD = 211.8). One reason for paying attention to the jigsaw activity was that students were expected to actively engage in creating new ideas by integrating their knowledge from three documents. To visualize and computationally investigate collective knowledge advancement, we conducted SSNA by the following procedure: We assumed that we could represent the state of collective knowledge as clusters of vocabulary that students used in their discourse. For investigating their collective knowledge, we selected words used to represent *structures* and *functions* of the immune system in the SBF. The socio-semantic network of vocabulary refers to meaningful links between words in exchanges. When students used words in their exchanges, we assumed that they were attempting to create meaningful links between words. We used 23 nouns representing *structures* and *functions* in the SBF framework of the immune system. We then used an application called KBDeX (<http://www.kbdex.net>) to SSNA to calculate the transition of the total



value of degree centralities of nodes in the network across discourse exchanges, following the method of previous research (Oshima et al., 2012).

Our SSNA approach requires complementary discourse analysis to examine how students exerted shared epistemic agency through regulatory processes of collaboration in their collective knowledge advancement. We therefore also conducted an in-depth analysis of discourse segments related to pivotal points of collective knowledge advancement as a drastic increase in the total value of degree centralities.

## Results and discussion

### Student learning outcomes after the jigsaw activity

We found that ten students fully integrated SBF understanding and eleven students did so partially. Eleven students indicated the understanding of a single part of a learned document and three did not sufficiently learn any piece of SBF. Chi-square analysis of student frequencies across three types of learning outcome (fully or partially integrated, the single document, or no understanding) showed significance ( $\chi^2 = 13.944, df = 2, p < .05$ ), and that the proportion of students having integrated SBF understanding was higher than students with no understanding. These results suggest that the jigsaw activity in practice facilitated student integration of knowledge through collaboration, but group differences remained in the learning outcomes.

### Group differences in collective knowledge advancement between high- and low-outcome groups

Based on the SBF framework evaluation of conceptual understanding, we categorized twelve groups as high or low learning-outcome groups. In the high learning-outcome groups, all members acquired fully or partially integrated conceptual understanding of documents. Low-outcome groups were mixed in their understanding. Figure 3 shows the transitions of total values of degree centralities. The total value of degree centralities represents how dense and structured a network (of words in this case) could be. This measure has been used as a typical index to detect collective knowledge advancement (e.g., Oshima et al., 2012). We found that the values quickly increased then finally exceeded 10.0 in the high-outcome groups, whereas the values stayed low and slowly increased across discourse exchanges in the low-outcome groups. The results suggest that students in the high-outcome groups engaged in collective knowledge advancement more quickly and sustainably.

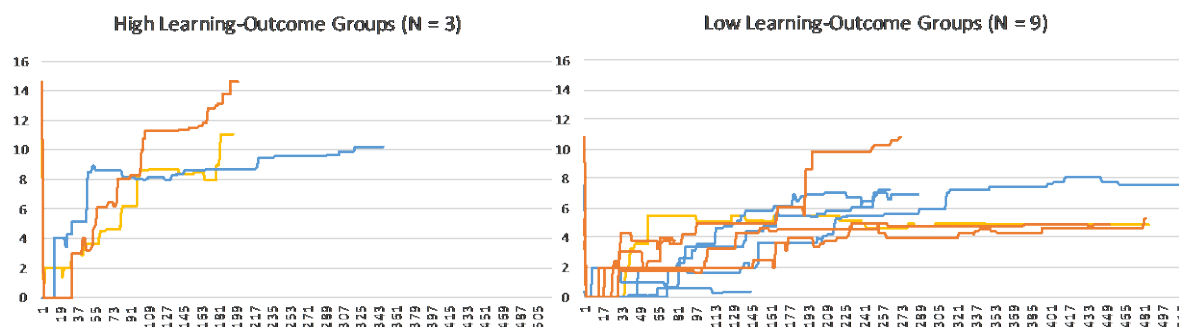


Figure 3. Transitions of Total Values of Degree Centralities Through Discourse Exchanges by High Learning-Outcome (left) and Low Learning-Outcome (right) Groups.

We further analyzed segments of student discourses for clarifying how they engaged in collective knowledge advancement. We first present one example of discourse segments by a high learning-outcome group (The original discourse was in Japanese and translated into English by the first author. SSNA vocabulary is in bold.).

Student A (156) So, what did we say? They are trapped and broken into smaller pieces, and their **antigenic** information is transmitted to **helper T cells**. Next, **helper T cells** emit a substance called **cytokine**.

Student B (157) **Cytokine**?

Student A (158) Yeah, **cytokine**. Oh, you [student B] put this [**cytokine**] down twice [on the worksheet].

Student B (159) Twice?

- Student A (160) Oh, never mind. Just put down “**cytokine**.” This is emitted. Then, draw an arrow from **T cell** to **B cell**, please. [Student A told B to draw an arrow on the worksheet.]
- Student B (161) T and B?
- Student A (162) I wonder how we can describe this... Well, **T cells** propagate. Would you [student B] shorten the space here [pointing at an area in the worksheet] a bit?
- Student B (163) Propagate?
- Student A (164) Yeah, **T cells** do propagate.
- Student C (165) Wait a minute. How about **memory T cells**? Are they part of the activated cells?
- Student A (166) **T cells** propagate. Then, how about these [**B cells**]? These [**B cells**] create **antibodies**.
- Student B (167) ... create **antibodies**.
- Student A (168) and, some will become **immunological memory cells**,
- Student B (169) **Immunological**?
- Student A (170) **Immunological memory cells**. Here, look [at a picture in their documents]. Some of them remain as **immunological memory cells**. Then, we go to **the secondary response**.
- Student B (171) **The secondary response**?
- Student A (172) Through **the secondary response**, when viruses come into our body...
- Student B (173) OK, they come into us.
- Student A (174) I wonder if we have to make two lines here [pointing an area in the worksheet], too. **Immunological memory cells** react to the viruses. **T cells** also react to them. **T cells** then become **killer T cells**. **Killer T cells** propagate. On no, we need more space [in the worksheet] to write this down... Then, this [**B cells**] emits **antibodies** to **antigens**. This is called **antigen–antibody response**.

From the perspective of shared epistemic agency, students mostly engaged in *creating shared understanding* of how the immune system works. Student A played a central role in externalizing shared understanding through monitoring confirmation by others (B and C). Student B had the role of recording their ideas on the worksheet, and so frequently *revoiced* student A’s externalizations (turn #159, #161, #163, #167, #169, #171, and #173). Student A interacted with B to co-create an external knowledge object in the worksheet. In contrast, student C had the different role of going beyond just *creating shared understanding* to *generative collaborative actions* by up-taking student A’s argument (turn #165). Student C might attempt to improve student A’s idea based on self-understanding by asking important questions from a different perspective (e.g., “Wait a minute. How about memory T-cells?”). Within this discourse segment, students A and C were more engaged in *generative collaborative actions*, whereas student B was engaged in *creating shared understanding*. Students A and C revealed fully integrated understanding in the post-test and B did partially integrated.

The next example is from a low-outcome group. They were discussing how to write their ideas on their iPad. They did not use their worksheet we introduced.

- Student E (63): So, what’s next? We acquire **immunity** systems. [looking at a picture in their documents]
- Student D (64): We learned two **immunity** systems.
- Student E (65): Yeah, **cell-mediated immunity** and **humoral immunity** around here? [pointing an area in a picture]
- Student D (66): **Humoral immunity**.
- Student E (67): **Humoral immunity**. We have to make a sentence.
- Student F (68): Yeah, we should.
- Student E (69): So, when something like **viruses** comes into our body, ...
- Student D (70): Yes.

Student E (71): **Macrophage** and **white blood cells** capture and decompose the **antigens** like **viruses**. This is the **innate immunity** system.

Student D (72): This is the most..

Student E (73): This is the most primitive and works a lot in our body. Sounds good [to others]?

Student D (74): Let's go on.

Student E (75): Uh,

Student D (76): Why do not we just follow this [picture in their studied document]?

Student E (77): I agree. So, we see "extinction" here. [pointing on an area in the document]

Student D (78): We do not mind how to say. How about "extermination"?

Student F (79): Shall we copy them? Just like them?

Student D (80): Can we use arrows in writing on the Pad? [typing on an iPad]

Student E (81): We can go from the left to the right. So, this should be the first.

Student D (82): OK, this is the first.

On the contrary to the high learning-outcome group, students in this low learning-outcome group could not sustain their shared epistemic agency during collaboration. In lines #63–#73, they engaged in *creating shared understanding*. They were collaboratively constructing sentences for their explanatory discourse. Their discourse, however, was digressed from their shared epistemic agency by student D's turn (#76 "Why do not we just follow this [picture in their studied document]?"). Student D proposed to transform the epistemic goal of their collaboration into the performance goal, and this proposal was quickly accepted by the other students (#77 and #79). Students D and E attained single-document understanding and F did partially-integrated in the post-test.

## Discussion

SBF assessment revealed that our design of student collaboration was partially successful in facilitating conceptual understanding. Significantly, most students succeeded in acquiring integrated SBF understanding, but more than ten students did not. At the group level, we found large group differences in learning outcomes. Only three groups revealed integrated conceptual understanding among all group members. In the other nine groups, some acquired integrated understanding while others attained single-document understanding or none. These results suggest group differences in collective knowledge advancement and individual differences in engagement in shared epistemic agency.

SSNA clearly demonstrated group differences in the collective knowledge advancement. In the high learning-outcome groups, students quickly engaged in advancing their knowledge and sustained it. In contrast, students in the low learning-outcome groups were slow starters. They did not discuss their ideas by linking one another and did not reach the high level of consolidation of ideas. Our analysis of group discourse in jigsaw activity supported the results from SSNA. For fully integrated SBF students in the high learning-outcome group, the knowledge object was a basis for *generative collaborative actions* by monitoring others' ideas. Neither directly operated the object, but instead monitored the inscription. In contrast, for another partially integrated SBF student, the same knowledge object was a product for *creating shared understanding*. That student was totally responsible for creating the product, and so had to devote mental power toward correctly inscribing ideas through frequent *revoicing*. This difference in knowledge object-oriented actions came from co-regulation and socially shared regulation. Collaboration by two fully integrated SBF students was socially shared, whereas a partially integrated SBF student was co-regulated by another student.

Through examination of student collaboration in jigsaw instruction from the knowledge-creation perspective, we found how students engaged in collective knowledge advancement when successfully acquiring deep conceptual understanding. As the preceding research (e.g., Damşa, 2014) suggests, epistemic agency played a key role in successful collaboration. A new finding suggested in this study is interaction between epistemic agency and multi-layered regulatory processes in collaboration mediated by knowledge objects. For every student to engage in productive collaboration, the knowledge object should be the basis for further epistemic actions. How to produce and share knowledge objects could be further designed from the perspective of regulatory process in collaboration for productively stimulating student epistemic agency.

## References

Bereiter, C. (2002). *Education and mind in the knowledge age*. Hillsdale, NJ: Erlbaum.

- Brown, A. L. & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education*. (pp. 289–325). Mahwah, NJ: Lawrence Erlbaum Associates.
- Damşa, C. I. (2014). The multi-layered nature of small-group learning: Productive interactions in object-oriented collaboration. *International Journal of Computer-Supported Collaborative Learning*, 9, 247–281.
- Damşa, C. I., Kirschner, P. A., Andriessen, J. E. B., Erkens, G., & Sins, P. H. M. (2010). Shared epistemic agency – An empirical study of an emergent construct. *Journal of the Learning Sciences*, 19(2), 143–186.
- de Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for social network analysis. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 87–103.
- Hmelo-Silver, C. & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, 28, 127–138.
- Jacobson, M. J. (2001). Problem solving, cognition, and complex systems: Differences between experts and novices. *Complexity*, 6(2), 1–9.
- Ma, L., Matsuzawa, Y., Chen, B., & Scardamalia, M. (2016). Community Knowledge, Collective Responsibility: The Emergence of Rotating Leadership in Three Knowledge Building Communities. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016*, Volume 1. Singapore: International Society of the Learning Sciences.
- Martinez, A., Dimitriadis, Y., Rubia, B., Gomez, E., & de la Fuente, P. (2003). Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education*, 41(4), 353–368.
- Miyake, N. & Kirschner, P. A. (2013). The social and interactive dimensions of collaborative learning. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences, the Second Edition*. (pp. 418–438). NY: Cambridge Univ. Press.
- Oshima, J., Oshima, R., & Knowledge Forum Japan Research Group. (2007). Complex network theory approach to the assessment on collective knowledge advancement through scientific discourse in CSCL. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Proceedings of CSCL 2007* (pp. 563–565). Mahwah, NJ: Lawrence Erlbaum.
- Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research & Development*, 60, 903–921.
- Paavola, S., & Hakkarainen, K. (2005). The knowledge creation metaphor—An emergent epistemological approach to learning. *Science & Education*, 14, 535–557.
- Reuven, A., Zippy, E., Gilad, R., & Aviva, G. (2003). Network analysis of knowledge construction in asynchronous learning networks. *Journal of Asynchronous Learning Networks*, 7(3), 1–23.
- Schaffer, D. W. et al. (2009). Epistemic network analysis: a Prototype for 21st Century assessment of Learning. *International Journal of Learning and Media*, 1(2), 1–21.
- Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Chicago, IL: Open Court.
- Scardamalia, M., Bransford, J., Kozma, B., & Quellmalz, E. (2012). New assessments and environments for knowledge building. In P. Griffin, B. McGraw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 231–300). New York, NY: Springer Science+Business Media.
- Suthers, D. D., Lund, K., Rosé, C. P., Teplov, C. & Law, N. (Eds.). (2013). *Productive Multivocality in the Analysis of Group Interactions*. New York: Springer.
- Wilensky, U. & Jacobson, M. J. (2013). Complex systems and the learning sciences. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences, the Second Edition*. (pp. 319–338). NY: Cambridge Univ. Press.
- Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge-building communities. *The Journal of the Learning Sciences*, 18, 7–44.

## Acknowledgments

We thank the participating teacher and students. This work was supported by JSPS KAKENHI Grant Number 24240105.

# Technology-Mediated Teacher Noticing: A Goal for Classroom Practice, Tool Design, and Professional Development

Janet Walkoe, University of Maryland, [jwalkoe@umd.edu](mailto:jwalkoe@umd.edu)  
Michelle Wilkerson, University of California, Berkeley, [mwilkers@berkeley.edu](mailto:mwilkers@berkeley.edu)  
Andrew Elby, University of Maryland, [elby@umd.edu](mailto:elby@umd.edu)

**Abstract:** We introduce *technology-mediated teacher noticing* (TMTN): a vision for the design and use of technology-mediated tools that takes seriously the need for teachers to attend to, interpret, and respond to their students' thinking. This vision is situated at the intersection of research on teacher noticing, and on technology to support student thinking. We synthesize that work to highlight specific ways that technology-mediated classroom tools can focus and stabilize teachers' attention on valuable aspects of student thinking emphasized by current reform efforts. We then illustrate TMTN with classroom examples in which technology supported or obstructed teachers' attention to student thinking, and consider implications for research on technology in teacher practice, professional development, and the design of technological tools for K-12 classrooms.

## Objectives

Educational technology has exploded over the past few decades, and many tools have been created to help students think deeply about disciplinary concepts. In many cases these tools have been shown to further student learning, conceptual understanding, efficacy and affect, and motivation. In short, if used productively within a broader classroom culture of inquiry these tools have the power to transform student learning.

Educators often argue that curricular materials are created for teachers as much as for students (e.g., Ball & Cohen, 1996). Yet this idea has not carried over to discussions of technological tools for classroom use. While some research has looked at how such tools can reveal and support student thinking in a given domain, the primary focus has been the students' experience. The teacher's contribution to student learning within such environments is often treated as an afterthought. More research is needed that explores teachers' role in facilitating student learning while using technologically mediated tools, especially in specific content domains; and more attention is needed in educational technology *design* to supporting teachers in noticing, attending, and responding to core disciplinary aspects of students' thinking.

Much work that explores teachers' use of technology (e.g. Mishra & Koehler, 2006) focuses on teachers' use, non-use, or competency with technology in general, or their beliefs about how and when to use technology. Other work explores how teachers design or modify curricula based on feedback from systems (e.g. Kali, McKenney, & Sagy, 2015). However, this focuses on evaluation—whether students answer correctly—and backgrounds the substance of student thinking, which can include productive ideas on which teachers can help students build. In contrast, we are interested in exploring how teachers can use technology-mediated tools to notice new and different aspects of their students' thinking in ways consistent with reform efforts and standards—for example, by supporting teachers' attention to reasoning about mechanism in science (NGSS, 2013), or pattern in mathematics (CCSS-M, 2010). In this paper, using classroom data for illustration, we argue that (i) designers and researchers of educational technology have not foregrounded supporting or studying the teachers' noticing of substance of student thinking as mediated by technological tools, but (ii) such *technology-mediated teacher noticing* (TMTN) should play a role in teacher professional development, research on classroom teaching, and the design of technological tools for classroom use.

## Related work

In mathematics and science instruction, teacher noticing of the disciplinary substance of student thinking is critical for student learning (Schifter, 1998; Franke, Carpenter, Levi, Fennema, 2001; Carpenter, Fennema, Peterson, Chiang & Loef, 1989). Therefore, both teacher professional development and curricular design have aimed to support teacher noticing (Sherin & van Es, 2009). However, teacher noticing has yet to influence the design, study, and implementation of technological tools in the classroom. Here, we briefly review the literature on teacher noticing/teacher responsiveness, and on the design and use of technological tools to support student thinking. Then we investigate the intersection of these two literatures, to situate and inform our notion of technology-mediated teacher noticing.

## Teacher noticing and responsiveness

A growing literature focuses on teachers' noticing of, attention to, and responses to the substance of student thinking (Sherin, Jacobs, & Philipp, 2011). In math and science, researchers and professional developers generally value noticing/attention that seeks to interpret rather than just evaluate students' ideas, and that attends to details of individual students' ideas rather than just general abstractions of "what the class was thinking." However, the seeds of productive disciplinary thinking that teachers can notice and nurture vary by discipline, e.g., productive intuitions about motion and causal reasoning in physical science; precursors to the concept of "variable;" and generalizing patterns from instances in algebra. Partly for this reason, both research and professional development focused on teacher noticing has generally been discipline- and even sub-discipline-specific (e.g., Star & Strickland, 2008). By contrast, work focused on teachers' use of technology explores teachers' general use of technology, not attending to the disciplinary context of its use (Voogt, Fisser, Pareja Roblin, Tondeur & van Braak, 2013).

## Existing tools that focus on student thinking or on teachers' tracking of student progress

A number of tools exist that allow teachers to analyze their students' performance and reflect on curricular and instructional interventions (Rich & Hannafin, 2009). Dashboards and ambient displays provide visualizations of student progress on activities (Clarke & Dede, 2009; Phillips & Popovic, 2012), and help teachers determine where to direct help (Alavi & Dillenbourg, 2012; Börner, Kalz, & Specht, 2011; Slotta, Tissenbaum, & Lui, 2013). Some environments use data mining and analysis to lend insight into student competencies and needs (Gobert, Sao Pedro, Raziuddin, & Baker, 2013), or to guide teachers in assessing student knowledge (mCLASS; Amplify, n.d.). Other technology-mediated tools for classroom use offer supports to guide teachers' attention to student learning (Williams, Linn, Ammon & Gearhart, 2004), and provide data on student performance to inform the adaptation of curriculum (Matuk, Linn, & Eylon, 2015). Though useful for tracking student progress toward correct understandings, these tools do not focus on highlighting the disciplinary substance of individual students' thinking.

Other tools, designed with student users in mind, are intended to amplify reasoning and make thinking visible to peers and researchers. Interactive galleries and collaborative tools allow students to share and build upon one another's work (Scardamalia & Bereiter, 1994). Interactive mathematics environments and scientific modeling tools provide students with new representational systems and modes of interaction for expressing and exploring ideas (e.g. SimCalc; Geogebra; Boxer; NetLogo; a long tradition of research has modeled student reasoning with such tools; Williams, Linn, Ammon, & Gearhart, 2004; diSessa 2001; Papert, 1980; Simpson, Noss & Hoyles, 2005). However, the majority of such work has focused on *student* knowledge and interactions, rather than on teacher practice.

Our interest in *technology-mediated teacher noticing* contributes to this existing work a complementary view of what counts as "successful use" of such technologies by teachers. For most teacher-directed tools, success is marked by successful implementation or improved student performance on activities. For student-directed tools, it is deep engagement with discipline-specific content and practices. What we are interested in is active teacher engagement, within the context of planned classroom activity, to those disciplinary aspects of student thinking that are amplified and made available for observation through the use of technology-mediated tools.

## Theoretical framework and driving questions

We argue there is untapped opportunity for technology to mediate teacher noticing in the classroom. Technology-mediated tools are increasingly a part of classroom practice, and can make student thinking visible by "...afford[ing] a view of the meaning-making process... a screen on which learners can express their thinking... the chance to glimpse the traces of their thought" (Noss & Hoyles, 1996, p. 6). Furthermore, the *types* of student thinking expressed in these media often reflect those that are emphasized by current educational reforms (Table 1) but that teachers often do not elicit and build upon. While technology is not a prerequisite for supporting these types of reasoning, research has shown that certain tools can foreground, stabilize, and highlight them.

Table 1: Examples of technology-mediated tools that emphasize disciplinary thinking in mathematics and science.

Aspect of Reasoning	Example & Related Research	Tools	Connection to Reform Efforts
Dynamicity	Students notice invariant relationships in a geometric construction and work to describe and explain it. (Jones, 2000; Mor et al., 2006)	Geogebra	CCSS-M “Look for and make use of structure”
Use of Linked Representations	Students coordinate information displayed across linked tables, graphs, algebraic expressions, and other representations to confirm/explore their understanding of a relationship. (Smith, diSessa, Roschelle, 1994; Hegedus & Kaput, 2003)	SimCalc, MiGen	NCTM “Select, apply, and translate among mathematical representations to solve problems”
Emphasis on Mechanism	<i>provided in Evidence &amp; Analysis section below</i> (Blikstein & Wilensky, 2009; Sherin, 2001; Wilensky & Reisman, 2006)	NetLogo, Scratch	NGSS “Constructing explanations”
Exploration of Complex Systems	Students conduct investigations of varying systematicity within simulation environments (Hmelo-Silver, Liu, Gray & Jordan, 2014; Jackson, Stratford, Krajcik & Soloway, 1994; Sao Pedro, Gobert, & Betts, 2014)	WISE, PhET	NGSS “Planning and Carrying Out Investigations”

Attending to technology-mediated forms of teacher noticing yields many questions ripe for exploration. For instance, what are features of technology-mediated tools that draw teachers’ attention to specific disciplinary aspects of student thinking important for a given domain of study? How can teachers learn to look for key student thinking practices, such as those outlined in the CCSS-M or NGSS, through the lens of technology-mediated student work? What are mechanisms that can be embedded in technology to allow aspects of student thinking, that otherwise might be hidden, to rise to the forefront?

### Evidence and analysis

Here we present two vignettes that exemplify productive and unproductive instances of TMTN, to illustrate its relevance for research, professional development, and technology design. The first episode comes from a whole-group discussion in a fifth grade science class in an urban rim public school. The school serves a diversity of students with respect to socioeconomic background, ethnicity, and special education status, and the school’s demographics were roughly represented in the classroom from which these vignettes come. The classroom teacher had attended a teacher certification program that was explicitly focused on noticing and responding to student thinking. Students had worked in small groups to create animations and simulations of evaporation. They were now sharing and critiquing their work. In the excerpt below, the classroom teacher encourages students to describe specific computational rules they used in their simulation, and what those rules represent about evaporation as a scientific phenomenon. He connects those rules and interpretations to conversations he observed among student groups earlier during the activity.

- Teacher** What do we think guys? What do we think about this, this simulation, this representation of it? Sheree?
- Sheree** I think it represents when the sun evaporates the water, um the clouds they start to make new ones because of the water vapor.
- Edgar** I think it represents because the water droplets are going up, and then the clouds are getting bigger and bigger because all the water's up, then when it gets full it [gestures down].
- Teacher** Ok, and that's the next step if this simulation were to keep going it would probably show that.
- Miles** I think it's just like the water droplets are going up, and then it's just gonna get bigger and bigger and then it's gonna like start getting ready to-

**Alan** I think they're trying to represent that the water vapor forms new clouds, like more clouds.

**Teacher** I'm even seeing something, I'm trying to remember if this came up in this class or the other class, like, when there's evaporation, and it goes into the air, does it form its own new clouds, or does it add on to the clouds that are already here? So it seems, from what we see here it seems to be adding on to clouds that are already there. That idea was kind of floating around in this room too.

In this excerpt, available functions within the simulation environment such as changing the size of an object or cloning an object focused both teachers' and students' attention on describing potential mechanisms within the represented scientific system (clouds "get bigger" when "full" with water, versus water vapor "forming" clouds by "making more"). These functions lent a shared language to the activity, and allowed the teacher to highlight and connect different student ideas about mechanism.

Our second episode features a small group of students working with the same teacher and tool, this time earlier during the unit to build their simulation of evaporation. However, this time the constraints of the tool blunted conversations about mechanism, focusing the teachers' attention on *what was possible to represent* in the simulation rather than students' ideas about evaporation.

**Ryan** Then when it [water droplet] hits it [cloud], the clouds are gonna like get bigger.

**Teacher** Oh wait sorry, say that again Ryan?

**Ryan** When it hits is, um, it's gonna get bigger

**Teacher** When it hits the cloud, the cloud should get bigger?

**Ryan** Yea. I don't know if we can do that

**Teacher** Yea, that might be, so let's think what's uh

**Luis** No, like when it gets like when it touches the cloud the water droplets like go away.

**Teacher** So they should disappear?

**Luis** [Nods]

**Teacher** So what commands, or sorry what rules do we have to give to this water droplet to have it disappear the way you want it to?

In this case, the teacher's preoccupation with which commands were available to use in the software impeded his noticing and drawing out students' conceptual ideas (clouds "containing" water and droplets being "absorbed" or going away).

We emphasize here that what the teacher is attending to is manifested in the moment and through the technological media. In the first case, the media help make evident the persistence and development of student ideas over time. In the second, noticing of student thinking is obstructed, in favor of attention to practical constraints within the software. In both cases, the teacher must interpret student thinking as mediated by the available tools, and choose what aspects of that thinking to elaborate and act upon.

## Scholarly significance

The work started in this paper helps shed light on the ways technologically-mediated tools can foreground or background student thinking. Moving forward, we will continue to explore cases that help us understand what features of these tools help expose student ideas to teachers and help teachers make sense of these ideas. Ultimately this insight will help inform the development of professional development, classroom tools, and research methods that can support teaching practice in technology-rich spaces.

## References

- Alavi, H. S., & Dillenbourg, P. (2012). An ambient awareness tool for supporting supervised collaborative problem solving. *IEEE Transactions on Learning Technologies*, 5(3), 264-274.
- Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is- Or might be- the role of curriculum materials in teacher learning and instructional reform? *Educational researcher*, 25(9), 6-8.



- Blikstein, P., & Wilensky, U. (2009). An atom is known by the company it keeps: A constructionist learning environment for materials science using Agent-Based Modeling. *International Journal of Computers for Mathematical Learning*, 14, 81-119.
- Börner, D., Kalz, M., & Specht, M. (2011). Thinking outside the box—a vision of ambient learning displays. *International Journal of Technology Enhanced Learning*, 3(6), 627-642.
- Carpenter, T.P., Fennema, E., Peterson, P.L., Chiang, C. & Loef, M. (1989). Using children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, 26, 499-531.
- Clarke, J., & Dede, C. (2009). Design for scalability: A case study of the River City curriculum. *Journal of Science Education and Technology*, 18(4), 353-365.
- diSessa, A. (2000). *Changing minds: Computers, learning, and literacy*. Cambridge, MA: MIT Press.
- Franke, M.L., Carpenter, T.P., Levi, L., Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal*, 38(3), 653- 689.
- Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2013). From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *Journal of the Learning Sciences*, 22(4), 521-563.
- Hegedus, S., & Kaput, J. (2003). *The effect of a SimCalc connected classroom on students' algebraic thinking*. Paper presented at the Psychology in Mathematics Education conference, Honolulu, HI.
- Hmelo-Silver, C. E., Liu, L., Gray, S., & Jordan, R. (2015). Using representational tools to learn about complex systems: A tale of two classrooms. *Journal of Research in Science Teaching*, 52(1), 6-35.
- Jackson, S.L., Stratford, S.J., Krajcik, J., & Soloway, E. (1996). Making dynamic modeling accessible to pre-college science students. *Interactive Learning Environments*, 4, 233-257.
- Jones, K. (2000). Providing a foundation for deductive reasoning: Students' interpretations when using dynamic geometry software and their evolving mathematical explanations. *Educational Studies in Mathematics*, 44(1), 55-85.
- Kali, Y., McKenney, S., & Sagy, O. (Ed.). (2015). Teachers as designers of technology enhanced learning. *Instructional Science*, 43(2), 173-179.
- Matuk, C. F., Linn, M. C., & Eylon, B. S. (2015). Technology to support teachers using evidence from student work to customize technology-enhanced inquiry units. *Instructional Science*, 43(2), 229-257.
- Mishra, P., & Koehler, M. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *The Teachers College Record*, 108(6), 1017-1054.
- Mor, Y., Noss, R., Hoyles, C., Kahn, K., & Simpson, G. (2006). Designing to see and share structure in number sequences. *the International Journal for Technology in Mathematics Education*, 13(2), 65-78.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington, D.C: Authors.
- National Research Council (NRC), (2013), *The next generation science standards*, Washington, D.C.: The National Academies Press.
- [NGSS] Lead States. (2013). *Next Generation Science Standards: For States, By States*. Achieve, Inc.
- Noss, R., & Hoyles, C. (1996). *Windows on mathematical meanings: Learning cultures and computers* (Vol. 17). Springer Science & Business Media.
- Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: Basic Books.
- Phillips, V., & Popović, Z. (2012). More than child's play: Games have potential learning and assessment tools. *Phi Delta Kappan*, 94(2), 26-30.
- Rich, P., & Hannafin, M.J. (2009). Video annotation tools: Technologies to scaffold, structure, and transform teacher reflection. *Journal of Teacher Education*, 60(1), 52-67. Doi:10.1177/002248710832848.
- Roschelle, J., Kaput, J., & Stroup, W. (2000). SimCalc: Accelerating students' engagement with the mathematics of change. *Innovations in science and mathematics education: Advanced designs for technologies of learning*, 47-75.
- Sao Pedro, M.A., Gobert, J.D., & Betts, C.G. (2014). Towards scalable assessment of performance-based skills: Generalizing a detector of systematic science inquiry into a simulation with a complex structure. In the *Proceedings of the 12<sup>th</sup> International Conference on Intelligent Tutoring Systems*. Honolulu, Hi. (pp. 591-600).
- Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences*, 3(3), 265-283.
- Shifter, D. (1998). Learning mathematics for teaching: From a teachers' seminar to the classroom. *Journal of Mathematics Teacher Education*, 1(1), 55-87.

- Sherin, M., Jacobs, V., & Phillip, R. (2011). *Mathematics Teacher Noticing: Seeing Through Teachers' Eyes*. New York, NY: Routledge.
- Sherin, M. G., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60*, 20–37.
- Simpson, G., Hoyles, C., & Noss, R. (2005). Designing a programming-based approach for modelling scientific phenomena. *Journal of Computer Assisted Learning, 21*(2), 143-158.
- Slotta, J. D., Tissenbaum, M., & Lui, M. (2013, April). Orchestrating of complex inquiry: three roles for learning analytics in a smart classroom infrastructure. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 270-274). ACM.
- Smith, J.P., III, diSessa, A.A., & Roschelle, J. (1993/1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences, 3*, 115-164.
- Star, J. R., & Strickland, S. K. (2008). Learning to observe: Using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education, 11*(2), 107–125.
- Voogt, J., Fisser, P., Pareja Roblin, N., Tondeur, J., & van Braak, J. (2013). Technological pedagogical content knowledge—a review of the literature. *Journal of Computer Assisted Learning, 29*(2), 109-121.
- Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories—an embodied modeling approach. *Cognition and Instruction, 24*(2), 171-209.
- Williams, M., Linn, M. C., Ammon, P., & Gearhart, M. (2004). Learning to teach inquiry science in a technology-based environment: A case study. *Journal of Science Education and Technology, 13*(2), 189-206.

# Girls' Interest in Computing: Types and Persistence

Michelle Friend, University of Nebraska Omaha, mefriend@unomaha.edu

**Abstract:** This paper examines interest development through a longitudinal study of young women who had extensive middle school computer science experience. A repeated measures survey was conducted at the end of high school and results compared from the end of middle school to the end of high school. For girls who had developed an interest in a computing career by the end of middle school, interest in computing increased. Aspirational expressions of interest, defined as stating an interest in computing, were highly correlated while embodied expressions of interest, defined as engaging in computing activities such as classes, clubs, or hobbies were generally not correlated. Participants appeared more definite in their attitudes towards computing by the end of high school, particularly interest in computing as a career and college major, than they had at the end of middle school.

## Introduction

It is well-documented that women are underrepresented in computing, as early as middle school through the workforce (NCWIT, 2016). Economists and computer scientists have long made a case that in an increasingly-technological world, there is an increased need for well-prepared high-tech workers; women could fill these jobs. Further, diverse teams create higher quality products (Ashcraft & Breitzman, 2012), speaking to not only a social justice motivation for increasing diversity but also an economic one.

Speculation on the role of early experience in developing students' interest in computing, particularly engaging underrepresented populations such as women, is rampant. On the one hand, it is widely perceived that early experience is important, and so clubs, camps, and programs to introduce children, especially girls, to computing proliferate (e.g. Adams, 2010; Ericson & McKlin, 2012). At the same time, little work has been done to understand the long-term effects of these early experiences. While some programs do engage in high quality program evaluation, longitudinal follow-ups are challenging for informal programs who may have no meaningful way to track students after the program ends. Further, while engaging workshops and camps certainly can stimulate situational interest described in the Four-Phase Model of Interest Development (Hidi & Renninger, 2006), a short experience may not be enough to sustain the transition to individual interest, though it may inspire a student to seek out learning opportunities (Barron, 2006). School environments may be better suited to support interest development due to the opportunity for repeated re-engagement with the topic required in a course.

At a time when the "CSForAll" movement is gaining traction in providing computer science courses in schools across the U.S. it is crucial to understand the implications for interest development. On the one hand, providing computer science classes in all schools can provide access and opportunities for students to discover a new passion (Ainley & Ainley, 2011). On the other hand, one lesson of "school science" is that class experiences disconnected from the inspirational features of "real science" may diminish students' interest (Osborne, Simon, & Collins, 2003). It is critical to understand the long-term implications of compulsory school-based computer science courses on students' long-term interest.

This paper reports on a longitudinal study of girls who attended a middle school where computer science courses were mandatory. Students were initially surveyed at the end of eighth grade, then re-surveyed at the end of high school/beginning of college. This paper is descriptive, examining the changes in students' attitudes about computing, as well as differences in their experiences. It examines the long-term effects of early experience and career interest on engagement and continued interest.

## Context

This study took place in Silicon Valley, among girls who had attended an all-girls school. Silicon Valley, home to Apple, Google, and myriad other tech companies, celebrates computing and technology. Further, this study took place as interest in computing took off, with record enrollment at the college level and increased funding for tech startups regularly making the news. While the stereotype of computing as nerdy and unrewarding persists, the temporal and geographical context of this study, in which computing was seen as financially and socially rewarding, make it unique.

The participants in the study were recruited from a girls' middle school (grades 6-8) where computer science was mandatory for all students in all years. The potential stigma of engagement in traditionally masculine disciplines is removed in a single-sex environment, as everyone doing math, science, and computer

science is a girl. Many of the confounding factors described in other research, such as boys taking over the keyboard or boys dominating class discussions to the detriment of girls are not issues in single-sex environments. Research has demonstrated that the positive effects of single-sex education persist after women return to coeducational settings, as they take on more leadership roles and have higher confidence than their peers who have attended only coeducational schools (Sax, Riggers, & Eagan, 2013).

The school provided female role models, as all computer science teachers during this study were women. One was a young, blonde engineer; one a pierced & tattooed young computer scientist; and one an older MIT graduate whose own daughter had previously attended the school. Further, the computer science curriculum was designed to be engaging. It was a breadth-based approach, including robotics, web design, programming with Python and Scratch, database design, animation, as well as deeply conceptual topics such as ethics, information flow, object-oriented design, etc. The school had a one-to-one laptop program, and thus technology use was spread across all classes, with students using word processing in all classes, Excel and other programs for science data collection and analysis, and various math applications to examine algebraic functions and other topics.

The setting should be ideal for sparking girls' interest in computing – project-based hands-on curriculum, positive stereotype-busting message about computing, female teachers to act as role models, extensive experience and message of competence and mastery, and the larger setting of Silicon Valley where technology careers are celebrated.

## Research questions

This work was guided by the following research questions:

- How does computing interest vary over time between girls who were open to computing careers at the end of middle school and those who were not?
- Are there differences between the expressions of interest as career interest, interest in computing generally, level of engagement in computing experiences, and plans to engage in future computing experiences between the two groups?

## Method

### Setting and participants

This paper reports on the findings from 40 young women who took surveys on computing attitudes and experiences at the end of middle school and again at the end of high school. Participants had attended a private girls' school in Silicon Valley, California, USA, where computer science is a required course for all students all years. This provided a baseline of unusually high computing experience.

Following middle school, the students dispersed to a variety of high schools. Most attended high school at area public ( $n=26$ ) or private ( $n=10$ ) schools. While the effects of high school configurations and offerings were of interest, the population was determined to be too small to make any claims, and this line of inquiry is left to future work.

For this longitudinal follow-up, participants from the first study were asked to participate in a follow-up survey approximately 3.5 to 5 years after the first data collection, during either participants' senior year of high school or freshman year of college. This paper reports on data from the 40 participants who completed surveys at both time points.

### Procedure and instrument

The survey was based on an existing survey on interest, access, and experience with technology (Barron, 2004; Friend, 2015). The survey distributed in eighth grade is lengthy and covers not only topics related to computing but to technology broadly. The middle school survey was distributed by paper in school during the students' final days of eighth grade. For the longitudinal follow-up, a shortened version was distributed through a web form. Most of the questions are repeated measures and were identical in both surveys. In a few cases, questions were included or updated to reflect participants' life position (e.g. middle school or high school) as described below.

### Measures

Measures were repeated between the two surveys, updated as necessary to reflect participants' changed context. In other words, while the middle school survey asked about high school plans, the high school survey asked about college plans. The following is a brief description of each survey construct analyzed below.

### Career interest

One question asked students if they could see themselves becoming “a computer programmer or engineer of some sort” and was used to measure interest in a computing career. This question was used to group the students into a “CS Career” group who were open to a computing career and a “No CS Career” group who were not. The groups were created following the middle school survey: those who indicated they “definitely” or “probably” could not see themselves as a programmer or engineer of some sort were put in the “No CS Career” group and those who indicated they “maybe”, “probably”, or “definitely” could see themselves as a programmer or engineer were in the “CS Career” group. Once assigned to a group, based on the middle school survey, group membership was maintained independent of a participant’s response to the question in the high school survey.

### Computing interest

Four questions measured each student’s express statement of interest in computing: “I would like to learn more about computers”, “Computers are interesting to me”, “Learning about what computers can do is fun”, and “I like the idea of taking computer classes.” Responses were averaged for a single measure of computing interest.

### Future plans

Two measures were developed to investigate participants’ future plans to engage in computing: interest in majoring in computer science, and interest in future learning about computing. To determine interest in a CS **major**, the survey asked to what extent participants could see themselves majoring in computer science in college. Interest in **future learning** was measured through several questions about to what extent participants could see themselves participating in a variety of experiences around computing. These included taking more classes about computers, taking a class on programming or web design in the next step of their academic career, enrolling in a computer summer program, or “learning about” programming, hardware, simulations, or robotics.

### Experiences

Participants were asked how much they had engaged in various computational experiences. In the middle school survey, the question asked how many times they had *ever* done the activities; in the high school survey, the question asked how many times they had done the activities in high school, in order to distinguish continued participation after middle school.

**Classes:** Students were asked how many computer science classes they had taken at school.

**Clubs:** Students were asked how many technology-based clubs such as First robotics they had participated in.

**Hobbies:** Participants were asked how much they had done each of thirteen computational activities. Examples of activities included making a robot, designing a 2-D or 3-D model, or making a web site. In the eighth grade survey, one of the activities was “created your own newsgroup, blog, or discussion site on the internet.” In the high school survey, this question was replaced with “used a makerspace at school or elsewhere” because of the increased prevalence of Maker Spaces in the area, and extent to which creating a newsgroup or blog is no longer a computational activity. “Hobbies” represented the number of activities a participant indicated she had engaged in six or more times. To have engaged in a single activity so many times would be more than required by a class, and therefore represents an act of volition such as joining a club or engaging in the activity in their free time, therefore it is considered a computational hobby.

### Analysis

Following the middle school survey, participants were grouped for analysis based on their response to the question about seeing themselves as a programmer or engineer. The “CS Career” group responded they definitely, probably, or maybe could see themselves as a programmer or engineer in the future. The “No CS Career” group responded that they could probably or definitely *not* see themselves as a programmer or engineer. For the purposes of this paper, participants were kept in the same group. Thus, the “CS Career” grouping represents whether participants had CS career interest *following middle school*.

Repeated measures analysis was used to compare outcomes between the two groups (O’Brien & Kaiser, 1985).

## **Findings**

### **Career interest**

Responses to the question about whether a participant could see herself becoming a programmer or engineer formed the main grouping variable for analysis, as described above. A chi squared test of independence was

highly significant ( $X^2 = 13.7, p < .001$ ), demonstrating that group membership in middle school was strongly associated with group membership in high school. Seven participants had high school responses that would have resulted in a different career group, as summarized in Table 1. The five participants who had been open to a CS career in middle school and could no longer see themselves as a programmer or engineer had all responded “Maybe” on the original study, so not a large change in attitude. Of the two participants who became open to a computing career, one moved from “probably not” to “maybe” while the other showed substantial increase in interest from “probably not” to “definitely yes.”

Table 1: Interest in a programmer/engineering job by time

		Middle School		
		CS Career	No CS Career	Total
High School	CS Career	8	2	<b>10 (25%)</b>
	No CS Career	5	25	<b>30 (75%)</b>
<b>Total</b>		<b>13 (33%)</b>	<b>27 (67%)</b>	

The distribution of responses on the five-point Likert scale are shown in Figure 1. Participants appear more decisive in their feelings about a computing career by the end of high school – the majority who had been neutral to mildly negative had moved into rejecting the idea more completely, while those who were open to a computing career had become more positive about it.

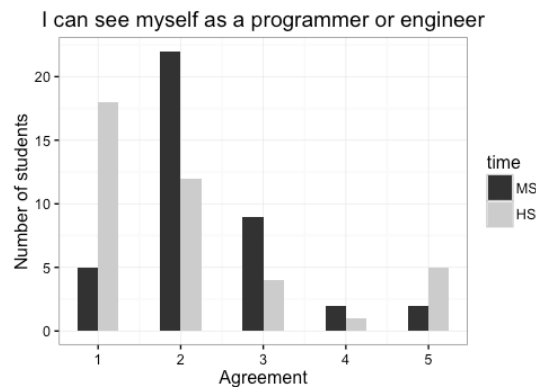


Figure 1. Programmer Career Interest.

### Computing interest

As described above, while the participants were grouped by their openness to a computing career, interest in computing more broadly was also interrogated. “Interest” measures each participant’s response to explicit questions about her interest in computing, such as “learning what computers can do is fun.”

As shown in Figure 2, there is a significant difference in interest between the two groups. A paired t-test on the CS Career group shows an increase in interest between middle and high school ( $t(12) = 2.16, p = .05$ ) while the No CS Career group does not ( $t(26) = .70, p = .49$ ). To further investigate, a repeated measures ANOVA was run, which shows a main effect of group ( $F(1, 38) = 15.59, p < .001$ ) and a main effect of time ( $F(1, 38) = 4.24, p = .046$ ), but no interaction ( $F(1, 38) = 1.52, p = .225$ ). This suggests that girls who were already interested in computing increased interest during high school, but for girls who were not open to a computing career, their interest in computing is relatively unchanged.

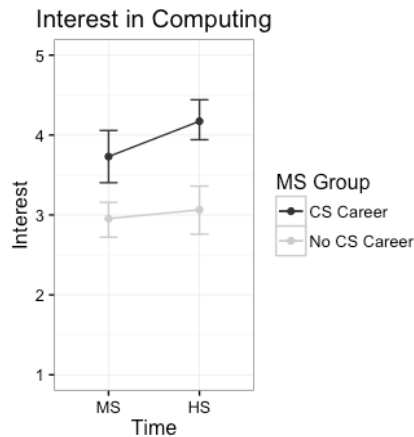


Figure 2. Change in interest by group.

### Future plans

Beyond asking directly how interested participants are in computing, the survey also indirectly measured interest through asking about participants' future plans to engage in computing. Future plans were measured both by asking to what extent participants could see themselves majoring in computer science, and also their plans to participate in computing in less formal ways, such as taking classes and learning independently.

Participants in the CS Career group were more interested in majoring in CS (HS  $M=2.85$ ,  $SD=1.46$ ) than participants in the No CS Career group (HS  $M=1.56$ ,  $SD=.89$ ), a trend that did not change over time: repeated measures ANOVA shows a main effect of group ( $F(1, 38) = 34.88$ ,  $p < .001$ ) but no effect of time or interaction between time and group. There was high variance in the responses, particularly in the CS Career group. Figure 4 disaggregates the CS Career group responses and suggests that individuals are becoming more decisive in their consideration of a CS major, whether positive or negative.

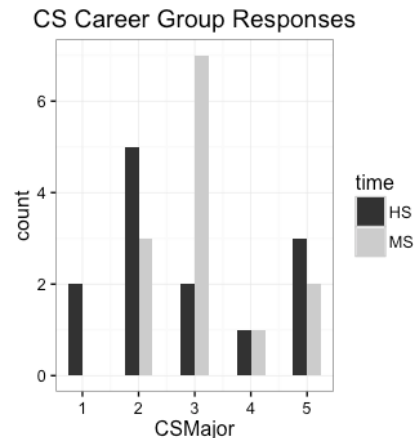


Figure 3. Openness to a CS Major for CS Career group.

In contrast to participants' inclination to *major* in computer science, their plans to continue learning more about computing and computer science continues to be high. Repeated measures ANOVA shows a main effect of group ( $F(1, 37) = 24.86$ ,  $p < .001$ ) but no interaction nor main effect of time. In other words, participants in the CS Career group have more intention to continue learning about computing than their No CS Career peers both in middle school and in high school, and neither group changes intention over time.

### Experiences

While stating that one is interested in a topic, even interested in pursuing it in the future, is one measure of interest, another important factor in interest development is the enactment of that interest through engaging in activities related to the interest. The high school survey inquired into whether students had taken CS classes or joined clubs relating to computing. The middle school survey had not asked about these topics because all students were required to take computer science classes all years and few clubs were offered due to the small

size. Both surveys inquired into the depth of participants experience with a variety of computational activities such as creating web sites, digital art, and programming.

Two participants in each group had taken three or more CS classes in high school. While most participants in the CS Career group *had* taken a CS class ( $n=8, 62\%$ ), most participants in the No CS Career group had *not* taken CS in high school ( $n=19, 70\%$ ). However, a t-test comparing the groups was not significant ( $t(20)=1.5, p=.14$ ).

The results were similar in terms of joining computing clubs. The vast majority of participants in the No CS Career group had not joined any computing clubs in high school ( $n=24, 89\%$ ), while a smaller majority of the CS Career group had not joined computing clubs ( $n=7, 54\%$ ). As before, a t-test indicated no significant difference between the groups' on joining computing clubs ( $t(15.9) = 1.79, p=.09$ ).

Participants were asked about how often they had engaged in each of thirteen activities, and the number of activities where they indicated they had re-engaged more than six times were counted and considered a digital hobby. For example, a student who had created more than six web sites, more than six pieces of digital art, and more than six robots would have three digital hobbies.

There was incredible variation in the number of participants' digital hobbies, with few trends. Repeated measures ANOVA showed no main effect of time or group and no interaction. A histogram of the responses is shown in Figure 4. In both groups, many participants have fewer hobbies in high school than they did in middle school. Although on average the number of hobbies demonstrated by participants in the CS Career group decreases from middle school to high school, the number of participants reporting *no* hobbies decreases, from four to one. It would be expected that participants who were interested in a computing career would have computational hobbies. By contrast, it would be expected that participants who were uninterested in a computing career may not have computational hobbies. Thus, it is notable that a substantial number of participants did have digital hobbies. Of particular notice is the number of people who had four or more hobbies in middle school – including the participant who had seven hobbies.

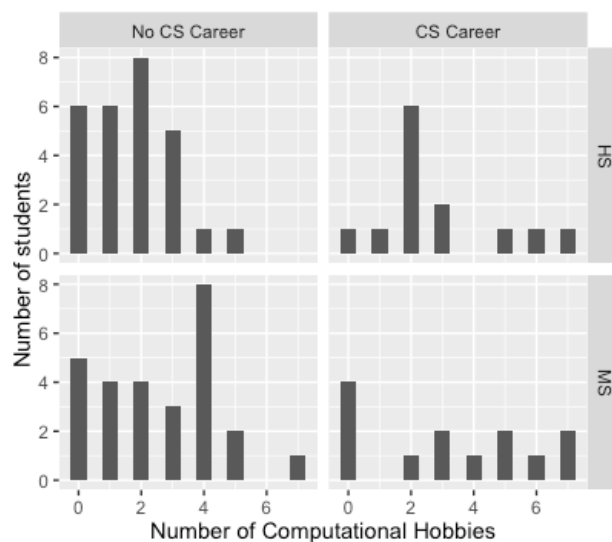


Figure 4. Number of computational hobbies by group and time.

### Relationships between expressions of interest

In order to understand the relationships between the different expressions of interest, not just through the lens of career interest, but across all expressions described above, correlations were calculated, as shown in Table 2. These are correlations of the results from the high school survey.

Table 2: Relationship between variables

	Career	Interest	Major	Learning	Classes	Clubs
Comp. Interest	.58***					
CS Major	.76***	.59***				
Future Learning	.73***	.79***	.78***			



HS CS Classes	.36*	.06	.09	.00		
Clubs	.50**	.27	.30	.41**	.29	
Hobbies	.36*	.14	.18	.14	.47**	.2

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

Aspirational expressions of interest – defined as being open to a computing career, expressing interest in computing, being open to a CS major, and being interested in learning more about computing in the future - were all very highly correlated. Embodied expressions of interest – defined as engaging in activities that would demonstrate that interest: taking classes, joining clubs, or having computational hobbies - were generally not strongly correlated with aspirational expressions of interest. One exception is computing career interest, which was moderately correlated with the embodied outcomes.

## Discussion

To the extent that “computing interest” can be broadly conceptualized, the prediction would be that students who express interest in computing would display that interest broadly, that they would not only agree with statements such as “I am interested in computing” but would also engage in computing activities and would be open to future opportunities. The results discussed here present a more complex picture of interest, in which aspirational expressions of interest – stating that one is interested or would be willing to engage in the future – are markedly different than embodied expressions of interest – actually engaging in the activity, whether by taking courses, joining clubs, or engaging in computing as a hobby.

Within the aspirational expressions of interest, one notable result is the change in computing interest expressed by the CS Career group, as shown in Figure 2. Not only did members of the group have a higher interest in computing generally at the end of middle school, but their interest continues to grow in high school – independent of whether they engaged in computing activities. It appears that once girls are “hooked” on computing, their interest may continue to increase even if they do not continue to engage in extensive computing activities. This could be seen as an expression of individual interest, and that once participants had developed individual interest their interest deepened from emerging to a more well-developed individual interest during high school. Further investigation is warranted, both through investigating how participants understand and conceptualize their interest in computing, but also whether this is an operationalization of individual interest, such as identifying the relationship between participants’ identity with computing and expressions of interest.

In terms of attitudes, participants appeared to become more decisive about their future plans during high school. While a substantial number of participants were “maybe” or “probably not” open to a computing career at the end of middle school, by the end of high school a much larger number were definite in their response (see Figure 1). A similar trend occurred with participants’ response to majoring in computer science (see Figure 3). This is not entirely surprising, as younger students may be open to more possible futures than adolescents who are close to having to choose a path (Eccles, 2007). The finding that participants who were more positive about computing careers at the end of middle school are more positive about majoring in CS at both times echoes Tai, Qi Liu, Maltese & Fan (Tai, Qi Liu, Maltese, & Fan, 2006) and suggests that despite their mixed engagement in computing experiences in high school they may go on to major in computing areas in college and even continue to computing careers.

The mixed results in terms of participants’ high school experiences warrants further investigation. One open question relates to the participants in the No CS Career group with high experience. The participants who took many courses, had many hobbies, and joined computing clubs yet reject a future in computing are of interest, to understand their choices. It is notable that this study took place in Silicon Valley, where computing is prevalent and privileged. Further their middle school computer science experience included messages about the utility of computing; these participants may have seen computing experiences as a vehicle to other goals. Follow-up study is required to understand the meaning of this finding.

Perhaps the greatest area for future research is not explanatory, but one of generalizability. The sample in this study is not only small but also quite unique, both in the general setting of Silicon Valley and also in the particulars of graduates of a girls’ school with a mandatory computer science curriculum. Future research would not only expand to schools in other settings outside of Silicon Valley but also other configurations of schools. As an increasing number of schools are offering and even requiring computer science courses, finding schools and students with high CS experience should become easier, allowing for comparison and generalization of the findings.



# Educational Technology Support for Collaborative Learning With Multiple Visual Representations in Chemistry

Martina A. Rau, Educational Psychology, University of Wisconsin – Madison, marau@wisc.edu  
Sally P. W. Wu, Educational Psychology, University of Wisconsin – Madison, pwwu@wisc.edu

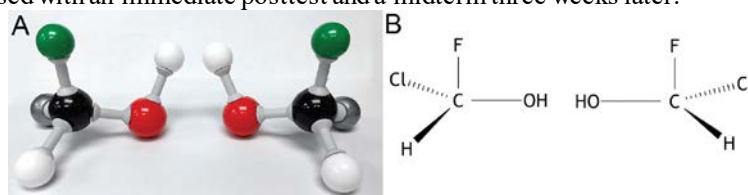
**Abstract:** Educational technologies have two features that can enhance collaborative learning. First, they can provide collaboration scripts that adaptively react to student actions and prompt them to engage in effective collaborative behaviors. Second, collaboration often involves multiple visual representations. But many students have difficulties in making sense of representations. Educational technologies can support students in doing so by adapting to how they construct, interpret, and connect representations. We conducted a quasi-experiment with 61 undergraduate chemistry students to test the effectiveness of an adaptive collaboration script that prompts students to discuss visual representations. A control condition collaboratively solved worksheet problems with multiple visual representations without a collaboration script. An experimental condition solved the same problems using an educational technology with the script. The experimental condition showed significantly higher learning gains on a transfer posttest and on complex questions on a midterm exam three weeks later.

## Introduction

Educational technologies play an increasingly important role in undergraduate instruction in science, technology, engineering, and math (STEM) (Freeman et al., 2014). One reason for this trend is that practice guides recommend engaging students in authentic problem-solving activities to help them reason about concepts in the same way as experts do (NRC, 2006). Educational technologies offer two key features that may make them particularly effective platforms for such problem-solving activities. First, because experts often solve problems collaboratively (Kozma, Chin, Russell, & Marx, 2000), STEM instruction often involves collaborative activities (Freeman et al., 2014). Educational technologies can provide adaptive support for collaboration, for example by providing collaboration scripts that adapt to student needs (Walker, Rummel, & Koedinger, 2009). Second, experts often use multiple visual representations to solve problems (Kozma et al., 2000). Therefore, STEM instruction often asks students to do the same. For example, chemistry students may collaboratively construct, interpret, and connect ball-and-stick models (Figure 1A) and wedge-dash structures (Figure 1B) when they learn about isomers (i.e., chemical compounds made of the same atoms that differ only in the spatial arrangement of their atoms, which can have dramatic effects on the properties of chemical compounds). Educational technologies can provide adaptive support for learning with visual representations, for example by grading student-generated representations automatically, by providing real-time feedback on students' interpretations of the representations, and by prompting them to connect multiple representations (Rau, 2016a; Seufert, 2003).

Consequently, combining adaptive support for collaboration with adaptive support for using visual representations may significantly enhance students' learning of content knowledge. The following brief review of prior research shows that this question remains open because (1) research on adaptive collaboration scripts has not focused on supporting students in making sense of visual representations, while (2) research on learning with visual representations has mostly focused on individual learning.

To address this limitation, we conducted a quasi-experiment within a 3-hour lab session in an undergraduate chemistry course. A control condition worked on a traditional version of an activity about isomers. Students collaboratively constructed ball-and-stick models (see Figure 1A) and drew wedge-dash structures (see Figure 1B) on a worksheet. Students in the experimental condition worked on the same activity, except that they drew wedge-dash structures using an educational technology that incorporated an adaptive collaboration script. The script prompted students to collaboratively discuss mistakes they made in their drawings. We tested effects on learning gains assessed with an immediate posttest and a midterm three weeks later.



**Figure 1.** Physical ball-and-stick model (A) and wedge-dash structure (B). Each shows two chlorofluoromethanol isomers that have the same molecular formula but different 3d arrangement of the atoms.

## Adaptive collaboration scripts

Collaboration can significantly enhance students' learning, but it is not always effective (Lou, Abrami, & d'Apollonia, 2001). The effectiveness of collaborative activities depends on the quality of interactions among students. They need to actively co-construct meaning, for instance by discussing divergent views and sharing information rather than splitting the work (Miyake & Kirschner, 2014). Students often fail to spontaneously engage in effective collaborative behaviors (Lou et al., 2001).

Collaboration scripts provide an effective means to support collaboration by suggesting sequences of interactions (e.g., analyze the problem, critique partner's analysis, respond to critiques), posing questions for students to discuss (e.g., do you understand the problem?), or prompting them to engage in particular behaviors (e.g., ask your partner to explain the rationale for the solution). Such collaboration scripts can significantly improve the quality of students' collaboration (Fischer, Kollar, Stegmann, & Wecker, 2013). However, results on students' learning of content knowledge are mixed. Several studies found null effects on content knowledge—even if collaboration quality was improved (e.g., Stegmann, Weinberger, & Fischer, 2007; Walker et al., 2009).

The lack of evidence for the effectiveness of collaboration scripts for learning of content knowledge has been attributed to the fact that they do not adapt to students' needs for support. That is, scripts may provide too much or too little support, or support at the wrong time (Rummel, Walker, & Aleven, 2016). Inadequate support can have negative effects on students' affect because they may perceive it as annoying or distracting (Rummel et al., 2016). In contrast, human instructors adapt the amount, timing, and type of support to students' state (e.g., current knowledge level) (Gweon, Rose, Carey, & Zaiss, 2006).

Educational technologies can make adaptive collaboration support scalable by tailoring collaboration scripts to the students' needs (Walker et al., 2009). At a technical level, adaptation is achieved by computational model that detects the students' needs in real time and formalizes the procedure for tailoring support to these needs. For example, the model may infer the students' current knowledge level from their action (e.g., an answer to a problem). Based on the inferred knowledge level, the model can dynamically adjust the amount, timing, and type of support the collaboration script provides (Magnisalis, Demetriadis, & Karakostas, 2011).

Thus far, evidence for the effectiveness of adaptive collaboration scripts for students' learning of content knowledge is mixed (Magnisalis et al., 2011). While some studies show that adaptive collaboration scripts enhance students' learning of content knowledge (Karakostas & Demetriadis, 2011), several studies have failed to show that activities with adaptive collaboration scripts are more effective compared to activities with non-adaptive collaboration scripts and compared to individual learning (e.g., Walker et al., 2009). We are not aware of studies that compared adaptive collaboration scripts to collaborative activities without scripts.

## Support for learning with visual representations

Many collaborative activities involve visual representations. Indeed, visual representations and collaborative activities may mutually enhance one another. On the one hand, visual representations can enhance the quality of collaboration. Visual representations allow students to externalize their reasoning, which can reduce cognitive load in the group (Kirschner, Paas, & Kirschner, 2010). Further, externalizing reasoning through visual representations can help the group reach a consensus about how to explain a complex concept or how to solve a task (Suthers & Hundhausen, 2003). On the other hand, collaboration can enhance students' ability to make sense of visual representations. When working individually, students often fail to spontaneously reflect on their understanding of visual representations (Ainsworth, Bibby, & Wood, 2002). When students collaborate with visual representations, they may realize that they hold divergent views on how to interpret, construct, or connect visual representations. This, in turn, may prompt students to engage more deeply in making sense of the representations (Gnesdilow, Bopardikar, Sullivan, & Puntambekar, 2010).

Helping students make sense of visual representations is a key goal of STEM instruction (Ainsworth, 2008; NRC, 2006). Because any individual visual representation shows only a particular aspect of the concepts, instruction typically uses *multiple* visual representations that depict complementary information (Ainsworth, 2008). Besides understanding how each representation depicts information, students need to make connections among the different representations to integrate this information into a coherent mental model (Rau, 2016a). Connection making is a major stumbling block that interferes with students' learning of content knowledge in many STEM domains (Ainsworth, 2008). For example, in chemistry, failure to make connections among representations can yield misconceptions that interfere with learning of crucial concepts (De Jong & Taber, 2014). In the example in Figure 1, if students fail to understand that the wedge-dash structure on the left is *not identical* to the ball-and-stick model on the right, they may incorrectly infer that the melting point of a sample that contains both isomers is equal to the melting point of a sample that contains only one of the isomers.

Much research shows that educational technologies can enhance students' learning of content knowledge by helping them make sense of visual representations (e.g., Ainsworth, 2008). Effective technology-based support

typically provides real-time feedback on student-generated visual representations (Rau, 2016b), asks them to map representations to concepts (Seufert, 2003), and prompts them to explain connections between representations (Rau, 2016b). Experiments show that such technology-based support can enhance students' learning of content knowledge compared to educational technologies without such support (Seufert, 2003).

Two limitations of research on learning with visual representations need to be addressed. First, the effectiveness of technology-based support over traditional activities with visual representations remains to be shown. We are not aware of a study that has systematically compared technology support for sense making of representations to traditional activities without an educational technology. Second, prior research has mostly focused on individual students in using visual representations. This stands in contrast to the fact that visual representations are often used collaboratively for problem solving in STEM instruction, as discussed above.

## Research question

In sum, educational technologies can enhance learning by prompting *individual sense making* of visual representations and by *scripting collaboration*. Prior research has not investigated whether an educational technology can enhance learning by prompting *collaborative sense making* of visual representations. Further, research has not compared educational technology support for visual representations or for collaboration to traditional activities without technology support. Therefore, we investigate the following question: Does a technology-based adaptive collaboration script that prompts students to collaboratively make sense of visual representations enhance learning of content knowledge?

## Methods

### Participants and setting

To address this question, we conducted a quasi-experiment with 69 students in an undergraduate chemistry course at a university in the U.S. Midwest. The course involved two weekly 50-minute lectures, two weekly 50-minute discussion sessions, and one weekly 3-hour lab session. The lecture was attended by all students. Lab and discussion sessions were held in smaller sections; namely four sections of about 18 students each. The lab and discussion sessions were led by two teaching assistants (TAs) who went through the same training program at the beginning of the semester. During the semester, students worked in small groups of 2-3 students during discussion and lab sessions. Our quasi-experiment took place in the lab session in week 5 of the semester.

### Experimental design

We assigned two of the four lab sections of the course to the control condition ( $n = 37$  students) and two to the experimental condition ( $n = 32$  students). Students selected lab sections at the beginning of the semester so that they fit well into their class schedule. We do not have any reason to believe that systematic differences exist between sections. In addition, we took the following steps to ensure equivalency of the conditions. To counterbalance potential effects of class period, each control session was held concurrently with an experimental session. To counterbalance TA effects, each TA led one control and one experimental session. We also counterbalanced the sequence in which the TAs led control and experimental sessions. Both conditions worked on problems collaboratively in the same small groups as in discussion and lab sessions throughout the semester.

#### Control condition

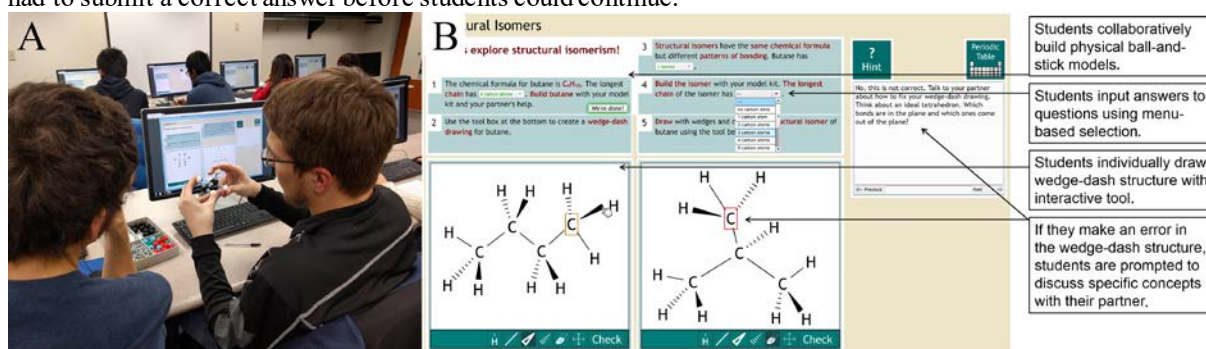
The control condition received the traditional version of the problem-solving activities: a worksheet that consisted of ten multi-step problems about isomers. In each problem, students had to construct physical ball-and-stick models that represent specific molecules. Students worked on this step collaboratively, using a shared modeling kit to construct these models. After constructing each model, they had to draw a wedge-dash structure of the same molecule. Students drew the structures individually on their own worksheet, but they were encouraged to consult with their partner. Each activity also required students to answer conceptual questions about the molecule. Students wrote down their answers individually, again while being encouraged to consult with their partner. At the end of the 3-hour lab session, students handed their worksheets to the TAs who provided written feedback on the problem solutions and on the wedge-dash drawings in the following week's lab session.

#### Experimental condition

The experimental condition received the technology-enhanced version of the same problems. To ensure equivalency to the worksheet version, the technology-enhanced problems contained the same steps, the same conceptual questions, and the same molecules. Problems were presented in the same order and required students to build the same physical ball-and-stick models. TAs led the sessions in the same way as for the control sessions (e.g., they

were available answer questions about the problems). The difference to the control condition was that problems were presented and answered within an educational technology, shown in Figure 2. Students used the educational technology to draw wedge-dash structures and to answer conceptual questions via mouse and keyboard. The technology incorporated an adaptive collaboration script that prompted students to discuss specific concepts when they made a mistake in their wedge-dash drawing. At a technical level, the script used a computational model that detects conceptual errors students often make when drawing a wedge-dash structure or answering conceptual questions. When the computational model identified an error and a misconception that may have led to this error, the educational technology highlighted the feature of the wedge-dash structure that students had drawn incorrectly and prompted students to discuss the concept with their partners while using the ball-and-stick model.

In sum, the only difference between experimental and control conditions was that students in the experimental condition drew wedge-dash structures using an educational technology with an adaptive collaboration script. The script changed the nature of the collaboration in several ways. First, the timing of feedback differed: while the control condition received written feedback on their worksheets in the following week, the experimental condition received immediate feedback from the technology. Second, the form of feedback differed: while the control condition received only correctness feedback, the experimental condition received feedback in the form of collaboration prompts to discuss concepts that students may have misunderstood. Third, the consequentiality of feedback differed: while the control condition did not have to revise their answers, the experimental condition had to submit a correct answer before students could continue.



**Figure 2.** Students in the experimental condition built physical ball-and-stick models (A) and drew wedge-dash structures in an educational technology (B).

## Assessments

To assess students' learning of content knowledge, we created a *pretest* and *posttest* on isomerism concepts. The test had two scales. The *reproduction* scale had six multiple-choice items that assessed students' ability to recall and understand the concepts (i.e., levels 1 and 2 of Bloom's taxonomy, as defined by Anderson & Krathwohl, 2001). The *transfer* scale had four multiple-choice items that assessed students' ability to apply and analyze the concepts (i.e., levels 3 and 4 of Bloom's taxonomy). Hence, the reproduction scale assessed simple concepts; the transfer scale assessed complex concepts. Two versions of the test were counterbalanced across pretest and posttest. The tests were optional, but students received course credit for completing them.

To assess students' long-term retention of content knowledge, we used data from two exams that were provided as part of the course. A *pre exam* in the second week of the semester assessed students' prior understanding of chemistry concepts that they may be expected to have covered in high school courses. A *midterm exam* in the eighth week of the semester (i.e., three weeks after the experiment) assessed students' understanding of the chemistry concepts covered in the course thus far. We focused on one question on the midterm exam that assessed the isomerism concepts covered in the lab session in which we conducted our quasi-experiment. This *isomerism question* was one of five advanced questions on the midterm exam, and students had to choose three of these five advanced questions. This question asked students to draw wedge-dash structures and to transfer their knowledge about isomers to novel tasks. We coded students' responses to this question by giving points for each of 20 aspects that were correctly drawn. In addition, we coded for errors that indicated students' difficulties in remembering the target chemistry concepts (level 1 in Bloom's taxonomy), to understand and apply the concepts (level 2 and 3 in Bloom's taxonomy), to analyze and evaluate the concepts (levels 4 and 5 in Bloom's taxonomy), and to make novel inferences (level 6 in Bloom's taxonomy).

## Procedure

Figure 3 shows how the experiment aligned with course activities in the entire semester (i.e., two weekly 50-minute lectures, two weekly 50-minute discussion sessions, and one weekly 3-hour lab session). In the second

week of the semester, students took a pre exam. A lecture in the fourth week of the semester covered stereoisomerism and related concepts. Our experiment took place in the fifth week. The pretest was made available online three days prior to the lab. Up to this point, all course activities were identical for students in the control and experimental conditions. Then, students attended the version of the 3-hour lab session that corresponded to their condition. All following activities were again identical for both conditions. On the following day, the posttest was made available online for three days. The following discussion and lecture sessions did not focus on isomers. The midterm exam was given in the eighth week of the semester.

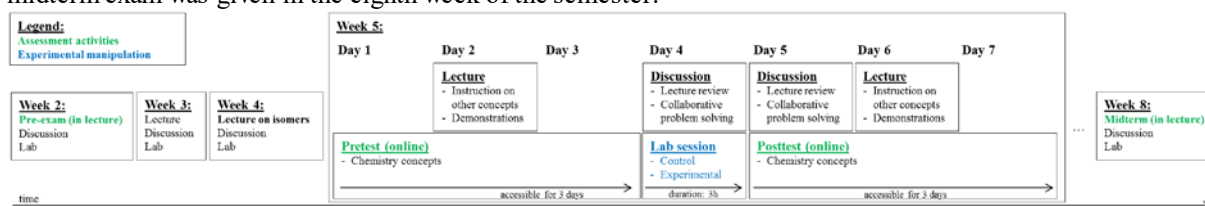


Figure 3. Timeline of assessment (green) and experimental manipulation (blue) in the chemistry course.

## Results

### Prior checks

As mentioned, students were free to choose whether or not to complete the pretest and posttest for extra course credit, and whether to choose the isomerism question on the midterm exam. Therefore, we first tested for differences between students who chose to complete the tests to those who did not. Eight students did not complete the pretest and posttest, yielding  $N = 61$  for these analyses ( $n = 30$  in the control condition,  $n = 31$  in the experimental condition). Students who did not choose to complete the pretest and posttest did not differ from included students on their pre-exam scores ( $F < 1$ ). Forty students chose to complete the isomerism question ( $n = 20$  in the control condition,  $n = 20$  in the experimental condition). Students who did not choose this question did not differ from students who chose it on reproduction pretest,  $F(1, 55) = 2.647, p = .109$ , or transfer pretest ( $F < 1$ ), but had significantly lower pre-exam scores,  $F(1, 55) = 4.383, p = .031, p. \eta^2 = .074$ .<sup>1</sup>

Because we developed the pretest and posttest specifically for this experiment, they had not been evaluated. Therefore, we conducted a factor analysis to evaluate the separation of the reproduction and transfer scales. A factor analysis showed that a two-factor model that separates the reproduction and transfer scales had a better model fit than a one-factor model. A reliability analysis showed that the reproduction scale had poor reliability (Cronbach's  $\alpha = .525$ ), whereas the transfer scale had good reliability (Cronbach's  $\alpha = .851$ ).

Next, we tested for differences between conditions prior to the experiment. There were no significant differences on pre exam ( $F < 1$ ), reproduction pretest,  $F(1, 59) = 1.190, p = .280$ , or transfer pretest ( $F < 1$ ).

Finally, we tested whether students' understanding of isomerism improved as a result of the interventions. To this end, we used a repeated-measures ANOVA with test time (i.e., pretest and posttest) as the repeated within-subjects factor. Pre-exam scores were not a significant predictor and were hence not used in this analysis. There was no significant effect of test time on the reproduction test ( $F < 1$ ). There was a significant effect of test time on the transfer test,  $F(1, 59) = 8.776, p = .004, p. \eta^2 = .128$ , showing that students' ability to transfer knowledge about isomers to novel tasks improved significantly from pretest to posttest.

### Differences between conditions on learning outcomes

To test whether the adaptive collaboration script enhanced learning of content knowledge, we used an ANCOVA with condition as independent factor, scores on the reproduction posttest and transfer posttest as dependent measures, and scores on the respective pretests as covariate. The pre exam was not included because it was not a significant predictor. Figure 4 shows the estimated marginal means on the posttests that control for pretest. There was no significant effect of condition on the reproduction posttest ( $F < 1$ ), suggesting that the adaptive collaboration script did not enhance knowledge reproduction. There was a significant effect of condition on the transfer posttest,  $F(1, 59) = 4.256, p = .044, p. \eta^2 = .068$ , such that the experimental condition outperformed the control condition. This suggests that the adaptive collaboration script enhanced knowledge transfer.

Next, we tested the effect of condition on overall midterm exam scores using an ANCOVA with condition as the independent factor, scores on the midterm exam as dependent measure, and scores on the pre exam as the covariate. We included scores on the pre exam as a covariate in this model because they were a significant

<sup>1</sup> We report effect sizes using  $p. \eta^2$ :  $p. \eta^2$  of .01 corresponds to small, .06 to medium, and .14 to large effects.

predictor of students' midterm exam scores. The reproduction pretest and transfer pretest were not included because they were not significant predictors. Results revealed no significant differences on the overall midterm exam scores ( $F < 1$ ). Using the same ANCOVA model to test for differences on the isomerism question for the 40 students who chose this question, we found no differences between conditions on this question ( $F < 1$ ).

A more fine-grained assessment was provided by the errors on the isomerism question, which indicated difficulties in using the isomerism concepts with respect to Bloom's taxonomy levels 1 (remember), 2-3 (understand/apply), 4-5 (analyze/evaluate), and 6 (novel inferences). The same ANCOVA model showed no effects of condition on level 1-5 errors ( $F_s < 1$ ), suggesting that the adaptive collaboration script did not enhance students' learning of concepts of simple to medium complexity. There was a significant effect on level-6 errors,  $F(1, 33) = 4.272, p = .047, \eta^2 = .115$ , such that the control condition made more level-6 errors (i.e., difficulties in making inferences about complex concepts). This result suggests that the adaptive collaboration script enhanced students' learning of complex concepts and that this effect persisted three weeks after our quasi-experiment.

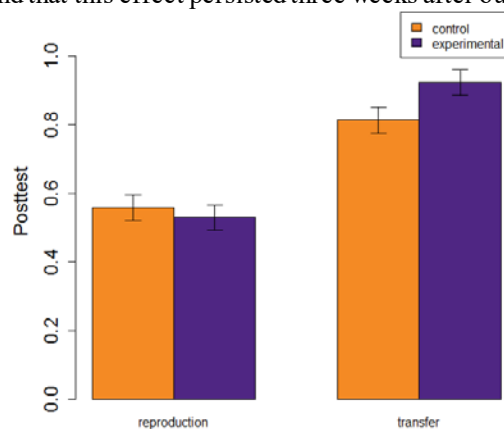


Figure 4. Estimated marginal means for control condition (orange) and experimental condition (purple) on reproduction and transfer posttest, controlling for pretest. Error bars show standard errors of the mean.

## Discussion

We conducted a quasi-experiment to test whether an adaptive collaboration script can enhance students' learning of content knowledge from problems that involve connection making among visual representations. Results on students' learning outcomes show a medium-size advantage of the adaptive collaboration script on the transfer posttest over the traditional worksheet version of the same activity. There were no effects on students' scores on the reproduction posttest. There were no effects on overall midterm exam scores or on the isomerism question on the midterm exam three weeks after our experiment. Yet, a fine-grained analysis of the isomerism question showed a medium-sized reduction of errors for the experimental condition on questions that required students to make novel inferences based on complex concepts. This suggests that an adaptive collaboration script can enhance learning with visual representations, but that this effect is confined to complex concepts.

These findings extend prior research on *individual sense making* of visual representations. There is abundant evidence that connection making among visual representations is a difficult but crucial mechanism through which students acquire content knowledge. There is also abundant evidence that educational technologies can enhance students' learning of content knowledge by helping them make sense of the connections. Even though much prior research suggests that collaboration can enhance students' connection making, a limitation of this research is that it has focused mostly on individual rather than collaborative learning. Our findings provide a first affirmation that prompting students to collaboratively make sense of connections when they encounter difficulties in making connections can enhance their learning of content knowledge.

Our findings also extend research on *collaborative learning*. Even though effects of collaboration scripts on the quality of students' collaboration are well established, few studies have found effects on learning of content knowledge. We show that an adaptive collaboration script can significantly enhance learning of content knowledge, compared to a traditional version of the same problems without a collaboration script. Specifically, adaptive collaboration scripts that focus students' collaboration on connection making among visual representations when they struggle with the connections may be effective.

We found effects on complex concepts (i.e., the transfer scale of the posttest and on level-6 concepts on the isomerism question on the midterm exam) but not on simpler concepts (i.e., the reproduction scale of the posttest and lower-level concepts on the isomerism question). The fact that we did not find effects on overall midterm exam scores is not surprising because the midterm exam contained questions about all content covered



up to the midterm, and not just on the content covered in the lab session in which we situated our quasi-experiment. The null effects on the reproduction scale of the test may result from the fact that we did not see significant learning gains on this test, which may in turn result from poor reliability of this scale. In future research, we plan to revise the reproduction scale of the test.

The fact that we found effects on the scales that assessed complex concepts can be interpreted in light of research on sense making of visual representations. Integrating information from multiple visual representations is more important for learning of complex concepts than for simple concepts. For this reason, it seems plausible that students are more likely to make mistakes when connection making involves complex concepts. Further, they may be more likely to hold divergent views on complex concepts. Hence, collaboration that yields deeper engagement in connection-making processes may pay off more for complex than for simple concepts.

The finding that effects of the adaptive collaboration script are confined to complex concepts can also be interpreted in light of research on collaborative learning. Discussing complex concepts is cognitively demanding. External representations can be used to off-load these cognitive demands (Kirschner et al., 2010). Hence, prompting students to focus collaborative interactions on the visual representations may benefit their learning of complex concepts more so than their learning of simple concepts. Thus, if complex concepts require connection making more so than simple concepts and if collaboration can help students make these connections, we expect adaptive collaboration scripts to be more effective for complex than for simple concepts.

## Limitations

Our findings should be interpreted in the context of the following limitations. First, quasi-experimental designs provide less stringent causal evidence than randomized control trials. Even though we found no differences between conditions prior to the experiment and took steps to ensure equivalency of conditions, unmeasured differences may have affected the results. Hence, a randomized control trial should replicate the results.

Second, while most students completed the pretest and posttest, eight students did not. Even though we did not find differences between these students, it is possible that they differed in unmeasured aspects. Further, students who chose not to complete the isomerism question had lower pre-exam scores, so we do not know whether findings on this question generalize to students with low prior knowledge. We suggest that future research should replicate our findings in a setting that allows for compulsory testing.

Third, our quasi-experiment investigated whether a carefully designed educational technology that contains an adaptive collaboration script is more effective than a traditional version of the same activity. We did not attempt to compare collaborative to individual learning and hence cannot conclude that adaptive collaboration scripts are more effective than individual learning with or without the technology. Likewise, we did not aim at comparing an educational technology with an adaptive collaboration script to an educational technology without a script. Similarly, we did not compare non-adaptive scripts to adaptive scripts. Therefore, we cannot conclude that an adaptive collaboration script enhances the effectiveness of educational technologies. Finally, we did not compare different versions of adaptive collaboration scripts. Hence, we cannot conclude that scripts that adapt to connection making are more effective than scripts that adapt to other aspects of collaboration.

Finally, although we consider the realistic context a particular strength of our study, it limits the conclusions we can draw. Of particular importance may be that students had worked in the same groups since the beginning of the semester and may have had an established collaboration routine. It is possible that the adaptive collaboration script was not maximally effective in altering this routine. Because we did not assess collaboration quality, future research should examine the effects of adaptive collaboration scripts on collaboration quality and examine if a script introduced before students establish a collaboration routine may be more effective.

## Conclusion

A quasi-experiment in an undergraduate chemistry course shows that an adaptive collaboration script that supports students in making connections among visual representations enhanced their learning of content knowledge more so than a traditional version of the same collaborative activity without a script. Effects were of medium size and were found immediately after and three weeks after the experiment. We extend research on learning with visual representations by showing that an adaptive collaboration script can support sense making of visual representations. We extend research on collaborative learning by showing that an adaptive collaboration script focused on visual representations can enhance learning of content knowledge.

## References

- Ainsworth, S. (2008). The Educational Value of Multiple-Representations When Learning Complex Scientific Concepts. In J. K. Gilbert, M. Reiner & A. Nakama (Eds.), *Visualization: Theory and Practice in Science Education* (pp. 191-208). Netherlands: Springer.
- Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the Effects of Different Multiple Representational Systems in Learning Primary Mathematics. *Journal of the Learning Sciences, 11*(1), 25-61.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York, NY: Longman
- De Jong, O., & Taber, K. S. (2014). The Many Faces of High School Chemistry. In N. Lederman & S. K. Abell (Eds.), *Handbook of Research on Science Education* (pp. 457-480). New York, NY: Routledge.
- Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educational Psychologist, 48*(1), 56-66.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active Learning Increases Student Performance in Science, Engineering, and Mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415.
- Gnesdilow, D., Bopardikar, A., Sullivan, S. A., & Puntambekar, S. (2010). *Exploring Convergence of Science Ideas through Collaborative Concept Mapping*. Paper presented at the 9th International Conference of the Learning Sciences.
- Gweon, G., Rose, C., Carey, R., & Zaiss, Z. (2006). Providing Support for Adaptive Scripting in an on-Line Collaborative Learning Environment *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 251-260). New York, NY: ACM Press.
- Karakostas, A., & Demetriadis, S. (2011). Enhancing Collaborative Learning through Dynamic Forms of Support: The Impact of an Adaptive Domain-Specific Support Strategy. *Journal of Computer Assisted Learning, 27*(3), 243-258.
- Kirschner, F., Paas, F., & Kirschner, P. A. (2010). Task Complexity as a Driver for Collaborative Learning Efficiency: The Collective Working-Memory Effect. *Applied Cognitive Psychology, 25*(4), 615-624.
- Kozma, R., Chin, E., Russell, J., & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *Journal of the Learning Sciences, 9*, 105-143.
- Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small Group and Individual Learning with Technology: A Meta-Analysis. *Review of Educational Research, 71*(3), 449-521.
- Magnisalis, I., Demetriadis, S., & Karakostas, A. (2011). Adaptive and Intelligent Systems for Collaborative Learning Support: A Review of the Field. *IEEE Transactions on Learning Technologies, 4*(1), 5-20.
- Miyake, N., & Kirschner, P. A. (2014). The Social and Interactive Dimensions of Collaborative Learning. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2 ed., pp. 418-438). New York, NY: Cambridge University Press.
- NRC. (2006). *Learning to Think Spatially*. Washington, D.C.: National Academies Press.
- Rau, M. A. (2016a). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review, 1-45*. doi: 10.1007/s10648-016-9365-3
- Rau, M. A. (2016). A framework for discipline-specific grounding of educational technologies with multiple visual representations. *IEEE Transactions on Learning Technologies*. doi: 10.1109/TLT.2016.2623303
- Rummel, N., Walker, E., & Aleven, V. (2016). Different Futures of Adaptive Collaborative Learning Support. *International Journal of Artificial Intelligence in Education*.
- Seufert, T. (2003). Supporting Coherence Formation in Learning from Multiple Representations. *Learning and Instruction, 13*(2), 227-237.
- Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating Argumentative Knowledge Construction. *International Journal of Computer-Supported Collaborative Learning, 2*(4), 421-447.
- Suthers, D. D., & Hundhausen, C. D. (2003). An Experimental Study of the Effects of Representational Guidance on Collaborative Learning Processes. *Journal of the Learning Sciences, 12*(2), 183-218.
- Walker, E., Rummel, N., & Koedinger, K. R. (2009). Ctrl: A Research Framework for Providing Adaptive Collaborative Learning Support. *User Modeling and User-Adapted Interaction, 19*(5), 387-431.

## Acknowledgments

This work was supported by the Wisconsin Alumni Research Foundation, the Wisconsin Center for Education Research, and by the University of Wisconsin – Madison Educational Innovation fund. We thank John Moore, Hannah Bowman, Judy Hines, Abe Wu, and Rachel Bain for their help.

# Whose Culture Is It?

## Modeling the Designs of Authentic Learning Environments and the Cultures They Mediate

Yotam Hod and Ornit Sagy  
yhod@edu.haifa.ac.il, ornit.sagy@gmail.com  
University of Haifa

**Abstract:** A major theme of educational research has focused on cultural practices that are learned within formal and informal settings. Many innovative approaches to classroom design come with the intention that practices of the people who are experts in a domain are enculturated by classroom students. This idea, known as authenticity, has been carefully conceptualized in a large variation of educational settings. Cognizant of the inherent gap between conceptualizations and their implementation, we used a constant-comparative method to analyze different variations of authentic learning designs. With the aim of bridging this research-practice gap, our analysis resulted in a model of cultural interaction within learning environments based on different configurations of participants and settings. The conceptual contribution of our research is a refined framework of authenticity that foregrounds the role of human interaction in cultural mediation. Practically, our model contributes new insights into the design of authentic learning environments.

**Keywords:** Authentic; CSCL; culture; enculturation; human interaction

### Introduction

The modern history of education has been shaped by two revolutions – the industrial revolution, which marked a transition between learning by apprenticeship in the agrarian age to traditional schooling in the industrial age; and the digital revolution, currently underway, from traditional schooling to learning in a networked society (Collins & Halverson, 2009). One of the most significant differences between learning by apprenticeship and traditional schooling has been a change in the interaction of the student with the practitioner. In the apprenticeship model, the student learned directly from the person who practiced a profession or skill. In traditional schooling, a teacher was introduced so that students could learn *about* one or more professions. This separation between the students and practitioner has been a long-standing criticism of traditional schooling as fundamentally inauthentic compared to the ‘real’ way society operates (Dewey, 1916; Sawyer, 2014).

The rise of the networked society has opened new opportunities to re-establish ties between the student and practitioner, addressing this year’s conference theme of prioritizing equity and access in CSCL. Such revelations were dominant forces in the establishment of the field, expressed in ideas such as cognitive apprenticeship (Collins, 2006) and authentic learning (Edelson & Reiser, 2006). Socioculturally-minded researchers broadened this view beyond just the direct interaction of student and practitioner, but as giving students access to communities of practice (Lave & Wenger, 1991) or the culture and norms of a particular community (Rogoff, 2003). In this way, authentic learning has been conceived of as enculturation of the practices in a relevant domain (Brown, Collins, & Duguid, 1989).

The application of these ideas to the design of CSCL learning environments has been far reaching, particularly in classroom learning communities (Bielaczyc, Kapur, & Collins, 2013; Hod & Ben-Zvi, 2015). With an eye on contributing to these important advances, this research focuses on human interaction within learning environments designed for authenticity. Specifically, in this study we looked back at the past two decades of research in the learning sciences and CSCL communities to analyze the way different variations of student-teacher-practitioner interactions have been designed to mediate authentic cultural practices.

### Designing for authenticity

The term authenticity has been taken up in disciplines both outside and inside education (Radinsky, Bouillion, Lento, & Gomez, 2001). In this paper, we refer to authenticity in the context of a large theme of CSCL research, informed by sociocultural perspectives of learning. While specific conceptions of authenticity vary, the motivation as it relates to the design of classrooms articulated by Brown, Collins, and DuGuid (1989) in their seminal paper, *Situated Cognition and the Culture of Learning*, is widely accepted. According to them, “Too often the practices of contemporary schooling deny students the chance to engage the relevant domain culture,

because that culture is not in evidence” (p. 34). Along these lines, CSCL environments have been designed to approximate the culture of the people who actually practice the domain – the authentic practitioners (Edelson & Reiser, 2006).

### The constraints of human interaction in traditional educational settings

The inability for students to have direct, continuous interaction with authentic practitioners over meaningful periods of time is a constraint of educational settings (Lim & Barnes, 2005; Timmis, 2014). For example, the ratio of newcomers to old-timers found in classrooms contrasts sharply with learning in professional communities, where cultural maintenance and evolution have a higher balance of old-timers versus newcomers (Roth, McGinn, Woszczyna, & Boutonne, 1999). These distinctions highlight how real-world professional practice comprises of a distinct ecology compared with educational programs (Table 1). As such, educational programs require different types of innovative designs to prepare students for life *outside* of school.

Table 1: Comparison of professional communities and educational programs

	Professional Communities	Educational Programs
Quantity and ratio	Large membership, making the old-timer-to-newcomer ratio high. For example, the ratio of a newcomer to a disciplinary community can be 1:1,000’s.	Small membership, making the individual-to-culture ratio low. For example, the newcomer (students) to old-timer (teacher) ratio in a classroom can be 30:1.
Continuity and duration	Membership changes rotationally. Members enter, often stay for a long period of time (e.g., career), then leave.	School membership changes rotationally. Members enter, often stay for several years, then leave; Classroom membership begins and ends together, for a greater part of a year.

### Approaches to designing for authenticity

Given the constraints of educational settings, authentic learning environments have been designed and conceptualized as taking either a *simulation* or *participation* approach, with the crux of the distinction based on whether or not students have direct interaction with the practitioner, as well as in what context or setting the interactions take place (Cho, Caleon, & Kapur, 2015; Radinsky et al., 2001). Simulations refer to formal educational programs that aim for their culture to more closely resemble, align with, or approximate the authentic culture (Hay & Barab, 2001; Hung et al., 2008; Bereiter & Scardamalia, 2003). In this approach, cultural mediators such as tools, discourse, and artifacts “map to the activity of some professional community” (Radinsky et al., 2001, p. 406).

In contrast to the simulation approach, the participation approach provides students with opportunities for direct interaction with practitioners of the culture that the designer intends for their students to enculturate, typically in the context of out-of-school communities. In such approaches, the cultural mediation is embedded within these interactions. Students learn cultural practices as an outcome of these apprenticeship-like interactions. Even though the term participation is useful to describe this approach, we emphasize that this isn’t full participation. These interactions are designed within the frameworks of the school setting and are typically regulated by a school instructor, may be limited to working on developmentally appropriate tasks, and/or have time restrictions. As such, we prefer to call them hybrids.

While the distinction between the simulation and hybrid approaches appear straightforward, a close look reveals its problematic nature. Hypothetically, if the teacher is a member of the intended authentic culture, then should the design be considered a simulation or hybrid? Alternatively, if a group of students learns in an authentic setting but doesn’t have direct interaction with its actual practitioners, how should this be categorized? This issue lays bare a fundamental problem with the way the design of authentic learning environments are conceptualized. Whose culture is really being enculturated? Stated differently, *whose culture is it?*

Given the primacy of human interaction and setting in existing conceptions of authentic designs and the problems they raise, we were interested to see if there was a way to refine these categorizations. Specifically, we asked: What are the different variations of simulation and hybrid approaches to authentic designs? How can a refined categorization help elucidate which culture is being enculturated?

## Methods

To answer our research questions, we conducted a review of research designs for authenticity from the perspective that we have previously elaborated upon. To find a representative data set of existing research, we turned to the two official journals of the International Society of the Learning Sciences: *The Journal of the Learning Sciences* (JLS) and the *International Journal of Computer Supported Collaborative Learning* (iJCSCL). The foundation of our data corpus was built upon an exhaustive search of the entire catalogues of these journals, from their inception through 2014. We limited our search to articles that explicitly included derivatives of the word authenticity (e.g., authentic, authentically). We similarly broadened our search to include derivatives of the term enculturation (e.g., enculturative, enculturate, enculturating) to be inclusive of research that may not have been explicitly identified with the authentic concept, but maintained the related ideas.

Following an initial review of the contents of the 39 articles that we found, we further limited our corpus for formal, systematic analysis to 21 articles that clearly articulated the components of a clearly elaborated design where the purpose was for students to enculturate authentic practices (a list of these articles can be found at <https://goo.gl/4l9ycz>). Due to the high yield of articles in our final corpus as well as their distribution over time, we concluded this was a sufficient and representative sample of articles for our analysis, consistent with other such reviews (e.g., Ellis & Goodyear, 2016).

We began our review by carefully examining the designs of each study within our final corpus of articles using a constant-comparative method (Glaser & Strauss, 1967). This involved going through stages of (1) collaboratively negotiating the meaning of a concept and design, which often required interpretation and contextual inference, until we reached a consensus view; (2) going back and forth between our emerging conception and subsequent articles to integrate categories; (3) defining our conceptualization until we developed the tools necessary to model the design within each paper we considered; and (4) going through our entire data corpus carefully to verify our findings.

## Findings

To show the categories and variations of the authentic designs which we reviewed, we start this section by explicating the refined conceptual framework and model that resulted from the analysis. Based on this framework, we describe the authentic design variations that we found. We note that although we present the refined conceptualization first, our analysis was recursive in that we went back and forth between our model and our conceptualization of the designs. We present the conceptual framework first because this provides us the language and symbolic tools necessary to communicate the model.

### Refined conceptual framework and model

Our analysis resulted in four dimensions necessary to distinguish between the designs from our data corpus, including the number of participants (individual, group, community); types of participants (learner, practitioner, teacher, designer, cultural representative); culture (actual, authentic, intended), and setting (classroom, practitioner). While most of these were unambiguous as they were described in the research we reviewed, the different cultures described required us to clarify certain definitions that were based on the authors varied conceptions. Ultimately, we settled on these definitions:

- An *actual culture* is a pattern of activities that is developed over time for a community to achieve its valued purposes (Nasir, Roseberry, Warren, & Lee, 2014). It can be found within and across educational, professional, or practitioner settings.
- An *intended culture* is a designer's vision of one or more actual cultures that establishes the goals of the learning environment.
- An *authentic culture* is the actual cultures upon which the intended culture is based.

During our analysis we recognized the importance of applying the notion of intended culture for the analysis of all our cases, even though it was seldom conceptualized as part of the designs we reviewed (e.g., Bielaczyc & Ow, 2013; Hay & Barab, 2001; O'Neill, 2001). The intended culture is important because it acts as a conceptual bridge between actual classroom cultures and authentic cultures. The intended culture is based upon the designer's experiences, knowledge of learning, interpretation of authentic cultures, etc., that may not even be clearly articulated (McKenney & Reeves, 2012; Sandoval, 2014). It is necessarily imagined, representing a combination of one or more authentic cultures that the designer(s) may be a part of. The teacher, who can be the designer or the enactor of others' designs (Kali, McKenney, & Sagy, 2015), can vary between being a central member of an authentic culture or can just have knowledge of it without ever being a participating member. We are not saying that one situation is better than the other, as oftentimes practitioners

are bad teachers, or the best teachers are not authentic practitioners. But certainly, a defining characteristic of authentic learning environments is that the teacher represents the culture that the designers intend to foster. Based on this conceptualization of these cultures, and in addition to the other relevant dimensions described in our data corpus, we generated a model that shows their relationships (Figure 1).

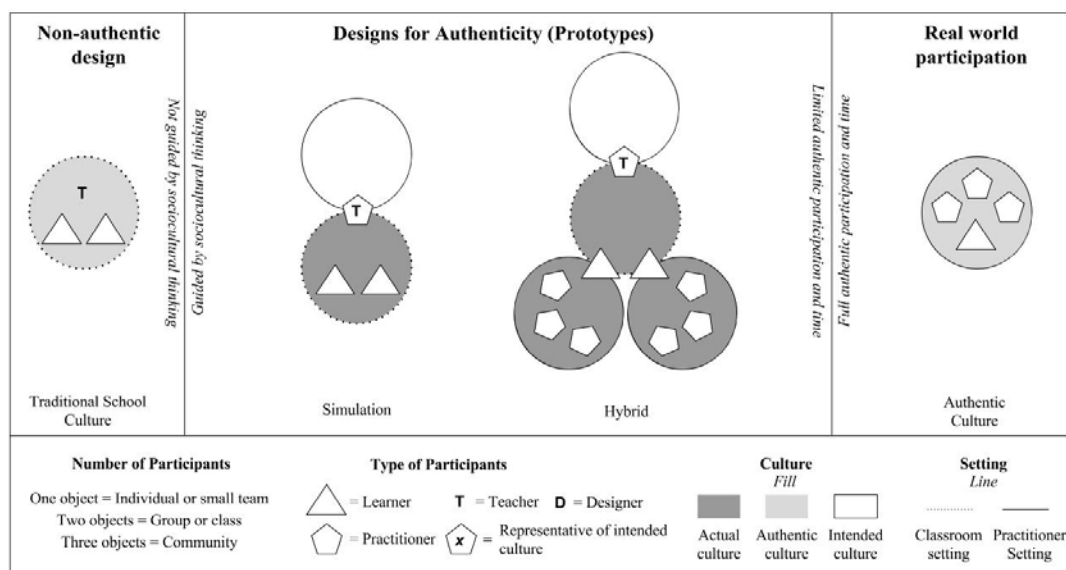


Figure 1. Refined model of authentic designs with simulation and hybrid prototypes.

### Modeling the designs of authentic learning environments

Table 1 summarizes the two categories of authentic learning environments. Within each category, we have found a prototypical version along with three variations. Prototypical versions represent the simplest case. In cases where there were multiple designs within one study, we labelled the Design# alphabetically in the order they appeared in their publication (e.g., 11a, 11b) as specified in the online list (<https://goo.gl/4l9ycz>).

Table 2: Categories and variations of authentic designs

Category	Variation	Description	Design#
Simulation	Prototype	Interaction with authentic practitioners limited to teacher	2, 5, 12, 15, 18, 24, 25, 26, 28, 30, 31, 33, 34, 37, 38
	Variation 1	Interaction with practitioners from a single actual culture with a focus on classroom activities	1
	Variation 2	Interaction with practitioners from several actual cultures with a focus on classroom activities	11a
Hybrid	Variation 3	Interaction with practitioners from the authentic culture with a focus on classroom activities	3
	Prototype	Participation in an educational setting (e.g., classroom) reflexively related to interaction with practitioners from an actual culture with a focus on the practitioners' culture	14
	Variation 1	Participation in an educational setting (e.g., classroom) reflexively related to interaction with authentic practitioners outside of class with a focus on their culture	10, 22b
	Variation 2	Participation in an educational setting (e.g., classroom) reflexively related to interaction with authentic practitioners in and outside of class with a focus on their culture	22a
	Variation 3	Participation in an authentic setting with a focus on the practitioners' culture	11b

## Simulations

Most of the cases (18/23; 78%) that we found were simulations. This is expected, as the simulation category requires the least dependency on and coordination with practitioners outside of the educational setting. In simulation approaches, interactions are predominantly in or about what is happening in the educational setting, contributing to its actual culture. This is true even when outsiders are involved in the designed learning environment, such as can be found in the variations. Figure 2 models the prototype and three variations that we found.

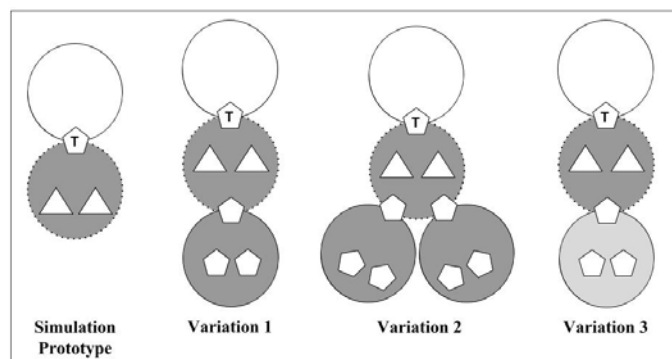


Figure 2. Simulation approach with variations.

The predominant case among all categories and variations that we found was the simulation prototype. While there are vast differences between these cases, all of them rely on one or more teachers being the representatives of an intended culture. For example, to get students to enculturate knowledge building practices, Zhang et al. (2009) adjusted the participatory structures within the classroom so that the students could opportunistically collaborate. In some cases, there may be outside experts, such as in one of Gordin and Pea's (1995) SciV models, where the teacher is also an expert member of the authentic culture. But this is in line with our view of the teacher as a representative of the intended culture and therefore can be modeled by the simulation prototype.

Variation 1 is a simulation where outside practitioners are involved in the classroom activities. Unlike in Gordin and Pea's (1995) design, these outsiders do not have roles as teachers. Roseberry, Warren and Conant's (1992) collaborative inquiry approach is an example of variation 1. The main focus of the design was for students to enculturate scientific discourse by planning and carrying out investigations in their local and home communities. As part of their investigations into the quality of water from their school fountains, students interacted with their local community to collect data and share their findings. Therefore, the design is a simulation that included interactions with practitioners from an actual culture.

Variation 2 is a simulation where outside practitioners are involved in the classroom activities. This variation represents Hay and Barab (2001)'s FC97 summer camp, where three groups of students were working closely with a pair of practitioners who were not part of the authentic culture. While the intended culture was that of disciplinary-based virtual-world designers (e.g., solar system virtual reality), the practitioners included one education and one technology-related graduate student. Therefore, the design was a simulation that included interactions with practitioners from two different actual cultures.

In variation 3, students in a classroom has direct interaction with authentic practitioners. This variation is exemplified by Magnusson, Templin, and Boyle's (1997) Dynamic Science Assessment. Specifically, the practitioners are researchers who practice the relevant domain culture. They come to the classroom to participate in doing dynamic science assessment, which is the basis for the intended culture that tries to approximate the scientific practice of continuously advancing conceptualizations. Therefore, the design was a simulation that included interactions with practitioners from the authentic culture.

## Hybrids

We found four variations of hybrid designs that are modeled in Figure 3. Shared among all the cases was that in addition to the simulation, students had direct interactions with actual or authentic practitioners. In comparison to the simulation variations, the outside-of-classroom interactions in the hybrid models focused upon the practitioners' culture. This can be seen visually in figure 3 in the placement of the learners on the border between the actual classroom culture and the actual or authentic practitioners' cultures.

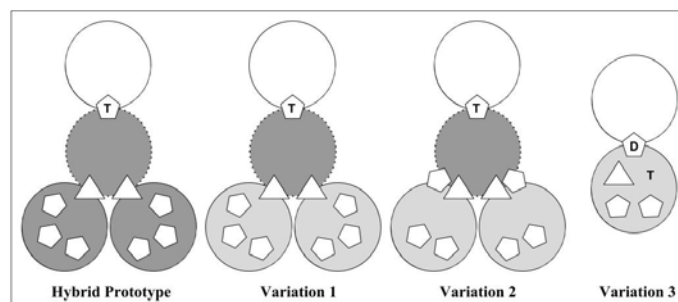


Figure 3. Hybrid approach with variations.

The hybrid prototype is exemplified in Barab, Barnett and Squire's (2002) community of teachers (CoT). In this case, the learners (who are teachers) not only participate in classroom activities (University seminars), but at the same time interact with the staff and students in an actual school where they have a chance to implement their ideas. As the purpose of the CoT is based on an intended culture of "expert teaching" (p. 491), the school is a setting that the teachers attempt to change. Therefore, the design has a simulation that is reflexively related to direct participation within an actual practitioner culture.

Fisher et al. (2007) and O'Neill (2001) provide two cases from different disciplines, age groups, and settings that both exemplify hybrid variation 1. In Fisher et al.'s study, students from the University of Seigen balance between "learning about" and "learning to be" as part of their practice-oriented education in information systems. Specifically, students learn to be by participating in local IT companies. They learn about by participating in a University-based community system that involves academic supervisors, guest lectures, and other students. O'Neill's (2001) design similarly involves a hybrid of simulation and participation approaches. High school students studying earth science develop self-directed research projects within the context of their classroom. Additionally, each student develops a long-term online relationship with a "telemotor" who is an authentic practitioners (graduate student or professionals in the discipline). The telemotor's role is to guide and provide critical feedback to the student on their research. In both cases, there are two settings that are reflexively related. In comparison to the hybrid prototype where interactions in the outside-the-classroom settings are in an actual culture, in variation 1 the interactions are with authentic practitioners.

Hybrid variation 2, exemplified in Fisher et al.'s (2007) University of Colorado Center for Lifelong Learning and Design Research Apprenticeship Program, is similar to hybrid variation 1 with an additional type of interaction. Each student works in a research team that includes doctoral students, post-doctoral researchers, and faculty. This 'vertical integration' provides interactions with authentic practitioners for the graduate students. At the same time, the graduate students enter into the 'horizontal integration', which is a course that consists of graduate students along with their colleagues from each research team. The goal of this hybrid is "crossing different knowledge spaces and nourishing a fertile middle ground between disciplines" (p. 19). Therefore, the learners (graduate students) are members of both a course (simulation) and authentic culture (participation) along with authentic practitioners.

Hybrid variation 3 involves designing to provide direct interaction with practitioners *without* an educational setting for the learners to convene as a group, such as in a classroom. In this design the focus is on the practitioners' culture. Because there is no classroom, there is no intended culture outside of where the student participates. This variation is exemplified in Hay and Barab (2001)'s SAC97 summer camp, where "apprenticeship was operationalized as simply putting students into a real laboratory with a practicing scientist" (p. 288). Their design consisted of small groups of students working directly with a mentor scientist (with guidance of a K-12 teacher) on authentic research problems in the settings where the research took place. Because there was no classroom, the teachers in this case were not representatives of an intended culture, but helped students enculturate the practices of the authentic culture. Still, there was a role of a designer (the camp director) who created this educational opportunity.

## Discussion and conclusion

This research examined the different ways that learning environments are designed to foster authentic learning. It is interesting to note that the most common design (65% of those we found) are the simulation prototypes, where the learners interact only amongst themselves and the teacher(s). This testifies to the creativity of designers who find unique ways to provide students access to authentic discourse, practices, tools, and alike given the common limitations of educational settings and the large investment of resources required. To be clear, we are not making any judgments regarding the quality of the enculturation that is the outcome of any



model or specific study. We even note others who purposefully stepped back from near complete participation by having teachers serve as intermediaries between experts and students, such as in Hay and Barab's *SAC97* (2001).

The project that we have taken upon ourselves to model these different designs is an attempt to rise above a broad array of educational designs which we hope provides clarity on the similarities and differences among them. In particular, we have focused a great deal on developing a parsimonious set of symbolic tools (bottom of Figure 1) which takes into account the relevant dimensions underlying the interactions between participants, cultures, and settings. Beyond the theoretical contribution, this research can be beneficial for designers by giving them a framework to identify the constraints of their programs and the aspects of their designs that can facilitate enculturation. Beyond this, using the tools we have identified can help designers imagine different variations that may not exist.

While our research is based upon widely held notions of authentic learning (e.g., Cho, Caleon, & Kapur, 2015; Radinsky et al., 2001), it has a different emphasis and conceptualization, exemplified by the definitions we have articulated. We recognize that authentic learning can happen in simulated or hybrid designs, and that there is no hard barrier or restriction on anyone, in any setting, to engage in forms of authentic practice. Even when there is no direct interaction with authentic practitioners, students in classrooms can have access to an authentic culture through the use of developmentally appropriate tools, discourse, participatory structures, or other culture mediators within the classroom (Edelson & Reiser, 2006). Having direct human interaction with experts or practitioners in the relevant domain is not a condition of authentic learning; it is just one potent way to foster authentic learning through apprenticeship, such as modeling, coaching, and reflection (Collins, 2006).

Where we conceptualize things differently is that we see simulation and hybrid approaches as two forms of guided participation of school students into authentic cultures. Instead of there being a single authentic culture that students in learning environments can enculturate, educational constraints, where there *is* a hard barrier between participation in schools and participation in actual cultures, require that designers provide interpretations or visions of the authentic cultures. Consequently, notions like the intended culture become a vital part of the conceptualization, as is the relationship between the role of the teacher and designer, actual and authentic cultures, and school and professional settings. We are not saying anything that hasn't been said before; our contribution is modeling these ideas within one integrated framework.

## References

- Barab, S. A., Barnett, M., & Squire, K. (2002). Developing an empirical account of a community of practice: Characterizing the essential tensions. *Journal of the Learning Sciences*, *11*(4), 489-542.
- Bereiter, C., & Scardamalia, M. (2003). Learning to work creatively with knowledge. In E. De Corte, L. Verschaffel, N. Entwistle, & J. van Merriënboer (Eds.), *Unravelling basic components and dimensions of powerful learning environments* (pp. 55-68). Chicago, IL: Open Court.
- Bielaczyc, K., Kapur, M., & Collins, A. (2013). Cultivating a community of learners in K-12 classrooms. In C. E. Hmelo-Silver, C. A. Zhang, C. K. Chan, & A. M. O'Donnell (Eds.), *International Handbook of Collaborative Learning* (pp. 233-249), New York, NY: Routledge.
- Bielaczyc, K., & Ow, J. (2014). Multi-player epistemic games: Guiding the enactment of classroom knowledge-building communities. *International Journal of Computer-Supported Collaborative Learning*, *9*(1), 33-62.
- Brown, A. L. & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.) *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229-272). Cambridge, UK: The MIT Press.
- Brown, J. S., Collins, A., Duguid, P., & Brown, S. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32-42.
- Cho, Y. H., Caleon, I. S., & Kapur, M. (2015). *Authentic Problem Solving and Learning in the 21st Century: Perspectives from Singapore and Beyond*. Singapore: Springer.
- Collins, A. (2006). Cognitive apprenticeship. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 47-60). New York, NY: Cambridge University Press.
- Collins, A., & Halverson, R. (2009). *Rethinking education in the age of technology: The digital revolution and schooling in America*. Teachers College Press.
- Dewey, J. (1916). Democracy and education: An introduction to the philosophy of education. In J. Boydston (Ed.), *The middle works of John Dewey* (Vol. 9). Carbondale: Southern Illinois University Press.
- Edelson, D., & Reiser, B. (2006). Making authentic practices accessible to learning: Design challenges and strategies. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 335-354). New York, NY: Cambridge University Press.

- Ellis, R. A., & Goodyear, P. (2016). Models of learning space: Integrating research on space, place and learning in higher education. *Review of Education*.
- Fischer, G., Rohde, M., & Wulf, V. (2007). Community-based learning: The core competency of residential, research-based universities. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 9-40.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Gordin, D. N., & Pea, R. D. (1995). Prospects for scientific visualization as an educational technology. *The Journal of the Learning Sciences*, 4(3), 249-279.
- Hay, K. E., & Barab, S. A. (2001). Constructivism in practice: A comparison and contrast of apprenticeship and constructionist learning environments. *The Journal of the Learning Sciences*, 10(3), 281-322.
- Hod, Y., & Ben-Zvi, D. (2015). Students negotiating and designing their collaborative learning norms: A group developmental perspective in learning communities. *Interactive Learning Environments*, 23(5), 578-594.
- Hung, D., Lim, K. Y., Chen, D. T. V., & Koh, T. S. (2008). Leveraging online communities in fostering adaptive schools. *International Journal of Computer-Supported Collaborative Learning*, 3(4), 373-386.
- Kali, Y., McKenney, S., & Sagy, O. (2015). Teachers as designers of technology enhanced learning. *Instructional Science*, 43(2), 173-179.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Lim, C. P., & Barnes, S. (2005). A collective case study of the use of ICT in economics courses: A sociocultural approach. *The Journal of the Learning Sciences*, 14(4), 489-526.
- Magnusson, S. J., Templin, M., & Boyle, R. A. (1997). Dynamic science assessment: A new approach for investigating conceptual change. *The Journal of the Learning Sciences*, 6(1), 91-142.
- McKenney, S., & Reeves, T. C. (2012). *Conducting educational research design*. London: Routledge.
- Nasir, Roseberry, Warren, & Lee, (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences, Second Edition* (pp. 687-706). New York, NY: Cambridge University Press.
- O'Neill, D. K. (2001). Knowing when you've brought them in: Scientific genre knowledge and communities of practice. *The Journal of the Learning Sciences*, 10(3), 223-264.
- Radinsky, J., Bouillion, L., Lento, E. M., & Gomez, L. M. (2001). Mutual benefit partnership: A curricular design for authenticity. *Journal of Curriculum Studies*, 33(4), 405-430.
- Rogoff, B. (2003). *The cultural nature of human development*. New York, NY: Oxford University Press.
- Rosebery, A. S., Warren, B., & Conant, F. R. (1992). Appropriating scientific discourse: Findings from language minority classrooms. *The Journal of the Learning Sciences*, 2(1), 61-94.
- Roth, W. M., McGinn, M. K., Woszczyzna, C., & Boutonne, S. (1999). Differential participation during science conversations: The interaction of focal artifacts, social configurations, and physical arrangements. *Journal of the Learning Sciences*, 8(3-4), 293-347.
- Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences*, 23(1), 18-36.
- Sawyer, R. K. (2014). Foundations of the learning sciences. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences, 2<sup>nd</sup> Edition* (pp. 1-20). New York, NY: Cambridge University Press.
- Timmis, S. (2014). The dialectical potential of Cultural Historical Activity Theory for researching sustainable CSCL practices. *International Journal of Computer-Supported Collaborative Learning*, 9(1), 7-32.
- Zhang, J., Scardamalia, M., Reeve, R. & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge-building communities, *Journal of the Learning Sciences*, 18(1), 7-44.

## Acknowledgements

This research was supported by the I-CORE Program of the Planning and Budget Committee and the Israel Science Foundation (1716/12)

# Learning Alone or Together? A Combination Can Be Best!

Jennifer K. Olsen, Human-Computer Interaction Institute, Carnegie Mellon University, [jkolsen@cs.cmu.edu](mailto:jkolsen@cs.cmu.edu)  
Nikol Rummel, Institute of Educational Research, Ruhr-Universität Bochum, [nikol.rummel@rub.de](mailto:nikol.rummel@rub.de)  
Vincent Aleven, Human-Computer Interaction Institute, Carnegie Mellon University, [aleven@cs.cmu.edu](mailto:aleven@cs.cmu.edu)

**Abstract:** Collaborative and individual learning are both frequently used in classrooms to support learning. However, little research has investigated the benefits of combining individual and collaborative learning, as compared to learning only individually or only collaboratively. With our study, we address this research gap. We compared a combined condition to individual-only or collaborative-only learning conditions using intelligent tutoring systems for fractions. The study was conducted with 382 4<sup>th</sup> and 5<sup>th</sup> grade students. Students across all three conditions had significant learning gains. However, the combined condition had higher learning gains than the individual or collaborative condition. This difference was more pronounced for 4<sup>th</sup> grade students than for 5<sup>th</sup> grade students. In addition, we found that students in the combined condition expressed higher situational interest in the activity compared to those working individually and the same as students working only collaboratively. Through a combination, we may support better student learning.

## Introduction

Although collaborative and individual learning are both frequently used in the classroom, little research has been done to investigate if and when their combination is more effective than either one alone. Computer-supported collaborative learning (CSCL) research has combined social planes (i.e., individual, collaborative, whole class) within learning activities using integrative scripts (Dillenbourg, 2004) that prescribe different social planes for different phases of a learning activity (Dillenbourg, 2004; Diziol, Rummel, Spada, & McLaren, 2007). For example, integrative CSCL scripts based on the Jigsaw method have people work individually to gain expertise in an area before working in expert groups and then mixed expert groups to share that expertise (Aronson, 1978). Although these scripts use a combination of collaborative and individual learning, they are often only compared to individual only interventions and not to collaborative only interventions when their effectiveness is investigated. Collaborative and individual learning may each have different strengths that influence the learning process in productive ways; each may be more beneficial than the other for certain types of knowledge (Mullins, Rummel, & Spada, 2011). One might hypothesize, therefore, that a combination of collaborative and individual learning can be more effective in supporting student learning than learning within either of the social planes separately (i.e., collaborative or individual learning), especially when the combination is set up in a way that plays to the strengths of each of the two learning modes. However, it is also possible that switching between social planes adds overhead to the learning process, which could have a negative impact on the student performance that outweighs the benefits of a combination, even if this combination is aligned to their particular strengths. Hence, it is important to understand whether combining individual and collaborative learning, in a way that aligns with their respective strengths, is more effective than individual or collaborative learning alone. In this paper, we investigate this question, using different versions of an intelligent tutoring system (ITS) for elementary school fractions learning as a platform.

Previous research that has compared collaborative and individual learning has found mixed results: some studies found that collaboration is more beneficial, whereas other studies found that individual learning is more beneficial (Lou et al., 2001). These mixed results may be due to how the collaboration and individual learning is being aligned with the learning activities and how the collaborative and individual learning phases are being combined, if at all, in the collaborative learning scenario. Collaborative learning may be beneficial by supporting students in giving and receiving explanations as well as the opportunity to co-construct knowledge with their partner (Hausmann, Chi, & Roy, 2004). In addition, discussions that happen during collaboration can potentially support the students' social goals (e.g., responsibility goals, popularity goals) and make them feel more connected to their group members, which can increase their motivation for the activity (Rogat, Linnenbrink-Garcia, & DiDonato, 2013) and increase the desire to continue working on the task. Specifically, situational interest in the task, which is interest that arises due to a response to the factors in the environment (Linnenbrink-Garcia et al., 2010), can increase when a task involves collaboration. On the other hand, for problem-solving practice, individual learning may be more beneficial than collaborative learning. Working individually may allow students to get more practice in the same amount of time and develop fluency (Mullins et al., 2011) since students are not sharing tasks with a partner and do not necessarily have to pause to explain

their actions. In light of these different strengths, collaboration may be better for conceptually oriented activities, such as working on erroneous examples, and individual work may be better for procedurally oriented activities, such as tutored problem solving.

To design a mixed collaborative/individual condition, we created learning activities that would play to the strengths of the given social plane; specifically, we used erroneous examples for collaborative learning and tutored problem solving for individual learning. Within research on example-based learning, both worked examples and erroneous examples have been shown to be successful for supporting learning (Renkl, 2005; McLaren, et al., 2012; Tsovaltzi et al. 2010). In addition, prior research shows that when students study worked examples collaboratively, they tend to avoid shallow processing, ask for fewer hints, and spend more time on explanations than when working individually (Hausmann, Nokes, VanLehn, & van de Sande, 2009). Further, erroneous examples can help to foster reflection and more fruitful explanations (Isotani et al, 2011; Siegler, 1995; Tsovaltzi et al., 2009). When students are able to collaborate around erroneous examples, they may benefit from engaging in sense-making with their partner, fostered both through the erroneous examples and the collaborative learning. On the other hand, for tutored problem solving, tutors often support student learning through step-by-step support. This step-by-step support focuses the attention of the student on one step at a time, which can lead to students entering an answer as soon as it is known instead of having a discussion around the problem (Mullins, Rummel, & Spada, 2011). When students are working individually, they do not have to divide tasks with another student, or stop often to discuss a problem step, which likely allows each student to get more practice with the problem-solving skills. In turn, more practice with the problems may allow the students to build more fluency and procedural knowledge (Anderson, 1983). When students are able to work individually around the tutored problem solving, they may benefit from the faster-paced practice that is fostered from both the step-by-step nature of the problems and the individual learning.

In this study, we investigated our hypothesis that a combination of collaborative and individual learning is more effective for student learning than the same tasks being performed only collaboratively or individually. Specifically, we investigated the combination of students working collaboratively on erroneous examples and individually on tutored problem solving. The study involved 382 students and ran over five class periods. To test our hypothesis, we assigned students to three different conditions (i.e., mixed, collaborative only, individually only). In addition, we measured the situational interest in the tutor for the students. We hypothesized that students who have a chance to work collaboratively (i.e., mixed and collaborative only conditions) will have more situational interest in the activity than students that only work individually.

## Methods

### Tutor design

As mentioned, we used a fractions ITS as a platform for our research. ITSs have been shown to be beneficial for student learning (Kulik & Fletcher, 2015; Ma, Adesope, Nesbit, & Liu, 2014) and are effective by providing cognitive support for students as they work through problem-solving activities. This cognitive support comes in the form of step-level guidance, namely, an interface that makes all steps visible, error feedback, and on-demand hints (VanLehn, 2006). Although the majority of ITSs have been developed for individual use, the integration of collaboration within an ITS, in prior studies, has effectively supported learning (Baghaei & Mitrovic, 2005; Diziol et al., 2010; Olsen Rummel, & Alevan, 2016). The support for the collaboration can be directly embedded into the tutor to support the students both cognitively and socially.

Informed by prior work on fractions tutors (Olsen, Belenky, Alevan, & Rummel, 2014a; Olsen, Rummel, & Alevan, 2016), we developed a new ITS for three fractions units: equivalent fractions, least common denominator, and comparing fractions. The ITS versions were built with the Cognitive Tutoring Authoring Tools (CTAT), extended to support collaborative tutors (Olsen et al., 2014b). For each of the three units, we created both tutored problem-solving activities (see Figure 1) and erroneous examples (see Figure 2). Further, we created both individual and collaborative versions of both types of activities, for use in different conditions. For each unit, there were eight problems. All of the problems within a unit were of the same type.

For the *tutored problem solving*, the students went through the steps needed to solve each problem. For example, for the unit on comparing fractions, students would first find the least common denominator for the fractions they were trying to compare (see Figure 1), then convert all of the fractions using this common denominator, and finally, put the fractions in order, from smallest to largest. For the *erroneous examples*, the students were asked to go through the process of finding the error that a fictional student had made in a problem (these were common errors that were made in the tutored problem solving problems), correct the error, and provide advice to the student for what they should do in the future. For example, for the least common denominator unit shown in Figure 2, students first needed to identify the error that Kaitie made. After

identifying Kaitie's error, the students were asked to correct the error in the original problem. The students were then given a space to write a message to Kaitie about what she could do differently the next time she encountered similar problems.

**Comparing Fractions**

**A Let's compare fractions**

1 What is the least common denominator of the below fractions?  
You and your partner can each move and enter only half of the fractions. Discuss the correct answers with your partner.

16

0 1 0 1 0 1 0 1 0 1

$\frac{1}{2}$   $\frac{3}{8}$   $\frac{3}{4}$

2 Using the least common denominator, write a new fraction for each of the provided fractions.

8 6 12 14 4  
16 16 16 16 16

3 Take the fractions below the number lines and reorder them from the smallest fraction to the largest fraction.

Smallest Largest

$\frac{4}{16} < \frac{3}{8} < \frac{1}{2} < \frac{3}{4} < \frac{7}{8}$  Submit

Figure 1. A collaborative tutored problem-solving problem for comparing fractions. The students go through the steps of converting fractions as a general mathematical procedure for solving this type of problem.

**Least Common Denominator - Erroneous Example**

**Kaitie made an error. Can you help her?**

What is the least common denominator for the fractions?

$\frac{5}{12}$   $\frac{4}{8}$

least common denominator

$12 \times 8 = 96 = 12 \times 8$

OK OK OK

**A What error did Kaitie make?**

1 Which answer best describes what Kaitie's mistake is? Drag an answer and then discuss with your partner.

Largest number that goes into both denominators  
Smallest number that goes into both denominators  
Largest multiple of both denominators

Your answer:  
Product of both denominators

Your partner's answer:  
Product of both denominators

Your group answer:  
Product of both denominators OK

**B Help Kaitie correct her error.**

1 Correct the errors on the problem to the left. When the step is correct, both you and your partner must press OK.

Hint

Previous Next

Figure 2. A collaborative erroneous example for least common denominators. The students are asked to find the type of mistake that the student in the problem has made (Panel A) and to then fix that mistake (top left).

The collaborative tutors were supported with embedded collaborative scripts for each tutor problem to provide social support for students (Kollar, Fischer, & Hesse, 2006). The collaborative tutors supported synchronous, networked collaboration, in which collaborating students sat at their own computer and had a shared (though differentiated) view of the problem state and different actions/resources available to them. The students sat next to each other and communicated through speech, which was recorded. The groups were preset in the system and after students signed into their account, the system was able to share their problem space. The embedded scripts supported collaboration through a distribution of responsibility to create accountability and

interdependence (Slavin, 1989) and cognitive group awareness (Janssen & Bodemer, 2013). In Figure 1, the screen shows that this student only received three of the five given fraction symbols in the problem (their partner had the second half). The students are responsible for sharing their fractions to be able to find the correct least common denominator and to convert the fractions. In Figure 2, panel A shows an example of support for cognitive group awareness where each student has to answer the question individually, the students are shown each other's answer, and then need to provide a group answer. When correcting the problem, the students each need to press the OK button to get feedback from the tutor. This prevents just one student doing the problem alone. Besides these collaboration script features, the collaborative and individual ITSs were identical.

## Design and procedure

The quasi-experimental study was conducted in a classroom setting with 382 4<sup>th</sup> and 5<sup>th</sup> grade students between 18 classrooms (7 fourth grade and 11 fifth grade), 12 math teachers, and five school districts. The study took place during the students' regular class periods. All students worked with the fractions ITS described above. At the class level, students were randomly assigned to one of three conditions: mixed, collaborative, or individual. Seven classes were assigned to the mixed condition, 6 classes to the collaborative only condition, and 5 classes to the individual only condition. In the mixed condition, the students worked collaboratively on the erroneous examples and individually on the tutored problem-solving activities to align with the strengths of the social planes. In the other conditions, students either worked collaboratively on both types of problems or individually on both types of problems. In all three conditions, the erroneous examples for a unit came before the procedural problems to allow the students to address errors before getting more instruction through the procedural problems sets (Renkl & Atkinson, 2003). Also, students in all conditions completed one unit each day; they switched from the erroneous examples to the tutored problem-solving activities half way through class. Within each class, all of the students were instructed to switch problem sets at the same time. Because the time-on-task was constant for all conditions within each unit, the students finished a different number of problems. Within each class, teachers paired their students based on who would work well together and had similar math abilities to avoid extreme differences that could hinder collaboration. Students worked with the same partner as much as possible and only changed partners due to absenteeism. If a student's partner was absent in the collaborative conditions, the student would be paired with another student working in the same condition for the remainder of the study. When students started with a different partner from the day before, they would begin on the problem set at the place of the student who had made less progress.

The study ran across five class periods of 45 minutes each. On the first day, the students took the pretest individually. At the beginning of the second day, the students took a short tutorial either individually or in groups (aligning with their social mode for the erroneous examples) that gave some instruction on how to interact with the tutor. The students then worked with the tutor for the next three days in their condition. On the fifth day, the students took a posttest individually and answered a short survey to gauge their situational interest when working with the tutors.

## Dependent measures

For the study, we collected pretest and posttest measures, tutor log data, and situational interest measures. We assessed students' fractions knowledge at two different times using two equivalent test forms in counterbalanced fashion. The tests targeted isomorphic problems for both the erroneous and procedurally oriented tutors and were administered on the computer. The tests also had transfer problems for naming, making, adding, and subtracting fractions. Each test had 15 questions, seven erroneous example, six problem solving, and two fractions explanations questions. For each question on the test, the students were able to get a point for each step completed correctly. On the tests there were 81 possible points for the 13 erroneous example and procedural knowledge questions. To assess the students' situational interest in the tutoring activity, we had the students answer a brief survey of 12 questions. The questions were adapted from the Linnenbrink-Garcia et al. (2010) situational interest scale. The questions were all written to ask about the time that was spent learning with the tutoring system. Each question was presented to the student on a Likert scale that ranged from one to seven. The total score could range from 12 to 84.

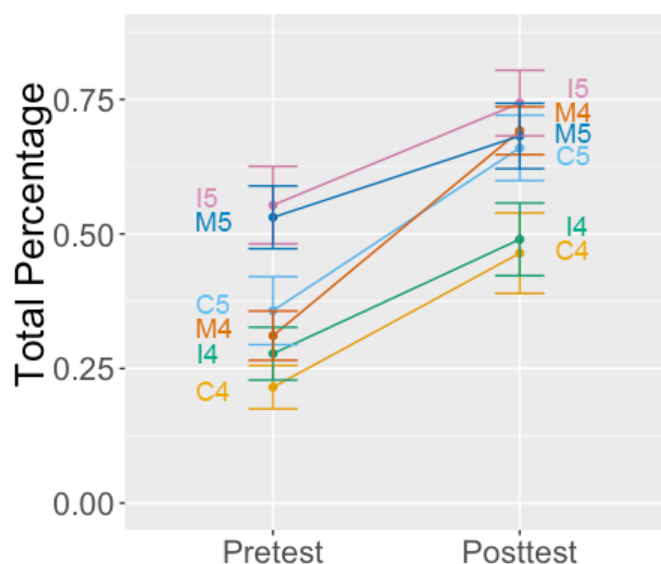
## Results

Out of the 382 students who participated in the study, 75 students were excluded from the analyses because of absenteeism during parts of the study, thus leaving us with a final set of 307 students. Out of the 307 students, 104 were in the collaborative only condition, 83 in the individual only condition, and 120 in the mixed condition. There was no significant difference between conditions with respect to the number of students excluded,  $F(379,2) = 0.59, p = .56$ . There was, however, a significant difference in the pretest scores across

conditions,  $F(2, 304) = 9.4, p < .05$ , with the collaborative only group being significantly lower than the other two conditions.

### Learning gains

To investigate whether students learned using our tutors and if there was a difference in learning between the students in the different conditions, we used a multilevel approach to take into account differences between school districts and the repeated measures of the pretest and posttest. We used a hierarchical linear model (HLM) with student at the first level and school district at the second level. At level 1, we modeled the pretest and posttest scores along with the student's grade (4<sup>th</sup> or 5<sup>th</sup>) and condition, and at level 2, we accounted for differences that could be attributed to the school district. For the different variables, we chose pretest for the test baseline, mixed condition for the condition baseline, and 4<sup>th</sup> grade for the grade baseline. For each variable, the model includes a term for each comparison between the baseline and other levels of the variable. We did not include dyads as a level because of the added complexity of some students working with no partner (i.e. individuals), some students having one partner, and some students having two partners because of absenteeism. We are aware of non-independence issues such as common fate and reciprocal influence within dyads that may have impacted our results (Cress, 2008). We measured the effect size with Pearson's correlation coefficient ( $r$ ) where 0.1 is considered a small effect size, 0.3 a medium effect size, and 0.5 a large effect size.



**Figure 3.** The students worked either collaboratively and individually (M), only collaboratively (C), or only individually (I) with the mixed condition having higher learning gains than the other conditions. This effect was more pronounced in the 4<sup>th</sup> grade students than the 5<sup>th</sup> grade students.

The results from the pretest and posttest analysis are shown in Figure 3. There was a significant difference between pretest and posttest scores,  $t(301) = 12.56, p < .05, r = .59$ , with the posttest scores being higher across all conditions. For the condition differences, there was a significant difference between collaborative only and mixed,  $t(297) = -3.12, p < .05, r = .18$ , and a marginally significant difference between individual only and mixed,  $t(297) = -1.83, p = .07, r = .11$ , with mixed condition having higher test scores than the other conditions. There was a significant interaction between pretest/posttest and collaborative/mixed conditions,  $t(301) = -2.78, p < .05, r = .16$ , and a significant interaction between pretest/posttest and individual/mixed conditions,  $t(301) = -3.56, p < .05, r = .2$ , with the learning gain slope being higher for the mixed conditions than the other conditions, supporting our hypothesis that the mixed condition would be more effective for learning. For the student's grade level (i.e., 4<sup>th</sup> v. 5<sup>th</sup> grade), there was a significant main effect of grade,  $t(297) = 2.93, p < .05, r = .17$ , with the 5<sup>th</sup> graders having higher test scores than the 4<sup>th</sup> grade students. Surprisingly, there was a significant interaction between grade and pretest and posttest,  $t(301) = -5.53, p < .05, r = .3$ , indicating that the 4<sup>th</sup> graders had higher learning gains than the 5<sup>th</sup> graders. There was not a significant interaction between grade and individual/mixed conditions or collaborative/mixed conditions,  $t(297) = 0.90, p = .37$  and  $t(297) = 0.80, p = .42$ . For the three way interactions, there were a significant interactions for both the

pretest/posttest, grade, and collaborative/mixed conditions,  $t(301) = 4.57, p < .05, r = .25$ , and the pretest/posttest, grade, and individual/mixed conditions,  $t(301) = 3.19, p < .05, r = .18$ , with the slope differences between the mixed conditions and the other conditions being more pronounced for the 4<sup>th</sup> grade students than the 5<sup>th</sup> grade students. These interactions indicated that the mixed condition, compared to the other conditions, was more beneficial for the learning gains of 4<sup>th</sup> grade students than those of 5<sup>th</sup> grade students.

### Situational interest

To investigate the impact that working with a partner may have had on the student’s situational interest in the tutoring activity, we used an HLM with student at the first level and school district at the second level. At level 1, we modeled the situational interest score and condition, and at level 2, we accounted for random differences that could be attributed to the school district. There was no significant difference between the collaborative only condition and the mixed condition,  $t(302.15) = -1.119, p = .26$  (see Table 1). There was a significant difference between the students working individually only and the students in the mixed condition,  $t(299.83) = -3.978, p < .05, r = .22$ , such that the students in the mixed condition had a higher situational interest score. These results indicate that the students who had an opportunity to work with a partner found the task more immediately motivating and that working individually for part of the activity did not lower this motivation, although it did for students only working individually.

Table 1: The situational interest score for each of the different conditions

Condition	Situational Interest Score Percentage (SD)
Collaborative Only	0.74 (0.20)
Individual Only	0.59 (0.19)
Mixed	0.75 (0.16)

### Discussion and conclusion

In our classroom study we found that students across all conditions and grade levels learned from working with a tutoring system that supported both studying erroneous examples and problem-solving practice. Thus, the results demonstrate the effectiveness of the instructional conditions. More interestingly, the results confirmed our hypothesis that a combination of collaborative and individual learning may be more beneficial than either alone, as was found across the 4<sup>th</sup> and 5<sup>th</sup> grades. Through a combination of collaborative and individual learning, we are able to align the strengths of the tasks with the strengths of the social planes to better support the student learning. We did find that this result was more pronounced with the 4<sup>th</sup> grade than the 5<sup>th</sup> grade students. This difference may indicate that the given combination of individual and collaborative learning is particularly effective early on in the learning process when students may need more support targeted at the skills they are trying to acquire. The 5<sup>th</sup> grade students may have been at a stage where fluency building was more productive for their learning and the collaboration that supported sense making was not as important. If so, the mixed condition would not have as much benefit for 5<sup>th</sup> grade students. These results resemble those from other research where the age of the students had an impact on the effectiveness of the learning intervention (Mazziotti, Loibl, & Rummel, 2015). However, future work is needed to develop an understanding of when collaborative and individual learning may be most effective. In addition, the 5<sup>th</sup> grade students in the collaborative only condition had higher learning gains than the other 5<sup>th</sup> grade conditions. However, this may be an effect of differences on the pretest. The 5<sup>th</sup> grade collaborative only condition did not have significantly different posttest scores than the other 5<sup>th</sup> grade students.

Additionally, in accordance with our hypothesis, we found that students who had a chance to work collaboratively (mixed and collaborative only conditions) had higher situational interest in the tutoring task than those only working individually. The results support the notion that a collaborative setting can be more motivating for students and that it can be so even when an individual component is added. The situational interest that arises from collaborating can influence the learning that happens around the domain knowledge. When students are more interested in a task, they are willing to put more time and effort into completing that task (Rogat et al., 2013). Allowing students to collaborate on tasks that, *a priori*, would align well with the strengths of collaboration (e.g., building conceptual knowledge) thus might be one way to both motivate students and to create a beneficial learning environment.

This paper opens up a broader line of inquiry of research in CSCL that focuses on the question of how collaborative and individual learning can most effectively be combined. In our study, we supported student learning through the use of erroneous examples and tutored problem solving. We chose these activity types because the strengths of collaborative and individual learning had related strengths to the learning activities so



that a combination may have built upon itself. Specifically, this combination may have been effective because it allowed the students to address misconceptions with a partner and thus develop a deeper understanding. After addressing misconceptions, the students then had an opportunity to build fluency with individual problem solving. This alignment of the learning activities with the hypothesized strengths of the social planes may have enhanced the support to the students more than either could provide alone. Although our results support that this combination of collaborative and individual learning with the learning tasks was more effective than either social plane alone, a limitation of our study is that we do not know what the most effective combination of collaborative and individual learning is and how the results from our combination would generalize. To be able to find what combinations of collaborative and individual learning can be effective for learning, additional research is needed. Our study indicates that this would be a promising direction for future research to explore. In this future exploration, it is important to consider how the switches between social planes are triggered. For example, we have explored switch points triggered by time on task. It may also be beneficial for students to switch social planes adaptively based on switch points triggered by student characteristics, such as repeated errors on a skill when working individually.

The results of our study are notable because of the complexity in supporting both collaborative and individual learning in the classroom and providing real-time support. This study adds to the CSCL literature by comparing a combination of collaborative and individual learning to both social planes alone, which is so far uncommon. By finding support for the effectiveness of combining collaborative and individual learning, this paper has opened a broader line of inquiry into how collaborative and individual learning can most effectively be combined to support learning. Within this space, we can begin to evaluate integrative scripts (Dillenbourg, 2004) to better understand what aspects of the scripts are proving to be effective for student learning or if any combination of social planes is enough to support students.

## References

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Aronson, E. (1978). *The jigsaw classroom*. Beverly Hills, CA: Sage.
- Baghaei, N. & Mitrovic, A. (2005). COLLECT-UML: supporting individual and collaborative learning of UML class diagrams in a constraint-based intelligent tutoring system. In *Int'l Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 458-464). Springer Berlin Heidelberg.
- Cress, U. (2008). The need for considering multilevel analysis in CSCL research—An appeal for the use of more advanced statistical methods. *International Journal of Computer-Supported Collaborative Learning*, 3(1), 69–84.
- Dillenbourg, P. (2004). Framework for integrated learning. Kaleidoscope Network of Excellence deliverable D23.5.1.
- Diziol, D., Rummel, N., Spada, H., & McLaren, B.M. (2007). Promoting learning in mathematics: Script support for collaborative problem solving with the Cognitive Tutor Algebra. In C.A. Chinn, G. Erkens & S. Puntambekar (Eds.) *Mice, minds and society. Proceedings of the Computer Supported Collaborative Learning Conference* (pp. 39-41). International Society of the Learning Sciences, Inc.
- Diziol, D., Walker, E., Rummel, N., & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review*, 22(1), 89-102.
- Hausmann, R. G., Chi, M. T., & Roy, M. (2004) Learning from collaborative problem solving: An analysis of three hypothesized mechanisms. In *26nd annual conference of the Cognitive Science society*, 547-552.
- Hausmann, R. G., Nokes, T. J., VanLehn, K., & van de Sande, B. (2009). Collaborative Dialog While Studying Worked-out Examples. In *International Conference on Artificial Intelligence in Education*. (pp. 596-598).
- Isotani, S., Adams, D., Mayer, R. E., Durkin, K., Rittle-Johnson, B., & McLaren, B. M. (2011). Can erroneous examples help middle-school students learn decimals? *Towards Ubiquitous Learning* (pp. 181-195). Springer Berlin Heidelberg.
- Janssen, J., & Bodemer, D. (2013). Coordinated computer-supported collaborative learning: Awareness and awareness tools. *Educational Psychologist*, 48(1), 40-55.
- Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts—a conceptual analysis. *Educational Psychology Review*, 18(2), 159-185.
- Kulik, J. A. & Fletcher, J.D. (2015). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. *Review of Educational Research*. Prepublished April 17, 2015. doi: 10.3102/0034654315581420
- Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., & Harackiewicz, J. M. (2010). Measuring situational interest in academic domains. *Educational and Psychological Measurement*.

- Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, 71(3), 449-521.
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901-918.
- Mazziotti, C., Loibl, K., & Rummel, N. (2015). Collaborative or Individual Learning within Productive Failure: Does the Social Form of Learning Make a Difference? In O. Lindwall et al. (Eds.), *Proceedings of the 11<sup>th</sup> International Conference on Computer Supported Collaborative Learning*, (570-575). Gothenberg, Sweden.
- McLaren, B. M., Adams, D., Durkin, K., Gogvadze, G., Mayer, R. E., Rittle-Johnson, B., Sosnovsky, S., Isotani, S., & Van Velsen, M. (2012). To err is human, to explain and correct is divine: A study of interactive erroneous examples with middle school math students. In *21st Century Learning for 21st Century Skills* (pp. 222-235). Springer Berlin Heidelberg.
- Mullins, D., Rummel, N., & Spada, H. (2011). Are two heads always better than one? Differential effects of collaboration on students' computer-supported learning in mathematics. *International Journal of Computer-Supported Collaborative Learning*, 6(3), 421-443.
- Olsen, J. K., Belenky, D. M., Aleven, A., & Rummel, N. (2014a). Using an intelligent tutoring system to support collaborative as well as individual learning. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Proceedings of the 12th International Conference on Intelligent Tutoring Systems*, (pp. 134-143). Berlin, Heidelberg: Springer.
- Olsen, J. K., Belenky, D. M., Aleven, A., Rummel, N., Sewall, J., & Ringenberg, M. (2014b). Authoring tools for collaborative intelligent tutoring system environments. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Proceedings of the 12th International Conference on Intelligent Tutoring Systems*, (pp. 523-528). Berlin, Heidelberg: Springer.
- Olsen, J. K., Rummel, N., & Aleven, V. (2016). Investigating effects of embedding collaboration in an intelligent tutoring system for elementary school students. In C. K. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Proceedings of the 12th international conference of the learning sciences, ICLS 2016* (Vol. I, pp. 338-345). Singapore: International Society of the Learning Sciences.
- Renkl, A. (2005). The worked-out example principle in multimedia learning. In R. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 229-246). Cambridge, UK: Cambridge University Press.
- Renkl, A. & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational psychologist*, 38(1), 15-22.
- Rogat, T. K., Linnenbrink-Garcia, L., & DiDonato, N. (2013). Motivation in collaborative groups. *International handbook of collaborative learning*, 250-267.
- Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive psychology*, 28(3), 225-273.
- Slavin, R. E. (1989). Cooperative learning and student achievement: Six theoretical perspectives. *Advances in Motivation and Achievement*, 6, 161-177. Greenwich, CT: JAI Press, Inc.
- Tsovaltzi, D., Melis, E., McLaren, B. M., Dietrich, M., Gogvadze, G., & Meyer, A. K. (2009). Erroneous examples: A preliminary investigation into learning benefits. In *Learning in the Synergy of Multiple Disciplines* (pp. 688-693). Springer Berlin Heidelberg.
- Tsovaltzi, D., Melis, E., McLaren, B.M., Meyer, A-K., Dietrich, M. & Gogvadze, G. (2010). Learning from erroneous examples: When and how do students benefit from them? In Wolpers, M., Kirschner P.A., Scheffel, M., Lindstaedt, S. & Dimitrova, V. (Eds.), *Proceedings of the 5th European Conference on Technology Enhanced Learning, Sustaining TEL: From Innovation to Learning and Practice*, Barcelona, Spain. (pp. 357-373). Springer-Verlag Berlin Heidelberg.
- VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16(3), 227-265.

## Acknowledgments

We thank Amos Glenn, Christian Hartmann, and the CTAT team. This work was supported by Graduate Training Grant #R305B090023 and by Award #R305A120734 from the US Department of Education (IES).

# Predicting Success in Massive Open Online Courses (MOOCs) Using Cohesion Network Analysis

Scott A. Crossley, Georgia State University, [scrossley@gsu.edu](mailto:scrossley@gsu.edu)  
Mihai Dascalu, University Politehnica of Bucharest, [mihai.dascalu@cs.pub.ro](mailto:mihai.dascalu@cs.pub.ro)  
Danielle S. McNamara, Arizona State University, [dsmcnamara1@gmail.com](mailto:dsmcnamara1@gmail.com)  
Ryan Baker, University of Pennsylvania, [ryanshaunbaker@gmail.com](mailto:ryanshaunbaker@gmail.com)  
Stefan Trausan-Matu, University Politehnica of Bucharest, [trausan@gmail.com](mailto:trausan@gmail.com)

**Abstract:** This study uses Cohesion Network Analysis (CNA) indices to identify student patterns related to course completion in a massive open online course (MOOC). This analysis examines a subsample of 320 students who completed at least one graded assignment and produced at least 50 words in discussion forums in a MOOC on educational data mining. The findings indicate that CNA indices predict with substantial accuracy (76%) whether students complete the MOOC, helping us to better understand student retention in this MOOC and to develop more actionable automated signals of student success.

## Introduction

Massive Open Online Courses (MOOCs) open a number of educational opportunities for traditional and non-traditional learning. However, the size of classes, which easily reaches into the thousands of students, requires educators and administrators to reconsider traditional approaches to instructor intervention and the manner in which student engagement, motivation, and success is assessed, especially since attrition rates in MOOCs is notoriously high (Ramesh, Godwasser, Huang, Daume, & Getoor, 2014). The uniqueness of MOOCs and the difficulties associated with them has opened new research areas, especially in predicting or explaining completion rates and general student success. Research has mainly focused on predicting success using click-stream data (i.e., student interactions within the MOOC software). Other recent approaches include the use of Natural Language Processing (NLP) tools to gauge students' affective states (Wen, Yang, & Rose, 2014b, 2014a), measure the sophistication and organization of students' discourse within a MOOC (Crossley et al., 2015; Crossley, Paquette, Dascalu, McNamara, & Baker, 2016, and a combination of click-stream and NLP data (Crossley et al., 2016). In this study, we examine new NLP approaches grounded in text cohesion and Social Network Analysis (SNA) to predict success in a MOOC related to educational data mining. Social interaction has long been recognized as an important component of learning (Vygotsky, 1978). However, while the relationship between language and social participation has been studied in MOOCs (Dowell et al., 2015), social interaction reflected through the language produced by MOOC students has not been investigated within large-scale, on-line learning environments.

The variables used in this study are based on Cohesion Network Analysis (CNA), which can be used to analyze discourse structures within collaborative conversations (Dascalu, Trausan-Matu, McNamara, & Dessus, 2015). CNA indices estimate cohesion between text segments based on similarity measures of semantic proximity. We hypothesize that students who produce forum posts that are on topic, are more related to other student posts, are more central to the conversation, and are more collaborative will be more likely to complete the MOOC than those that are not. We focus specifically on student completion rates because they are an important component of student success within the course, as well as after its completion (Wang, 2014). We assess links between completion and CNA indices because CNA indices afford a wide array of opportunities for better understanding student success in terms of collaboration. Using CNA indices to better understand student completion rates has the potential to inform pedagogical interventions that provide individualized feedback to MOOC participants and teachers regarding social interactions such as collaboration. Ultimately, our objective is to enhance participation and active involvement, to increase completion rates, as well as to increase our understanding of the factors associated with MOOC completion.

## MOOC analysis

MOOCs have become an important component of education research for both instructors and researchers because they have the potential to increase educational accessibility to distance and lifelong learners (Koller, Ng, Do, & Chen, 2013). Researchers examine links between click-stream data in MOOCs and academic performance because MOOCs provide a tremendous amount of data via click-stream logs containing detailed records of the students' interactions with the course content. The measures typically computed from click-stream data that have been used in MOOC analyses include variables related to counts of the different possible types of actions, the timing of

actions, forum interactions and assignments attempts among others (Seaton, Bergner, Chuang, Mitros, & Pritchard, 2014).

More recently, researchers have applied NLP tools to MOOC data (Chaturvedi, Goldwasser, & Daume, 2014; Wen, Yang, & Rose, 2014a, 2014b; Crossley et al., 2015; Crossley et al., 2016). Traditional usage of NLP tools in this context focus on a text's syntactic and lexical properties. The simplest approaches count the length of words or sentences, or use pre-existing databases to compare the word properties in a single text to that of a larger, more representative corpus of texts. More advanced NLP tools measure linguistic features related to the use of rhetorical structures, syntactic similarity, text cohesion, topic development, and sophisticated indices of word usage. Such tools have been used to examine text complexity (e.g., cohesion, lexical, and syntactic complexity) in forum posts and the degree to which these indicators are predictive of MOOC completion. For instance, Crossley et al. (2015) found that language related to forum post length, lexical sophistication, situational cohesion, cardinal numbers, trigram production, and writing quality were significantly predictive of whether a MOOC student completed the course (reporting an accuracy of 67%). In a followup study, Crossley et al. (2016) combined click-stream data and NLP approaches to examine if students' on-line activity and the language they produced in the on-line discussion forum was predictive of MOOC completion. They found that click-stream variables (e.g., weekly lecture coverage and how early students submitted their assignments) were the strongest predictors of MOOC completion but that NLP variables (e.g., the number of entities in a forum post, the post length, the overall quality of the written post, the linguistic sophistication of the post, cohesion between posts, and word certainty) significantly increased the accuracy of the model. In total, click-stream and NLP indices predicted which students would complete the course with 76% accuracy. Combined, these findings indicate that students who are more involved in the course and demonstrate more advanced linguistic skills, are more likely to complete a MOOC.

## Current study

The goal of the study is to test new indices that measure social integration and collaboration using Cohesion Network Analysis in order to examine student success in a MOOC. Thus, we perform a longitudinal analysis on the weekly timeline evolution of CNA indices to predict MOOC success and examine if students who engage in greater social interaction, that is on topic and central to the MOOC, are more successful (i.e., complete the course).

## Method

### The MOOC: Big data in education

In this paper, we evaluate course completion in the context of the Big Data in Education MOOC (BDEMOOC), using the data from the first iteration on this course, offered through the Coursera platform in 2013. This is the same MOOC investigated by Crossley et al. (Crossley et al., 2015; Crossley et al., 2016). The course was designed to support students in learning how to apply a range of educational data mining (EDM) methods to conduct education research questions and to develop models that could be used for automated intervention in online learning, or to inform teachers, curriculum designers, and other stakeholders. This course was targeted to the postgraduate level, and covered material comparable to a graduate course taught by the instructor. The MOOC ran from October 24, 2013 to December 26, 2013, and included several lecture videos in each of the 8 weeks, and one assignment per week.

In each of the weekly assignments, students conducted a set of analyses on a given data set and answered questions about the analyses. All assignments were automatically graded, and students had up to three attempts to complete each assignment successfully. Students received a certificate by obtaining an overall average grade of 70% or better on at least 6 of the 8 assignments. The course had an official enrollment of over 48,000 at the time of the course's official end. 13,314 students watched at least one video, 1,242 students watched all videos, 1,380 students completed at least one assignment, and 710 made a post in the discussion forums. Of those with posts, 426 students completed at least one class assignment while 638 students completed the online course and received a certificate. As such, some students earned a certificate for BDEMOOC without ever posting to the discussion forums.

### Student completion rates

We selected completion rate as our variable of success because it is one of the most common metrics used in MOOC research (He, Bailey, Rubinstein, & Zhang, 2015), and correlates to future career participation (Wang, 2014). For this study, completion was based on a smaller sample of forum posters as described below. "Completion" was pre-defined as earning an overall grade average of 70% or above. The overall grade was calculated by averaging the 6 highest grades extracted out of the total of 8 assignments.

## Discussion posts

Discussion posts are of interest within research on student participation in MOOCs because they are one of the core methods that students use to participate in social learning (Ramesh, Goldwasser, Huang, Daume, & Getoor, 2014). Discussion forums provide students with a platform to exchange ideas, discuss lectures, ask questions about the course, and seek technical help, all of which lead to the production of language in a natural setting. Such natural language can provide researchers with a window into individual student motivation, linguistics skills, writing strategies, and affective states. This information can in turn be used to develop models to improve student learning experiences (Ramesh, Goldwasser, Huang, Daume, & Getoor, 2014). In BDEMOOC, students and teaching staff participated actively in weekly forum discussions. Each week, new discussion threads were created for each week's specific content, including both videos and assignments under sub-forums, each with corresponding discussion threads. Forum participation did not count toward student's final grades. For this study, we focused on the forum participation in the weekly course discussions. For this study, we extracted all forum posts and corresponding comments from the MOOC environment for all 426 students who both made at least a forum post and completed an assignment. We removed all data from instructors and teaching assistants. We analyzed data from those students who produced at least 50 words in their aggregated posts ( $n = 319$ ). Fifty words was used as a cut-off to ensure sufficient linguistic information. Of these 319 students, 132 did not successfully complete the course while the remaining 187 completed the course.

## Cohesion network analysis

In Computer Supported Collaborative Learning (CSCL) environments, Cohesion Network Analysis analyzes discourse structure by combining NLP approaches with SNA (Dascalu, Trausan-Matu, McNamara, & Dessus, 2015). In CNA, cohesion is computationally represented as an average value of similarity measures (or an aggregated score) between semantic distances (Budanitsky & Hirst, 2006) using *WordNet* (Miller, 1995) *Latent Semantic Analysis* (Landauer & Dumais, 1997) and *Latent Dirichlet Allocation* (Blei, Ng, & Jordan, 2003). We used the Touchstone Applied Science Associates (TASA) corpus (approximately 13 million words; <http://lsa.colorado.edu/spaces.html>) together with a collection of articles extracted from the Learning Analytics & Knowledge dataset (652 Learning Analytics and Knowledge and Educational Data Mining conference papers and 45 journal papers; <https://www.w3.org/TR/REC-rdf-syntax/>) to train dedicated LSA and LDA semantic models. The resulting corpora covered both the curricula of the MOOC course and provided also a general knowledge background. Before training, the texts were preprocessed such that stop-words were removed and all words were lemmatized.

A *cohesion graph* (Dascalu, Trausan-Matu, McNamara, & Dessus, 2015) was generated using cohesion values in order to determine connections between discourse elements. This graph represents a generalization of the utterance graph (Trausan-Matu, Stahl, & Sarmiento, 2007) and can be used as a proxy for the semantic content of discourse. The cohesion graph is a multi-layered structure containing different nodes (Dascalu, 2014) and the links between them. A central node, representing the conversation's thread, is divided into contributions, which are further divided into sentences and words. Links are then built between nodes in order to determine a cohesion score that denotes the relevance of a contribution within the conversation, or the impact of a word within a sentence or contribution. Other links are generated between adjacent contributions, which are used to determine changes in the topics or of the conversation's thread. These changes are reflected by cohesion gaps between units of texts. Explicit links, created using an interface functionality such as the "reply-to" option, are contained within the cohesion graph as well. In addition, cohesive links determined using semantic similarity techniques are added between related contributions within a timeframe of maximum 20 successive contributions, which can be considered the maximum span for these type of cohesive links (Rebedea, 2012).

## Cohesion scoring mechanism

The cohesion graph determines the active engagement in terms of participation in the MOOC. This is computed quantitatively based on relations established between nodes from the cohesion graph. The contributions are analyzed to determine their importance in relation to the discussion's thread, coverage of topics, and their relatedness to other contributions. The relevance score of a node in the cohesion graph is based on the relevance of underlying words and on its relation to other components. For example, a contribution's relevance score is computed as the sum of its constituent words based on *statistical presence* and the *semantic relatedness* (Dascalu, Trausan-Matu, Dessus, & McNamara, 2015). Statistical presence represents the word frequency within the text, while semantic relatedness refers to semantic similarity between the word and the entire conversation thread that contains it. Keywords for the whole conversation are determined by considering the aggregated score of the two factors.

Afterwards, the cohesion scoring mechanism assigns contribution scores by multiplying each word's previously determined score with its normalized term frequency (Dascalu, 2014), estimating an on-topic relevance of the utterance. Links with other contributions, stored within the CNA are further used to improve contribution scores. Each contribution's local relevance is then calculated with regards to related contributions. Thus, each textual element's score can be viewed as its importance within the discourse, covering both the topic and the semantic relatedness with other elements.

### Collaboration assessment

Social knowledge-building (KB) processes (Bereiter, 2002) are derived through collaboration (i.e., scores calculated on the inter-animation of interactions between different participants). Social KB refers to the external dialog between at least two participants supporting collaboration, while inner dialogue is reflected by the continuation of ideas or explicit, referred contributions belonging to the same speaker.

Each contribution has a previously defined importance score and an effect score in term of both personal and social KB. The personal score is initially assigned as each utterance's importance score, while the social score is initially assigned a zero. By analyzing the links from the cohesion graph, these scores are augmented. If a link is established between contributions belonging to the same speaker, the knowledge (personal and social) from the referred contribution is transferred to the personal dimension of the current contribution through the cohesion score. If the link is established between different users, only the social dimension of the currently analyzed contribution is increased by the cohesion measure. This enables a measurement of collaboration perceived as a sum of social KB effects that consist of each contribution's score, multiplied by the cohesion value to related contribution (Dascalu, Trausan-Matu, McNamara, & Dessus, 2015).

### Interaction modeling and integration of multiple CNA graphs

The *sociogram* reflects information exchanges between users and represents the central structure for modeling interaction and information transfer between participants (Dascalu, 2014). The nodes represent users, while the edges represent interchanged contributions. This graph considers not only the number of exchanged contributions, but weights each utterance as a sum of social KB effects to other MOOC participants. Specific SNA metrics are further computed starting from the sociogram in order to measure centrality or involvement (Dascalu, 2014). Some examples include the number of links to (out-degree) and from (in-degree) other participants for a specific user. Betweenness centrality (Bastian, Heymann, & Jacomy, 2009) is computed to determine central nodes and highlights the information exchange between participants who, if eliminated, would highly reduce communication. The participant's connection to other nodes, called closeness centrality (Sabidussi, 1966), is computed as the inverse distance to all other nodes. A higher value represents a participant's stronger connection to all other discussion thread participants. The maximal distance between a node and all other nodes, called eccentricity (Freeman, 1977), shows the closeness of a user to other participants. These models were extended to facilitate the evaluation of not only a single discussion, but of an entire MOOC by considering the aggregation of multiple discussion threads. Such a global analysis was used to build a social network consisting of all involved participants and their contributions, thus enabling the evaluation of participation at a macroscopic level, not only for specific discussions, but for the entire MOOC. The sociogram between all participants was generated considering the sum of contribution scores per discussion thread within the forum. The overview of different user goals, distributions, and interactions provides a broader perspective of a participants' evolution within the MOOC.

### Longitudinal analysis

We performed a longitudinal analysis by measuring the distribution of each participant's involvement throughout the duration of the MOOC which enabled us to quantify the evolution of learners' participation, collaboration and interaction patterns across time. In order to generate each participant's time distribution, specific sociograms were built for incremental weekly timeframes and CNA-derived quantitative indices were evaluated, covering the following elements, as discussed above: a) cumulative utterance scores per participant (i.e., the sum of individual contribution importance scores that were uttered by a certain participant), b) social KB effect as the cumulative effect of a participant's contribution in relation to other speakers, and c) specific SNA metrics (i.e., in-degree, out-degree, betweenness, closeness and eccentricity centrality measures) computed on the CNA interaction graph.

As expected due to attrition, a large discrepancy was observed in terms of the density of the interaction graphs found between the first and last week of the course, denoting a significant decrease in density. The values of each CNA index per timeframe were used to create individual time series reflecting each participant's evolution throughout the course. Afterwards, the longitudinal analysis indices presented in Table 1 were used to model the trends of the time series generated per participant and per CNA quantitative index. This approach creates an in-depth NLP-centered perspective of our longitudinal analysis built on top of CNA.

## Statistical analysis

CNA indices that yielded non-normal distributions were removed. A multivariate analysis of variance (MANOVA) was conducted to examine which indices reported differences between students who completed or did not complete the MOOC. The MANOVA was followed by a stepwise discriminant function analysis (DFA) using CNA indices that were normally distributed and demonstrated significant differences between students who completed the course and those who did not. CNA indices were also checked for multicollinearity ( $r > .90$ ). In the case of multicollinearity between indices, the index demonstrating the largest effect size in the MANOVA was retained in the analysis. The DFA was used to develop an algorithm to predict group membership through a discriminant function coefficient. A DFA model was first developed for the entire corpus of student forum posts. This model was then used to predict group membership (completers v. non-completers) for the student forum posts using leave-one-out-cross-validation (LOOCV) in order to ensure that the model was stable across the dataset.

Table 1: Longitudinal analysis indices applied on students' social media contributions across time

Name	Description
Avg. & St. Dev.	Average and standard deviation of the considered CNA quantitative index within all timeframes
Slope	The degree of the slope corresponding to the linear regression applied on the time series. The slope indicates whether students became more actively involved (slope > 0), had a uniform involvement (slope = 0), or lost their interest throughout the semester (slope < 0).
Entropy	Considering the probability of posting within each timeframe, Shannon's entropy formula (Shannon, 1948) grasps the discrepancies or inconsistencies in participation patterns. For example, if students are active in only one timeframe, their entropy is 0, whereas if they have a constant activity throughout the course, their entropy converges towards the maximum value of $\log(n)$ , where $n$ is the number of timeframes
Uniformity	Degree of uniformity is measured using Jensen Shannon dissimilarity (JSD) (Manning & Schütze, 1999) to a uniform distribution of $1/n$ . The JSD is a symmetric function based on the Kullback–Leibler divergence and is used to measure the similarity between two distributions, in our case the student's time series and an ideal, uniform participation in each week
Local extreme points	The number of local extreme points determined as the number of timeframes for which the inflection or the direction of the evolution of the CNA index changes. This reflects the monotony degree of the evolution or inconsistency in participation or collaboration - if multiple spikes are encountered, these will be identified as local minimum or maximum points; therefore, more local extreme points will be identified within the time series evolution
Average & standard deviation of recurrence	Recurrence is expressed as the distance between timeframes in which the learner had at least one contribution in the time series. This is useful for identifying and quantifying pauses as adjacent weeks without any activity. If each timeframe has at least one event, recurrence is 0, whereas if students take long pauses that inherently generate timeframes with 0 events, recurrence increases (e.g., if they post every 2 weeks, recurrence becomes 1, and so forth).

## Results

A MANOVA was conducted using the CNA indices as the dependent variables, and whether the student completed or did not complete the MOOC as the independent variable. Of the 56 indices, 15 indices were not normally distributed and were removed. Of the remaining 41 indices, 27 indices did not demonstrate multicollinearity and were retained. Of these 27 indices, 26 of them demonstrated significant differences between students who completed the MOOC and students who did not complete the MOOC (see Table 2 for details). These indices demonstrated that MOOC completers produced posts that were on topic, were more related to other posts, demonstrated greater collaboration, and were more central to the conversation. These indices were used in the subsequent DFA.

A stepwise DFA using the 26 indices selected through the MANOVA retained three variables: Standard deviation of recurrence (Overall Score), Slope degree (Closeness), and Average (Closeness). The results demonstrate that the DFA using these three indices correctly allocated 243 of the 319 forum posts in the total set,  $\chi^2(df=1) = 86.325, p < .001$ , for an accuracy of 76.2%. For the leave-one-out cross-validation (LOOCV), the discriminant analysis allocated 242 of the 319 students for an accuracy of 75.9%. See Table 3 for recall, precision,

and F1 scores for this analysis. The Cohen's Kappa measure of agreement between the predicted and actual class label was .518, demonstrating moderate agreement.

Table 2: Longitudinal analysis indices applied on students' social media contributions across time

Index	Did not complete: Mean (SD)	Completed: Mean (SD)	F	$\eta^2$
Standard deviation of recurrence (Overall score)	2.433 (0.839)	1.395 (0.994)	95.666**	.232
Local extremes (Overall score)	2.106 (1.134)	3.401 (1.550)	66.842**	.174
Slope degree (Closeness)	0.006 (0.011)	0.024 (0.022)	71.637**	.184
Slope degree (Eccentricity)	0.084 (0.115)	0.281 (0.252)	69.91**	.181
Local extremes (Out-degree)	1.864 (1.247)	3.198 (1.678)	60.045**	.159
Degree of uniformity (Overall score)	0.639 (0.099)	0.518 (0.169)	54.739**	.147
Entropy (Overall score)	0.277 (0.349)	0.634 (0.542)	44.412**	.123
Standard deviation of recurrence (In Degree)	2.113 (0.949)	1.338 (0.996)	48.713**	.133
Standard deviation of recurrence (Out Degree)	2.207 (1.117)	1.434 (1.062)	39.325**	.110
Average (Closeness)	0.063 (0.056)	0.118 (0.093)	36.965**	.104
Local extremes (In-degree)	2.265 (1.313)	3.166 (1.492)	31.108**	.089
Entropy (Closeness)	0.309 (0.432)	0.702 (0.654)	36.333**	.103
Average recurrence (Overall score)	2.628 (0.949)	1.856 (1.456)	28.548**	.083
Local extremes (Betweenness)	1.409 (1.266)	2.369 (1.709)	29.997**	.086
Degree of uniformity (Closeness)	0.606 (0.123)	0.494 (0.198)	33.153**	.095
Degree of uniformity (In-degree)	0.613 (0.110)	0.522 (0.165)	30.736**	.088
Entropy (Out-degree)	0.162 (0.290)	0.416 (0.469)	30.461**	.088
Entropy (In-degree)	0.319 (0.372)	0.598 (0.534)	26.909**	.078
Average recurrence (Out-degree)	3.181 (1.593)	2.232 (1.724)	24.941**	.073
Degree of uniformity (Out-degree)	0.646 (0.098)	0.574 (0.143)	24.844**	.073
Standard deviation of recurrence (Betweenness)	1.905 (1.394)	1.318 (1.129)	17.236**	.052
Entropy (Betweenness)	0.123 (0.287)	0.284 (0.416)	14.893**	.045
Standard deviation (Closeness)	0.121 (0.083)	0.155 (0.087)	12.586**	.038
Average recurrence (In-degree)	2.449 (1.392)	1.889 (1.640)	10.219*	.031
Degree of uniformity (Betweenness)	0.626 (0.110)	0.583 (0.128)	9.909*	.030
Average recurrence (Betweenness)	3.999 (1.931)	3.320 (2.239)	7.956*	.024

\*  $p < .010$ , \*\*  $p < .001$

Table 3: Recall, precision, and F1 scores for LOOCV DFA

Count	Did not complete	Completed
Recall	.687	.820
Precision	.765	.754
F1-score	.724	.786

## Discussion and conclusion

Previous MOOC studies have investigated completion rates through click-stream data or NLP techniques or a combination of both. Our interest in this study was to focus on language indices related to social interaction and collaboration, which are important components of learning, both inside and outside the classroom (Vygotsky, 1978). This study examined MOOC completion rates using novel Cohesion Network Analysis indices to estimate connections between discourse elements in order to develop models of the underlying semantic content of the MOOC forum posts. The findings from this study indicate that CNA indices are important predictors of student completion rates with students who produce more on-topic posts, posts that are more strongly related to other posts, or posts that are more central to conversation. Thus, the results support the notion that students who collaborate more are more likely to complete the MOOC. These findings have important implications for how students' interactions within the MOOC in reference to collaboration and social integration can be used to predict success.

The results indicate that overall contribution scores showed the strongest differences between those that completed the MOOC and those that did not (see MANOVA results in Table 2). In addition, overall contribution



scores, which reflect an estimate of on-topic relevance for each utterance made by each participant, were a significant predictor in the DFA model. The mean scores (see Table 2) show that participants who produced a greater number of on-topic posts (i.e., were more engaged with the topic of the MOOC) were more likely to complete the course. The next strongest predictors of whether students completed or did not complete the course were related to *closeness* and *eccentricity* applied on weekly CNA interaction graphs. These indices reflect how strongly a student's posts are related to other posts made by other students (i.e., strength of connection to other posts). The results indicate that students are more likely to complete the MOOC if their posts share semantic commonalities with posts made by other students. Two indices related to closeness were included in the final DFA model. After closeness and eccentricity indices, the next strongest indices were related to *in-degree* and *out-degree*. These indices are also computed based on interaction graphs and measure the number and the semantic strength of links to and from other students. The findings show that students who complete the MOOC have a greater number of semantically related links to and from other students in the MOOC. Lastly, a number of *betweenness* indices demonstrated significant differences between students who completed the MOOC and those that did not. Betweenness is a measure of how central a node is to communication in terms of the information exchanged between participants. Importantly, betweenness indices indicate how much information would be reduced if participants were eliminated from the conversation. The findings from this study indicate that participants who were more critical to forum discussion threads were more likely to complete the MOOC.

In terms of comparison to previous findings, our CNA indices alone are as powerful as the ones employed in previous studies that combined both NLP and click-stream data (Crossley et al., 2016) with accuracies of 76% in both cases, and more powerful than using NLP indices alone (67% with NLP indices compared to 76% with CNA indices used in the longitudinal analysis; (Crossley et al., 2015). More importantly, the indices indicate that patterns of collaboration and social interaction are important for understanding success, going beyond individual linguistic differences and click-stream patterns. Thus, the findings help to provide support to the basic notion that cognitive engagement during learning is a key component of learning and success (Corno & Mandinach, 1983) and that cooperative work may lead to greater learning gains (Johnson & Johnson, 1990). More importantly, these theories of collaboration within learning environments can be extended to large scale on-line classrooms, such as MOOCs. Even in MOOCs, it appears that those students who deviate less from the expected content (Standard deviation of recurrence [Overall Score]), and have higher and stronger connections to other participants (Slope degree and Average [Closeness]) are more likely to be successful. Other CNA indices that were not included in the DFA, but demonstrated significant differences between students who completed the course and those that did not, indicated that more successful students had more links to and from other students (in- and out-degree), were central within the community (low eccentricity) and facilitated conversation among students (betweenness).

The models presented in this paper could be employed to monitor and support students less likely to complete the course by providing timely and personalized feedback in order to increase MOOC engagement and long-term completion. However, much of this depends on the availability of textual traces, which are not always available in many MOOCs. While we focused on forum posts in this study, the employed mechanisms should generalize and, as such, could be applied on other text traces such as participation in collaborative chats, written assignments that are scored in terms of effectively summarizing course lectures, responses to open answer questions which are automatically assessed. In all cases, the results reported here need to be substantiated in follow up studies that evaluate the applicability of the introduced CNA indices in the analysis of MOOCs from other domains and on MOOCs built on other platforms. The LSA and LDA spaces developed for this study may need to change based on new domains, although this needs to be tested. In addition, the CNA indices introduced here could be combined with more traditional NLP indices, click-stream variables, and individual difference measures to further enhance our understanding of student success in on-line classes.

## References

- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *Int. AAAI Conf. on Weblogs and Social Media* (pp. 361–362). San Jose, CA: AAAI Press.
- Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(4-5), 993–1022.
- Budanitsky, A., & Hirst, G. (2006). Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics*, 32(1), 13–47.
- Chaturvedi, S., Goldwasser, D., & Daume, H. (2014). Predicting instructor's intervention in MOOC forums. In *52nd Annual Meeting of the ACL* (pp. 1501–1511). Baltimore, MA: ACL.
- Corno, L., & Mandinach, E. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist*, 18, 88–100.

- Crossley, S. A., McNamara, D. S., Baker, R., Wang, Y., Paquette, L., Barnes, T., & Bergner, Y. (2015). Language to completion: Success in an educational data mining massive open online class. In *8th Int. Conf. on Educational Data Mining* (pp. 388–392). Madrid, Spain.
- Crossley, S. A., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining Click-Stream Data with NLP Tools to Better Understand MOOC Completion. In *6th Int. Conf. on Learning Analytics & Knowledge (LAK '16)* (pp. 6–14). Edingurgh, UK: ACM.
- Dascalu, M. (2014). *Analyzing discourse and text complexity for learning and collaborating*, *Studies in Computational Intelligence* (Vol. 534). Cham, Switzerland: Springer.
- Dascalu, M., Trausan-Matu, S., McNamara, D. S., & Dessus, P. (2015). ReaderBench – Automated Evaluation of Collaboration based on Cohesion and Dialogism. *International Journal of Computer-Supported Collaborative Learning*, *10*(4), 395–423. doi: 10.1007/s11412-015-9226-y
- Dowell, N., Skrypnik, O., Joksimovic, S., Graesser, A., Dawson, S., Gasevic, D., de Vries, P., Hennis, T., & Kovanovic, V. (2015). Modeling Learners' Social Centrality and Performance through Language and Discourse. In *EDM 2015* (pp. 130–137). Madrid, Spain: International Educational Data Mining Society.
- Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*(1), 35–41.
- He, J., Bailey, J., Rubinstein, B. I.P., & Zhang, R. (2015). Identifying at-risk students in massive open online courses. In *AAAI 2015* (pp. 1749–1755). Austin, Texas: AAAI Press.
- Johnson, D. W., & Johnson, R. T. (1990). Cooperative learning and achievement. In S. Sharan (Ed.), *Cooperative learning: Theory and research* (pp. 23–37). New York, NY: Praeger.
- Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and Intention in Massive Open Online Courses. *EDUCAUSE Review*, *48*(3), 62–63.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*(2), 211–240.
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- Miller, G.A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*(11), 39–41.
- Ramesh, A., Goldwasser, D., Huang, B., Daume, H., & Getoor, L. (2014). Understanding MOOC Discussion Forums using Seeded LDA. In *9th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 28–33). Baltimore, MA: ACL.
- Rebedea, T. (2012). *Computer-Based Support and Feedback for Collaborative Chat Conversations and Discussion Forums*. (Doctoral dissertation), University Politehnica of Bucharest, Bucharest, Romania.
- Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, *31*, 581–603.
- Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., & Pritchard, D. E. (2014). Who does what in a massive open online course? *Communications of the ACM*, *57*(4), 58–65.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27*, 379–423 & 623–656.
- Trausan-Matu, S., Dascalu, M., & Dessus, P. (2012). Textual complexity and discourse structure in Computer-Supported Collaborative Learning. In S. A. Cerri, W. J. Clancey, G. Papadourakis & K. Panourgia (Eds.), *11th Int. Conf. on Intelligent Tutoring Systems (ITS 2012)* (pp. 352–357). Chania, Greece: Springer.
- Trausan-Matu, S., Stahl, G., & Sarmiento, J. (2007). Supporting polyphonic collaborative learning. *E-service Journal*, *Indiana University Press*, *6*(1), 58–74.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Wang, Y. (2014). MOOC Learner Motivation and Learning Pattern Discovery. In J. Stamper, Z. Pardos, M. Mavrikis & B. McLaren (Eds.), *7th Int. Conf. on Educational Data Mining* (pp. 452–454). London, UK.
- Wen, M., Yang, D., & Rose, C. P. (2014a). Linguistic Reflections of Student Engagement in Massive Open Online Courses. In *Int. Conf. on Weblogs and Social Media*.
- Wen, M., Yang, D., & Rose, C. P. (2014b). Sentiment Analysis in MOOC Discussion Forums: What does it tell us. In J. Stamper, Z. Pardos, M. Mavrikis & B. M. McLaren (Eds.), *7th Int. Conf. on Educational Data Mining* (pp. 130–137). London, UK.

## Acknowledgments

This research was partially supported by the FP7 2008-212578 LTfLL project, by the 644187 ECH2020 *Realising an Applied Gaming Eco-system* (RAGE) project, by University Politehnica of Bucharest through the “Excellence Research Grants” Program UPB–GEX 12/26.09.2016, as well as by the NSF grant 1417997 to Arizona State University.

# Anchored Annotations to Support Collaborative Knowledge Construction

Justin Plevinski, Jennifer Weible, and Michael DeSchryver,  
pleviljs@cmich.edu, j.weible@cmich.edu, desch2m@cmich.edu  
Central Michigan University

**Abstract:** Online discussion forums have been shown to support collaborative reflection, critique, and construction of knowledge. However, when these discussions are reading-centered and hosted in traditional threaded discussion forums, it is often difficult to maintain focus on the readings themselves, as well as navigating the difficulty of attributing ideas to just one discussion post within a thread. This study demonstrates how an anchored annotation system was used in a first semester online, asynchronous doctoral course to support more effective reading-centered knowledge construction. We investigated the discussion activities of 12 online doctoral students as they explored seven articles during two non-consecutive weeks. Through analysis of 591 student comments and replies, we examined how students used the anchoring system as a support to help make sense of the articles. We found that using anchored annotation reduced coordination activities and supported knowledge construction, particularly interpretation and elaboration of ideas.

## Introduction

Asynchronous discussion is a common collaborative activity in both undergraduate and graduate online courses. These discussion forums are useful for reflection, communication, and knowledge construction (Gao, Zhang, & Franklin, 2013). They may also facilitate deeper understanding of concepts (Gilbert & Dabbagh, 2005). Concerns with the quality of interactions have been related to the design of typical threaded forums, leading to development of innovative environments intended to support more effective discussions (van der Pol, Admiraal, & Simons, 2006). To better support online doctoral students' construction of knowledge, we explored a tool, Framebench (Agarwal, 2015) that allowed for online, collaborative annotations to be anchored within the text of assigned readings.

In this paper, we synthesize contributions from previous work in computer supported collaborative learning to frame our study. Then, we describe Framebench (Agarwal, 2015), an anchored system that we used in the context of reading-centered academic discussion, and explore its affordances as compared to prior research on linked annotation systems. Thereafter, we present findings regarding the knowledge building patterns of 12 online doctoral students over the course of first semester readings using this anchored system as the underlying tool to support asynchronous discussion focused on the text. Finally, we discuss the implications of these findings for future research and practice.

## Conceptual framework

Readings, and discussion about them, are a common activity in most graduate seminars. In its simplest form, such conversation involves teacher-asked questions, followed by student answers, followed by teacher feedback or responses. However, this linear, teacher-directed model does not reflect the generative discourse patterns in effective knowledge building communities (Scardamalia & Bereiter, 1994). In this way, knowledge building results from collective understanding through interaction, often supported by technology within the learning environment. As such, students must articulate their ideas, questions, and responses to demonstrate insights and understandings about the text in online discussions focused on course readings (Gao, Zhang, & Franklin, 2013). In online classes, studies of this type of socially constructed knowledge have often focused on discussion forums, or threaded discussions (e.g., Hew & Cheung, 2010). The content of these discussion forums may center on various intellectual artifacts such as topics, problems, or case studies drawn from the readings. However, because of the various affordances and constraints of systems designed to support these discussions, the actual dialogue within the forum is typically physically separated from the text of interest. For instance, a common method utilized when engaging students in a browser-based threaded discussion is to open one browser window for the threaded discussion while a PDF of the reading is available in another window or tab. In this way, the reading itself is distal to the related discussion. Eryilmaz et al. (2013) found that this distance may increase coordination efforts (cognitive expenditures needed to align the text and the discussion) and decrease energy and effort available for knowledge construction as students navigate back and forth between the reading and the

forum discussion. This leads to lower numbers of references to the related readings and reduced understanding of the content. As such, many threaded forum discussion posts do not demonstrate a rich understanding of content, nor do they fully integrate the key ideas from the related readings (Hewitt, 2001).

For this study, we adopted a constructivist approach to learning in which the individual's knowledge construction is based on their individual and social practices (Pena-Shaf & Nicholls, 2004). In so doing, we mapped knowledge construction activities onto six cognitive processes including remembering, understanding, applying, analyzing, evaluating, and creating (Krathwohl, 2002). We were particularly interested in how remembering and understanding supported the comprehension of assigned readings. That is, remembering involves retrieving and recalling what a learner knows, and understanding is the act of "[d]etermining the meaning of instructional messages" (Krathwohl, pp. 215). Included in the understanding category are cognitive actions of interpreting, classifying, summarizing, inferring, comparing, and explaining (Krathwohl).

We also situated this study in the computer supported collaborative learning environment (CSCL) of anchored discussions, which emphasize social interaction (Gao, Zhang, & Franklin, 2013) and "offer creative activities of intellectual exploration" (Stahl, Koschman, and Suthers, 2006, p. 410). Using a collaborative, computer-based environment encourages students to make their thinking visible to others by articulating their learning socially (Chu & Kennedy, 2011). These visible demonstrations help students expose their understanding to other students, explore different views, and contribute to shared understanding, which may deepen learning processes (Stahl, Koschman, and Suthers).

Previously Guzdial and Turns (2000) used an anchoring system called CaMILE as a way of improving online course discussions. They anchored specific notes onto web pages that were linked to the CaMILE system. By clicking on a note anchored within a web page, threaded discussion comments appeared on a different web page. Although the notes and discussion comments were connected, they were linked on two separate web pages. Using the CaMILE system, Guzdial and Turns' research focused on improved sustained on-topic forum discussions. Other authors have used more advanced features of the web for different types of anchored discussions in order to reduce the cognitive load due to distance between the reading and the discussion. One way of reducing the distance between readings and discussion is the use of linked discussions; to integrate the discussion forum and reading together in an environment and "link" them to each other (e.g., Eryilmaz et al., 2013; van der Pol et al., 2006). These previous studies of "anchored" discussion systems explore what are termed "parallel linked systems" where the discussion forum and artifact are presented side-by-side on the screen, and then linked visually via hyperlink. In these parallel linked systems, the distance between discussion and text is less than in traditional threaded discussion forums, but coordination activities that represent extraneous cognitive load still exist. However, using document-mediated systems such as linked discussions may provide a stronger collaborative context and "direct users' collaborative intentions towards the processing of that text" (van der Pol et al., p. 344), while reducing the need for coordination of the collaborative process.

Building on linked discussions, other formats supporting collaborative discussion from the anchored perspective also exist. These systems have been studied through the lenses of constrained environments, visualized environments, and anchored arguments (e.g., Gao, Zhang, & Franklin, 2013). The latter demonstrate truly anchored discussions through online annotation systems that allow for questions, comments, and replies to be embedded directly on the text of readings, which may reduce cognitive load and related coordinative activity (e.g., Annotate, Hypothesis, Diigo). However, across the annotation systems currently available, many do not cleanly accommodate the discussion and interaction patterns desired for academic discussion of the text. For instance, some systems allow for anchored comments within the text, but multiple replies to that comment result in a long window on the screen that extends on to the next page, rendering the anchoring more distal and less useful. Some other systems allow anchoring, but the notes cannot be collapsed. As such, students who encounter the reading after other classmates have annotated the document find a document covered with notes, disrupting their ability to read the original text. All of these activities may have an "interaction cost" that is either social or coordinative (van der Pol et al., 2006), and may reduce the effectiveness of a CSCL environment if the costs are too high. Coordinative costs (mental capacity used to reference evidence) are particularly persistent in traditional asynchronous text based online discussions (Eryilmaz et al., 2013). For instance, some systems leave too much coordination and structuring to the students, particularly for maintaining a shared frame of reference (Häkkinen & Järvelä, 2006). This is often in the form of specific references to the learning artifact from within the discussion contributions (Herrmann & Kienle, 2008).

Because of this problem, Eryilmaz et al. (2013) explored the coordination costs of establishing and maintaining shared focus in proximal online readings using a linked artifact-centered discourse system, finding that decreasing the need for coordination costs in an online discussion environment increased knowledge construction potential. They utilized the Annotation Tool (Van der Pol et al, 2006), which presents both discussion threads and the readings in the same window in two separate frames, using links between the image

of a text page and the discussion threads for two-way reference between them. Previous research indicated this sort of “anchoring” may stimulate sustained on-topic conversations (Guzidal & Turns, 2000), encourage messages on specific points in the reading (Häkkinen, et al. 2002), and help students better engage with complex ideas (Suthers, et al., 2006). Eryilmaz et al. found participants demonstrated less coordination activities and increased knowledge construction activities. However, although linked artifact systems like the Annotation Tool significantly reduce the navigational distance between discussion and artifact, they do still demonstrate spatial distance across the screen and window.

Framebench software (Agarwal, 2015) uses a more advanced anchoring system whereby the annotation system is integrated directly within pdf documents. The threaded discussions appear as a collapsible floating window on the pdf document itself rather than being linked through a separate webpage. We focused on the students’ collaborative knowledge construction through threaded discussions anchored directly at the point of reference within the pdf document text. As such, given that linking discussion and text across the screen reduced coordination activity, we explored whether proximally anchored annotations might reduce it even more, leaving greater energy and effort to expend in knowledge construction activities.

## Methodology

### Research questions

The goal of this research was to examine the affordances of anchored annotations for both reducing coordination activities and supporting construction of knowledge. In this analysis, we answer the questions: (1) How do anchored annotations support student coordination activities?, and (2) How does using anchored annotations in a learning environment support student collaborative knowledge construction activities?

### Selection of the online system

We evaluated options for anchored discussion that connect the discussion in even closer proximity or directly on the text itself than linked systems. The tool used for this study was Framebench (Agarwal, 2015) that afforded two features integral to supporting the proximal anchoring. First, initial comments are embedded with only a small icon so that the text is “clean” when readers return to it, but comments are readily available with a single click. Second, Framebench uniquely encapsulates comments and replies in a single window where readers can scroll through them, still anchored to the text, using very little screen real estate. In most other systems, the four replies demonstrated by the open comment window (See Figure 1) would extend on to the next page. Therefore, this environment provided several key affordances of anchored discussion to support reading-focused knowledge construction that are not available in other currently available tools.

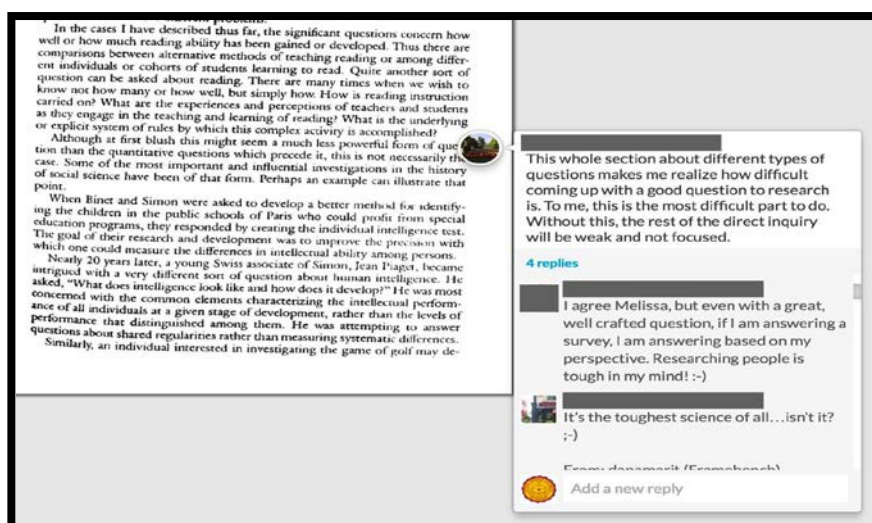


Figure 1. Screenshot of a comment and responses encapsulated within Framebench’s anchored annotation format.

## Participants and settings

For this study, participants (n=12) were students enrolled in an entirely online doctoral program in educational technology. Ten were from states within the United States, and two were international students. The setting for this study was a first semester course intended to provide an overview of educational technology research and the scope of the field. Students were assigned four sets of readings to be discussed in Framebench (Agarwal, 2015) during weeks two, six, ten, and 14 (other synchronous and asynchronous options were utilized in the remaining weeks of the course for weekly discussion), of which we selected week six and fourteen for analysis. Week six was chosen to ensure that all students were already familiar with the Framebench system and were able to access it without difficulty, and week 14 was chosen to analyze for changes in interaction patterns across the semester.

## Data, findings, and analysis

Data examined for this analysis were 591 anchored annotations (both initial posts and replies) written by participants across nine articles assigned during the two selected weeks. Each comment and all replies were logged into a spreadsheet in their entirety, as well as word counts, number of comments and replies, and number of highlights and other written annotations from within the documents. Each comment and reply were segmented into activities (Erilmaz et al, 2013), which were coded based on *a priori* codes for coordination (Erilmaz et al.) and knowledge construction activities (Krathwohl, 2002; Pena-Shaff & Nicholls, 2004). We also allowed for emerging codes (see Table 1). These emergent codes were developed, identified, and defined by all researchers before being utilized within the coding scheme. In particular, we modified the framework of Pena-Shaff & Nicholls to include a more nuanced understanding of clarification, interpretation, and elaboration that incorporated ideas from Krathwohl. Clarification was used to denote when ideas and thoughts were identified, while elaboration was added to differentiate between clarification and allow for coding of other activities (see Table 1 for definitions and examples). Interpretation was defined as “inferences, conclusions, summaries, generalizations, problem solution suggestions, or hypotheses” and elaboration was defined as expanding the scope of the current discussion by adding additional information or examples (Pena-Schaff & Nichols, pp. 257). Aligning with Krathwohl, elaboration as a cognitive skill falls into both the remembering and understanding categories: elaboration relies on long-term knowledge (remembering) to provide relevant, related examples that demonstrate understanding the content of the text. The cognitive actions in understanding, according to Krathwohl, are the same cognitive actions used to define interpretation. Descriptive statistics for each code category as well as percentages of the coded activities were calculated. Two researchers coded all segmented annotations separately and discussed differences to reach consensus to provide validity.

Table 1: Conceptual Framework, Definitions, Examples, and Percentage

<b>Categories</b>	<b>Operationalization</b>	<b>Example of the applied code</b>	<b>Percent</b>
<i>Knowledge Construction Activities</i>			
Interpretation	Inferences, conclusions, summaries, generalizations, problem solution suggestions, hypotheses	<i>“This supports the idea that active learning is beneficial to students”</i>	21%
Question	Seeking to find additional information pertaining to the discussion; prompt further discussion about the current topic; a question that reflects upon the current discussion	<i>Are there other content areas that may benefit also, why science (other than the video was available)? Couldn't there be an option to learn from books, videos, games, whatever best matches the student with perhaps formative assessments along the way for students to see if they are learning?</i>	14%
Conflict	Disagreeing with another student; mentioning a different point of view in direct reply	<i>“Perhaps I misread this first sentence, but I'm not sure I agree with this statement”</i>	2.5%
Consensus Building	Discussion of misunderstandings; reaching an	<i>“I too believe that these assumptions are more vital than just to label them "assumptions.””</i>	6%

	agreement on an idea, fact, or interpretation; negotiating the definition, interpretation, or truthfulness of a claim or fact		
Support	Empathizing; statements of acknowledgement; providing direct feedback	<i>Tammi, I also see a big problem here</i>	12.4%
Clarification	Listing main ideas, facts from the reading, assumptions	<i>"aka - a student-teacher scenario."</i>	1.8%
Elaboration	Connecting ideas with examples, defining terms, causes or consequences, listing advantages or disadvantages, using analogies to explore ideas, making connections, comparing and contrasting	<i>"Gee's background in sociolinguistics is shining through here in this section. :)" "This happens with adult learners who tend to be focused on learning just what is needed to achieve the task at hand"</i>	21%
<i>Coordination Activities</i>			
Distal	References to information outside of the text.	<i>"I think it's in Steinkuelher's article that she questions why relate gaming to TV or radio, etc."</i>	5.9%
Proximal/Far Proximal	References to information either within the same paragraph or within the text.	<i>"This paragraph rings a bell about student empowerment &amp; motivation that comes when students can take control of their learning-actively."</i>	1.43%
<i>Other Emerging Codes/categories</i>			
Social Interactions	Comments directed at participants but not at the task at hand.	<i>"Also, it's my 3 year old who watches Caillou. My daughter would be mortified if she knew I mixed that up :-)"</i>	2%
Tangential	Comments peripherally related to the topic.	<i>"J---, I read a great article this week, also by Gee, talking about games and libraries."</i>	3%
Directions	Directions or guidance from the instructor.	<i>"This is the kind of annotation that is helpful to everyone else, and demonstrates the kind of reading I'd like to see you all practice. If you don't know something, it can be easily explored by our friend Google.."</i>	.23%
Confusing	Comments that did not make sense within the context.	<i>"My english-speaking peer mentoring partnership based on the research findings, recommendation of need and lack of english-speaking and transitioning support for Domestic students and recommendations."</i>	.3%

Table 2: Data Grouped by Category

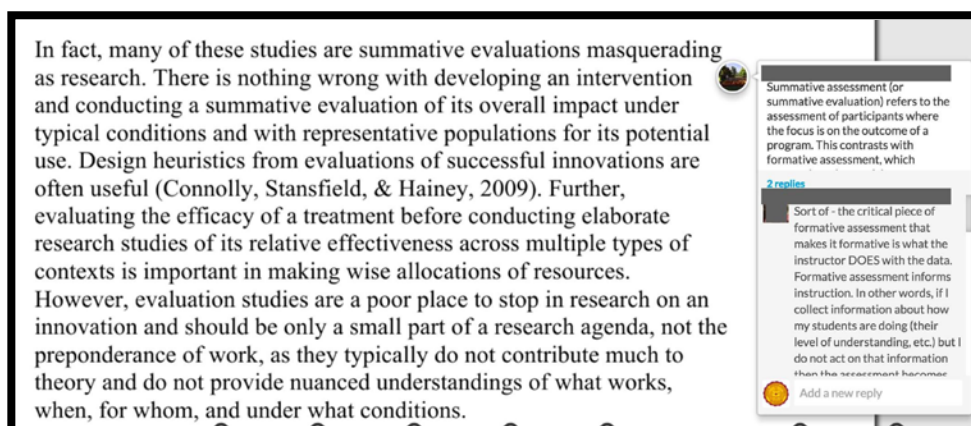
<i>Activities by Category</i>	<i>Week 6</i>	<i>Week 14</i>	<i>Total</i>	<i>Percentage</i>
<b>Coordinating (focusing on shared topic of discourse)</b>	<b>40</b>	<b>39</b>	<b>79</b>	<b>6.7%</b>
<b>Knowledge construction (interpretation and elaboration)</b>	<b>221.5</b>	<b>234.5</b>	<b>456</b>	<b>38.6%</b>
Knowledge construction (other)	284.5	290.5	575	48.6%

Other	26	26	52	4.4%
Total number	578	604	1182	100%

## Discussion

### Anchored annotation reduces coordination activities

Within our study, coordination activities accounted for 6.7% of the overall activities (see Table 2). Considering that Erilmaz et al. (2013) found that between 17% and 26% of all activities were considered to be coordination activities in a parallel linked annotation system, our findings indicate that a within-text anchored annotation system (such as Framebench) may decrease coordination effort and allow more focus on construction of knowledge. This may have resulted from the proximal nature of the anchored comment. That is, when a comment is embedded in the text of a reading, the content of the comment is clearly referencing the text near or on which the comment is connected. For instance, a student comments “Summative assessment (or summative evaluation) refers to the assessment of participants where the focus is on the outcome of a program. This contrasts with formative assessment...” (see Figure 2). Another student replies, “Sort of - the critical piece of formative assessment that makes it formative...” Due to the proximity of the user’s icon to this paragraph, and even more precisely, aligning with the first two sentences of the paragraph, there is a seamless discussion of the authors’ contentions about summative evaluations in a research context. As indicated, there is little need for specific references to the text on the part of the initial commenter, and subsequent repliers need not search for intended references.



**Figure 2.** Screenshot of anchored annotation of student discussion in Framebench in which they are developing collaborative understanding of summative vs. formative evaluations.

In Eryilmaz et al., participants used specific coordinative references such as “the author’s argument on page 8” six percent of the time, and another 11% of comments were used to maintain that focus. We propose two possibilities for why the anchored system within Framebench (Agarwal, 2015) may reduce such references in comparison to the linked system. First, the anchored annotations are located immediately within the text, whereas in the linked system, the screen is covered in multiple threaded comments on the left side of the browser and the document itself can be covered in several colored highlights. Within the threaded systems, students may find that in spite of all of these cues connected across the document and browser, references to the text are still necessary or useful.

Second, instructions provided to the learners in each context may have primed them for the functionality of the system. That is, in this study, given that the students were doctoral students, the instructor, was transparent upon introducing Framebench about the apparent and intended affordances of the system. Students were told that they were using a discussion system that might help them closely explore the text of assigned readings with embedded comments. Specific language in those instructions may have indicated, explicitly or implicitly, that the proximity of the comments reduced the need to make specific reference to the related text. Eryilmaz et al. indicated that students were “briefed about the functionality of the utilized



interaction environment” (p. 125). But, it is unclear if this briefing may have underrepresented the connective affordances of the systems, or if identical instructions were given to the treatment and control groups, which would not have highlighted the potential benefits of linked-parallel artifact systems.

### **Anchoring annotation supports remembering and understanding of text**

In our study, anchoring annotations within the text supported multiple types of knowledge construction activities, primarily elaboration and interpretation. The number of activities within our data set coded as elaboration and interpretation statements within the anchored annotations accounted for 42% of all comments or replies (21% of comments coded elaboration and 21 % of comments coded interpretation). Finding that the use of anchored annotations may foster knowledge construction activities that align with remembering and understanding (Krathwohl, 2002) more than other types of knowledge construction activities, such as consensus building and synthesis, may be due to the manner in which the tool handles instances of annotations. Each annotation can be anchored immediately beside the text referenced, and each article contains numerous opportunities for threaded annotated discussions near the text that spurred the initial comment for the threaded discussion. In alignment with Gao, Zhang, and Franklin (2013) who found that anchoring annotations reduced the localization effect (the given distance of text’s proximity to the location of the annotation is low, thus moving from one space to another changes the qualitative nature of the annotation), our students changed focus with every post. Within our study, students created multiple, shorter length comments and replies, often within the same paragraph, that focused on a small portion of the text instead of constructing a longer post that synthesized across several paragraphs or ideas. We posit that this ease of “targeted” annotations facilitated students’ remembering and understanding of the content, but hindered synthesis across texts and consensus building activities between the students.

Furthermore, the way in which Framebench (Agarwal, 2015) was implemented within the course may support the knowledge construction activities of remembering and understanding more so than other knowledge construction activities because of the length and the number of readings. Each week there were four to five scholarly readings of extended length (well over 100 pages per week), and students could annotate anywhere in the documents. That is, in addition to the focused nature of the annotations within Framebench, the sheer volume of pages to read and annotate may have encouraged the knowledge construction activities of understanding and remembering, key steps for comprehending the assigned readings, in place of synthesis or evaluation.

### **Conclusion and implications**

This study explored how first semester, online doctoral students utilized a text anchored annotation system to make sense of and discuss their weekly readings. Just above, we outlined how our data indicated that coordination activities were relatively minimal and that this system supported knowledge construction activities closely aligned with the cognitive processes of remembering and understanding. Some might interpret these findings to be disconcerting in the context of knowledge construction. That is, given remembering and understanding are often considered “lower-order” skills in the context of Krathwohl’s (2002) presentation of cognitive processes, it would be easy to dismiss the value of in-text anchored annotation systems, since the current study did not promote the more generative “higher-order” processes often associated with knowledge construction. In practical terms, however, the online students used the affordances of Framebench’s in-text, anchored annotation system as a support while they processed broad, foundational readings needed to make sense of the field. As such, using anchored annotations in online learning environments may have positive implications for providing students with a social method to explore and understand complex readings.

This being the case, it is also important to emphasize the specific goals of the course in which this study was conducted. Situated in the first semester of an online doctoral program, the primary objective in this survey course was to ensure that students absorbed the key ideas in the assigned readings from across 14 different genres of education and educational technology research. Absent the more traditional conversational style of a face-to-face doctoral survey course that affords the instructor ample opportunities to assess the extent to which students are internalizing assigned readings, we specifically endeavored to find a system that would support such discussions and assessment in an asynchronous manner. Given the tendency of threaded discussion to stray away from the readings, we utilized Framebench as a way to focus students on the specific text elements within the readings through visible discussion anchored at that point.

There are currently several styles of systems available to support discussion in online learning: embedded anchors, standard threaded discussion forums, and parallel linked anchors. We posit that in choosing a discussion system for online learning, it is important to consider (a) the goals of the online discussion (e.g. understanding vs. knowledge construction), and (b) the design affordances that undergird the system (e.g.

reducing coordination costs vs. creating connections). Future research may provide continued exploration of the tradeoffs between and among these considerations. Furthermore, this study may also inform design principles for the creation of additional systems that utilize anchored annotation within text.

## References

- Agarwal, R. (2015). Framebench.com [computer software]. India.
- Chu, S. K-W., & Kennedy, D. M. (2011). Using online collaborative tools for groups to co-construct knowledge. *Online Information Review*, 35(4), 581-597.
- Eryilmaz, E., van der Pol, J., Ryan, T., Clark, P. M., & Mary, J. (2013). Enhancing student knowledge acquisition from online learning conversations. *Journal of Computer-Supported Collaborative Learning*, 8, 113-144.
- Gao, F., Zhang, T., & Franklin, T. (2013). Designing asynchronous online discussion environments: Recent progress and possible future directions. *British Journal of Educational Technology*, 44(3), 469-483.
- Gilbert, P. K. & Dabbagh, N. (2005). How to structure online discussions for meaningful discourse: A case study. *British Journal of Educational Technology*, 36(1), 5-18.
- Häkkinen, P., & Järvelä, S. (2006). Sharing and constructing perspectives in web-based conferencing. *Computers & Education*, 47(4), 433-447.
- Herrmann, T., & Kienle, A. (2008). Context-oriented communication and the design of computer-supported discursive learning. *International Journal of Computer-Supported Collaborative Learning*, 3(3), 273-299.
- Hew, K.F., & Cheung, W.S. (2010). Fostering higher knowledge construction levels in online discussion forums: An exploratory case study. *International Journal of Web-Based Learning and Teaching Technologies*, 5(4), 44-55.
- Hewitt, J. (2001). Beyond threaded discourse. *International Journal of Educational Telecommunications*, 7, 207-221.
- Jonassen, D., Davidson, M., Collins, M., Campbell, J., & Haag, B. B. (1995). Constructivism and computer-mediated communication in distance education. *The American Journal of Distance Education*, 9(2), 7-26.
- Krathwohl, D. R. (2002). A revision of Bloom's Taxonomy: An overview. *Theory Into Practice*, 41(4), 212-218.
- McGinley, W. (1992). The role of reading and writing while composing from sources. *Reading Research Quarterly*, 27(3), 226-248.
- Pena-Shaff, J. B. & Nicholls, C. (2004). Analyzing student interactions and meaning construction in computer bulletin board discussions. *Computers & Education*, 42 (3), 243- 265.
- Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 97-118). New York: Cambridge Press.
- Stahl, G., Koschmann, T., & Suthers, D. D. (2006). Computer-supported collaborative learning. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 409-425). New York: Cambridge Press.
- van der Pol, J., Admiraal, W., & Simons, P. R. J. (2006). The affordance of anchored discussion for the collaborative processing of academic texts. *International Journal of Computer-Supported Collaborative Learning*, 1(3), 339-357.

# Collaborative Game Design: A Bounded Case Study of Undergraduate Students in a Capstone Course

Helen Fake, George Mason University, hfake@gmu.edu  
Jennifer Crist, George Mason University, jwhitema@gmu.edu  
Brenda Bannan, George Mason University, bbannan@gmu.edu

**Abstract:** In this deductive qualitative case study, researchers observed the interactions of undergraduate students in a Capstone Game Design Class at a mid-Atlantic University. Referencing the literature, the researchers found that all facets of the collaborative framework to include governance, administration, organizational autonomy, mutuality, and norms were represented in ethnographic observations and focus group sessions. Specific findings regarding the collaborative game design process as well the social and cultural dimensions of game design are discussed.

## Game based learning

According to a research report from Ambient Insight Research, an eLearning and mobile predictive analytics firm, the Game Based Learning (GBL) industry is projected to hit revenues of \$2.6 billion in 2016 and grow to \$7.3 billion by 2021 (Ambient Insight Research, 2016). Coupled with the research firm's prediction of a 22.4% industry growth rate from 2016 - 2021, it is evident, from a business perspective, why there has been considerable buzz and excitement surrounding GBL. With concerns about shortening attention spans and how the internet is impacting our collective ability to focus, GBL and its potential to engage 21st Century Learners has garnered enthusiastic responses from both Academic and Organizational communities (Carr, 2011; Duncan, 2013).

From a pedagogical standpoint, projections of growth in the GBL industry are not unfounded, especially given the increased interest in GBL in literature. Recent research about GBL suggests playing games as an instructional strategy has positive cognitive and emotional benefits. Research indicates that GBL supports statistically significant improvements in three areas of problem-solving (Akcaoglu & Koehler, 2014), feelings of self-efficacy (Ke, 2014), and higher levels of motivation, critical thinking, achievement and engagement (Robertson & Howells, 2008; Yang & Chang, 2013). In addition to these encouraging outcomes, Mayo (2009) found that when games are well designed, they can improve performance from 7 to 40 percent compared to students who attended lectures.

These findings beg the question what "well designed" games look like. While there is considerable excitement surrounding the GBL movement, research firms have cautioned about the zealous fervor surrounding GBL initiatives, emphasizing the need for solid designs to back these pedagogical and organizational efforts. For example, Gartner, a leading technological research company, anticipates that while 70% of companies will implement a gamified solution, 80% of those initiatives will fail to meet business objectives primarily due to poor designs (Gartner, 2014). Projections from leading Research and Advisory firms, as well as known gamification failures like Google Badges, have encouraged increased interest and focus on the game design process (Lunden, 2012).

While there is a tremendous amount of attention given to the efficacy of GBL as an instructional method, and focus on the process of designing games from an industry and academic perspective, there have been limited studies that specifically focus on collaboration, a critical component to effective design, within game based learning design contexts. We explore collaborative frameworks in more detail below.

## Collaboration

Disparate entities working together to achieve one purpose or a shared vision that could not be attained individually is recognized as the defining characteristic of collaboration (Gadja, 2003; Woodland and Hutton, 2012). This complex concept can be examined from a variety of perspectives. The research available on the topic reflects this complexity while also contributing to the challenge of measuring collaborative practices. Our research suggests there is a gap in the current literature regarding collaboration during the game design process. The subsequent review of the literature aims to frame our study within the existing understanding of collaboration while addressing additional gaps and difficulties in researching this topic. Additionally emerging collaborative theories were explored as potential frameworks for our study.

The highly competitive nature of the game design industry lends itself to an expectation that students in a game design program will have an understanding of common practices and knowledge of the field. One way to bridge gaps between these expectations and the realities of student experience is for academic institutions to partner with established industry experts. De Frietas, Mayer, Arnab, and Marshall (2014) present a theory by which these partnerships, in addition to the inherent learning support, can support research funding while accelerating commercial development efforts. Findings from this model are further reinforced by research from Marcketti and Karpova (2014). Their study explores student perceptions of an industry partnership between design teams and an apparel company. Students perceived that they developed highly relevant knowledge and skills by tackling authentic problems.

While there are many instruments in the literature that measure collaboration, few feature instruments generalizable to a broad population or small groups. In our search for a study that would generalize to game based design environments we identified a study from Thomson, Perry, and Miller (2007). Featuring the responses from a survey sent to 1382 directors of organizations who participate in large national service programs, Thomson et al. (2007) synthesized the existing research on collaboration to test their multifaceted model. Using a higher order confirmatory factor analysis, they found that their collaboration framework was supported empirically. The collaborative framework proposed by the researchers highlights the importance of the following five factors in the process of collaboration: governance, administration, organizational autonomy, mutuality, and norms. Governance was defined as “understanding how to jointly make decisions about rules that will govern group members' behavior and relationships” (p. 3). While arguably an overlapping term, administration was explained as times where the “focus is less on institutional supply and more on implementation and management” (p. 4). Organizational autonomy is referred to as the “intrinsic tension between organizational self-interest” (p. 4) and achieving and keeping agents accountable for meeting their collaborative goals. In contrast, mutuality is described as the “mutual beneficial exchange based on difference in skillsets, resources, etc.” (p. 5). Finally, the researchers explain norms as an “I-will-if-you-will” mentality based on perceived degrees of reciprocal obligations each have toward the others” (p. 6). These categories provide a robust framework for capturing the overlapping complexities of collaborative efforts.

Research on collaboration within groups or teams designing games was not found. Only one study was identified that addressed collaboration at the team or intragroup level (Colbry, Hurwitz, & Adair, 2014). While the context of the Colbry, Hurwitz, and Adair (2014) study focused on small groups, it was exploratory and lacked the validity and reliability of the Thomson, et al. (2007) framework.

Given the gaps in the literature regarding collaboration in game design, we came up with the following research questions:

1. How do game design teams collaborate internally?
2. How do game designers collaborate externally with university & industry partners?
3. How do game designers create opportunities for collaboration within the game itself?

We hope to increase the understanding of how these three levels of collaboration interact with each other to impact final results in game based designs.

## Method

### Participants

Eight George Mason University undergraduate students over the age of 18 participated in this study. GAME 490 is the Senior Game Design Capstone course for the Computer Game Design Program at a mid-Atlantic University. This interdisciplinary class is open to students from a range of degree programs in the arts and sciences, providing an excellent opportunity to observe collaboration between multiple operating paradigms in action. Informed consent was obtained from all participating students following a description of the study.

While there were 8 individuals that submitted their consent, only four were a part of the same game design team. In an effort to uphold our commitment to ethical research and mitigate any potential unauthorized observational recordings, we decided to focus our observational efforts towards the members of the intact team. By focusing our observations on the single participating team, we were also able to record the interactions and collaborations between team members.

### Procedures

The study was introduced to the potential participants by the student researchers at the beginning of a class meeting. The first observation occurred midway through the semester; therefore, many procedural aspects had been established. The impact of this starting point is addressed in our findings and discussion.

The presenting researcher described her interests in cultivating a better understanding of how the students collaboratively design games. The study introduction was less than 15 minutes in duration and allowed students to take the time to consider their participation in the study as well as a chance to ask questions. Willing participants submitted their consent forms prior to the end of class. Following the attainment of consent, forms were collected and stored in a secured location.

In order to better understand the collaborative practices of students in the game design process ethnographic observations and focus group data were collected and analyzed. We will go into further detail about these data collection methods in the next few paragraphs.

### Observations

An ethnographic observational approach was pursued in recognition that, "...there is an intimate connection between a culture and its designed objects leads us to advocate for an ethnographic approach to research design where cultural practices are the focus of inquiry." (Crouch & Pearce, 2012, p. 84). In total, the collaborative practices of students were observed in three classroom sessions in the Arts Building at the mid-Atlantic University campus in Fairfax, Virginia during three sequential class meetings. Ethnographic notes were taken throughout the duration of these sessions during the class's normal hours which spanned a three hour period. The classroom setting consisted of five pods of computers organized in hexagonal arrangements. Screens on the desks consisted of large format tablets. The observed group congregated in a pod near the windows and frequently shared their space with one or two other smaller groups. Occasionally, a group member in the observed group would shift to the most proximal seat at the central pod due to lack of space. To limit interference in the group's dynamics, we sat two to three feet away from their space in unused chairs from the central pod. Due to the hexagonal nature of the pods, we did occasionally change perspective by sitting nearer to the central pod, allowing us to observe different group members in action.

### Focus groups

A semi-structured focus group was held following the first classroom observation. The focus group was in a small cramped study room within the student community center. Three attendees participated in the focus group. A second focus group planned for after the last observation was cancelled due to lack of participants.

After reading the protocol aloud, participants were instructed to respond to questions posted on the walls of the meeting room. Using post-it notes, the participants responded to the research questions during the first 20 minutes of the session. The questions included the following:

1. What challenges of collaborating on a design project have you faced?
2. In your opinion, what have been the advantages of collaborating with a team for this game design project?
3. How has your team integrated collaboration into your game design? If applicable, have you collaborated with any external stakeholders in this project?
4. How does your team collaborate?
5. What is collaboration?

At the end of the allotted twenty minutes, the participants were asked to rejoin the group and discuss their answers question by question. During this facilitated discussion, observations and notes were taken. This debrief allowed the researchers to ask clarifying questions and dig deeper into the collective meaning-making of the group as a whole.

## **Results**

Given our small sample size, we pursued a bounded deductive case study approach to our data analysis process (Glesne, 2011). The data collected during the ethnographic observations and focus group was coded line by line while referencing a code manual based on the Thomson et al. (2007) theoretical framework of collaboration. The predetermined categories we looked for included governance, administration, organizational autonomy, mutuality, and norms. Data was aggregated into an Excel spreadsheet and then each researcher individually coded the observational and focus group data using a line by line coding approach. Following our individual coding, we compared codes for emergent themes and to also to ensure inter-rater reliability. Findings from our data analyses are presented in the paragraphs below. Addressing our multi-tiered questions, these are framed within three levels of context: the group, the course, and the university.

## **Governance**

As stated previously, governance is “understanding how to jointly make decisions about rules that will govern group members' behavior and relationships.” (Thomson, et al, 2007) Our interpretation of this terminology combined with the analysis of observation data resulted in three findings at the group level during observations. The focus group revealed additional constructs provided within the course structure and game design profession that support decision-making processes.

### Observations

The first finding was that governance within the group was fluid. Decision-making appeared to occur naturally, with one group member eliciting feedback at the moment of need. When the activity appeared critical to the overall game design or was significant in scope, such as the addition of lighting effects, all group members that were present would stand up to view the screen in question, review the information, and provide input. Decisions with less impact such as minor edits to graphics elicited the formation of dyads or triads. Occasionally, seemingly random comments inspired movement around the pods so that members could view each other's computer screens.

There was no evidence of clear leadership within the group, although it appeared that the individual in charge of the task being reviewed would implement or reject recommended changes. This supports the second finding that the team as a whole respected the domain expertise of the individual members. The artists and programmers made the final decisions in their respective specialities. Decisions in areas which crossed domains were made collaboratively by members of both domains. For example, the implementation of graphic tilesets within the programming software occurred through a dyadic dialogic interaction in which a programmer moved the mouse while an artist used the stylus to draw on the tablet screen.

The final finding was that governance was driven by anticipated results. The observed group maintained a focus on the course objective as evidenced in the course stand-up procedures in which progress and effort were reported. Besides the stand-up reporting procedures, the course provided additional constructs that guided governance within the group, such as the professor's process for circulating and checking in with groups and the scheduled deadlines of specific assignments.

The environment established by the professor was casual. At this level, an overlap between governance and norms was identified. As such, this theme is further explored within the analysis of norms.

### Focus group

During the focus group, additional constructs provided by the course and the group were identified. Group members expressed that game design documents are a critical governing process in the industry. Created at the beginning of a project, the group revealed that they relied heavily on the game design documents to guide their processes and progress. They shared that the game design document helped them establish group consensus on a shared final goal early in their game development process. This documentation also served to guide the creation of a portfolio piece with a defined scope (the vertical slice required for the course) and narrowed the team's focus on their end goal which also drove their decision-making processes. Participants referred to experiences in previous classes in which differences in group members' goals or intentions led to difficulties or challenges. They felt that these experiences contributed to their current practices in collaboration.

### Administration

The administrative processes which set the stage for implementation and management were heavily guided by the course construct. This was evident during observations and the focus group. Technology also played a critical role.

### Observations

In a sense, the professor acted as an occasional team member through the stand-up procedures, check-ins, assignment definition, and deadline scheduling. The casual atmosphere and interactions helped create a seamless environment between the dynamics of the group and the expectations of the course. Evidence of the casual environment included instances in which group members arrived late to class bearing gifts of food for the professor. Laughter and humor ensued.

In another example, in one class session the professor, looking to conduct a check-in with the team, found that a majority of the of the group's members had left for a smoke break. Instead of expressing concern or frustration, the professor engaged in a relaxed conversation with the remaining teammate until the students returned. This interaction may indicate that a mutual understanding between the team members and professor had been reached. Some evidence was observed that students had previously demonstrated their ability to meet implementation and management expectations. There was clear evidence during check-ins with the professor that

the students had been highly successful in their efforts to date and were exceeding assignment expectations and deadlines.

At the group level, administration was highly procedural. Communication from absent or late group members was prompt, with reports of illness being shared early in the classroom session. Edits and revisions to the game design document provided a tasking system in which progress was checked against the proposed design. Future tasks and iterations were assigned by the group as deemed appropriate, including to absent members. It is possible that a more robust understanding of administration would be obtained through a more thorough research study, in which work and efforts outside of class could be observed.

### Focus group

The results of the focus group reinforced the findings from the observations. Participants highlighted procedural activities and the technology used to carry out those procedures. Team Members relied on the capabilities of Google Drive to work collaboratively on a document as well as email and texting for communication. Again, much of their reflection on this theme was in terms of past experience in which administration was lacking. Examples were cases in which group members were non-communicative or did not contribute to the implementation of the project.

## Organizational Autonomy

As stated earlier, organizational autonomy refers to the “intrinsic tension between organizational self-interest” (Thomson, et al., 2007, p. 4) and the obligations to the group. This tension was apparent in the interactions between members as they completed individually tasked portions of their work.

### Observations

Throughout the game design and development process, team members worked on their own segments of the project. Also, possibly due to the intensive nature of development, individuals were given the personal latitude to come and go as they pleased with several team members disappearing from the classroom session for several minutes at a time.

The tensions inherent in organizational autonomy were observed when team members questioned each other about the status of certain game assets and deliverables. In one instance, a team member was developing music and sound effects for the game. It was clear the group member’s commitment to creating high quality sound effects, specifically in the quest for the perfect rat sound. Despite this dedication to quality, another group member indicated their anxiousness for incorporating sound edits as soon as possible. The other team member felt that the sound effects chosen were sufficient for meeting the project’s needs. The group member in charge of sound effects quickly acquiesced to provide the audio files needed to move the project forward.

### Focus group

Our ethnographic observations regarding the game design process were reinforced during our focus group conversations. Group members indicated they were extremely selective in forming their groups and looked for specific characteristics which are discussed in more detail in the norms and mutuality sections.

Previous negative interactions and experiences from prior classes flavored group member expectations in relation to organization autonomy. Challenges brought about by organizational autonomy included working with “lone rangers” or group members who were “not doing work”. Organizational autonomy, therefore, could result in unfavorable outcomes with participants who did not pull their weight or those who were not open to working with others.

## Mutuality

In contrast to organizational autonomy, mutuality refers to the “mutual beneficial exchange based on difference in skillsets, resources, etc.” (Thomson, et al. 2007, p. 5). Lachmann, et al. (2013) referred to the concept of “flow” in which people become absorbed in their activities to a high degree. The observed group demonstrated both mutuality and flow in their ability to work together toward a common goal, achieving their collaborative goals.

### Observations

During the ethnographic observations, we witnessed several times when group members completed each other’s sentences. This level of sync between group members was inherent in their interactions. In one instance, two group members with different specialties discussed adding new features to the artwork that would later be integrated into the game. During this interaction, one group member stated that it was “hard when you have...”. The other responded immediately, “very few pictures to work off of”. Both quickly agreed and moved

to the next task. This interaction suggests that conversations between the two members had progressed to a level of mutual understanding wherein the development challenges and landscape was well defined and known despite the differences in each team member's specialty.

Another example in which disciplines combined to reach a mutually beneficial solution appeared when the group explored new features within their game development software platform. In one instance, one of the artists discovered a new "sprites" or graphic element that could bring light to objects within the game design. It was unclear, however, if their game would support this feature. Once this feature was discovered, the artist and one of the programmers worked together to investigate whether or not they could integrate the "sprite" into the game design. Each team member recommended different potential buttons or settings that might garner the hoped for results. After a short while, both stopped to research existing literature on the internet. This challenge eventually became a group wide endeavor. Check-ins with the professor in the following class revealed that the team was embracing new methods of resolving the issue to include trial and error tinkering and by submitting a help request to the software company.

A factor that appeared to contribute to and enable mutuality is the innate socialization that occurred within the group. Members consistently looked to each other for affirmation and humor. Side conversations and laughter were frequent. At the end of our final observation as students were wrapping up their conversation with the professor, one member explicitly stated that "everyone is engaged and helping out", "it's been a good experience", and "I'm happy with the team we have" followed by an invitation to the other group members to go get pizza and hang out. This is in line with evidence gathered during the focus group which we examine next.

### Focus group

During the focus group, clear evidence emerged regarding the social factors at play in mutuality. Participants expressed that they perceived the group formation and selection process had a strong effect on their collaborative success. Students clearly defined their criteria for group selection, with one student expressing that they were "really really really careful about who we picked". "Attitude", previous collaboration, social dynamics, "work ethic", and "communication skills" were all cited as critical attributes of potential team members. Additionally an off topic conversation on the perceived importance of socializing beyond the demands of class emerged. Team members felt it was important to spend time together, to stay friends, and keep it "fun and comfortable" as opposed to other groups who "only meet and talk during class". It was evident that the commitment to the design process and the final product were driven by the social experience of the group. Thomson, et al. (2007) acknowledge that "commitment is unlikely without the presence of the final defining dimension of collaboration: norms of reciprocity and trust" (p. 6).

### Norms

Like Gajda (2004) noted, it is quite possible that by the time we began our study, that the team of interest was far along in their process of assembling/forming, storming/order, norming/performing, and transforming/adjourning. By the time we began our study, the project teams were in the "performing" stage, actively editing and adding elements, functionalities, and details to their game design. Therefore, our observations may be less robust than if we began observations earlier in the game design process. Since we did not enter until the "performing stage" of team development, it is unclear if the team's established norms were naturally emergent or if they were explicitly established through conversations with team members. Norms were additionally influenced by the structure of the class, the professor, and the aforementioned development environment (Hartson & Pyla, 2012).

### Observations

As stated earlier, our observations within the classroom environment suggested that the development of game assets was a fluid process. Multiple times throughout the allotted class meeting, team members would leave for smoke breaks, food, or to take phone calls. Given the flexibility associated with the team members and their movement, it appeared that there were no formal requirements to be in certain places at certain times.

Norms additionally stemmed from the technological choices team members made. Team members knew where to find documents based on mutually agreed upon locations. For instance, we witnessed several exchanges of team members directing others to working files in Google Drive.

Fluidity of movement extended to the game development process. Group members tended to get up and observe their team members' new developments and make comments. On several occasions, team members were called over to comment on a new design, game asset, or interaction. This type of iterative development characterized by continually feedback is consistent with commonly documented agile development methodologies referenced in the Agile Manifesto (Beck et al., 2001).



### Focus group

The focus group built upon our ethnographic observations of norms within the game design context. In response to the focus group session questions, participants described what characteristics they found desirable when selecting fellow team members. Focus group participants were interested in working with team members who had varying skillsets, but similar goals. The participants looked for teammates who exhibited “empathy”, “discipline”, “attention to detail”, flexibility, and who had “strong communication skills”. They also looked for team members who “communicated challenges early on”, were “interested in more than just one artistic style”, were constantly “testing things”, adhered to a “schedule and project documentation” and were open to constructive criticism.

During our focus group, “crunchtime”, a term used to describe an all-night or an intensive development session, emerged as a norm for the game designers. When describing these sessions, it was clear that these were not only times that yielded productive results, but also incorporated social elements like ordering pizza, playing other games, and drinking beers. Participants described “crunchtime” as being both challenging, fun, and a bonding session between teammates.

Another norm that emerged was the lack of a clear decision-making process. Participants stated they had an equal say in “decision-making, scheduling, and house-keeping”. The flexibility associated with the organizational hierarchy was exemplified in our ethnographic observations when the team members worked to set a time to meet during the weekend. It was clear that each team member was free to come in when they liked, however, there was an expectation that each team member would do their part in meeting team goals.

### **Overlapping codes**

In the validation processes of their framework, Thomson et al. (2007) were able to define the five categories of collaboration through statistical analysis. The need for this clarity was due to the inherently complex and interwoven nature of collaboration. Overlapping codes were anticipated given that the Thomson et al. (2007) analysis showed high interrelationships between the terms in a higher order confirmatory factor analysis. We feel it is important to acknowledge observed overlaps between the codes with some occurring more frequently than others. The observed ebb and flow between these different categories warrants a deeper exploration that is beyond the scope of this paper.

### **Validity**

As Maxwell (2013) states, “the researcher is the instrument of the research” (p. 45). Researcher experiences were seen as an asset in the analysis of the data, however, we attempted to be as aware of our inherent biases by using a code manual, employing interrater reliability during our coding process, and incorporating formative respondent validation during the focus group sessions. We also ensured that all observation sessions were conducted at the same times during the day.

### **Limitations**

Given the nature of the participant’s consent, we were only able to observe the collaborative dynamics of one group. It was clear that other participants and their groups faced collaborative challenges and successes, however, we were not able to observe why and did not feel comfortable documenting those trials and tribulations without the consent of the group as a whole.

The timing of the experiment also served as a limitation in a variety of ways. To begin with, we began our study late in the group’s development, missing rich information about the formation of the group’s dynamics, their governance structures, and their formation of norms. It was also clear that a majority of the design work occurred outside of the classroom and during all times of the day and night. Future, in-depth studies could benefit by increasing the number of observations, specifically during “crunchtime” sessions or other group meetings outside of the designated class time.

Domain knowledge was yet another arena in which we struggled. As instructional designers, we did not have the same lexicon and vocabulary as the participants we studied. Words like “baked”, “sprite”, and “crunchtime” all required further clarification. It is quite possible that without a baseline knowledge of game design and game design culture, we may have missed many rich observational opportunities, particularly during the focus group sessions.

### **Discussion**

Like Gajda and Koliba (2007), we found that group dynamics and interpersonal relations play a large role in collaborative game design. It was apparent from our focus groups that group members were concerned with who should be included as a team member from the onset of the design process.

Also, similar to Gajda and Koliba (2007) concerns regarding individual confidentiality and consent made this study difficult to implement. As researchers, we found it challenging to observe consenting individuals involved in a collaborative context since their contributions and interactions were tied to the larger group. Therefore, we were limited to observing a single team and their interactions. Future studies may wish to consider alternative approaches to observing collaborative practices in light of the difficulties associated with securing individual consent.

While we were able to observe all attributes of the Thomson et al. (2007) framework in our limited observational and focus group sessions, future research could be amplified by following game designers throughout their game design process, from group formation to project completion. Specifically, it would be important to observe game designers during “crunchtime”, late night sessions, and more frequently over the design and development timeframe. Observations conducted outside the classroom times could yield rich findings into game design as well as its organizational culture.

## Conclusion

“Practice is not just doing, but also thinking about actions” (Crouch & Pearce, 2012, p. 39). As designers, it is important to be reflective about practice especially in consideration of different design contexts. Our review of the literature revealed that although game design was a critical component to the success or failure of an organizational game based initiative, few studies focused on the game design process and specifically the role of collaboration in design. Continued design research will be critical in enhancing our understanding of how successful game design teams collaborate to achieve and fulfil the promising potential of new approaches to teaching and learning.

## References

- Akcaoglu, M., & Koehler, M.J. (2014). Cognitive outcomes from the Game Design and Learning (GDL) after-school program. *Computers & Education, 75*, 72-81.
- Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R. C., Mellor, S., Schwaber, K., Sutherland, J. & Thomas, D. (2001). Manifesto for Agile Software Development *Manifesto for Agile Software Development*.
- Colbry, S., Hurwitz, M., & Adair, R. (2014). Collaboration theory. *Journal of Leadership Education, 13*(4), 63-75.
- Gajda, Rebecca. (2004) Utilizing collaboration theory to evaluate strategic alliances. *American Journal of Evaluation, 25*(1), 65. doi: 10.1016/j.ameval.2003.11.002
- Gajda, R., & Koliba, C. (2007). Evaluating the imperative of intraorganizational collaboration. *American Journal of Evaluation, 28*(1), 26-44. doi: 10.1177/1098214006296198
- Glesne, C. (2011). *Becoming qualitative researchers: An introduction*. Boston, MA: Pearson.
- Hartson, R., & Pyla, P. (2012) *The UX book: Process and guidelines for ensuring a quality user experience*. Waltham: Elsevier.
- Ke, F. (2014). An implementation of design-based learning through creating educational computer games: A case study on mathematics learning during design and computing. *Computers & Education, 73*, 26-39.
- Lunden, I. (2012). Badges beware: 80% of gamification apps will end up being losers says Gartner, *Tech Crunch*. Retrieved from: <http://techcrunch.com/2012/11/27/badges-beware-80-of-gamification-apps-will-end-up-being-losers-says-gartner/>
- Maxwell, J.A. (2013) *Qualitative research design: An interactive approach* (3rd Ed.). Thousand Oaks, CA: SAGE.
- Mayo, M. J. (2009) Video Games: A Route to Large-Scale STEM Education? *Science 323*(5910), 72–89.
- Robertson, J., Howells, C. (2008) Computer game design: Opportunities for successful learning. *Computers & Education, 50*, 559 – 578.
- Thomson, A.M., Perry, J.L., & Miller, T.K. (2007). Conceptualizing and measuring collaboration. *Journal of Public Administration Research Theory, 19*(1), 23-56.
- Woodland, R. H., & Hutton, M. S. (2012). Evaluating organizational collaborations: Suggested entry points and strategies. *American Journal of Evaluation, 33*(3), 366-383. doi: 10.1177/1098214012440028
- Yang, C.Y., & Chang, C.H. (2013). Empowering students through digital game authorship: Enhancing concentration, critical thinking, and academic achievement. *Computers and education, 68*, 334-344.

# Which Visualization Guides Learners Best? Impact of Available Partner- and Content-Related Information on Collaborative Learning

Melanie Erkens, University of Duisburg-Essen, melanie.erkens@uni-due.de  
Daniel Bodemer, University of Duisburg-Essen, daniel.bodemer@uni-due.de

**Abstract:** A large body of research covers the positive impact of cognitive group awareness tools on implicitly guiding learners. However, the impact of visualizing partner-related information and content-related information in such tools is barely investigated separately. Thus, we compared the impact of both types of information on collaborative learning in an experimental study ( $N = 120$ ) by systematically varying partner-related information (given or not) and content-related information (given or not) in a 2x2 design. Concerning communication behavior, we found no effect of content-related information, but a main effect of partner information indicating that the visualization of a learning partner's knowledge deficit implicitly guides learners to give longer explanations. A qualitative comparison contrasting learners with the highest and lowest knowledge gain indicate that content-related information could still be relevant for implicit guidance, since successful learners seem to use this information for addressing topics in their explanations.

## Introduction

The positive impact of implicit guidance on collaborative learning was widely demonstrated (cf. Janssen & Bodemer, 2013). Implicitly guiding learners means to only suggest them certain ways of thinking, communicating, and behaving, without directly instructing them to perform specific activities (cf. Hesse, 2007; Dillenbourg, 2002). Such guidance can be implemented with the help of cognitive group awareness tools. Following Bodemer and Scholvien (2014), these tools utilize three functions for suggesting specific behaviors: First, they can cue and list essential information, e.g. specific content, which helps learners to organize their knowledge and structure their communication with the learning partner. Second, they provide information on the whole group or single members of the group to facilitate important grounding and partner modeling processes. Third, they facilitate comparison processes to trigger discussions that are focused on relevant topics. In order to automatize these functions, the "Grouping and Representing Tool" (GRT) was developed (Erkens, Bodemer, & Hoppe, 2016). This tool extracts different information from written text (e.g. students' homework or essays): on the one hand, it transforms content-related information by forming clusters of concepts that are interpreted as topics and visualized as a list. On the other hand, it provides partner-related information, since the clustering results in quantitative values on the extent of how intensively these topics are debated in each text. These values can be visualized as bars per topic illustrating a learning partner's cognitive information separate or side-by-side with a learner's own cognitive information to facilitate a comparison, which is particularly useful in collaborative learning scenarios applying complementary group formation. The GRT was evaluated beneficial for classroom learning (Erkens et al., 2016), but due to their confounding both types of information still need to be investigated separately to know about their single effects. Since said confounding is given for most group awareness tools, the current study is guided by the pending question: What and how much impact has partner- and content-related information on learning behavior? In the following, we present results from a current study combining quantitative and qualitative measures to answer this question.

## Instructional purposes of cognitive group awareness tools

Cognitive group awareness tools fulfill several instructional purposes by providing learners with relevant cognitive information. Visualizations with content-related information can unfold their effects by information cueing (e.g. highlights) or by visualizing only selected information (e.g. topics). Information cueing refers back to the signaling principle which means that learners focus their attention and learn more deeply, when cues are added that highlight the organization of essential material (Mayer, 2005). In the context of collaborative learning, Scholvien and Bodemer (2013) found that learners supported by a cognitive group awareness tool prioritized content that was highlighted blue to reveal dissimilarities between them and their learning partner. Concerning the visualization of selected information, the effect of information clustering is based on the idea that visual representations that include organized labels, names, and graphics group relevant information (Hyerle, 2000), thereby focus learners' attention (Mayer, 1979) and provide better orientation, especially in cooperative learning scenarios (Purdom & Kromrey, 1992). Regarding this, Erkens, Bodemer and Hoppe (2016)

found indicators that a list of contents created by the GRT might influence learners in selecting topics to be discussed: one observed strategy was that learners scanned the list of content to check for topics about which they did not write in their essays (independent of the information given on a learning partner's knowledge) and finally added concepts visualized to their essays.

The visualization of partner-related information, either visualized separately or side-by-side with the information of a learner, supports comparison processes. The comparison between learning partners is associated with relevant learning mechanisms such as partner modeling, conflict solution, and (self-)explication (Dillenbourg, 1999). The concept of cognitive conflict goes back to Piaget (1977) and is based on the idea that interactions with physical or social environments can lead to a disequilibrium. In the case of a partner model contradicting the model of a learner's own knowledge, intrapersonal (Piaget & Inhelder, 2008) or interpersonal cognitive conflicts (Doise & Mugny, 1984) might occur. A learner can solve such conflicts by selecting specific topics and discussing them with the learning partner. Bodemer and Scholvien (2014) found that learners provided with partner-related information by a group awareness tool search for and prioritize topics visualizing cognitive dissimilarities to their learning partners before discussing less relevant topics. The concept of explication can be associated with audience design which means that speakers design each utterance for specific listeners (Clark & Murphy, 1982). In the case of visualizations supporting learners to estimate their partner as less knowledgeable, they might explain more to this learning partner. Regarding this, Dehler, Bodemer, Buder, and Hesse (2011) found that learners supported by a comparative visualization of learners' own and their respective partner's knowledge explained more in a discussion than learners without support, when a learning partner's deficit was visualized.

Overall, we can summarize that the visualization of combined content- and partner-related information can determine learners' focus and induce them to explain more to a learning partner. However, while both types of learner support have been evaluated separately, they have not been tested yet in one experiment that allows for comparing the impact of both. Thus, it is our objective to separate content- from partner-related information and investigate their single effects on learning behavior exemplarily for the GRT. We can conclude from the above that learners knowing more than their learning partner give longer explanations, if partner-related information is given. Further, we could assume that this effect might be especially strong, if content-related information is also given, since this combination was used in most tools that showed an effect on explaining behavior. Further, it is plausible that a list of several sub topics might evoke more detailed explanations than only being aware of the main topic, since a summary of principal issues could remind learners of relevant aspects that might be forgotten while explaining without such support. However, Erkens, Schlottbom and Bodemer (2016) found that learners gave more explanations without the visualization of content-related information. So it is still an open question, if there is an interaction effect of both partner- and content-related information on the length of explanations. Finally, we ask, which visualization of the given information supports elaboration or rather selection of topics best.

## Method

To conduct our laboratory experiment lasting one and a half hours we recruited a total of 120 university students (42 men; 78 women; mean age:  $M = 23.84$ ,  $SD = 4.40$ ). The participants were either paid 12 Euro, or received a certification on their contribution, and were randomized in a 2x2 factorial design to ensure evenly distributed knowledge on the topic to be learned prior to the collaboration ( $p = .71$ ). As can be seen in Figure 1, we deployed four different visualizations to support the four groups during the collaboration: (1) one visualization neither displaying content- nor partner-related information, (2) another one displaying no content- but partner-related information, (3) one displaying content- but no partner-related information, and (4) one displaying both content- and partner-related information. Consistent with the GRT's functions, content-related information stands for displaying a list of topics (bold text listed in (3) and (4)) with related concepts (normal text listed in (3) and (4)) generated from the texts used in the study. Partner-related information stands for visualizing partner's topical extent(s) as a grey bar. Own topical extent(s) was visualized as black bar. Said topical extents (bars) were based on the knowledge that was imparted to the participants by a text on climate change.

As dependent variable, we captured communication behavior that was operationalized as the length of explanations given to a partner and we controlled for time of text re-reading as a covariate. Further, we qualitatively examined the behavior of particularly successful or unsuccessful learners more closely and contrasted the best and worst learner of the most contrasting experimental groups, (1) and (4), to learn more about the quality of explanations.

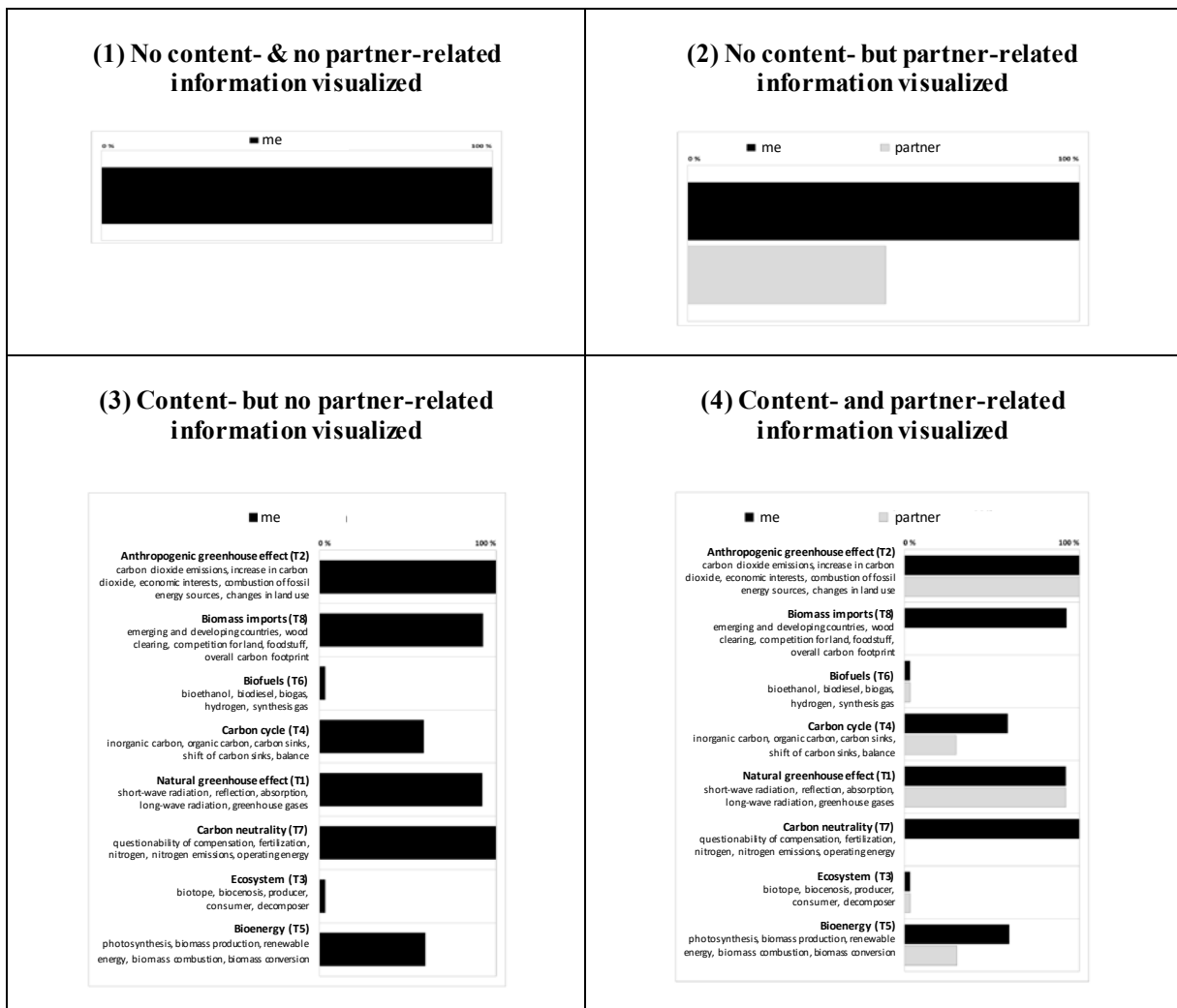
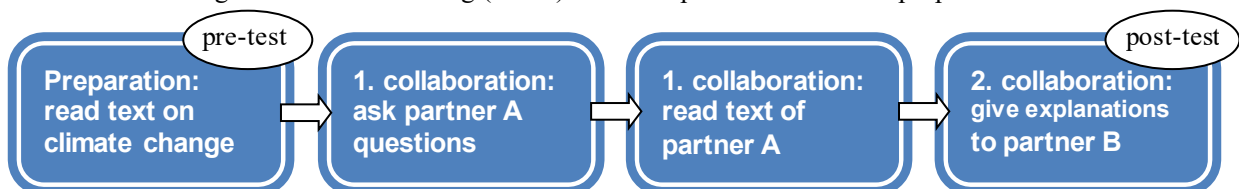


Figure 1. Four visualizations with different content- and partner-related information. Content-related information is given in (3) and (4) with topics in bold text and concepts in normal text. Partner-related information is given in (2) and (4) with the partner's knowledge visualized as grey bar(s).

## Procedure and instructions

The idea of our study was to reconstruct a real collaborative scenario with learners asking and answering questions instead of triggering a real discussion in order to ensure controlled framework conditions. Although this article presents only results related to the phase of answering questions or rather explicating, we outline the whole procedure in the following (cf. Figure 2) for better comprehension. The experiment was conducted with each participant working on a single computer. After welcoming and declaration of consent, the participants were informed that they had to read and memorize a text on climate change within 15 minutes forming a basis for their later explanations and to simulate a collaboration with two other participants, of whose non genuine nature they were aware. Following a pre-test to evaluate their knowledge, they were instructed to ask their first learning partner (A) questions. During the simulated collaboration they were provided with a visualization illustrating a scenario in which they had less overall knowledge than their learning partner A (similar to Figure 1 but with other values or rather shorter bar lengths). The participants were supported by a representation showing their own knowledge status and visualizing (or not) a list of topics and main concepts presented in their own and



**Figure 2.** Graphical overview of the experiment’s procedure.

their learning partner’s text and visualizing (or not) bars that represent their learning partner’s extent of knowledge allowing for comparison to the partner status. They should write down then a minimum of three questions as open text. After this first learning phase, they were informed that learning partner A had not the task to ask questions but to write a text about his knowledge on climate change and that they should carefully read this contribution within 10 minutes. Following this phase, participants were requested to give explanations to questions of their second learning partner (B) (cf. Figure 3) to simulate a discussion with a partner that might disclose a possible adaption to the available partner’s knowledge. This time, the participants were supported by a visualization based on a scenario in which they had a higher overall knowledge than their learning partner B (cf. Figure 1), which means that they had the same availability of content- and partner-related information as in the first collaboration, but the lengths of bars was mainly longer in the respective visualization. The learning environment offered to make their contributions as open text again. Finally, participants had to take a post-test.

## Instruments and variables

### Length and topical scope of explanations

During the second collaboration, we presented the participants three questions of the bogus learning partner: (1) “What are the decisive advantages of energy from biomass against the background of global warming?“, (2) “What are the decisive disadvantages of energy from biomass against the background of global warming?“, and (3) “What are your conclusions regarding energy from biomass against the background of global warming?“. Under each question, we presented the respective visualization on the left, a possibility to access the text on climate change on the right, and a scalable input window for answering the question with an explanation of any chosen length in the middle (cf. Figure 3). Subsequent to this bogus collaboration, we counted the number of words for each answer and cumulated them to calculate the total length of explanations. Further, the elaborations were investigated concerning included topics to contrast the best and worst learners in the experimental group with no visualization of content- and partner-related information and in the experimental group with partner- and content-related information given. Therefore, we counted the occurrences of each topic within the three explanations per participant and summed them up.

### Time spent re-reading the text on climate change

On each of the three pages for answering the learning partner’s questions, learners had the possibility to access the text on climate change (cf. Figure 3) and read it as long as they wanted. We herewith wanted to create a situation with all participants being able to fall back on the same content knowledge. We measured the time between opening and closing the text each time it was opened. Finally, we summed up all values per person to establish the total time of re-reading while explaining. We chose this total time as control variable, since it is probable that more time of re-reading the material might come along with less time for writing explanations, and might also be predictive for the length of explanations.



**Figure 3.** Participants’ view on the learning environment during the collaboration.

### Score of knowledge test pre and post the collaboration

The knowledge test consisted of 36 statements related to four topics in the text on climate change and to two topics of the first bogus learning partner's text. We presented the statements (e.g. "Photosynthesis ensures that plants gain energy to produce dead matter" or "The total carbon content of the earth is constant) prior to the collaboration (pre-test) and afterwards in randomly rotated order (post-test). Participants had to state whether the answer was true or false. In order to calculate the total score of the knowledge test, we checked for each item, whether the given answer was correct or not. Finally, we summed up all points for both times of measure (pre and post). Thus, both total scores could range from 0 to 36 points. They were needed to identify the learners with the highest and lowest knowledge gain in the different experimental groups.

### **Results**

For answering the question of how partner- and content-related information affects learning behavior, we observed the impact of different visualizations on the length of explanations and on learning results. All effects are reported as significant at  $p < .05$ .

### **Impact of partner-related and content-related information on explanations**

We hypothesized that learners supported by the visualization of partner-related information explain more about subject matters than learners without partner-related information. Further, we investigated if this effect is particularly strong, if additional information on the content is given. For investigating these assumptions, we used a two-factorial ANCOVA with availability of partner-related information and content-related information as between-subject factors and the number of total words in explanations per person as dependent variable. Further, we controlled for the time of re-reading the learning material to keep the prior knowledge constant across learners. This covariate was suitable to be included in the analysis, since it fulfilled all preconditions.

The covariate *re-reading time* was significantly related to the count of words in explanations,  $F(1, 115) = 12.16, p < .001, \eta_p^2 = .096$ . The longer the time of re-reading, the shorter the explanations. Regarding the assumption that available partner-related information triggers longer explanations, results indicated a significant main effect of partner information on the number of words used in explanations,  $F(1, 115) = 3.97, p = .049, \eta_p^2 = .033$ . Further, results yielded no significant main effect of content-related information on the number of words used in explanations,  $F(1, 115) = 0.12, p = .725, \eta_p^2 = .001$ . Finally, there was no interaction effect of partner- and content-related information on the number of words in explanations,  $F(1, 115) = 0.13, p = .716, \eta_p^2 = .001$ . Table 1 shows related descriptive statistics.

Table 1: Number of words in explanations

	content-related information given ( $n = 61$ )		content-related information not given ( $n = 59$ )		overall ( $N = 120$ )	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
partner-related information given ( $n = 61$ )	256.00	95.05	249.90	78.07	253.00	86.45
partner-related information not given ( $n = 59$ )	225.03	98.65	224.97	85.75	225.00	91.74
overall ( $N = 120$ )	240.77	97.29	237.64	82.19	239.23	89.82

### **Contrasting cases: Comparison of successful and unsuccessful learners**

To learn about the quality of elaborations, we examined the behavior of particularly successful or unsuccessful learners more closely and contrasted the best and worst learner of experimental group (1) with neither partner-related nor content-related information and experimental group (4) with both types of information available. Against the theoretical background, we assumed that the availability of content- and partner-related information might improve the selection of topics used in explanations, since the visualization of selected information from

text should focus learners' attention and the scenario displaying the learning partner to be less knowledgeable might evoke deeper elaborations. To investigate whether learners in the experimental group with both awareness information available might be better supported concerning the selection of topics to be discussed than the group with less information, our special interest were the matches between knowledge test items that learners improved from pre- to post-test and the topics used in their explanations. We observed, how often they were related to each other in two ways of matches: (a) topics addressed within explanations (T1 to T8) did match topics addressed in knowledge test items (topical matches), and (b) concepts addressed within explanations did match concepts addressed in knowledge test items (conceptual matches). In Table 2, you can see an overview of our findings including knowledge gain and information on improvements and worsening.

Table 2: Characteristics of learners in contrasting cases.

	(1) No content- & no partner-related information visualized		(4) Content- and partner-related information visualized	
	<i>learner ID 102</i>	<i>learner ID 27</i>	<i>learner ID 3</i>	<i>learner ID 62</i>
score pre-test <sup>1</sup>	26	18	22	24
score post-test <sup>1</sup>	19	26	18	29
knowledge gain	-7	+8	-4	+5
# improvements	2	8	5	5
# worsening	9	0	9	0
improved topics from pre- to post-test	T2 <sub>(1x)</sub> , T7 <sub>(1x)</sub>	T1 <sub>(2x)</sub> , T2 <sub>(2x)</sub> , T5 <sub>(2x)</sub> , T7 <sub>(1x)</sub> , T8 <sub>(1x)</sub>	T2 <sub>(3x)</sub> , T4 <sub>(1x)</sub> , T5 <sub>(1x)</sub>	T1 <sub>(1x)</sub> , T2 <sub>(1x)</sub> , T5 <sub>(2x)</sub> , T7 <sub>(1x)</sub>
worsened topics from pre- to post-test	T1 <sub>(1x)</sub> , T2 <sub>(2x)</sub> , T4 <sub>(2x)</sub> , T5 <sub>(1x)</sub> , T7 <sub>(1x)</sub> , T8 <sub>(2x)</sub>	-	T1 <sub>(1x)</sub> , T4 <sub>(1x)</sub> , T5 <sub>(2x)</sub> , T7 <sub>(1x)</sub> , T8 <sub>(3x)</sub>	-
# topical matches	2	5	1	4
# conceptual matches	0	1	0	3

T<sub>x</sub> = topic (cf. Figure 1). Numbers in brackets = count of occurrences. <sup>1</sup>scoring from 0-36.

### Comparison of the best and worst learner without content and partner information

To contrast cases in the experimental group in which learners had neither partner-related nor content-related information we chose the learner with the highest and the lowest knowledge gain in the knowledge test. We found that learner 27, who improved her/his score by 8 points, showed a huge amount of topical matches. In five cases she/he wrote about topics having errors in the pre-test, but only in one case the concept was actually congruent: Concerning the test item "The chemical transformation of carbonaceous compounds runs completely independent of the nitrogen cycle." (which is wrong, since both cycles depend on each other), learner 27 who checked this answer as "right" in the pre-test explained in the second collaboration to her/his learning partner: "I do not consider this [generating energy from biomass] as beneficial, if the cultivation and growing of biomass entails the warming of the atmosphere caused by nitrogen und carbon dioxide." Thus, she/he understood that both nitrogen and carbon cycle interact in the context of global warming. In contrast to learner 27 with the highest improvements in this experimental group, learner 102 dropped off 7 points in the score of the post-test or to be more precise she/he dropped 9 points and gained 2 points. In this case, the learner wrote about 5 topics that were improved from pre- to post-test, but the concepts did not match to concepts in the improved test items. To give an example: Learner 27 stated "Likewise, during the production of biomass, dead matter is bound and decreases in the atmosphere..." what is part of topic 4 (concept: dead matter) and topic 5 (concept: biomass production). On the one hand, this utterance is wrong, since it is CO<sub>2</sub> that is bound and reduced in the atmosphere. On the other hand, the concept does not match the concepts of the improved items in the post-test that instead were related to the topics T1, T5 and T7. Thus, learner 102 seemed to decrease her/his knowledge caused by misconceptions that emerged during the explanation process.



### Comparison of the best and worst learners with content and partner information

In this case, we contrasted the best and the worst learner with a visualization available that combines partner- and content-related information. We found that learner 3, who dropped off 4 points from pre- to post-test, had no matches of concepts in his/her explanation and concepts in the test items, only once a topical match was given. In contrast, we found that learner 62 who improved his/her score by 5 points had a good topical match between test items and explanations. She/he met 4 times the topical scope of test items and even though mentioned 3 times concepts in her/his explanations that were part of improved test items. For instance, learner 62 scored better at the test item “The photosynthesis enables plants to produce energy and dead matter” (which is wrong, since the plant produces organic matter). Accordingly, she/he explained to his/her learning partner “Renewable energy using biomass is energy that stems from non-fossil algal or vegetable biomass.” and “Biomass first needs to be produced by photosynthesis”. Both utterances show that she/he understood that photosynthesis does not produce dead matter. Precisely because learner 62 improved his/her knowledge after giving explanations to the learning partner, it is noteworthy how the learners evaluated the visualization in an open question following the collaboration: while learner 3 reported that she/he used the visualization “partially” which is in line with his/her behavior, learner 62 reported that the “visualization was not helpful to answer the questions”. However, from the five improvements of learner 62 (in T1, T2, 2xT5, T7), she/he consequently gave only explanations when it was displayed that she/he is more knowledgeable than her/his learning partner (in T5 and T7). Since she/he further reported that she/he used the visualization to scan for cases in which his/her knowledge was displayed being less than the learning partner’s knowledge and to ask questions exactly in this, we can assume that she/he (implicitly) did the same concerning explanations (although described differently by learner 62).

### **Conclusion**

We conducted this empirical study to systematically investigate the impact of partner- and content-related information on collaborative learning, or rather on explaining behavior and on the selection of topics to be discussed. Consistent with previous research (Dehler et al., 2011) we found that the visualization of partner-related information implicitly guides learners’ communication behavior and makes them give longer explanations if they are more knowledgeable than their learning partner. In contrast, content-related information that is inherent to most such tools seem to have no effect on adapting explanations to the learning partner’s knowledge in terms of explaining more to a less knowledgeable learning partner. Erkens, Schlottbom, and Bodemer (2016) assumed that too many topics might overstrain learners’ cognitive system. Further, visualizing partner-related information without content-related information might suggest learners assuming the role of an expert and explaining as much as possible. Both these assumptions need to be further investigated. Also it is still unclear, whether shorter explanations could nevertheless show a better quality. Learners supported by the visualization of both partner- and content-related information might better elaborate the contributions of a learning partner than learners without this support, since their prior knowledge could be better activated. After all, activation of prior knowledge is known as the most important antecedent for learning (Glaserfeld, 1984; Resnick, 1983; Ausubel, 1960), since it supports learners to successfully add new concepts learned from the partner to the existing knowledge base. Finally, learners only provided with content-related information are most likely unable to precisely detect their learning partner’s knowledge gaps and might thus not adapt their behavior to the knowledge of their learning partner.

In a second step, we investigated qualitatively how improvements from pre- to post-test were supported by implicit guidance. Here, our focus was on the usage of given information to explain learners’ improvements. We found that successful learners across experimental groups frequently discussed relevant topics, while unsuccessful learners did not discuss such topics. But only the successful learner provided with partner- and content-related information also wrote more often about concepts relevant to get better in a test item than the successful learner with none of this information. Based on this, we can first assume that content-related information might have helped to be reminded of concepts relevant to understand a topic. Thus, displaying content-related information still seems to be helpful to implicitly guide the learner in terms of influencing learning behavior. Although said learner mentioned that the visualization was not needed for answering questions, we further found that this learner especially explained relevant topics, when it was displayed that she/he is more knowledgeable than her/his learning partner (in T5 and T7). Thus, we can further assume that partner-related information might have helped her/him to select topics to be discussed. However, both these aforementioned assumptions were not captured with the length of explanations and further qualitative research is needed to tell more about the quality of elaborations within the explanations.

Overall, the support by visualized partner- and content-related information seems to fulfil different functions. Partner-related information leads to longer and possibly more detailed explanation. Concerning the

role of content-related information, we have to further investigate whether their visualization influences the quality of explanations by activating prior knowledge or by determining the selection of topics. This research might offer answers on how to appoint cognitive information in a task-oriented way and on how to even better support collaborative learning processes.

## References

- Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, 51(5), 267–272. <https://doi.org/10.1037/h0046669>
- Bodemer, D., & Scholvien, A. (2014). Providing knowledge-related partner information in collaborative multimedia learning: Isolating the core of cognitive group awareness tools. In *Proceedings of the 22nd International Conference on Computers in Education ICCE 2014* (pp. 171–179). Nara, Japan.
- Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In *Advances in Psychology* (Vol. 9, pp. 287–299). Elsevier. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0166411509600595>
- Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Guiding knowledge communication in CSCL via group knowledge awareness. *Computers in Human Behavior*, 27(3), 1068–1078. <https://doi.org/10.1016/j.chb.2010.05.018>
- Dillenbourg, P. (1999). What do you mean by collaborative learning? *Collaborative-Learning: Cognitive and Computational Approaches*, 1–19.
- Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. *Three Worlds of CSCL. Can We Support CSCL?*, 61–91.
- Doise, W., & Mugny, G. (1984). *The social development of the intellect*. (W. Doise & G. Mugny, Eds.) (Vol. 10). Amsterdam, Netherlands: Pergamon. Retrieved from <http://www.sciencedirect.com/science/article/pii/B978008030215750010X>
- Erkens, M., Bodemer, D., & Hoppe, H. U. (2016). Improving collaborative learning in the classroom: text mining based grouping and representing. <https://doi.org/10.1007/s11412-016-9243-5>
- Erkens, M., Schlottbom, P., & Bodemer, D. (2016). Qualitative and quantitative information in cognitive group awareness tools: Impact on collaborative learning. In U. Cress (Ed.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016*. Singapore: International Society of the Learning Sciences.
- Glaserfeld, E. v. (1984). An introduction to radical constructivism. In P. Watzlawick, (Ed.), *The invented reality*. New York, NY: W.W. Norton.
- Hesse, F. W. (2007). Being told to do something or just being aware of something? An alternative approach to scripting in CSCL. In *Scripting computer-supported collaborative learning* (pp. 91–98). Springer. Retrieved from [http://link.springer.com/chapter/10.1007/978-0-387-36949-5\\_6](http://link.springer.com/chapter/10.1007/978-0-387-36949-5_6)
- Hyerle. (2000). *A Field Guide to Using Visual Tools*. Alexandria, Va: Association for Supervision & Curriculum Deve.
- Janssen, J., & Bodemer, D. (2013). Coordinated computer-supported collaborative learning: Awareness and awareness tools. *Educational Psychologist*, 48(1), 40–55.
- Mayer, R. E. (1979). Twenty years of research on advance organizers. Technical report series in learning and cognition. Report No. 79-1. Retrieved from <http://eric.ed.gov/?q=Mayer+advance+organizer&id=ED206691>
- Mayer, R. E. (2005). *The Cambridge Handbook of Multimedia Learning: Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles*. Retrieved from </chapter.jsf?bid=CBO9780511816819A020&cid=CBO9780511816819A020>
- Piaget, J. (1977). *The development of thought: equilibration of cognitive structures*. Viking Press.
- Piaget, J., & Inhelder, B. (2008). *The psychology of the child*. New York, NY: Basic Books.
- Purdom, D. M., & Kromrey, J. D. (1992). A comparison of different instructor intervention strategies in cooperative learning groups at the college level. Retrieved from <http://eric.ed.gov/?id=ED344898>
- Resnick, L. B. (1983). Mathematics and science learning: A new conception. *Science*, 220(4596), 477–478. <https://doi.org/10.1126/science.220.4596.477>
- Scholvien, A., & Bodemer, D. (2013). Information cueing in collaborative multimedia learning. In N. Rummel, M. Kapur, M. Nathan, & S. Puntambekar (Eds.), *To See the World and a Grain of Sand: Learning across Levels of Space, Time, and Scale: CSCL 2013 Conference Proceedings Volume 2 — Short Papers, Panels, Posters, Demos & Community Events* (Vol. 2, pp. 149–152). Madison, WI.

# Articulating Uncertainty Attribution as Part of Critical Epistemic Practice of Scientific Argumentation

Hee-Sun Lee, Amy Pallant, Sarah Pryputniewicz, and Trudi Lord  
hlee@concord.org, apallant@concord.org, spryputniewicz@concord.org, tlord@concord.org  
The Concord Consortium

Ou Lydia Liu, Educational Testing Service, lliu@ets.org

**Abstract:** Models are important in discovering trends, developing and testing theories, and making predictions about complex systems. Since models cannot represent all known and unknown aspects of how nature operates, claims based on model-based data inevitably contain uncertainty. This study explores (1) how high school students attribute sources of uncertainty when prompted as part of an argumentation task and (2) how intelligent feedback may guide them to become more cognizant about deep uncertainty associated with model-based data. Phenomenological analyses of students' uncertainty attributions (N = 840) identified five distinct patterns: self-introspection, personal theories and experiences, data source acknowledgement, scientific description based on singular causal accounts and frequency of observations, and deep uncertainty based on epistemic or ontic accounts. Discourse captured on video illustrated how intelligent feedback enhanced uncertainty attribution.

## Topic introduction

*The Next Generation Science Standards* (NGSS Lead States, 2013) encourage students to engage in practices through which scientific ideas are originated. Integrating scientific argumentation into science teaching is recommended because it allows students to participate authentically in sense making with data during investigation (Duschl & Osborne, 2002) and through communication (Kuhn, 2010). During investigation, students make claims based on data in light of their understanding of established knowledge (Bricker & Bell, 2008). In communication, students compare and contrast the strengths and weaknesses of evidence-based arguments (Erduran, Simon, & Osborne, 2004). Scientific argumentation as an epistemic practice generates two types of discourse: "theoretical discourse, pertaining to what theories of the world best fit the data, and practical, deliberative discourse, regarding how to apply those theories to reach practical goals" (Nussbaum, Sinatra, & Owens, 2012, p. 17). For example, students can use carbon dioxide data captured in ice cores as part of theoretical discourse concerning how changes in greenhouse gas concentrations contribute to atmospheric temperatures over time. Students also can engage in "deliberative" arguments about whether to impose carbon taxes based on various types of data. This study addresses the former, i.e., knowledge generation discourse rather than the latter dealing with arguments in sociocultural decision making.

Challenges to engaging students in argumentation are well documented in the literature: (1) students have difficulties in differentiating among claim, evidence, and reasoning (Berland & Reiser, 2009), (2) students lack experience in interpreting evidence in terms of theory (McNeill, Lizotte, Krajcik, & Marx, 2006), and (3) students lack epistemic commitment (Sandoval, 2003). In constructing a scientific argument, students need rhetorical support on how to write a convincing argument as well as content support on what knowledge should be used to interpret evidence. While scaffolding individual students is needed, it is unrealistic to expect a teacher to systematically intervene with every student in her class (McNeill & Pimentel, 2010). To address this issue, we developed an intelligent feedback system that (1) diagnoses students' arguments through an automated scoring engine based on machine learning algorithms developed for natural language processing and (2) provides students with immediate feedback matching their current performance.

This study addresses high school students' written as well as spoken discourse that occurred when they used a computer-based groundwater model to determine whether water that infiltrates is trapped underground. Evidence from the groundwater model for students to make claims was constrained due to the fact that groundwater systems in the real world cannot be modeled exactly. Students needed not only to select and use data from this imperfect model to make a knowledge-based claim, but also to recognize that their claim was constrained. This study explores two research questions: (1) how high school students attributed sources of uncertainty when prompted as part of model-based argumentation and (2) how intelligent feedback might guide students to become more cognizant about deep uncertainty associated with model-based data.

## Theoretical framework

Argumentation is carried out through written or spoken discourse often accompanied by symbols, representations, and visualizations (Walton, Reed, & Macagno, 2008). The field-independent structure of arguments is well recognized (Toulmin, 1958) such as a *claim to answer* a driving question, *data* that support the claim, *warrants* that explain how data support the claim and how established scientific knowledge *backs* the warrants, *qualifiers* that indicate the strength of the claim given evidence and knowledge, and *conditions of rebuttal* where the claim may not be held true. Most research on written scientific argumentation has focused on characterizing and improving students' coordination between theory and evidence embodied in claim, data as evidence, and warrants and backing as knowledge-based reasoning (Sampson & Clark, 2008). Students' uses of qualifiers and conditions of rebuttal have mostly been studied as counterarguments or rebuttals in written argumentation (Erduran et al., 2004) or in social settings (Sampson & Clark, 2009).

However, qualifiers and conditions of rebuttal can play a different but still important role in written argumentation because they invoke the notion of uncertainty, i.e., the extent to which knowledge claims are bounded by evidence generated from particular investigation contexts. As Bogen and Woodward (1988) pointed out, scientific knowledge explains facts related to a phenomenon but not necessarily facts related to raw data that represent aspects of the phenomenon. Data to which students have access are "dependent upon the peculiarities of the particular experimental design, detection devices, or data-gathering procedures" (Boumans & Hon, 2014, p. 2) and are just one of many incidences that can exemplify some but not all aspects of the phenomenon. Drawing unwavering knowledge claims from data is almost impossible and thus involves a great degree of uncertainty. Staley (2014) characterized two modes of reasoning by scientists when working with data. In the *use* mode, scientists use theoretical and methodological assumptions to arrive at substantive conclusions from the data. In the *critical* mode, scientists carefully examine those assumptions. The current emphasis on promoting scientific argumentation in the classroom through claim-evidence-reasoning may not take full advantage of an instructional opportunity where students can also learn about the *critical* use of data.

Environmental science topics such as climate change have been commonly used in classrooms for scientific argumentation (Nussbaum, Sinatra, & Owens, 2012) based on publicly available data and simulation models by scientists (Spiegelhalter, Pearson, & Short, 2011). Since environmental systems are complex, there exists epistemic uncertainty due to fundamental limitations with investigators' theoretical and methodological abilities to understand and predict how nature works. There exists also ontic uncertainty because "the physical world has an element of irreducible elusiveness. The result of an experiment is not determined by the conditions under the control of the experimenter. The lack of control is not the experimenter's deficiency, but rather nature's indeterminism" (Ben-Haim, 2014, p. 165). As models are the main means of investigating and understanding environmental systems, both epistemic and ontic uncertainty sources associated with model-based data should be examined explicitly by a person who is presenting an argument.

Some uncertainty associated with scientific investigation or modeling can be quantified in probabilistic terms. However, there is *deep uncertainty* that "results from myriad factors both scientific and social, and consequently is difficult to accurately define and quantify" (Kandlikar, Risbey, & Dessai, 2005, p. 444). Kahneman and Tversky (1982) pointed out that uncertainty is part of everyday life because we act without full knowledge, information, or understanding of any encountered stimulus. They noted that uncertainty can be captured as confidence on "a prediction, estimate or inference to which one is already committed" (p. 150). Kahneman and Tversky (1982) also reported that people often attribute uncertainty in natural language to the external world to seek more objective criteria or to personal state of knowledge due to internal ignorance. Those who seek external attribution use either frequencies of occurrence across multiple similar cases (distributional) or causal propensities that explain a typical or exemplary case that proves their point (singular mechanistic). Internal attributions can be based either on personal theories and experiences irrespective of external criteria or on introspective confidence. We use Kahneman and Tversky (1982)'s framework as a starting point to categorize students' uncertainty attribution when they construct an argument.

## Methods

### Scientific argumentation task

The scientific argumentation task we analyzed in this study was embedded in an online curriculum module entitled "Will there be enough freshwater?" This water module consists of six activities that guide students to explore the distribution and use of fresh water on Earth. Students experiment with models to explore Earth's groundwater system. They are also introduced to scientific data about freshwater distribution and use on Earth. The module is designed for five 45-minute class periods. Throughout the module, students are engaged in the practice of scientific argumentation as they work with scientific data, observations, and computer-based simulations. There are eight argumentation tasks that are structured similarly to scaffold students: (1) multiple-

choice claim, (2) open-ended explanation of claim using “Explain your answer,” (3) five-point Likert scale uncertainty rating from not at all certain to very certain, and (4) open-ended explanation of uncertainty rating “Explain what influenced your certainty rating.” Figure 1 illustrates the first argumentation task related to the movement of water in and out of the Earth’s surface. This task structure was validated to measure uncertainty-infused scientific argumentation (Lee et al., 2014). The activity starts with the importance of the topic and asks students to tinker with the model to make observations about how precipitation moves through the various sediment layers in the ground with different degrees of permeability. Students take a snapshot of the model run and draw the longest path a water droplet may take. Students then answer a multiple-choice question asking which layer of the model blocks water from flowing through. When students choose an answer, the module immediately answers whether the answer is correct. These two questions are designed to help students interpret the data and the representation of the model and elicit the knowledge necessary for students to successfully complete the scientific argumentation task that follows. The driving question for the groundwater model reads, “When water is absorbed by the ground, is it trapped in the ground?” Students respond to the four argumentation prompts, which are placed within a blue border called “arg-block.” For each of the four argumentation prompts, students can expand a set of hints. For instance, hints for uncertainty rating explanation are (1) A good certainty explanation will explain why you are certain or uncertain about your response; (2) Some topics are more certain than others; (3) Consider the completeness of the evidence, biases in the evidence, and changes that could affect the trends over time.

(a) Investigation context setup

(b) Simulation model showing how water moves around Earth's surface

(c) Questions to elicit data and understanding necessary for argumentation that follows

(d) Arg-block

(e) Submit

After submitting, intelligent feedback is provided to students with explanation and uncertainty rationale scores

Expandable hints that can be hidden

Checking answers

Feedback

Claim explanation score

Feedback on how to improve

Uncertainty explanation score

Feedback on how to improve

**Figure 1.** Groundwater Argumentation Task with Natural Language Processing-based Intelligent Feedback System.

## Data collection and analysis

This study took place in two phases: identification of uncertainty attribution patterns based on students' written responses to argumentation prompts and uncertainty discourse impacted by intelligent feedback. In the first phase, we used written responses from 840 students to the groundwater argumentation task. These students were taught by 15 teachers in 8 states across the U.S. These teachers found the water module from various outreach efforts and voluntarily participated. They attended a summer workshop prior to implementing the module in their classrooms. Students were in eighth to twelfth grades. Both genders were equally represented; 15% spoke English as second language; 80% used computers regularly for science learning prior to the module. We used the uncertainty attribution framework proposed by Kahneman and Tversky (1982) that listed four different types of uncertainty attribution in natural language: introspective confidence vs. personal theories under internal attribution and frequency-based distributional vs. singular causal under external attribution. Once we identified patterns in students' open-ended responses to explanation and uncertainty attribution prompts in the four-part scientific argumentation task, we conducted natural language processing (NLP) of these students' responses based on machine learning algorithms through c-rater-ML™ developed by Educational Testing Service. We also developed a feedback statement unique to each pattern of uncertainty attribution to address shortcomings.

A new version of the water module was developed by integrating automated scoring models into all eight arg-blocks so that students could receive feedback immediately after submitting their arguments. As soon as students clicked the submit button at the end of each arg-block, students' open-ended responses to the explanation and uncertainty attribution prompts were processed by the automated scoring engine developed to recognize whether they included a scientifically valid claim, evidence, and reasoning for an explanation (score ranging from 0 to 7) and how they articulated their uncertainty rationale (score ranging from 0 to 4). For the uncertainty attribution prompt analyzed in this study, the human-machine agreement was measured at 0.83. The whole process of submitting, autoscoring, finding feedback matching the score, and displaying the score and feedback to the student took two to five seconds in real time.

The second phase of this study was based on the intelligent feedback system-enabled water module. The module was implemented by four high school teachers in two suburban and two rural schools located in KY, MA, NH, and PA. Approximately 156 ninth to twelfth grade students used the water module. We recorded videos of the computer screens of 14 student groups, including student voices, as they worked through the water module. Uncertainty discourse was examined with these 14 student groups who worked on the groundwater model argumentation task shown in Figure 1. A total of 214 minutes of the videos were transcribed verbatim. We selected the groundwater argumentation task in this study because the task was the first time when students had to articulate uncertainty associated with model-based argumentation.

## Findings

### Uncertainty attribution patterns

From the uncertainty responses we analyzed, five different types of uncertainty attribution emerged. When asked to explain their uncertainty rating of their claim based on evidence they used, some students appeared to reiterate their uncertainty rating such as "We are not completely positive," "I'm kind of guessing but it makes sense and I'm pretty sure it's right," "I tried to think with common sense," and "I am almost certain about my answer." These statements describe introspectively the students' knowledge state about the claim they were making without any external reference to the phenomenon or personal justification rationale. On the other hand, we noticed a sizable number of responses referring to personal rationale based on their knowledge, abilities, or experiences relevant to understanding the driving question, processing and interpreting data, and conducting investigation. Examples of personal knowledge and ability references included "I didn't quite understand the rain example [shown in the model]," "I am familiar with the water cycle," "I didn't understand the question," "I cannot see the images very well," and "Based off my knowledge of what I learned about precipitation." This category also included personal experiences such as "Because my teacher told me," "We saw a movie about water being pumped," and "If the water were trapped, we wouldn't have enough water to live." We also categorized misconceptions about the groundwater topic ("Because most water travels through lakes and rivers") as personal rationale.

We found three general types of external uncertainty attribution. The first type simply acknowledges a scientific data source such as "Background information and the model," "The graph clearly, obviously, and very blankly shows this idea," and "After observing the diagram for a few moments, I managed to reach a conclusion as to what each dot represented and when it would change into the other dot." In these cases, students mentioned the source without providing details that explain the phenomenon or describe what happened in the data they were citing. The second type relates to external scientific disposition that either explains a mechanism for why

water is not trapped (“We know that the water can pass every sediment layer but the black layer. The water moves through the layers easily on its way down to Earth until it hits the black layer which will stop the water from proceeding”) or describes multiple outcomes related to water droplets (“The water moves through the layers easily on its way down to Earth until it hits the black layer which will stop the water from proceeding. Then after the water piles up and is overflowing to the top, evaporation occurs for some of them”). The former relates to singular mechanistic attribution and the latter relates to distributional attribution as multiple outcomes are acknowledged. We grouped both singular mechanistic and distributional attribution accounts under the external scientific disposition because these attribution types address the investigation at hand. The third type shows scientific limitations beyond the current investigation based on the groundwater model. Both ontic and epistemic uncertainty attributions were observed. The ontic attribution statements include “earth has many layers and not all of them can stop the water flow and some layers absorb the water and dispose the unwanted use of the water.” The epistemic attribution statements focus mostly on model limitations such as missing factors (“because if they were they would stop flowing. However layers further down may be able to stop it from flowing”).

Table 1 summarizes five categories we identified from students’ responses to uncertainty attributions. In order to incorporate all that emerged from our analysis, we modified the framework of Kahneman and Tversky (1982) by adding a new category of external scientific source acknowledgement and external scientific limitation. Based on these characterizations, we created an ordinal progression by assigning scores from 0 to 4. The order of progression moves (1) from not mentioning to mentioning attribution, (2) from internal to external attribution, and (3) from vague external description to external scientific disposition then to scientific limitation. Based on this scoring method, we developed an intelligent feedback statement for each score.

### Uncertainty discourse aided by intelligent feedback

In order to study uncertainty attribution discourse, we examined the discussions of 14 student groups, which were captured on video. Each group consisted of two or three students who responded to argumentation prompts in the arg-block together. These 14 groups made an average of 1.71 argument revisions after receiving real-time, intelligent feedback. Four groups did not make revisions; five groups revised once; three groups twice; two groups three or more times. Three groups claimed that groundwater was trapped while the other 11 groups claimed it was not. The claims did not change throughout revisions. When first submitted, explanations of nine groups included scientific reasoning that showed scientific mechanisms regarding whether the water was trapped or not. For example, one group wrote, “Water that is in the ground does not stay trapped in the ground because roots from plants suck up the water and through transpiration it evaporates from the plants to the atmosphere.” Four groups included data they observed from the groundwater model: “It is trapped in the ground because the black layer won’t let the water seep through that layer.” Only one group’s explanation did not explicitly mention data or reasoning: “The water stays always flowing.” Intelligent feedback prompted those groups who included reasoning to also include data from the model and those who included data to also include reasoning. Six groups revised their explanations so that all 14 groups included at least data and/or reasoning at the end. Students’ uncertainty rating ranged from 2 to 5 with an average of 3.8. Uncertainty ratings were rather stable before and after revisions as only one group changed their uncertainty rating to be higher (i.e. more certain) after revision. In initial submissions, six groups used internal attributions. Among the eight groups who used external attributions, four acknowledged a scientific data source without elaborating while four groups used external disposition based on singular mechanistic or distributional frequency-based accounts. Nine groups opted to revise uncertainty attribution: five groups’ revisions resulted in external scientific disposition. Three groups reached the external scientific limitation level that discussed factors not currently represented in the groundwater model and that could alter their claims about whether the water was trapped in the ground.

From the video analysis, we identified several ways in which intelligent feedback supported students’ recognition of uncertainty sources. First, feedback helped students frame uncertainty, which was understandably a novel task. Receiving low scores in uncertainty attribution was an eye opener to most students, which led to discussion and planning for what to do. For example, a group of students submitted uncertainty attribution by writing, “We are fairly certain of our answer because we watched many droplets come down and the path we chose was the fastest. We were also able to think out our answers reasonably.” They received a score of 1, which meant personal attribution.

S1: A one? And a one?

S2: Huh? How’d we get a one for that?

S1: We were fairly certain...*[after reading the feedback]* what are you certain about from the groundwater model?

Table 1. Categories of uncertainty attribution

Source of Uncertainty	Categories	Description of categories	Intelligent feedback statement
No information (score 0)	No response	<ul style="list-style-type: none"> <li>• Did not respond to the related uncertainty item but answered the linked claim and explanation items</li> </ul>	You haven't explained your certainty rating. Have you compared the strengths and weaknesses of the evidence that you used to support your claim?
	off-task response	<ul style="list-style-type: none"> <li>• "I do not know" or similar answers</li> <li>• Provided off-task answers</li> </ul>	
<b>Internal: Introspective confidence</b> (score 0)	Restatement	<ul style="list-style-type: none"> <li>• Restated the uncertainty rating</li> </ul>	
<b>Internal: Rationale</b> (score 1)	Question	<ul style="list-style-type: none"> <li>• Did/did not understand the question</li> </ul>	Your personal beliefs, experiences, and attitudes can influence your certainty rating. How do the strengths and weaknesses of the scientific evidence affect your certainty rating?
	General knowledge/ability	<ul style="list-style-type: none"> <li>• Did/did not possess general knowledge or ability necessary in solving the question</li> <li>• Did/did not learn the topic (without mentioning the specific topic)</li> <li>• Can/cannot explain/estimate</li> </ul>	
	Lack of specific knowledge/ability	<ul style="list-style-type: none"> <li>• Did not know specific scientific knowledge needed in the item set</li> </ul>	
	Difficulty with data	<ul style="list-style-type: none"> <li>• Did not make sense of data provided in the item</li> </ul>	
	Authority	<ul style="list-style-type: none"> <li>• Mentioned teacher, textbook, and other authoritative sources</li> </ul>	
<b>External: Scientific source acknowledgment</b> (score 2)	Nominal data source	<ul style="list-style-type: none"> <li>• Just acknowledged the existence of "data," "model," "chart," etc.</li> </ul>	You mentioned that either the data or the model affected your certainty rating. Can you be more specific about how the data or model influenced your rating?
<b>External: Scientific disposition</b> (score 3)	<b>Singular mechanism</b>	<ul style="list-style-type: none"> <li>• Referred to/elaborated knowledge or data that can explain the claim with data</li> </ul>	You mentioned specific evidence and knowledge that influenced your certainty rating. Have you also considered the strengths and limitations of the data and models related to this question?
	<b>Distributional frequency</b>	<ul style="list-style-type: none"> <li>• Compared and contrasted multiple cases</li> </ul>	
<b>External: Scientific limitation</b> (score 4)	<b>Ontic uncertainty</b>	<ul style="list-style-type: none"> <li>• Elaborated why the scientific phenomenon addressed in the item is uncertain</li> </ul>	You recognized strengths and limitations of knowledge and evidence related to the current investigation. Excellent!
	<b>Epistemic uncertainty</b>	<ul style="list-style-type: none"> <li>• Recognized the limitation of data provided in the item and suggested a need for additional data.</li> <li>• Mentioned that not all factors are considered</li> <li>• Mentioned that current scientific knowledge or data collection tools are limited to address the scientific phenomenon in the task</li> </ul>	

S2: We were certain that the other one runs slower. We did well on the first one [*explanation score 4*].

S1: I don't understand why we got through that [*explanation prompt*] really well.

S2: Then, we said, we said something else. We said that we...

S1: Hey, we're answering a completely different question than what it asked. That's why. [*Went back to the model to re-examine their evidence.*] It is asking us why we are certain about how the groundwater can get back up and be evaporated if it's not trapped!

S2: Yes!

S1: Are you certain of your answer?

S2: Oh, okay we figured it out.



S1: We are fairly certain, uh, wait, hold on....maybe it is trapped.

S2: No, it's not. We got a good score on explanation.

S1: Yeah, I know but it's not trapped.... This is what it is asking. It's asking, we answered, why we thought it is not trapped.

S2: Then, how do you explain it?

S1: Okay, we are certain or we are fairly certain because... um, 30% of the water we get is groundwater. And in the model it showed that water was being evaporated afterwards. Also, in the model, it showed the water evaporating from sediments.

*[Resubmitted uncertainty attribution and received a higher score related to external singular mechanistic attribution.]*

Second, students became more deeply engaged with interpreting data in light of their knowledge after receiving feedback. Students in the example above went back to the model to reexamine their claim; they also elicited a piece of knowledge that could be useful in interpreting data. The information that "30% of the water we get is groundwater" was learned earlier in the water module. Students voluntarily elicited this piece of knowledge to justify that water could not be trapped forever in the ground if they were to use groundwater in their life.

Third, feedback helped students revise their uncertainty attribution. We illustrate an example of a student group who first claimed that the water is trapped because "the water moved slowest through the black layer, so slow that you might think it blocks the water movement." With that explanation, the group chose an uncertainty rating of 4. The following sequence of uncertainty attribution occurred:

1. Initial attribution: "because we had an activity that backed up our reasoning."
2. Since this attribution was based on personal experience, feedback was given to students: "Your personal beliefs, experiences, and attitudes can influence your certainty rating. How do the strengths and weaknesses of the scientific evidence affect your certainty rating?"
3. First revision of uncertainty attribution: "because we had an activity that showed water sitting on top of the black layer which caused [us] to make the conclusion that the black layer absorbed water and it could not absorb any more so the water just sat on top."
4. The revision included their description of one outcome they observed in the groundwater model, the automated scoring engine recognized this statement as external, scientific disposition based on singular mechanism, thus the following feedback was provided: "You mentioned specific evidence and knowledge that influenced your certainty rating. Have you also considered the strengths and limitations of the data and models related to this question?"
5. In responding to this new feedback, the second revision was made: "because we had an activity that showed water sitting on top of the black layer which caused [us] to make the conclusion that the black layer absorbed water and it could not absorb any more so the water just sat on top. There are however limitations to the groundwater model because there are no plant life or roots in the ground that would help absorb some of the water." Note that students used a factor that was not represented in the current model to illustrate limitations in the model in making a claim about the groundwater trapping.
6. The second revision was recognized as external scientific limitation related to epistemic uncertainty. As such, congratulatory feedback was given to students: "You recognized strengths and limitations of knowledge and evidence related to the current investigation. Excellent!"

This revision sequence illustrates how students used intelligent feedback to turn their attention from internal to external sources of uncertainty and from focusing on finding an exemplar from the current data to limitations in the model where the data were generated.

## Significance

While the pressure for implementing the Next Generation Science Standards is mounting, integration of science practices into current teaching and learning in science classrooms appears difficult. As students engage in science practices independent of one another, how to support students' diverse needs becomes an important issue in the design of instructional support systems. When students engage in argumentation with model-based evidence, uncertainty is prevalent as data and evidence are not fully understood by them or are not fully representing the phenomenon under investigation. The intelligent feedback system we tested to promote scientific argumentation delivers immediate, tailored supports for individual students commensurate with their

progress on argumentation. Preliminary findings indicate that this automated feedback system can be seamlessly integrated into an online curriculum module to support students' uncertainty articulation about complex systems as part of written argumentation tasks.

## References

- Ben-Haim, Y. (2014). Order and indeterminism: An info-gap perspective. In M. Boumans, G. Hon, & A. C. Petersen (Eds.), *Error and uncertainty in scientific practice: History and philosophy of technoscience* (pp. 157–176). London: Routledge.
- Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, 93, 26–55.
- Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review*, 97, 303–352.
- Boumans, M., & Hon, G. (2014). Introduction. In M. Boumans, G. Hon, & A. C. Petersen (Eds.), *Error and uncertainty in scientific practice: History and philosophy of technoscience* (pp. 1–12). London: Routledge.
- Bricker, L. A., & Bell, P. (2008). Conceptualizations of argumentation from science studies and the learning sciences and their implications for the practices of science education. *Science Education*, 92, 473–493.
- Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, 38, 39–72.
- Erduran, S., Simon, S., & Osborne, J. (2004). TAPPING into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education*, 88, 915–933.
- Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, 11(2), 143–157.
- Kandlikar, M., Risbey, J., & Dessai, S. (2005). Representing and communicating deep uncertainty in climate-change assessments. *Comptes Rendus - Geoscience*, 337(4), 443–455.
- Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, 94(5), 810–824.
- Lee, H.-S., Liu, O. L., Pallant, A., Roohr, K. C., Pryputniewicz, S., & Buck, Z. E. (2014). Assessment of uncertainty-infused scientific argumentation. *Journal of Research in Science Teaching*, 51, 581–605.
- McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, 15, 153–191.
- McNeill, K. L., & Pimentel, D. S. (2010). Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education*, 94(2), 203–229.
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press.
- Nussbaum, E. M., Sinatra, G. M., & Owens, M. C. (2012). The two faces of scientific argumentation: Applications to global climate change. In M. S. Khine (Ed.), *Perspectives on scientific argumentation: Theory, practice, and research* (pp. 17–38). New York: Springer.
- Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education*, 93, 448–484.
- Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *The Journal of the Learning Sciences*, 12(1), 5–51.
- Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science*, 333(6048), 1393–1400.
- Staley, K. W. (2014). Experimental knowledge in the face of theoretical error. In M. Boumans, G. Hon, & A. C. Petersen (Eds.), *Error and uncertainty in scientific practice: History and philosophy of technoscience* (pp. 39–56). London: Routledge.
- Toulmin, S. (1958). *The uses of argument*. New York: Cambridge University Press.
- Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes*. New York: Cambridge University Press.

## Acknowledgments

This material is based upon work supported by the National Science Foundation (NSF) under Grant Nos. DRL-1220756 and DRL-1418019. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

# Creating Socially Relevant Mobile Apps: Infusing Computing into Middle School Curricula in Two School Districts

Lijun Ni, University at Albany, [lni@albany.edu](mailto:lni@albany.edu)  
Fred Martin, University of Massachusetts Lowell, [fredm@cs.uml.edu](mailto:fredm@cs.uml.edu)

**Abstract:** In this paper, we share our experiences implementing a professional development program in two school districts with middle school teachers who integrated an introductory computer science curriculum into their teaching. The 15 to 20-hour curriculum was based on students collaboratively creating mobile apps for socially relevant purposes with MIT App Inventor. Eleven teachers infused the curriculum into technology, math, engineering, library and art courses. We investigated how teachers modified the curriculum to fit their respective standards and students' needs. We discuss the challenges they faced and propose ways of addressing these barriers. We found that the teachers were successful in combining digital literacy skills with computer science—not only to facilitate students' learning, but also to connect with their diverse ethnic backgrounds and their contemporary passions.

## Introduction

This study is part of a larger project, Middle School Pathways in Computer Science (CS Pathways), aimed at establishing a sustainable, institutionalized computer science curriculum at the middle school level in two school districts. Working with the districts' existing teachers, the project is infusing computer science into the districts' existing curriculum, with an explicit focus on creating mobile apps that address local community needs. When fully implemented, the project will provide an introductory computer science experience to all middle school students in these two districts.

The funding program supporting the project is named Innovative Technology Experiences for Students and Teachers (ITEST). This program “promotes PreK-12 student interests and capacities to participate in the science, technology, engineering, and mathematics (STEM) and information and communications technology (ICT) workforce of the future”(NSF, 2016a). While the program is focused on student interest and pursuit of STEM careers, it explicitly recognizes the crucial role of teachers in this process. Our project, and this study, focus on this research question: “What instructional and curricular models can effectively engage teachers to utilize and integrate technologies so as to enhance student understanding of STEM-related occupations?” (NSF, 2016b). This project demonstrates the challenges and opportunities in working with teachers who have a range of backgrounds, including technology, library, engineering, and mathematics, understanding what is necessary for them to feel supported and be effective in teaching computer science to all students.

## Background

There is much evidence showing a need for greater high-tech skills in today's workforce (e.g., Olson & Riordan, 2012). Critically, there is substantial under-representation by women and ethnic minorities in technical fields, including computer science (Jackson, Starobin, & Laanan, 2013). This is a matter of social justice and international competitiveness (Leggon et al., 2015). Addressing this, since 1999 NSF has spearheaded a series of funding programs to “broaden participation in computing” and other STEM fields (Aspray, 2016). Most recently, the White House announced Computer Science For All, which strives to “empower a generation of American students with the computer science skills they need to thrive in a digital economy” (Smith, 2016).

In order to reach all students, it is necessary to have a curriculum and pedagogical approach that will engage all students. The Exploring Computer Science (ECS) project has demonstrated how to bring values of equity and inclusiveness into the classroom (Goode, Chapman, & Margolis, 2012). Furthermore, research has shown the power of engaging students from underrepresented groups with culturally relevant examples of computing (e.g., Eglash's culturally situated design tools) and a clear social purpose to computing with real-world applications (Buckley, Nordlinger, Subramanian, 2008; Eglash, Gilbert, & Foster, 2013; Fisher, & Margolis, 2002). The CS Pathways project's focus is to engage students in learning computing through creating mobile apps for socially beneficial purposes.

We use MIT App Inventor, a blocks-based programming system, as the design environment for students to create mobile apps that address local community needs. In App Inventor, students can simply drag and drop blocks of code to create an app, which can be downloaded to an Android mobile phone or tablet. Prior work has demonstrated its success in providing positive computing experience to students with its expressive power (Wagner et al., 2013; Sherman & Martin, 2015; Ni et al., 2016). We believe that App Inventor can be

used to democratize computing—to provide the expressive power of computing to all learners, not only the “small group of technically elite” (Wolber, et al., 2015). The CS Pathways computing curriculum was based on students collaboratively creating apps for socially relevant purposes with App Inventor.

## Partnership and teachers

The project partnership is with the school districts of Medford and Everett in MA. Medford has 63% white, 15% African-American, and 10% Hispanic students with 44% high needs students. This is approximately equivalent to state averages. Everett has 31% white, 18% African-American, and 44% Hispanic students with 62% high needs students. This is more diverse and low-social economic status than state averages. The districts have a history of collaborating on technology initiatives via a non-profit that was created with the districts’ support. Technology is a required subject at the middle school level in both districts; by infusing computer science into this course, all students would get an introduction to computer science. In the project design, all of the districts’ middle school technology teachers would participate. Other teachers were invited to participate too.

In the project’s first year, we recruited Cohort 1 including: one of the two technology teachers, an engineering teacher, and an art teacher in Medford; two of the five technology teachers in Everett. In the second year, Cohort 2 included the other technology teachers from both districts, a librarian, and a math teacher. In its final year, Cohort 3 will consist of a replacement technology teacher and a science teacher (see Figure 1).

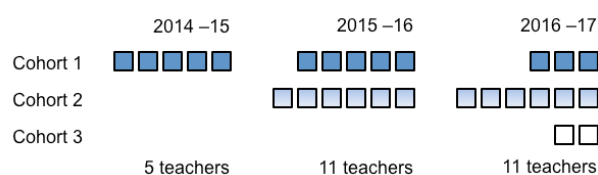


Figure 1. Teacher Cohorts

## Professional development

The professional development (PD) was designed as an ongoing, multi-year process to introduce teachers to computer science content and pedagogy, support them in classroom implementation, and encourage them to share their learning with each other. The PD was led by the Teacher Learning Center Director, a full-time staff member hired for the project, and the second author, with contributions from the whole project team.

### PD goals and structure

The professional development encompassed 38 hours of instructional, meeting, and homework time per teacher per year. Our project was funded to begin in Fall 2014; thus a summer PD with the first cohort of teachers was not possible. Instead, we organized a series of ten 2-hour afterschool meetings, beginning in October 2014 and ending in January 2015. For the second cohort, we ran a one-week summer camp. In both cases, for a given teacher’s first year, the content of the PD was the same, and included these elements:

1. Introduction to building apps in MIT App Inventor, including use of digital media (images, sound) and blocks-based programming;
2. Computer Science pedagogy of equity and inclusiveness (e.g. contesting the “geek gene”, using pair programming, understanding the values of the ECS project);
3. Engaging teachers in an original app design process, from brainstorm to planning to completion, so they could facilitate this with their students;
4. Lesson planning, including integration of new material and addressing standards.

For the Cohort 1 teachers in their second year, we included differentiated PD. We conducted a series of four mini-workshops. Each included readings, a homework assignment, and debrief conversation that was conducted via video conference. The monthly topics included assessment of student work, pedagogical content knowledge as it applies to computer science, social impact of computing including career opportunities, and advanced programming in App Inventor (e.g., using index variables and lists).

## Methods

Our project’s external evaluator administrated baseline surveys and end-of year surveys with the 11 teachers to assess the quality of the PD program and teachers’ experience and attitudes. The Teacher Learning Center Director regularly visited project classrooms and took observation notes. The last PD session of the school year

focused on group reflection on how the curriculum was implemented. The project researchers took meeting notes and discussed afterward. This informed the development of a teacher interview protocol.

We use interviewing as our qualitative method to explore teachers' curriculum implementation experiences (Seidman, 2005). Interviews were conducted to further understand the variation in their curriculum design and implementation, and barriers or opportunities for implementation in different classroom contexts. Teachers were asked to describe how they used/modified the model lesson plans, to what extent the curriculum was implemented, what challenges they encountered, and the most helpful things supporting their teaching. We interviewed eight of the eleven teachers in Spring 2016, representing six schools and both school districts.

The next section presents the stories teachers reported about their curriculum implementation. These teacher reflections have been triangulated with our analysis of group reflection, classroom observations, and interviews.

## Findings

Teachers gave high ratings to the professional development. Most teachers reported that the overall quality of the PD, app development support, and overall usefulness as "good" (44%) or "excellent" (44%). Teachers also reported increased confidence in using apps, creating apps, and using computer terms ( $p < .05$ ). In addition, Cohort 2 teachers reported an increased frequency of creating apps ( $p < .05$ ).

During the past school year, all the 11 teachers implemented the CS Pathways curriculum into existing courses, reaching an aggregate total of over 750 students. Next sections report stories from the eight teachers participated in our research interviews.

## Curriculum implementation

The eight teachers we interviewed include three teachers from Cohort 1 and five teachers from Cohort 2. The five teachers (P1-P5) came from one district and P6-P8 were from the other district. These teachers taught the CS Pathways curriculum within technology, engineering, library, or mathematics classes with 6th, 7th or 8th graders (see Table 1). The curriculum was implemented with different class schedules. Teachers from one school district met students more often while the other teachers saw students once every 6 days, 8 days or 9 days. The research team also investigated the project's impact on students' attitudes and computational thinking skills through surveys and analysis of student work. The focus of this paper is on teacher PD and curriculum implementation, and results on student learning are being prepared for separate publication.

Table 1: Interviewed Teacher Information

Teacher	Cohort	School	Area/Course	Schedule	Grade
P1	1	A	Technology	6-day cycle	7 <sup>th</sup> 8 <sup>th</sup>
P2	1	B	Technology	8-day cycle	8 <sup>th</sup>
P3	2	C	Library	8-day cycle	6 <sup>th</sup>
P4	2	C	Technology	9-day cycle	8 <sup>th</sup>
P5	2	D	Technology	8-day cycle	7 <sup>th</sup> 8 <sup>th</sup>
P6	1	E	Engineering	Twice/week	6 <sup>th</sup>
P7	2	F	Math	3 times/week	6 <sup>th</sup>
P8	2	E	Technology	Twice/week	8 <sup>th</sup>

## Model lesson plans

During the 1st year of the project, Cohort 1 teachers developed and implemented a 15–20 hour computing curriculum with App Inventor. Prior to the start of the project's 2nd year, a series of model lesson plans were created from the teachers' work and shared with all the project teachers. These lesson plans were drawn from the work of three Cohort 1 teachers. Two were from the same school district and had collaborated closely on creating each lesson. The third teacher's lesson plans used a faster pace with varied starter apps. The Teacher Learning Center Director edited the curriculum materials from these three teachers by linking CSTA curriculum standards (Seehorn et. al, 2011), checking overall sequence, and using a standardized lesson plan template.

Students were introduced to App Inventor and then had three to five classes to create original final apps. They had opportunities to share their work to the class or the whole school community. The curriculum included one lesson introducing students to computing and career pathways, and another lesson inviting local computing professionals to visit the classrooms. Table 2 shows the overall curriculum sequence.

The model lesson plans provided a map for Cohort 2. All teachers were encouraged to modify the lessons based on their own classroom needs. They infused the lessons into their existing curricula with a variety

of strategies, including using different starter apps to motivate and inspire students, integrating digital literacy, and changing sequence.

Table 2: Model Curriculum Sequence

Day	Overview of Lessons	Day	Overview of Lessons
1	Pre-survey; using your Google login	9	Media app with if-then-else programming
2	Overview of computing and career fields	10	Apps for social good / final project brainstorming
3	Code.org block programming intro / CS Unplugged	11	Storyboarding final project app
4	Introduction to tablet and App Inventor	12	Working on final project app
5	First app: Talk To Me (text to speech)	13	Working on final project app
6	Talk To Me part 2 and sharing	14	Finish final app; export source / installer files
7	Media player app using w/preloaded starter code	15	Computing professional visit
8	Digital literacy: gathering own media for apps	16	Post-survey and app showcase

## Adapting the model lesson plans

### Integrating with existing curriculum

The project teachers integrated the CS Pathways computing curriculum into technology/computer, engineering, library, math, and art classes with students from grade 6 to 8. All the eight interviewed teachers reported how they felt this computing curriculum could fit into the existing middle school curricula.

### **Computer/Technology Course**

The technology teachers recognized that the curriculum included digital literacy components, which were consistent with their existing technology curriculum. For example, students used Google search to find pictures to make slides. They searched pictures and presented them in apps too. A few teachers explicitly added relevant digital literacy content into the lesson plans: e.g., how to download and upload images and sounds to apps. In addition to seeing the overlapped content of digital literacy, the technology teachers also understood that the CS Pathways curriculum taught more computer science beyond applications. They saw the values of teaching computational thinking within the technology curriculum as “high-level things”. One technology teacher said:

“You have an idea, then you start thinking, that’s the computational thinking. I introduced to them you have to solve a problem; you have to breakdown each steps, you have to use blocks... These are the higher level things.”

### **Engineering Course**

The engineering teacher felt that it was challenging to cover the required engineering content with 15–16 hours focused on computer science of the 40 hours total of class time. As a compromise, he experimented with an integration approach by asking students to develop final apps around one of the six engineering topics in this course. Students developed engineering-focused apps, such as informational apps on transportation, machine tools or game apps on futurist flights. The engineering teacher explained how this change worked for his class:

“I started with social good apps for communities, but there is an issue for me. If I have 15–16 hours as computer science, that’s a huge chunk of class time. So we came to a compromise: the final apps have to do with engineering. They spent 3–5 classes on the final apps. They were thinking about engineering and focusing on engineering. It did make me feel a lot better because obviously that’s part of my curriculum... I want them to see what kind of apps they could make and give them some tools, and then they can make their own engineering apps.”

### **Library Course**

The library teacher was excited about finding ways of integrating the computing curriculum into a library course. She changed the media player sample lesson, which was based on showcasing Martin Luther King’s speeches, to bibliography apps on book authors. She identified relevant curriculum components in library curriculum standards. She reported this as a good opportunity of integrating computing into the library course.

“Here is all about the author, title... the students can choose a book to make a game, or informational app, some type of interactive app... I found the American Association for School Library has a framework that is more technology-oriented. The copy of that is almost

of the CSTA standards, and there was some overlap, something like digital citizenship; there were also things like acceptable use of technology... I was able to tie with a lot of that.”

For students’ final apps, the teacher relaxed the constraint of book authors only. She allowed a student to create a bibliography app about pop star Justin Bieber. The student recorded and uploaded his own voice about Bieber’s biography and using the media player component to play the audio. We support this approach of connecting to students’ interests.

### **Mathematics Course**

The math teacher saw the value of teaching logical thinking through this curriculum and felt it fit well with mathematics standards. He adopted the starter lesson on “if-then-else,” which was the media player using content from Martin Luther King’ speeches. Students were provided with a starter app and asked to change the app to control the media player to play and pause a media file. The math teacher explained his choice as below:

“I did the if-then-else. I think it’s very useful structure. I try to teach programming, following sequences. This is very good basic logic structure that particularly adds its lens to the implementation. I probably focus more on this than other teachers, with a goal of building computational thinking. It is perfect for the math—how to model, structure... they need to structure their thinking in order to do the apps as they want to do. There is a huge need for the students to develop to be successful.”

This teacher also relaxed the content requirement and allowed students to choose material based on their own interests. For him, this gave the students motivation to engage with the important if-then-else logic.

### Customizing starter apps and examples

Teachers used starter apps to scaffold learning. They modified or created new starter apps to meet a variety of students’ interests. The CS Pathways curriculum was implemented with a diverse group of students, including English Language Learner (ELL) students and students with special needs. One teacher had a group of ELL students creating educational apps to share their cultural heritage. A group of students created a whimsical “Cheese Around The World” app. Each of five students recorded themselves saying the word for cheese in the language they spoke at home. They incorporated these into an app with flags from each country. Clicking on a flag will make the app say “cheese” in that country’s language. This was an example of modification that supported a culturally diverse group of students’ self-expression and creativity—a simple but still powerful version of Eglash’s culturally situated computing. Another teacher had her special-needs students make a zoo menagerie—instead of the media player app based on the Martin Luther King’s speeches. This teacher assigned each of her students a different animal. Each student located a photo and sound of their animal, and made a player app from those digital media components. This teacher had previously shown her students the basic HelloPurr app, which was much less complicated.

Centered on the theme of social-good/community-needs, another teacher created and presented a school introduction app to her students. She recorded her own voice introducing the school and welcoming students and families as the principle with a map of the school buildings included in the app. Students felt this app was very cool and got inspired and motivated to learn about creating apps, including how to record audios, upload media, using media player component and buttons.

### Integrating digital literacy with computer science

Through the first year of PD program, we identified the need of teaching digital literacy to our participating teachers while they were learning computer science content. Over the process of implementing the CS Pathways curriculum, teachers also found that students were not well equipped with digital literacy skills. They needed to teach relevant digital literacy to support students’ learning of computing concepts, especially with 6<sup>th</sup> graders, who had not learned most of the computer literacy involved in this curriculum. In particular, the App Inventor curriculum requires using Gmail accounts and working on media resources. Teachers identified the need of explicitly teaching these skills. While following with the model lesson plans, teachers integrated relevant digital literacy content within the lessons. This work, which arose naturally from projects involving App Inventor, dovetailed with both the technology teachers’ existing curriculum standards and Massachusetts’ new Digital Literacy and Computer Science K-12 standards (Chester, 2016).

### **Facilitating implementation**

Teachers reported three aspects of the projects were most helpful for facilitating their teaching of the new curriculum: the model lesson plans, regular PD meetings, and ongoing support from the project team.

### The impact of PD

Teachers reported that the PD sessions were very helpful. They especially saw the value of being able to communicate with other teachers regularly for sharing, reflecting on their teaching and supporting each other. As one teacher said:

“It’s good for the teachers to collaborate together. I like to meet face to face because there is always someone in the group comes up and offer something for the conversion. Say, that sounds a great idea. I have got some ideas from [teacher name]. He has lots of great ideas... And someone might say, I haven’t thought about that. I want to try that too... Especially, when you first start, it’s nice to see others have maybe intimidation, make you feel a little better. The roundtable sessions are really instrumental. It offers a good start.”

### Ongoing support for teachers

This project offered ongoing support for teachers in implementing the CS Pathways curriculum. The Cohort 1 teachers participated in another four online mini-PD sessions. Cohort 2 teachers received regular classroom visits from the Teacher Learning Center Director, as well as computer science undergraduate and graduate students to help with troubleshooting technical issues and answering curriculum-related questions. Teachers all reported these visits as very “helpful.” Teachers realized the need of ongoing learning, and they also felt comfort with seeking support from the project team and other teachers.

“I need support for some more advanced things... Sometimes students want to make things more game like, maybe things such as a clock timer... I need more practice myself, I would need support in such things, but I know where to find resources. If I feel really stuck, I can email [project team] or anyone involved in the project. They are more than happy to help.”

### Model lesson plans and handouts

The three Cohort 1 teachers felt proud of their own contributions to the model lesson plans. Meanwhile, the Cohort 2 teachers used the lesson plans to start their own explorations. Even if they modified the lessons, all teachers appreciated the availabilities of these lessons. One teacher referred to the lessons as a “life saver.” The Teacher Learning Center Director also worked with teachers to create a series of class handouts. These handouts outlined specific apps or introduced technical settings (e.g., how to use the tablets as audio recording devices). Teachers felt these handouts were very helpful in reducing their anxiety and supporting their teaching. One teacher said:

“The model lesson plans are life saver... For me, I really like to have cards on that, for students to use to refresh memory, or just check things where it is that is working in the blocks, to clarify things. And have handouts for them to see... If you rename a button, you may spend 10 minutes to find it.”

### Barriers to implementation

Teachers were also asked to reflect on the major challenges and difficulties they encountered in implementing the curriculum. During the interviews, they reported four major challenges: access to labs and technology, the need of ongoing learning, motivating students, and class scheduling.

### Access to labs and technology

Teachers reported it was very challenging to access labs and set up technical settings ready for the class. Some lost access to their labs for significant periods of time because the computers were needed to administer standardized tests. Teaching this curriculum required access to devices (computers and tablets), WiFi, and student Gmail accounts to log into App Inventor. Any problem with these technical components could consume class time and prevent learning. For example, one teacher was struggling with an unstable network. Another teacher described another situation she encountered and felt very frustrated:

“The kids were all assigned Gmail accounts and they were broken into pairs... After they got all the introduction done and started implementing easier apps, they started getting better at it... One of our technicians changed the kids’ email accounts. They couldn’t access their original App Inventor accounts where their apps were in. That was very frustrating to my students, which robbed them off enthusiasm. Then my lab was shut down for over a month due to PARCC testing and special-ed testing... It’s hard to bring the kids back to the flow... I did bring them back with new accounts... but I lost lab for 5 weeks, very frustrated.”



The project team sent staff to help with setting up devices, which offered great support for the teachers to start with working devices. This kind of support addressed part of the technical issues. Teachers needed to prepare lessons that would advance their goals in the face of technology failures. Meanwhile, there was still a great need of technical support at the school level for the curriculum implementation.

### Need of ongoing learning

Although teachers reported an increase of their self-confidence in teaching the curriculum through post-school year surveys, they felt a strong sense of in need of ongoing learning. Four teachers reported this as a challenge for their teaching. First, the curriculum was new for them, including technology/computer teachers. Teachers felt the curriculum content was new and difficult. One technology teacher said:

“It’s difficult for me. I would like to have more support... I still don’t feel I have known enough to run this all by myself... For me, coding is like learning a new language... It’s difficult. I know definitely I have learned a lot and the kids have learned a lot too.”

Another teacher further explained that she also felt challenged when students started making their own apps. She needed to learn more to be able to help the students.

### Social good and student motivation

Two teachers reported another challenge related to motivating students. The project had a major focus on creating apps for social good, which was designed to motivate students’ learning. However, teachers found some of the students were not excited about creating social-good apps. Therefore, they needed to find other ways of motivating students. One teacher allowed her students to create apps on other topics of their interest.

“When they were told they could make their own apps for something social good, to solve a problem, they were not that thrilled about it. They want to make something interesting to them... I want to find some middle ground for them. Making apps is so exciting, but making social good apps are not so to them.... I want them to do something. A lot of them go towards games. I’m ok with that... I’m starting, I keep thinking of different ways of doing things.”

Looking ahead, we will encourage teachers to consider any apps that connect to students’ personal interests as socially relevant, supporting a culturally diverse group of students’ self-expression and creativity.

### Scheduling and standards

Teachers reported time/schedule-related issues as the biggest challenge. Teachers from one district had a long-cycle schedule: teachers met students every 6 days, 8 days or 9 days. Ideally, teachers would like to see students more often. Teachers felt this long-cycle schedule made it hard to teach the curriculum. On the other hand, the math teacher and engineering teachers felt the pressure of competing academic requirements together with teaching the new computing curriculum. The math teacher implemented the curriculum at the second half the school year after completing the required math requirements. He saw the challenge of getting students ready for math tests while providing sufficient time for students in exploring and creating final apps. He reduced class time for students to work on final apps. The Engineering teacher had similar worries. He changed the lesson plans to include engineering-related apps as students’ final projects. He was hoping the new state standards on digital literacy and computer science would be included in the engineering curriculum.

“It’s not in my standards... Computer science is just another thing. There is no room to fit with everything. I hope the new state standards become [part of] what I am supposed to teach.”

## **Discussion and conclusions**

In this study, teachers shared their experiences in implementing the computing curriculum with a community focus into their existing curricula. First, their stories have demonstrated how socially relevant computing curriculum could serve a diverse group of students. The students’ “Cheese Around The World” app was a playful way for them to share their cultural heritage with each other. Other apps, such as the Justin Bieber bio app, allowed students to share their passions with each other. We see parallels between Eglash’s culturally responsive computing (Eglash et al., 2013) and our teachers’ support of these kinds of projects. Further, building apps that exercise digital literacy skills is a comfortable on-ramp to projects that involve more complex coding, for teachers and students alike. With Massachusetts’ new standards for digital literacy and computer science (Chester, 2016), we see the opportunities for this approach to become widespread.

This study identifies both facilitators and barriers for teachers’ implementation of this computing curriculum. We see teachers made significant efforts and worked closely with the project team to address a

variety of challenges. Our experiences suggest a few strategies for infusing such a computing curriculum into middle school curricula. First, it can be challenging to work with diverse groups of students with different needs. Teachers can experiment with different strategies for motivating students and scaffolding learning, such as using starter apps, customizing app themes, modifying paces for students with special needs. Second, districts must commit to technology support and access. Teachers are demoralized when their instructional labs are appropriated for computer-based standardized testing. We saw a teacher gaining necessary access to her lab only because the district temporarily reverted to paper-based testing. Third, teaching (new) computing curriculum requires ongoing learning. Teachers value learning, reflection and conversations with their peers.

Our project's immediate goal is to establish an institutionalized computer science curriculum at the middle school level in the two districts. Our 3rd-year PD program is focused on supporting teachers with curriculum revision and institutionalization. Our experiences indicate that it is a long way for teachers to become self-sufficient. It also takes systemic support from the school district and above in terms of administrative support and standards.

## References

- Aspray, W. (2016). Opening computing careers to underrepresented groups. *Participation in Computing*. Springer International Publishing, 13-52.
- Buckley, M., Nordlinger, J., Subramanian, D. (2008). Socially relevant computing. *ACM SIGCSE Bulletin*, 40(1), 347-351.
- Chester, M. (2016). Digital Literacy and Computer Science (DLCS) Curriculum Framework. Massachusetts Department of Elementary and Secondary Education.
- Eglash, R., Gilbert, J. E., & Foster, E. (2013). Toward culturally responsive computing education. *Communications of the ACM*, 56(7), 33-36.
- Fisher, A., & Margolis, J. (2002). Unlocking the clubhouse: the Carnegie Mellon experience. *ACM SIGCSE Bulletin*, 34(2), 79-83.
- Goode, J., Chapman, G., & Margolis, J. (2012). Beyond curriculum: The Exploring Computer Science program. *ACM Inroads*, 3(2), 47-53.
- Jackson, D. L., Starobin, S. S., & Laanan, F. S. (2013). The shared experiences: Facilitating successful transfer of women and underrepresented minorities in STEM fields. *New Directions for Higher Education*, 2013(162), 69-76.
- Leggon, C., McNeely, C. L., & Yoon, J. (2015). Advancing women in science: Policies for progress. *Advancing Women in Science*. Springer International Publishing, 307-340.
- Ni, L., Schilder, D., Sherman, M., Martin, F. (2016). Computing with a community focus: Outcomes from an App Inventor summer camp for middle school students. *Journal of Computing Sciences in Colleges*, 31(6), 82-89.
- NSF (2016). Innovative Technology Experiences for Students and Teachers (ITEST) program synopsis. [https://www.nsf.gov/funding/pgm\\_summ.jsp?pims\\_id=5467](https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5467). Retrieved August 19, 2016.
- NSF (2016). Innovative Technology Experiences for Students and Teachers program solicitation #15-599.
- Olson, S., & Riordan, D. G. (2012). Engage to excel: Producing one million additional college graduates with degrees in STEM. *Report to the President*. Executive Office of the President.
- Seehorn, D., Carey, S., Fuschetto, B., Lee, I., Moix, D., O'Grady-Cunniff, D., ... & Verno, A. (2011). *CSTA K-12 CS Standards*. <https://csta.acm.org/Curriculum/sub/K12Standards.html>. Retrieved Nov. 4, 2016.
- Seidman, I. (2005). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. Teachers College Press.
- Sherman, M. & Martin, F. (2015). The assessment of mobile computational thinking. *Journal of Computing Sciences in Colleges*. 30(6), 53-59.
- Smith, M. (2016). Computer Science For All. <https://www.whitehouse.gov/blog/2016/01/30/computer-science-all>. Retrieved November 4, 2016.
- Wagner, A., Gray, J., Corley, J., Wolber, D. (2013). Using App Inventor in a K-12 summer camp. *SIGCSE '13*, 621-626.
- Wolber, D., Abelson, H., Friedman, M. (2015). Democratizing computing with App Inventor. *GetMobile: Mobile Computing and Communications Review*, 18(4), 53-58.

## Acknowledgments

We thank our project team members (Akira Kamiya, Molly Laden, Diane Schilder, Farzeen Harunani and Mark Sherman) for their collaboration and all the participating teachers. This material is based upon work supported by the National Science Foundation under Grant No. 1433592.

# Exploring a Text-Mining Approach as Rapid Prototyping Tool for Formative Assessments in Inquiry-Based Online Learning

Alejandro Andrade, Chris Georgen, and Michael Stucker  
laandrad@indiana.edu, cgeorgen@indiana.edu, mstucker@indiana.edu  
Indiana University Bloomington

**Abstract:** We make a preliminary case for a computational method intended to facilitate real-time formative assessment in online inquiry-based learning environments. With a focus on talk and text as disciplinary, we aim to address how learning analytics, in this case, text mining, can provide learners and instructors with meaningful information in rapid and real-time to support learning and engagement. Our results show that in measuring the distance between the expert and learners' discourse from forum posts and verbal discussions, resulting similarity values can offer stakeholders evidence of student learning trajectories. Moreover, similarity values provide teachers with an automated measure of students' progress toward disciplinary discourse, and also reveal critical moments during the collaborative activity where more alignment with disciplinary ways of talk are being enacted.

## Introduction

In recent years, there has been increasing recognition of the synergies between Computer-Supported Collaborative Learning (CSCL) and Learning Analytics (LA) (Jeong & Hmelo-Silver, 2016; Ludvigsen, 2016; Rodríguez-Triana, Martínez-Monés, Asensio-Pérez, & Dimitriadis, 2015). For example, LA has been used to analyze large amounts of student-generated data to explore and refine learning trajectories during collaboration (e.g., Xing, Guo, Petakovic, & Goggins, 2015). Another avenue of research uses LA to provide rapid or real-time feedback on student performance to instructors as they organize and orchestrate collaboration (e.g., Berland, Davis & Smith, 2015; Vatrappu, Teplovs, Fujita, & Bull, 2011). With this in mind, exploring the use of automated or semi-automated techniques is important in online, semi-structured learning environments where teachers are required to parse large amounts of information in order to make quick inferences about student learning. To make such inferences, teachers need support in developing both summative and formative assessments of the learning that takes place during collaboration. The purpose of this paper is to explore the affordances of a computational approach to facilitating real-time formative feedback in CSCL environments. With a focus on the discursive practices of collaboration, we aim to address how learning analytics—in particular text-mining—can capture whether a particular collaborative activity has an effect on the quality of learning. Moreover, we see text-mining and automation as capable of delivering rapid, iterative assessment prototypes that provide stakeholders with meaningful, albeit coarse-grained, data that may support collaborative learning.

## Theoretical background

Our focus is on the effect of collaboration in fostering alignment between student and expert discourses as students progress through inquiry-based learning activities. Discourse alignment serves as an indicator of sociocultural learning, because it is assumed that the way in which learners engage in dialogue is evidence of how they engage with knowing and reasoning in a particular field or discipline (De Liddo, Shum, Quinto, Bachler, & Cannavacciuolo, 2011). From a sociocultural perspective of learning and knowing (Brown, Collins, & Duguid, 1989), it is expected that the more students learn in a particular field, the more they are to adopt the ways of expert talk—as in the process of enculturation (Lave & Wenger, 1991). Here, we focus on a circuitry course wherein students discuss current and voltage in a particularly disciplinary way. For example, initially, a student may underspecify disciplinary principles when referring to the elements in a circuit: “The switch is open which means no *electricity* is flowing.” Later, as she advances through the class and develops her disciplinary discourse (Mercer, 2008), she might say: “Since the switch was open, this *series* circuit is *incomplete* and the *current* couldn't flow around the circuit any more.” Compared to the former, the latter statement is more disciplined to circuitry. To understand the development of, and to explore automated ways for, assessing students' disciplinary discourse, we explored a computational approach to automatically analyze text-based data. Subsequently, we applied this analysis to transcripts of student inquiry in order to target particular moments of disciplinary talk in collaboration. This text-mining approach provides measures of learners' discursive alignment to expert benchmarks throughout online, inquiry-based learning environments.

## Methods

### Participants and data sources

We collected textual and video data from undergraduate students ( $n=21$ ) enrolled in a sound engineering course focused on the mathematical and engineering principles of analog electronics. The course is taught by two engineers, who serve as our expert benchmarks, each with considerable technical, professional, and teaching experience. In general, the course is structured around six modules that culminate in collaborative inquiry around unintuitive, yet foundational, concepts of audio engineering. This inquiry is conducted in Peer Investigation Groups (PIGs) composed of stable groups of 3-5 students. PIGs are structured by four sequential activities: 1) individual answers (text), 2) preliminary discussion and group answers (text), 3) online collaborative inquiry (video), 4) final group answers (text). In phases 1, 2, and 4, each expert individually scored student textual responses [0,5]. In phase 3, video from synchronous online inquiry was collected using screen capture tools embedded within the videoconference software (Figure 1).

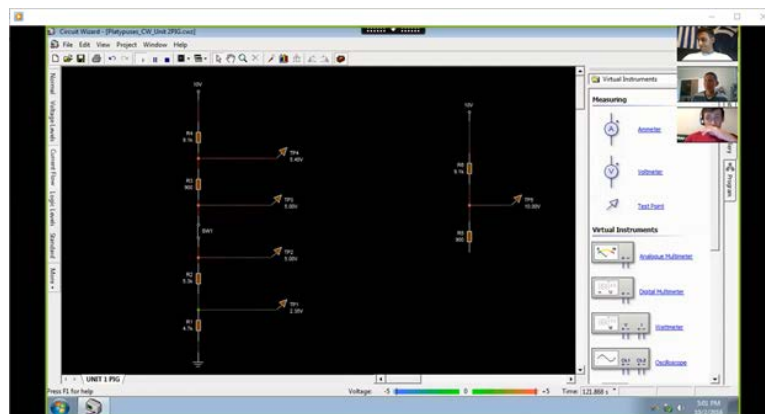


Figure 1. Online synchronous videoconference meeting with the circuitry simulator.

Our analysis will focus on student work in the second of six PIGs (PIG 2). PIG 2 targets interactions between voltage and current through the exploration of resistors and dividers. Students were given a simple circuit (Figure 2.a) and asked to calculate voltage at each test point. Next, the circuit was opened between two test points (Figure 2.b) and voltages were calculated again. The common misconception is based on a misinterpretation of Ohm's Law: you cannot have voltage without current. Since the circuit is open and current cannot flow through the entire circuit, students incorrectly predict the voltage as zero at each test point. Finally, students were asked to "describe why your answers for the open switch are what they are." The goal here was to require students to articulate the engineering principles behind their calculations.

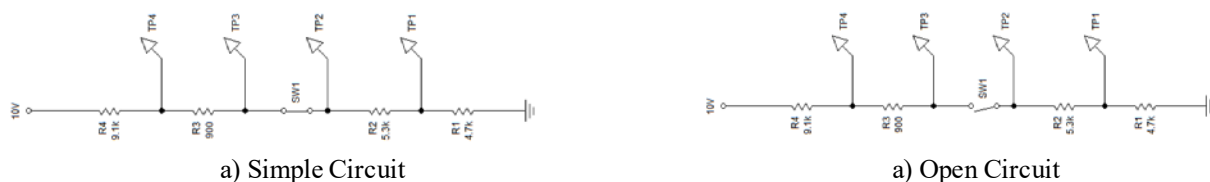


Figure 2. PIG 2 circuitry problem.

### Textual data

To examine the effect of collaboration we collected student individual, group, and final group answers to the question "describe why your answers for the open switch are what they are." We focused on this question due to prevalence of disciplinary discourse. Moreover, mathematical calculations of voltage are quite simple, whereas the majority of errors are based in misapplications of engineering principles. An example of these forum posts can be seen in Table 1.

Table 1: Examples of pre-, during-, and post-collaboration answers

Pre-Collaboration	During Collaboration	Post-Collaboration
-------------------	----------------------	--------------------

<p>“The switch is open which means no electricity is flowing. Therefore, no voltage is going through the circuit either.”</p>	<p>“Since the switch was open, this series circuit is incomplete and the current couldn't flow around the circuit any more. Therefore, there would be no voltage at any point of the circuit except for the battery.”</p>	<p>“Once the switch is open, TP3 and TP4 would connect to ground individually and it becomes a parallel circuit. In this parallel circuit, R4 and R3 would have the same voltage as the source, which is 10V.”</p>
---	---	--

### Video data

To investigate the process of collaboration we recorded online collaborative inquiry. The video data was transcribed clean verbatim.

### Expert benchmarks from course instructors

The course instructors provided their own answers to the questions and were used as the expert benchmarks. An example of a benchmark response is shown in Table 2.a.

### Analysis of student-expert discursive alignment in pre- and post-collaboration responses

In order to get a similarity value from each student response to the expert’s benchmark response, we constructed a document-term-frequency matrix. This approach is sometimes referred to as a “bag-of-words” approach (Bird, Klein, & Loper, 2009). The rows in this matrix are text-documents such as the students’ and instructor answers to the PIG question; columns are terms (relevant words); and entries are term frequencies. Thus, each row is a vector of term frequencies for a particular document. Following Kopainsky, Pirnay-Dummer, and Alessi (2012), we extracted only nouns and names by using part-of-speech (POG) tags. We only took names and nouns as they might represent concepts (Kopansky et al., 2012), but we also could have also accounted for verbs and adverbs that represent relationships between those concepts. In future iterations of the project, we will address these methodological variables by systematically comparing the inclusion of various parts of speech to the text-mining algorithm.

We illustrate the procedure with the instructor answer shown in Table 2.a. We used Python’s Natural Language Processing NLTK package to produce POS tags, as shown in Table 2.b. In keeping only those terms with ‘NN’, ‘NNP’, or ‘NNS’ tags, we stripped the document of everything but nouns and names. The following step was to get rid of capitalization and stem the words so that ‘circuit’ and ‘circuits’ are not counted as different terms, as shown in Table 2.c. Note that though the word ‘resistance’ was stemmed as ‘resist,’ ‘resistor’ was not. This list also contains some non-relevant terms such as ‘word,’ ‘end,’ and ‘matter,’ but for the most part the automated word stemmer works relatively well. Then, we converted the list into a term-frequency vector where entries are term frequencies (see Table 2.d). Although the term-frequency vector (Table 2.d) is a high-level representation of the original text (Table 2.a), the claim we make here is that it contains all the relevant features to represent a viable approximation of enacted discourse.

Table 2: Example of Text-Mining Process and Analysis

<p>(a) Expert benchmark example:          “You can have Voltage without current, but not current without voltage. Since there is an open switch we know that current equals zero. To find the voltage drop of each resistor we insert that 0 amps of current into Ohm’s law to get <math>V=0 \cdot R</math>. We know that V will always equal zero (no matter what the resistance is). If each resistor drops 0 volts, then that means there is 0 volts difference between the two ends of each resistor, in other words the voltage is the same at both ends of each resistor.”</p>
<p>(b) Vector with Part-of-Speech tags for each term in expert answer:          [(‘You’, ‘PRP’), (‘can’, ‘MD’), (‘have’, ‘VB’), (‘Voltage’, ‘NNP’), (‘without’, ‘IN’), (‘current’, ‘JJ’), (‘,’), (‘,’), (‘but’, ‘CC’), (‘not’, ‘RB’), … (‘in’, ‘IN’), (‘other’, ‘JJ’), (‘words’, ‘NNS’), (‘the’, ‘DT’), (‘voltage’, ‘NN’), (‘is’, ‘VBZ’), (‘the’, ‘DT’), (‘same’, ‘JJ’), (‘at’, ‘IN’), (‘both’, ‘DT’), (‘ends’, ‘NNS’), (‘of’, ‘IN’), (‘each’, ‘DT’), (‘resistor’, ‘NN’), (‘,’), (‘,’)]</p>
<p>(c) Vector with only stemmed nouns and names:          [‘voltage’, ‘voltage’, ‘switch’, ‘equal’, ‘zero’, ‘voltage’, ‘drop’, ‘resistor’, ‘amp’, ‘ohm’, ‘law’, ‘v=0*r.’, ‘v’, ‘zero’, ‘matter’, ‘resist’, ‘resistor’, ‘volt’, ‘volt’, ‘differ’, ‘end’, ‘resistor’, ‘word’, ‘voltage’, ‘end’, ‘resistor’]</p>

(d) Term-Frequency Vector:

FreqDist({'voltage': 4, 'resistor': 4, 'end': 2, 'zero': 2, 'volt': 2, 'ohm': 1, 'differ': 1, 'law': 1, 'word': 1, ...})

We followed the same procedure with each student response, and by stacking these vectors, we produce a document-term matrix, with as many rows as there are students, and as many columns as there are terms contained in the forum posts. For instance, an excerpt of the document-term-frequency matrix for the pre-collaboration PIG looks like the one shown in Table 3.

Table 3: Example of a Document-Term-Frequency Matrix

Document/Term	+10v	account	act	amp	chang	...	circuit
Expert	4	0	0	1	0	...	1
Student 1	0	1	0	0	0	...	2
...	...	...	...	...	...	...	2
Student 21	0	0	1	0	1	...	2

The distance from each student document to the benchmark document was computed using the cosine similarity metric (Leydesdorff, 2005), which measures the cosine of the angle between two vectors in a multi-dimensional space. Cosine similarity values range between 0 and 1, where 0 represents two totally different and 1 represents two totally identical texts. We selected the cosine similarity metric because of its robustness in the presence of sparse vectors (vectors with many zeros, Leydesdorff, 2005), as is the case with our dataset. To study the effect of collaboration in the improvement of individual students' disciplinary discourse, group mean similarity values were computed between the expert benchmark and pre- and post-collaboration phases and compared using a dependent-measures t-test at 5% significance level.

#### Validity and reliability of similarity values

To analyze the validity of the similarity values for capturing the forum post accuracy/correctness and student discursive alignment, quantitative scores for each PIG text were obtained from the instructor and correlated with the similarity values. To analyze the reliability of the similarity values, a correlation analysis was conducted between the similarity values computed from each expert benchmark.

#### Analysis of student-expert discursive alignment during synchronous collaboration

To study the development of each PIG's collaboration quality, similarity values were computed between the expert benchmark and a 100-word sliding-window of the transcripts. A sliding-window is a segment of certain number of words (100 in this case) in which the segment, or window, moves forward 1 word at a time. For instance, suppose you want to create a 3-word sliding-window for the text: "There can be voltage without a current", thus, the first three windows would be: ["There", "can", "be"], ["can", "be", "voltage"], ["be", "voltage", "without"]. A 100-word window includes enough information as to the dialogue that is taking place during that window of collaboration, and it is short enough as to produce a smooth measure of the ongoing flow of the conversation over time. The values computed per sliding window are then plotted on a line-over-time chart, which shows changes in student-expert discursive alignment over the span of the inquiry. The goal is to produce a visualization to show critical moments during the collaboration where students are engaging in disciplinary discourse. After identifying these moments, one can compare the prevalence of such moments across groups, or go back to the video data and conduct more in-depth qualitative analyses of these identified collaboration moments.

We also computed similarity values between each sliding window and each student's final texts. Although this is not possible to do in real-time while collaboration develops, because students are yet to produce the final text, it can be done as a post facto analysis. The goal is to provide a vantage into relevant moments during the collaboration that helped students orient their writing for their final texts. We hypothesize that sometimes expert's and final texts' similarity values would align, indicating that students were able to capture these productive conversations in their final texts. But we also anticipate that sometimes students can have some

productive conversations and yet fail to capture these reflections in their final texts. When this happens, a lack of alignment would be observed between the two similarity values trajectories.

## Results

### Effect of collaboration on individual learning

Results show that, according to our discursive alignment measure, collaboration had a significant effect in helping students align their discourse toward more disciplinary ways. In comparing the mean groups between the pre- ( $M_{pre} = 0.295$ ,  $SD_{pre} = 0.176$ ) and post-collaboration ( $M_{post} = 0.434$ ,  $SD_{post} = 0.133$ ) texts, there is a statistically significant increase of .139 points in the similarity values,  $t(17) = 2.59$ ,  $p = .019$ ,  $d = 0.61$ . This means that the forum posts became more similar to the way an expert would answer the question after students had an opportunity to interact in the collaborative forum and the videoconference activity.

### Validity and reliability of similarity values

Two pieces of information provide validity evidence that our measure of disciplinary discourse, represented by the similarity values, actually attest an improvement in students' conceptual understanding. First, there is a positive moderate-to-large association between the similarity values and the teacher-assigned forum post scores,  $r = .58$ ,  $p = .011$ , 95% CI [0.157, 0.824]. Second, the teacher-assigned scores also show a significant increase of 1.66 points from pre- to post collaboration ( $M_{pre} = 1.167$ ,  $M_{post} = 2.833$ ),  $t(17) = 3.58$ ,  $p = .002$ ,  $d = 0.84$ . This means that both teacher-assigned scores and similarity values covariate in the same direction and equivalent magnitudes.

We also found that this similarity value approach seems to be reliable at finding students' discursive development. Again, there are two pieces of reliability evidence. First, there is a positive moderate-to-large association between the two ratings provided by the two expert benchmarks,  $r = .54$ ,  $p = .016$ , 95% CI [0.117, 0.799]. Second, the second expert benchmark also shows a significant increase in students' similarity values,  $t(17) = 8.04$ ,  $p = .078$ ,  $d = 5.68$ . These pieces of evidence indicate that results are very similar regardless of which expert benchmark is used, and also imply that similarity values can serve as proxies for conceptual understanding because of their association with teacher-assigned scores.

### Time-sensitive analysis of collaboration quality

In this section, we explore the affordances of the discursive-alignment-over-time visualization chart. Figure 3 shows the chart for the group called "Koalas," which is a high-achieving group, according to the teacher-assigned scores ( $M = 4$ ) and similarity values ( $M = 0.63$ ). The chart displays two visible lines over time, though there are four lines in total. The solid bar represents the similarity distance to the expert benchmark, whereas the dotted line represents the distance to the final text, at every minute of the conversation. It is apparent that all students posted the same final text, because all the student lines overlay. The graph shows that this was a relatively long conversation of approximately 24 minutes (compared to other groups' videoconferences). From a quick read at the ebbs and flows of the lines, it is apparent that the discussion was really on-target around minutes 4, 6, 13 and 17, and that there was a drop in on-target talk between minutes 7 and 10. The slow decline at the end of the chart shows that productivity slowly went down after 17 minutes into the activity. The similarity values against both expert and final post benchmarks overlap for the most part, implying that students' final texts' ideas came from moments aligned with the expert's discourse.

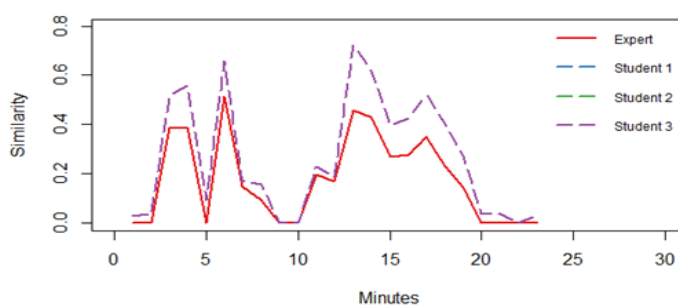


Figure 3. Koalas group collaborative discursive alignment with expert and final text over time.

As a validity check, we took a look at various points in time of the conversation to see if high similarity values represent good disciplinary discourse and low similarity values far off disciplinary discourse. For instance,

during the first two minutes of the activity the similarity values were very low. As Excerpt 1 shows (see Table 4), students were settling in and thus their talk was not about circuitry, instead, their conversation revolved around technical issues with the videoconferencing software. On the other hand, one of the most productive moments in the conversation seemed to have occurred between minutes 11 and 14, where there is a noticeable spike in the similarity values. As Excerpt 2 shows (see Table 5), at this point in the conversation there is an interesting exchange of ideas around the question of whether there can be voltage without current and how switches affect current flow. These two excerpts provide good evidence that similarity values extracted from the group's dialogue can provide a valid approach for measuring how discursive alignment develops throughout a collaborative activity. Then again, this measure only captures discursive alignment with an expert benchmark and nothing more, this is why other important aspects of collaborative learning do not seem reflected by the chart of similarity values over time. For instance, Excerpt 3 (see Table 6) shows a very important moment between minutes 8 and 10. This point in time shows a sharp drop in discursive alignment, and yet the dialogue reflects an interesting collaborative exchange where students try to find common ground around their interpretations of what the question is asking. We believe that, however, in the future, we might explore ways to capture other relevant aspects of the collaborative activity by systematically examining and developing distinct collaboration benchmarks.

Table 4: Excerpt 1: No disciplinary talk

[00:00:04]	<b>Student 1:</b>	So, I'm just going to connect the rest of these real quick.
[00:00:18]	<b>Student 2:</b>	Wait, this share screen's kind of weird.
[00:00:21]	<b>Student 1:</b>	Is it?
[00:00:25]	<b>Student 2:</b>	I can't go to like my actual...
[00:00:29]	<b>Student 3:</b>	Go up to options and say exit full screen.
[00:00:34]	<b>Student 2:</b>	Okay, thanks.

Table 5: Excerpt 2: High disciplinary talk

[00:11:59]	<b>Student 2:</b>	But did we ever say specifically what the voltage would be without current?
[00:12:02]	<b>Student 1:</b>	Exactly, we had never been in a situation to apply that until right now.
[00:12:16]	<b>Student 2:</b>	I think we have to kind of figure out when there's no current and when you're also dealing with a switch, how does that affect the voltage. Rather than whether there is voltage or not, it was more how it was affected.
[00:12:35]	<b>Student 3:</b>	I thought that this question was like asking like we had to find total current first before we could find any of the voltage for the test points. I thought that was the question.
[00:12:58]	<b>Student 2:</b>	There is no current, because it's open. I'm pretty sure they're all talking about the open one.

Table 6: Excerpt 3: No disciplinary talk, yet good collaborative talk not captured by our measure

[00:08:24]	<b>Student 1:</b>	I don't really know what else I'd put for question four. I mean, our first answer was right.
[00:08:36]	<b>Student 3:</b>	It was saying if you have to go back and find a new answer, how would you do that?



[00:08:47]	Student 1:	Yes, but are you saying that's what we did?
[00:08:51]	Student 3:	No, it's saying like even though we did prove it on our first group answer, [...] So, it's saying if we were wrong, what would we do in that situation.
[00:09:19]	Student 1:	I'm actually interpreting that a little differently. I'm reading it as if you did have to go back, then you need to answer the second part of the question.
[00:09:30]	Student 3:	I see what you're saying. It's saying, or did you have to go back and do this?

Finally, it seems important to show that our measures of discursive alignment can show interesting individual dynamics within collaboration groups. For instance, Figure 4 shows the chart over time for the group called Otters, which is also a high-achieving group (Teacher-assigned score = 4.67, similarity values = 0.45). This was a relatively short conversation, 8 minutes approximately, where the similarity values against expert benchmark show that the alignment was higher during the second half of the collaborative activity than the first half. However, students' final forum posts reflect different trajectories; student's 1 similarity values reflect a higher similarity during minute 3, whereas students' 2 and 3 trajectories show higher values during the second half of the activity. This can be interpreted in the following way: student's 1 final forum post seemed to have come from the ideas discussed during minute 3, whereas for students 2 and 3, their ideas came from the second half of the discussion. Although all three students got a high grade (5, 4, and 5, respectively) in the teacher assigned score, similarity value for student 1 (0.37) is lower than for students 2 and 3 (0.46 and 0.51, respectively). We think that student 1 deserves further analysis in order to understand why she is getting good grades but her discourse is not yet aligned with that of the expert.

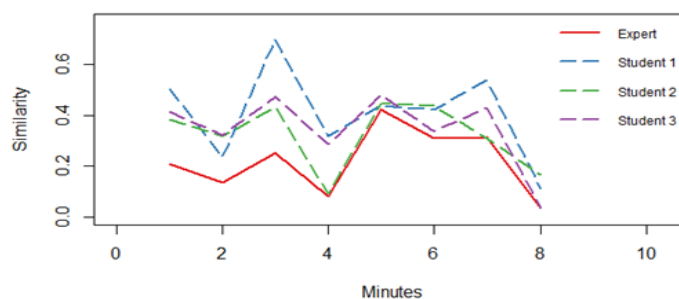


Figure 4. Otters group collaborative discursive alignment with expert and final post over time.

## Conclusions and implications

Over the course of PIG2, students made significantly closer alignments to expert discourses when articulating the practical application of engineering principles. Although similarity values serve only as an approximation of full disciplinary discourses, they appear to be a useful measure of the positive effect of collaboration on learning. Moreover, from this analysis it is clear that similarity values can be used to provide rapid or real-time feedback to students or groups of students as they work toward expert discourse. Over the course of the videoconferencing collaborative activity, students' dialogue exhibited several shifts in its alignment to expert discourse. Similarity values prove to be a valid approach to display a simple visualization of the changes in the content of students' conversations. These shifts reveal some critical points where the discussions were more or less aligned with disciplinary discourse instructors would expect to see in these sorts of activities. Furthermore, we showed how similarity values can evidence whether students are able to document, within their final texts, the productive ideas that emerge during their discussion.

This initial implementation of LA in a CACL environment serves to demonstrate two potential approaches for delivering rapid assessment prototypes to students and instructors. By exploring student-expert similarity values in both textual and video data, we focus on how students negotiate and construct discourses pre-, during-, and post collaboration, align to expert answers, and map to learning trajectories. Moreover, these promising results indicate where we can continue to refine our learning analytical approach to deliver finer-grained feedback. For example, the inclusion of other parts of speech such as adverbs in document-term-frequency matrices deserves a systematic exploration. Additionally, an exploration of other kinds of possible benchmarks with the purpose of capturing other valuable aspects within collaborative learning warrants further

investigation. Finally, we are currently exploring what we call misconception benchmarks, which would help provide distance measures to possible misconceptions students may be falling prey to.

All in all, we believe that this preliminary case serves as a fruitful proof-of-concept that text-mining in online, inquiry-based learning can be used to provide rapid feedback to students and instructors during and after collaboration. This has the potential to foster learning and engagement in complex or semi-structured learning environments where students construct, negotiate, and implement disciplinary concepts. While powerful learning spaces, making timely sense of broad fields of disciplinary discourse may be inherently difficult in online CSCL environments. Text mining and the production of similarity values may provide a snapshot of student-expert alignment pre-, during-, and post-collaboration. However, an important question remains: How might stakeholders differently use similarity values as rapid prototypes of feedback?

Here, similarity values were used to assess student learning and collaboration. A vital next step in research is to investigate how similarity values are taken up and used by students, instructors, and researchers. For example, how do students reorient their discourse when provided with similarity values or qualitative representations of similarity values (e.g., word clouds)? How can instructors make use of similarity values to reorient delivery of instruction and understand the performance of collaborative student groups? Based on this work, similarity values can provide significant utility to instructors in cases where multiple groups of students are collaborating synchronously online (or in-person) or when making sense of large amounts of textual information. Finally, researchers can use the similarity values as a lens to investigate the collaboration, pinpointing moments of interaction worthy of deeper investigation. Practically, this may help organize trends in large amounts of data, providing the resources and convergent support for more nuanced analyses of discourse (e.g., conversation analysis or discursive psychology) in CSCL studies.

## References

- Berland, M., Davis, D., & Smith, C. P. (2015). AMOEBA: Designing for collaboration in computer science classrooms through live learning analytics. *International Journal of Computer-Supported Collaborative Learning, 10*(4), 425-447.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. Sebastopol, CA, USA: O'Reilly Media, Inc.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42. doi:10.3102/0013189x018001032
- De Liddo, A., Shum, S. B., Quinto, I., Bachler, M., & Cannavacciuolo, L. (2011). *Discourse-centric learning analytics*. Paper presented at the Proceedings of the 1st International Conference on Learning Analytics and Knowledge.
- Jeong, H., & Hmelo-Silver, C. E. (2016). Seven Affordances of Computer-Supported Collaborative Learning: How to Support Collaborative Learning? How Can Technologies Help?. *Educational Psychologist, 51*(2), 247-265.
- Kopainsky, B., Pirnay-Dummer, P., & Alessi, S. M. (2012). Automated assessment of learners' understanding in complex dynamic systems. *System Dynamics Review, 28*(2), 131-156.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York, NY: Cambridge University Press.
- Leydesdorff, L. (2005). Similarity Measures, Author Cocitation Analysis, and Information Theory. *JASIST, 56*(7), 769-772.
- Ludvigsen, S. (2016). CSCL towards the future: The second decade of ijCSCL. *International Journal of Computer-Supported Collaborative Learning, 1*(11), 1-7.
- Mercer, N. (2008). Talk and the Development of Reasoning and Understanding. *Human Development, 51*(1), 90-100. doi:10.1159/000113158
- Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2015). Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology, 46*(2), 330-343.
- Vatrapu, R., Teplovs, C., Fujita, N., & Bull, S. (2011). *Towards visual analytics for teachers' dynamic diagnostic pedagogical decision-making*. Paper presented at the Proceedings of the 1st International Conference on Learning Analytics and Knowledge.
- Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior, 47*, 168-181.

# Beyond Demographic Boxes: Relationships Between Students' Cultural Orientations and Collaborative Communication

Nishan Perera, Kwantlen Polytechnic University, nishan.perera@kpu.ca  
Alyssa Friend Wise, New York University, alyssa.wise@nyu.edu

**Abstract:** This study investigated relationships between students' cultural orientations and the ways in which they communicated with their peers in collaborative online discussions. 211 undergraduate business students from diverse backgrounds completed questionnaires directly assessing their cultural orientations along four dimensions (individualism, collectivism, power-distance and cultural-context). Scales were input into mixed multi-level models to predict 12 aspects of the way students communicated in their course-based collaborative discussions of business cases. Results based on a sample of 1565 posts showed that students with weaker context-based orientations posted messages that showed greater levels of reasoning, hard evidence use, autonomous tone and linear argument structures. In addition, the local discussion group context moderated the relationship between students' degree of collectivistic orientation and how they referred to and agreed with each other, as well as their expressions of social presence. Findings highlight the group/individual interplay in understanding relationships between cultural orientations and collaborative communication.

## Introduction

This paper addresses a critical deficit in the current CSCL literature in attending to the diverse racial, ethnic and cultural background of today's students. While the field abounds with high quality studies of group processes, there is markedly less attention paid to the non-cognitive characteristics of the individuals who come together to engage in these interactions. Certainly, individuals are considered in the cognitive and behavioural sense. For example, scripts are designed to compose groups with complementary knowledge (Chan, 2009) or differences in opinion (Jermann & Dillenbourg, 2003) and analytics have been created to track imbalances in participation between students within a group (Janssen et al., 2007). However, attention to the personal factors that might influence *why* a student tends to participate less, hold a particular opinion or agree with others has been limited (c.f. Prinsen et al., 2007; Popov et al. 2014). Moreover, that work which has attempted to consider student diversity has often treated culture monolithically, using demographic proxies such as citizenship, race or gender (e.g. Prinsen et al., 2009) which assume a degree of uniformity across a group, thus precluding the examination of individual variations or the presence of multiple cultural influences within an individual. The myth of monolithic cultural groups is both seductive and pernicious in its oversimplification (Vatrapu & Suthers, 2010). Importantly, it simply does not match the reality of today's classrooms. For example, in 2011 the university participation rate within second-generation immigrants in Canada (which make up almost 20% of the population) was 53% (McMullen, 2011). Similarly, as of 2011 39% of the entire Canadian population came from ethnically diverse households (Dobson, Maheux & Chui, 2011). In fitting with the CSCL 2017 theme of improving equity and access, this paper takes an important step towards culturally responsive pedagogies by both offering a flexible and nuanced way to assess critical aspects of students' "culture" and by examining the relationship between these characteristics and ways students communicate in collaborative activities.

## A need to understand the (whole) student taking part in collaboration

Studies of CSCL can coarsely be divided in two based on whether an individual or collective epistemology underlies the work. Studies taking a collective epistemology tend to consider language as constructing social reality and examine the joint construction of knowledge through processes of intersubjective meaning-making (Suthers, 2006). There is a long tradition of such work in the CSCL community including the study of knowledge building (Scardamalia & Bereiter, 2006), group cognition (Stahl, 2006) and many other related constructs. Studies taking an individual epistemology draw on a different set of assumptions, taking language to represent an individual's inner cognition and focusing on the different contributions made by learners (Wise & Paulus, 2016), but evaluations of these contributions are still often made in aggregate. For example, a particular script may be found to elicit more argumentative moves on average from a collaborating group (Scheuer et al., 2013). These two conceptual categories align roughly with the four methodological clusters of CSCL work identified by Jeong and Hemlo-Silver (2014): socio-cultural classroom / eclectic descriptive which tended to describe group processes based on qualitative data and constructivist classroom / eclectic experimental which generally used inferential statistics based on individuals aggregated quantitative data. What neither group of studies does well is consider the personal characteristics of the individuals involved and their impact on how

they take part in the collaboration. This is important because the students that make up a collaborative group are, by no means, all alike. They bring in different perspectives influenced by their cultural background, family upbringing, individual skills, motivation levels, life experiences etc. that influence the ways they communicate in online learning contexts (Chase et al. 2002). These differences are further compounded by the fact that online learning designs are often implicitly based on values held by the English-speaking Western world which may not be readily recognized or understood by learners from other backgrounds (Hannon & D'Netto, 2007). Yet such characteristics have rarely been discussed in the CSCL literature (c.f. Vatrappu & Suthers, 2010, Sugimoto & Suthers, 2002) and even less frequently addressed empirically (c.f. Zhu, 2013). The extremely limited work that has been done suggests not only do these factors matter for how students engage in collaboration but also the learning that results from it (Popov et al., 2014). If we are to seriously take up the challenge of developing culturally responsive pedagogies to increase equity and access in CSCL, we must first understand the cultural characteristics that are relevant to consider and how they affect collaborative dynamics.

### Limitations in the use of demographic labels to index culture

While the personal characteristics of students has received limited attention within CSCL, the larger online learning literature is replete with studies attempting to link online communication behaviours with age, gender, ethnicity and citizenship differences of students (e.g. Rovai & Baker, 2005; Warden, Chen & Caskey, 2005). Such use of demographic labels as proxies is a widespread practice. For example, Schaffer and Riordan (2003) reported that 79% of studies in the organizational research literature from 1995 to 2001 examining culture used nationality as a proxy. While broad demographic labels may provide some useful insights into differences in behaviour overall, they are unlikely to be useful in predicting or explaining the activity of particular students. This is because conclusions from such studies necessarily make the assumption that all students that belong to a given "culture" (i.e. nationality, ethnicity, race) to be monolithic, i.e. identical to all those reported by the category at large. The limits of such dichotomous thinking with respect to learner activity has been shown time and time again, most recently in the debunking of the digital natives and immigrants myth (Bennett & Maton, 2010). The limited work in CSCL looking at cultural orientations (e.g. Popov et al., 2014) has also used demographics as a proxy. A better alternative to demographic labels is to directly identify and report results based on individual characteristics that are hypothesized to directly influence collaborative behaviours. Cultural orientations are one good example. Culture has been defined in many ways, but is generally taken to refer as a set of attitudes, values, beliefs and behaviours shared by a group of people, but different for each individual, communicated from one generation to the next (Matsumoto, 1996). These specific values, beliefs and attitudes are then the orientations towards interacting with the world that are influenced by being a member of the cultural group. For example, those who come from western cultures are thought, on the whole, to hold stronger individualistic orientations than those who come from eastern cultures (Warden et al., 2005). Of course, a particular person's individualistic orientation will be influenced not only by their ethnic origin but also by other factors such as their country of residence, familial dynamics and life experiences. By conceptualizing and measuring a collection of orientations for each individual we allow for the multiplicity of proclivities which can exist simultaneously within a person and the plurality of influences (multiple cultures, languages, geographies) on the orientations each one holds. Cultural orientations are also a more useful way to look at culture because they are more causally proximate to collaborative behaviours than demographic labels (they more precisely describe the characteristic of an individual thought to lead to a behaviour) and because they allow for more nuanced measurement (Vatrappu & Suthers, 2010; Perera, 2016).

### Conceptualizing students' cultural orientations

Of the many cultural frameworks in existence, that introduced by Hofstede in 1984 (Hofstede et al., 2010) is one of the most frequently cited. Many subsequent frameworks were influenced by this seminal model (e.g. Shulruf et al., 2011; Singelis et al, 1995). The original framework included four major cultural orientations to characterize differences in values, beliefs, norms and behaviours across countries: individualism/collectivism, power distance, masculinity/femininity and uncertainty avoidance (Hofstede et al., 2010). Of these four, individualism / collectivism and power distance have special relevance for CSCL (Zhu, 2013). Individualism / collectivism refers to an individual's tendency to situate themselves as part of, and identify with, a group and their orientation towards self-expression focusing on the projection of a unique identity for themselves and others (Hofstede et al., 2010). While individualistic and collectivistic orientations were originally conceptualized as a bipolar scale, the co-linearity of these scales has been questioned as over-essentializing the multiplicities of today's learners. Empirical evidence suggests that they are more usefully be considered to be orthogonal (Perera, 2016) such that an individual can hold a strong individualistic and a strong collectivistic orientation simultaneously. Power-distance refers to the degree to which an individual is willing to accept or

reject differences of equality and authority within a group (Hofstede et al., 2010): those with a higher power-distance orientation are more likely to look to a teacher for answers while those with a lower power-distance orientation may be more likely to value the ideas of their peers (Zhu, 2013). Another cultural orientation described in the literature with particular relevance for CSCL is the degree to which someone is context-based: this refers to the degree to which an individual uses context to create, communicate and interpret meaning (Hall & Hall, 1990). Those with lower context-based orientations tend to be more explicit in their communication, explaining their meaning precisely with the words they use in their messages; those with higher context-based orientations expect listeners to read between lines and rely on surrounding talk and the situation in which it occurs to understand the message fully. Both scales have been empirically verified as bipolar (Perera, 2016).

## The current study

The overall purpose of this study was to investigate relationships between students' cultural orientations and how they communicate with their peers in collaborative online discussions. Four relevant dimensions of cultural orientations were outlined above. The following sections address the issues of what aspects of students' collaborative communication merit examination and how to account for collaborators' lack of independence.

### Examining individuals' communication in collaborative contexts

Two common dimensions of communicative acts frequently referenced in literature: the extent to which a student attends to the substantive content of the discussion task and the extent to which their comments are attended to by others (Wise et al., 2014). Attention to the task encompasses issues related to being on/off topic, features of evidence use and reasoning, and the form in which ideas are argued while attention to others addresses issues of responsiveness, (dis)agreement and message tone (Hsiao, 2012). Previous studies (Salleh, 2005; Vatrapu & Suthers, 2010) have shown how various factors along each of these dimensions can be influenced by student's cultural characteristics. For example, the amount of evidence and first person pronouns (autonomous tone) used in messages were found to be influenced by low context-based orientation of students.

### Using multilevel models to account for nested group effects

Students' comments in collaborative activities are by definition not independent of each other (on the contrary if they are then we need to reconsider whether the activity can truly be called collaborative). This raises both conceptual and statistical issues for CSCL researchers when attempting to examine the activities of individuals distributed across multiple groups. Specifically, phenomena such as common fate and reciprocal influence make it likely that individuals participating in the same small group will act more similarly to each other than to those in other groups by virtue of their joint participation (Cress, 2008). *Conceptually* this implies that the local group context may moderate how particular cultural orientations become expressed in the comments students make (e.g. someone with a strong individualistic orientation may tone down (or amp up) the degree to which they project their unique identity if they find themselves in a group where no one else is doing so). This adds a layer of complexity to understanding the ways in which individuals' background affect how they communicate in collaborative contexts as both direct and indirect effects are possible. *Statistically*, when the assumption of independence is violated, it can lead to distortion in the final model (Cress, 2008) and the use of multilevel/hierarchical linear models (HLM) are required. These models are capable of handling observations that are not independent, as they take into account clustering effects of data by one or more grouping factors (Cress, 2008; Garson, 2013).

## Research questions

1. To what extent are how students attend to (a) others and (b) the task in their online discussion posts predicted by their levels of (i) individualistic, (ii) collectivistic, (iii) power distance and (iv) context-based orientations?
2. Are any of these relationships moderated by the local group in which students participate in the discussion?

## Methods

### Study context

There were 221 participants from 280 students enrolled in one of two introductory-level marketing courses taught at a small Canadian university. The elements of the courses pertaining to the study were identical. Data was collected from eight course sections taught by the same instructor during summer and fall 2014.

### Participants

58% of participants were female. While the majority of students (70%) were of standard university-age (17-22), a sizable proportion (30%) were mature learners (23+). Two-thirds of students were Canadian citizens; however, within this category there was great ethnic diversity with 41% of Caucasian descent, 27% of East Asian descent, 21% of South Asian descent and the remaining 11% from diverse backgrounds. Among the third of the class who were not Canadian citizens, the largest groups of ethnic origin were East Asian (57%), South Asian (22%) and Caucasian (5%), with the remaining 16% coming from diverse backgrounds.

### Cultural orientations questionnaires

Students completed previously-developed questionnaires about their levels of individualistic and collectivistic orientation (Shulruf et al., 2011) and power-distance and context-based orientations (Richardson & Smith, 2007). All instruments asked students to respond to how often they thought or acted in a certain way (e.g. I think of myself as competitive) on a 7-point frequency scale (1=never 7=always). The internal consistency of all scales as indexed by Cronbach's alpha was satisfactory: individualism (15 items;  $\alpha = .79$ ); collectivism (11 items,  $\alpha = .68$ ); power-distance (11 items;  $\alpha = .71$ ); context-based (17 items,  $\alpha = .74$ ).

### Collaborative case discussion activity

As part of their coursework, students were required to take part in three collaborative case-based online discussions worth 21% of their total course grade. Collaborative discussions took place in 24 small groups of 6-12 students across the eight sections. Students were assigned to groups at random. Due to non-participation and attrition, final group sizes were not equal. Each group included both males and females and students from different ethnic backgrounds. The discussion lasted for 10 days spaced across the semester, and were conducted in an installation of the asynchronous open-source discussion tool *Phorum*. For each activity, students were given an open-ended authentic business case (e.g. a local hotel introduced a fake brand name to sell pizza made internally within the hotel's restaurant) to address as a group. Students were asked to first identify the important issues pertaining to the case (in this situation relating to ethics). They were then asked to brainstorm possible ideas and finally come to consensus about how the characters in the case should proceed (in this situation how the hotel could increase their restaurant profitability without violating ethical principles).

### Communication data collection and coding

Table 1: Variables Indexing How Students Attended to Others in their Posts

Variable	Definition (student level)	Coding Operationalization (post level)
Ref to Others	% of posts making references to others.	Binary code for presence/absence of reference
Some Disagree	% of posts with full/partial disagreement	Categorical variable with neutrality/full agreement/partial disagreement/full disagreement
Full Agreement	% of posts with full agreement	
Social Presence	% of posts with social presence.	Binary code for social presence/absence
Autonomous Tone	Avg. level of first person singular pronouns use by post.	# of uses of pronouns (e.g. I, me, my, mine).
Connected Tone	Avg. level of first person plural pronouns use by post.	# of uses of pronouns (e.g. we, us, our, ours).

Table 2: Variables Indexing How Students Attended to the Task in their Posts

Variable	Definition (student level)	Coding Operationalization (post level)
Contextual Structure	% of posts made with contextual structure	Categorical variable for post structure as contextual (thoughts expressed circuitously) / linear / other
Reasoning	Avg. reasons level in a post	# of reasons coded into bins (0, 1, 2/3, 4/5, 6/7, 8+)
Evidence Refer	% of posts made referring to evidence	Categorical variable for no evidence / evidence mentioned / evidence explained
Evidence Apply	% of posts made applying evidence	
Hard Evidence	Avg. level hard evidence use by post	# of uses of independent evidence (e.g. citations)
Soft Evidence	Avg. level soft evidence use by post	# of uses of personal evidence (e.g. anecdotes)

221 consenting students generated a total of 4694 posts across the three case discussions. To keep post coding manageable, the second discussion was targeted. Tasks asked students to take an individual stance then negotiate with others to consensus, using evidence to support their claims. This invited students to agree / disagree with others, work collectively or individually to generate a shared solution, express (power) relations to others in the group and be more or less explicit in how they supported their claims with evidence. 211 of the 221 consented students took part in the second discussion, contributing a total of 1565 posts. All post data including author, title, content, and time/date stamp was extracted via MySQL query. All posts were coded by two researchers for twelve aspects of the way the student comments attended to others and to the task; codes were then rolled-up into student-level variables (see Tables 1 and 2). The post was taken as unit of analysis as it was the medium by which students communicated with the group and involved no challenges for segmentation (Schellens & Valcke, 2006; Wise & Paulus, 2016). 150 messages generated from the first online discussion were used as training data to refine the coding scheme and prepare the two coders. Inter-rater reliability as indexed by Krippendorff's alpha was high ( $\alpha > 0.80$ ) across all elements of the coding scheme.

### Statistical analysis

Multilevel models were set up and run with the student as the lowest-level unit. Examination of the Intra-class Correlation Coefficient (ICC) showed a significant and a substantial effect of the local discussion group for 10 of 12 variables, thus student-group was included as a second level in the models. ICC revealed little to no effect for course-section. HML7 was used to run Random Intercept and Random Slope (RIRS) models for each of the communication variables based on predicted relationships with cultural orientations. Expected relationships are indicated in Table 3; space precludes inclusion of the rationale but see Perera (2016). Log likelihood ratios (deviance statistic) were compared with those for the null models (baselined without predictors; Garson, 2013) to ensure predictor addition improved overall model fit. RIRS models produce estimates of fixed and random effects. Fixed effects include an intercept and a series of slopes estimated across the sample. This confirmatory portion of the model indicates presence or absence of predicted global relationship between cultural orientations and communication variables. Random effects also include an intercept and a series of slopes; however, in this exploratory portion of the model, a significant slope indicates an interaction between the cultural orientation variable and the local context of the discussion group in affecting the communication variable.

### Results

Sixteen statistical assumptions required to run RIRS models (Garson, 2013) were verified. Six cases were dropped due to uni-/multi-variate outliers; final sample was 205. Apply Evidence outcome variable was dropped due to poor model fit. Results of the remaining 11 models are reported in Table 3. There was a negative fixed effect for Context-Based orientation on students' level of Autonomous Message Tone, level of Reasoning, use of Hard Evidence and Linear Argument Structures. No fixed effects of Individualistic, Collectivistic or Power Distance orientations were found. For random effects, the intercepts for Some Disagreement, Social Presence, Autonomous Tone, Refer to Evidence, and use of Hard and Soft Evidence were significant. This indicates some groups engaged in high levels of each of these communicative acts than others during the discussion. There were also significant random slope effects indicating interaction between levels of Collectivistic cultural orientation and students' local discussion groups on Reference to Others, Some Disagreement, Full Agreement and Social Presence. Three of the four accounted for large portions of the variance explained. In addition, significant interaction effects were found between the local group context and Individualistic, Context-Based orientations. No significant interaction effects were found for Power Distance orientation.

Table 3. Full Model Results

Speaking Variables	Fixed (Global) Effects Slopes				Random (Group) Effects Intercept & Slopes				
	IND	COL	POWR	CONTX	Intercept	IND	COL	POWR	CONTX
Refer to Others	0.041 (-)	-0.028 (+)	0.033 (-)		0.008	0.059 3%	<b>0.172***</b> 27%	0.084 6%	
Some Disagree	0.032 (+)	-0.005 (-)	0.029 (-)		<b>0.001*</b>	<b>0.003**</b> 7%	<b>0.003*</b> 7%	0.003 7%	
Full Agree		0.011 (+)			0.002		<b>0.011*</b> 20%		
Social Presence		-0.027 (+)		0.037 (+)	<b>0.017***</b>		<b>0.019**</b> 18%		<b>0.011*</b> 10%
Autono. Tone	0.063 (+)			<b>0.327**</b> (-)	<b>0.485***</b>	0.052 3%			0.151 1%

Connect. Tone		0.065 (+)			0.035		0.040 7%		
Context Structure	-0.020 (-)			-0.218*** (+)	0.003	0.002 4%			0.011 20%
Reason	0.034 (+)	0.026 (-)	-0.131 (-)	0.212* (-)	0.116	0.081 6%	0.212 17%	0.080 6%	0.152 12%
Refer to Evidence		0.013 (+)			0.007*		0.000 0%		
Hard Evidence	-0.038 (+)		-0.140 (-)	0.414** (-)	0.431***	0.018 1%		0.046 3%	0.111 7%
Soft Evidence		-0.026 (+)		0.004 (+)	0.006**		0.008 13%		0.011 20%

\*p<.05 \*\*p<.01 \*\*\*p<.001; (+) / (-) = direction of predicted relationships; % = percent of total random effect variance explained

## Discussion

The results of the study confirmed certain cultural orientations to be useful predictors of students' collaborative communication. Specifically, students' degree of context-based cultural orientation was useful in explaining multiple aspects of how they attended to the task in terms of use of reasoning, hard evidence and systematic message structure, as well as the tone with which these were communicated. These findings align with prior work showing that people with low context-based orientations tend to use direct, precise and logical expression of ideas in their communications (Salleh, 2005) and expand the relationship to include the use of reasoning and evidence. This is important in online collaborative learning as providing reasons and evidence are elements of strong argumentation (Chinn & Osborne, 2010) which can encourage others to think deeply about the ideas presented and potentially change their own perspective in response. It may also draw out elaborated responses which offer further learning opportunities. The question of how to address the issue of students who have high context-based cultural orientations and are less likely to use reasoning and hard evidence in their collaborative communications is an open one. From a rationalist perspective, these students could be identified and encouraged to include these elements in their messages. This would address the issue of implicit western / anglo values embedded in online learning designs which may not be apparent to learners from other backgrounds (Hannon & D'Netto, 2007). However, such an approach also potentially raises the spectre of cultural imperialism. A more nuanced way to address the issue could draw on culturally conscious approaches that explicitly recognize value multiple style of communication and encourage students to code-switch between them appropriately (Delpit, 1988). In addition, instructors might consider ways to structure collaborative activities that present a need for or benefit to contextual styles of communication. The other major finding of this study was the presence of multiple effects of cultural orientations on students' collaborative communication that were moderated by the local group context. The most notable of these was for collectivistic cultural orientations which showed interaction effects on the degree of reference to others, the level of agreement /disagreement and the expression of social presence. This finding is noteworthy for two reasons. First, this finding provides empirical evidence to suggest that cultural orientations do not always affect behaviour directly, but in the context of the social situation in which students are placed in (in this case the group). It may also be the case that students who are multi-national, multi-cultural and multi-lingual themselves hold contradictory orientations which they bring to bear differently in different situations. Second, while collectivism is a commonly studied culture orientation (Oyserman et al., 2002), it was not a significant global predictor of any of the collaborative communication characteristics studied. It may be that this is because it is particularly susceptible to moderation by the local group context. This makes sense logically as the degree to which a student who is oriented towards identifying with the group (Hofstede et al., 2010) chooses to express social presence, refer to others and seek harmony through their messages may depend on characteristics of the group.

## Implications for research and practice

The results of this study have two important implications for research. First, past research (including the current study) focused on commonly studied cultural orientations due to widespread use and the availability of measurement instruments. The popularity of Hofstede's (1984) cultural model and its use across multiple settings is indicative of this. However, the only significant global effects were found for the less commonly studied context-based cultural orientation (Hall & Hall, 1990). This suggests that those orientations most often cited in the general literature on culture may not be best suited to differentiate and predict behaviours that take place in CSCL situations. In corollary, there may be other orientations (similar to context-based) more relevant for measuring differences in culture for CSCL; e.g. Parrish and Linder-VanBerschoot (2010). Second, the results of this study reaffirm the importance of using multi-level models to account for the non-independence of



collaborating learners (Cress, 2008). Without the random effects part of the model, it would have (incorrectly) appeared that collectivistic cultural orientations had no effects on students' collaborative communication behaviours. Equally importantly, the finding of such interaction effects opens a new important area for inquiry into the characteristics of collaborating groups that moderate the way individual orientations are expressed in their communications.

The findings of this study also have two important implications for practice. The significant global effects highlight some of the ways in which cultural orientations directly predict students' online discussion behaviours. Knowing these characteristics of students by having them complete self-assessments prior to engaging in collaborative activities can be beneficial. Further, instructors can support students with particular orientations or integrate elements of these into activities in the ways described earlier. Another important implication for practice is heightened attention to the manner in which students are grouped for collaborative activities. While group composition has been considered in CSCL in terms of cognitive factors (Jermann & Dillenbourg, 2003), it has not yet attended to personal characteristics such as cultural orientations. Importantly, the assessment of individual cultural orientation offers a more fine-grained basis for determining group composition than demographic factors. As future work investigates specific group characteristics that moderate the effects of particular cultural orientations, more nuanced group formation guidance can be generated.

## Conclusion

As we move into a complex, globalized and multicultural age, it is important to take students' cultural background and orientations into account as factors affecting computer supported collaborative learning. This study offered individually assessed cultural orientations as a more flexible and nuanced tool than previously used demographic labels. Results verified individual cultural orientations generally as a useful lens into students' online collaborative communication, with context-based and collectivism as specific orientations useful for prediction. The field needs to continue and build on this work by exploring a more diverse range of cultural orientations that may be relevant to collaborative learning contexts. In addition, results clearly documented the important role played by the local discussion group in moderating such effects and the need to better understand characteristics of the collaborating groups that can explain these interactions. This work thus adds to a growing chorus of calls for more CSCL work that connects levels of learning (Suthers et al., 2011) by examining both the group and the individual, and most critically the complex relationship between them.

## References

- Bennett, S., & Maton, K. (2010). Beyond the 'digital natives' debate: Towards a more nuanced understanding of students' technology experiences. *Journal of Computer Assisted Learning*, 26(5), 321-331.
- Chang, C. C. (2009). Using jigsaw collaborative learning strategy in online discussion to foster a project-based learning community on the web. *International Journal of Instructional Media*, 36(2), 221-234.
- Chase, M., Macfadyen, L., Reeder, K. and Roche, J. (2002). Intercultural challenges in networked learning: hard technologies meet soft skills. *First Monday* 7(8). [firstmonday.org/ojs/index.php/fm/article/view/975](http://firstmonday.org/ojs/index.php/fm/article/view/975)
- Chin, C., & Osborne, J. (2010). Supporting argumentation through students' questions: Case studies in science classrooms. *Journal of the Learning Sciences*, 19(2), 230-284.
- Cress, U. (2008). The need for considering multilevel analysis in CSCL research: An appeal for the use of more advanced statistical methods. *ijCSCL*, 3(1), 69-84.
- Delpit, L. (1988). The silenced dialogue: Power and pedagogy in educating other people's children. *Harvard Educational Review*, 58(3), 280-299.
- Dobson, J., Maheux, M. & Chui, T. (2011). Generation status: Canadian-born children of immigrants. Statistics Canada's Social and Aboriginal Statistics Division.
- Garson G. D. (2013). *Hierarchical linear modeling: Guide and applications*. Thousand Oaks, CA: Sage.
- Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education*, 49(4), 1037-1065.
- Jeong, H., Hmelo-Silver, C. E., & Yu, Y. (2014) An examination of CSCL methodological practices and the influence of theoretical frameworks 2005-2009. *ijCSCL* 9(3), 305-334.
- Jermann, P., & Dillenbourg, P. (2003). Elaborating new arguments through a CSCL script. In *Arguing to learn* (pp. 205-226). Springer: Netherlands.
- Hall, E., & Hall, M. (1990). *Understanding cultural differences: Germans, French and Americans*. Yarmouth, ME: Intercultural Press
- Hannon, J., & D'Netto, B. (2007). Cultural diversity online: Student engagement with learning technologies. *International Journal of Educational Management*, 21(5), 418-432.

- Hoftede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind: intercultural cooperation and its importance for survival*. New York: McGraw-Hill.
- Hsiao, Y. T. (2012). *The impact of task type on learners' argumentation, participation and collaboration in online discussions* (Masters Thesis).
- Matsumoto, D. (1996). *Culture and psychology*. Pacific Grove, CA: Brooks/Cole Publishing Company.
- McMullen, K. (2011). Postsecondary Education Participation among Underrepresented and Minority Groups. *Education Matters: Insights on Education, Learning and Training in Canada*, 8(4).
- Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, 128(1), 3–72
- Parrish, P., & Linder-VanBerschot, J. (2010). Cultural dimensions of learning: Addressing the challenges of multicultural instruction. *The International Review of Research in Open and Distance*, 11(2), 1-19.
- Perera (2016). *Students' cultural and personality factors as predictors of their asynchronous online discussion behaviours*. (Doctoral Thesis).
- Popov, V., et al. (2014). Perceptions and experiences of, and outcomes for, university students in culturally diversified dyads in a computer-supported collaborative learning environment. *CHB*, 32, 186-200.
- Prinsen, F., Volman, M. L., & Terwel, J. (2007). The influence of learner characteristics on degree and type of participation in a CSCL environment. *BJET*, 38(6), 1037-1055
- Prinsen, F., Volman, M. L. L., Terwel, J., & Van den Eeden, P. (2009). Effects on participation of an experimental CSCL-programme to support elaboration: Do all students benefit? *Computers & Education*, 52(1), 113-125.
- Richardson, R. M., & Smith, S. W. (2007). The influence of high/low-context culture and power distance on choice of communication media: Students' media choice to communicate with professors in Japan and America. *International Journal of Intercultural Relations*, 31(4), 479-501.
- Rovai, A. P., & Baker, J. D. (2005). Gender differences in online learning. *Quarterly Review of Distance Education*, 6(1), 31-44.
- Salleh, L. M. (2005). High/low context communication: The Malaysian Malay style. In Proceedings of the 2005 Association for Business Communication Annual Convention (pp. 1-11).
- Singelis, T. M., Triandis, H. C., Bhawuk, D., & Gelfand, M. J. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-Cultural Research*, 29(3), 240-275.
- Scardamalia, M. & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In *Cambridge handbook of the learning sciences* (pp. 97-116). New York: Cambridge Press.
- Schaffer, B.S., & Riordan, C.M. (2003). A review of cross-cultural methodologies for organizational research: A best practices approach. *Organizational Research Methods*, 6 (2), 169-215.
- Schellens, T., & Valcke, M. (2006). Fostering knowledge construction in university students through asynchronous discussion groups. *Computers & Education*, 46(4), 349-370.
- Scheuer, O., McLaren, B. M., Weinberger, A., & Niebuhr, S. (2013). Promoting critical, elaborative discussions through a collaboration script and argument maps. *Instructional Science*, 41(3), 2.
- Shulruf, B., et al., (2011). Measuring collectivism and individualism in the third millennium. *Social Behaviour and Personality: An International Journal*, 39(2), 173-188.
- Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge: MIT.
- Sugimoto, M., & Suthers, D. D. (2002). Multicultural issues in the design, evaluation and dissemination of CSCL systems. In *Proceedings of CSCL 2002* (pp. 677-678). Boulder, Colorado: ISLS.
- Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *ijCSCL*, 1(3), 315-337.
- Suthers, D., Teplovs, C., De Laat, M., Oshima, J & Zeini, S. (2011) Connecting levels of learning in networked communities. In *Proceedings of CSCL 2011* (pp 1200-1201). Hong Kong: ISLS.
- Vatrapu, R. K., & Suthers, D. D. (2010). Cultural influences in collaborative information sharing and organization. In *Proceedings of the 3rd ICIC* (pp. 161-170)., Denmark: ACM.
- Warden, C. A., Chen, J. F., & Caskey, D. (2005). Culturally related values and communication online: Chinese and Southeast Asian students in a Taiwan international MBA class. *Bus. Comm. Qu*, 68(2), 222-232.
- Wise, A. F., Hausknecht, S. N. & Zhao, Y. (2014). Attending to others' posts in asynchronous discussions: Learners' online "listening" and its relationship to speaking. *ijCSCL*, 9(2), 185-209.
- Wise, A. F., & Paulus, T. M. (2016). Analyzing learning in online discussions. *The SAGE handbook of e-learning research*, 2e, 270-290
- Zhu, C. (2013). The effect of cultural and school factors on the implementation of CSCL. *BJET* 44(3), 484-501.

# Collaborative Scientizing in *Pokémon GO* Online Communities

Jason C. Yip, Travis W. Windleharth, and Jin Ha Lee  
jcyip@uw.edu, travisw@uw.edu, jinhalee@uw.edu  
University of Washington, The Information School – GAMER Lab

**Abstract:** Finding and applying science practices in everyday contexts (scientizing) is a powerful way for people to engage in science learning. This paper examines how people collaboratively scientize through a massively multiplayer mobile game called *Pokémon GO*. For three months, we conducted observations of online communities around *Pokémon GO* and examined how crowdsourcing engagements in these communities can lead to science inquiry. We adhered to the standards of a comparative case study to examine three exploratory cases of scientizing in online discussion about *Pokémon GO*. These examples include crowdsourcing collaborations to compare multiple data sets, countering misinformation, and creating mapping sets. We demonstrate how aspects of these online engagements follow authentic and simple science inquiry. Furthermore, the implications of our research for informal science learning are especially meaningful for understanding how *Pokémon GO* can be used to motivate scientizing practices in families and children.

## Introduction

One goal of science education for youth is to produce scientifically literate citizens that are able to actively apply science in their everyday lives (Rutherford, & Ahlgren, 1990). Yet, for many learners, science learning is abstract and connecting science learning to everyday practice is challenging (e.g., Lee & Fradd, 1998). Researchers have attempted to understand how learners engage in everyday science practices, through interest-driven learning (Edelson & Joseph, 2004), connected learning (Ito et al., 2013), and the tailored practice of hobbies (Azevedo, 2013). One concept we have adopted to understand how people engage in science learning and practices on their own terms is the notion of “scientizing.” Clegg and Kolodner (2014) define scientizing as the ability to ask questions about how the world works, searching to understand knowledge in the world, recognizing the gaps in one’s own understanding to accomplish goals, and investigating personal scientific questions. Today, we live in digital age of crowdsourcing of knowledge and information; people create complex projects and can break it down to simpler tasks that many other individuals can contribute to. Within different online communities, people can connect and communicate with each other over distances and time using information communication technologies (ICT). However, we do not yet have a deep understanding of how online crowdsourcing can promote scientizing practices.

In this study, we are exploring scientizing and online crowdsourcing in the context of *Pokémon GO*, a massively multiplayer online (MMO) augmented reality game (ARG). We are specifically examining *Pokémon GO* because the location-based nature of the game encourages collaborative and participatory efforts for players to collect data in the world. Additionally, because *Pokémon GO* has a wide mainstream appeal, we believe it has a strong potential to motivate children and families in scientizing practices. For three months, we observed a series of discussions in online communities in how players engaged in problem solving tasks around *Pokémon GO*. The goal of *Pokémon GO* is for players to capture virtual creatures (Pokémon). Despite the simplicity of the game, the developers (*Niantic*) are often not forthcoming about the game mechanics. Because of the lack of transparency, players have come together online to crowdsource and reverse engineer how the game works. Based on our online observations, these collaborative efforts present a compelling case for using ARGs to engage people in scientizing practices via their own interests.

This case study addresses multiple gaps in knowledge in research. First, we lack a clear understanding of how scientizing takes place on an open massive online scale. Second, little is known about how scientizing takes place as people reverse engineer a sociotechnical system of their own interest. In this study, we explore the following research questions: **RQ1:** How do these massive, online, crowdsourced projects inform us about how people can engage and openly collaborate in everyday science practices?; and **RQ2:** What science inquiry practices exist as game players try to collaboratively investigate a sociotechnical system of their own interest? From our observations, we present three case studies of *Pokémon GO* and online crowdsourcing. We analyze each case study using a framework of scientizing and Chinn and Malhotra’s (2002) scientific inquiry practices. Finally, we conclude with a discussion on how online crowdsourcing for scientizing can be used to motivate for families and youth in science learning.

## Background

### Augmented Reality Games and *Pokémon GO*

*Pokémon GO* is usually called an ARG, a term which has become intertwined with mixed reality (or hybrid) game (i.e., a game which has both digital and real-world components) and location-based mobile game (i.e., LBMG: a digital game available on mobile platforms dependent on player location) (Rashid, Mullins, Coulton, & Edwards, 2006). Many ARGs are also labeled as pervasive games—taken from the field of pervasive (or ubiquitous) computing—which are “digital games that move beyond the traditional computer interfaces and into the physical world to occupy time and place on a human scale” (Falk & Davenport, 2004, p.127). *Pokémon GO* is also a MMO game, as all players are simultaneously online participating in one global game environment. In *Pokémon GO*, players use the game software to locate and catch Pokémon that appear in real-world locations. Three player factions use captured Pokémon to battle for control of virtual Gyms, which are mapped onto real-world locations. Other locations, termed PokéStops, allow for the collection of game resources. While Gyms and PokéStops are permanent virtual features, “wild” virtual creatures are subject to game mechanics that govern the location and frequency of their availability. While commercial digital games have been available since the 1970s, ARGs take advantage of consumer devices and GPS satellites affording highly accurate positioning information. *Pokémon GO* has achieved popularity unlike any of its predecessors (i.e., early ARGs like *Ingress*), and introduced millions of new players to the genre.

### Scientizing in online gaming communities

Scientizing in online gaming worlds can be a specific designed experience in virtual worlds. For instance, Asbell-Clarke et al. (2012) developed *Martian Boneyards*, a MMO game focused on using tools for scientific inquiry within the storyline. Educational multi-user virtual environments, such as *Quest Atlantis* (Barab, Thomas, Dodge, Cardeaux, & Tuzun, 2005) and *River City* (Ketelhut, 2007) have demonstrated evidence that learners can engage in collaborative scientific practices and activities in online gaming environments. A growing body of research is also examining how scientific practices take place in commercial digital games and their related online communities. Similar to gaming, online communities around games are contexts in which people work together in knowledge building communities (Scardamalia & Bereiter, 1994), engage and learn together in affinity spaces (Gee & Hayes, 2011), and distribute information and knowledge throughout a larger sociotechnical system (Hutchins, 1995). In online gaming communities, novice learners can come together to increase their knowledge by working with other highly skilled players. Online gaming discussion forums themselves allow for knowledge generation of how to solve complex problems in virtual worlds, such as how to conquer an in-game boss battle, determine the most efficient equipment, and figure out solutions and strategies for optimum play. Online collaborations can serve to foster the zone of proximal development among novice and expert players, that is the difference between what a learner can do on their own compared to what a learner can do with guidance and support from others (Vygotsky, 1978).

The most prominent example of scientific practice in discussion forums around gaming is Steinkuehler and Duncan's (2008) study of scientific habits of mind and disposition in online discussions of the massively multiplayer online (MMO) game of *World of Warcraft* (WoW). Using Chinn and Malhotra's (2002) theoretical framework for evaluating inquiry tasks and Kuhn's (1992) framework for epistemology, they noted that 86% of random forum discussions (nearly 2,000 posts) focused on social knowledge construction and 65% of evaluative epistemology in which knowledge is treated as open and evaluative. Practices such as peer-review, collaboration, data sharing, argument generation and counterargument, and the use of evidence to warrant one's claims took place in these forums. Many of these discussions took on similar activities to how professional scientists share data, challenge each other's interpretations of evidence, and confirm claims through their own peer review system. While Steinkuehler and Duncan's (2008) evaluation of WoW demonstrates clear empirical evidence of scientific practice in discussion forums, we need deeper dives into how players use and integrate scientific practices to *collaborate together*. Steinkuehler and Duncan note that while some may believe that these gaming discussion environments allow a large number of posters to make one to two contributions, the WoW online communities solve problems together as large groups of thinkers. In the case of *Pokémon GO*, large collaborative efforts exist because the game (like many games) is not very transparent in its mechanics. As such, misinformation can occur due to “trolling” (when players intentionally post deceptive information) and speculations mistakenly being conveyed as accurate information. Additionally, because of the location-based nature of game overlaying the virtual world over the real world, people are much more likely to collaborate in various scientizing projects as data collection typically requires efforts beyond what a single or a few players can offer. We are interested in understanding how players collaborate together in scientific practice to develop deeper understanding of the game in organized distributed efforts. Finally, we are interested in understanding

the shift from desktop gaming towards mobile gaming, and how learners also use their own locations and contexts to solve specific needs and questions.

## Theoretical frameworks

We use the concept of scientizing (Clegg & Kolodner, 2014) to explain how and why people use science inquiry practices in their everyday life. Scientizing is often not the norm of science education in young people's lives. However, it is an important concept to encourage people to pursue science learning. When young people scientize, they take on new roles in science learning and feel empowered to use science practices to improve their world. Scientizing focuses on the development of scientific dispositions. First, conceptual and procedural understanding focuses on how people understand scientific concepts and principles. In other words, how people know when to use science practices, why they use them, and how science is used in practice. Interest refers to people's desires and personal reasons to engage with science practices. Social interactions refer to people's engagements in science practice together. Finally, personal connections emphasize connecting science disciplinary practices to their own knowledge and value systems. We are using the scientizing framework because we are trying to understand the roles of scientific practices that come out of everyday practices of playing *Pokémon GO* and how these practices are driven by participants' interest, curiosity, and social relationships, rather than in formalized academic settings.

To understand the scientizing practices of *Pokémon GO* players in online discussion forums, we use Chinn and Malhotra's (2002) complex science inquiry framework. *Authentic scientific inquiry* refers to the complex activity and research that scientists carry out. Example processes include scientists generating their own research questions, selecting and even inventing variables to investigate, creating complex procedures to address questions of interest, devising models to address research questions, employing multiple controls, incorporating multiple measures of different variables, judging against bias, constantly questioning whether their results are accurate and correct, employing multiple forms of argument, and constructing theories and resolving inconsistencies. In contrast, *simple inquiry tasks* attempt to capture the core process of scientific practices, but are mostly basic processes. For instance, in a simple experiment, learners are usually looking at a single independent variable and single dependent variable. In simple observations, learners carefully observe, measure, and describe objects and phenomenon. For simple illustrations, learners might follow step-by-step a specified procedure and observe the outcome. Authentic and simple inquiry tasks are extreme ends of a spectrum. For this study, we are using Chinn and Malhotra's authentic and simple scientific inquiry to examine how players of *Pokémon GO* engage in these practices as they pursue their interests in online discussion forums.

## Methods

We adhered to the standards of a comparative case study (Merriam, 2009) to examine three exploratory cases of scientizing in online discussion about *Pokémon GO*. We used ethnographic research methods (Guba & Lincoln, 2005) to conduct a content analysis of the online forums. We took an emic perspective for this research (Guba & Lincoln, 2005); to understand *Pokémon GO* players, we joined and participated in *Pokémon GO* online groups and also actively played the game ourselves. In the third week of July 2016 of the release of *Pokémon GO*, the second author of this paper joined 215 online social media groups, Facebook, and Reddit forums. At the time of joining, each of these groups had between 500 and 15,000 members. To spread ourselves out evenly across the groups, we picked groups for each of the 25 most populous USA cities and focused on at least two cities per USA state. We also joined special focus *Pokémon GO* groups (e.g. New England, Disneyland). Once we joined and participated, we monitored the online feed in aggregate from July to October 2016. During this time, we observed and recorded 1) general content and types (topics) of posts; 2) very active posts; 3) themes of the posts; 4) behaviors of the discussants; 5) special events; and 6) collaborative projects about the game. To capture the data, we took notes of these observations and screenshots of the types of topics and significant events. We wrote up field notes that aggregated all the observations and screenshots around a set of user-generated projects and activities (e.g., mapping, egg hatching, determining statistics).

To build the exploratory case studies around the scientizing practices of the *Pokémon GO* online discussants, we set up a criterion for selection. Based on Chinn and Malhotra's (2002) scientific inquiry framework, we chose comparative cases that demonstrated varied scientizing practices, such as hypothesis testing, checking against anomalous data, and social construction of knowledge. We chose three online collaborative projects as cases (July to September 2016) from *The Silph Road*, a large public grassroots network of *Pokémon GO* players with an emphasis on researching various aspects of the gameplay, as they were often cited in many of the *Pokémon GO* communities. *The Silph Road* is a Reddit subgroup of *Pokémon GO* players focused on creating an in-person community network. There are at least nine core team members, with thousands of contributors. They engage the *Pokémon GO* player base to crowdsource large amounts of data to

answer questions about game mechanics and other aspects of play. Typically, these projects result from questions or information gaps that developer *Niantic* has not commented on. Once we selected these three focal cases for review, the first and second author coded the observations and screenshots using Chinn and Malhotra's framework. Example codes included cognitive processes (i.e., generating research questions, designing studies, selecting variables, planning procedures, controlling variables, planning measures, making observations, transforming observations, finding flaws, generalizations, types of reasoning, developing theories, coordinating results from multiple studies, and studying research reports) and dimensions of epistemology (i.e., purpose of research, theory-data coordination, theory ladenness of methods, responses to anomalous data, nature of reasoning, and social construction of knowledge). From these codes, we determined whether they were examples of authentic inquiry, simple inquiry, or mixed. After the coding process, we discussed and resolved issues and created narratives that are reflective of the players and the online discussion around *Pokémon GO*.

## Findings

We present three comparative case studies of players engaging in discussions around *Pokémon GO*. For each case study, we provide a description of the case and an analysis of the scientific inquiry practices.

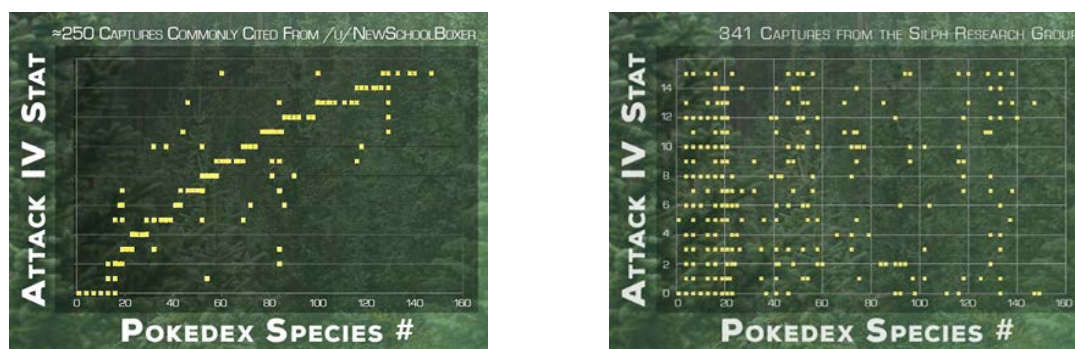
### Case 1: Crowdsourcing to compare multiple data sets to find glitches

Some players in *The Silph Road* suspected that there was a glitch present in the game for the first few months after launch regarding the calculated metric unique to each individual Pokémon called the "Individual Value". Individual Value (IV) is an aggregate score of statistics that pertain to unique Pokémon and is a ratio of the sum of three hidden attributes of each individual Pokémon to the maximum value possible for those attributes. The three hidden scores for each Pokémon are its "attack" score (ATT), "defense" score (DEF), and "stamina" score (STA)<sup>(1)</sup>. The value for each of these three scores for any given Pokémon is an integer between 0 to 15. The sum of these scores, divided by the maximum possible aggregate score (45) is the ratio that is referred to as IV-Individual Value, and it is always expressed as a percentage. The IV can also be expressed with the following equation:  $IV = (ATT + DEF + STA) / 45 * 100$ . For advanced *Pokémon GO* players, IV are an important "hidden" measure of Pokémon strength in battle.

*The Silph Road*, along with players contributing data, organized and conducted their own research to confirm the suspected glitch was indeed real. This glitch speculation emerged at first from anecdotes from players, with many people complaining in online forums that some Pokémon rarely had a high attack value and others almost always had a very high attack value. This seemed to be correlated, they claimed, with the "Pokédex Number" of each Pokémon<sup>(2)</sup>. The Pokédex Number of a Pokémon is simply that creature's permanent index number. For instance, the Pokédex Number for Bulbasaur is 1, for Ivysaur it is 2, and this goes up to 149 in the current version of *Pokémon GO*, as of November 18, 2016. Community contributors to *The Silph Road* decided to collect data on the IV numbers for caught Pokémon, as well as the attack attribute, to engage in hypothesis testing to see if the anecdote held in the face of an empirical test. Nine main participants led the design, management, and data analysis for the project. This group organized and collected data on July 23<sup>rd</sup>, 2016 from 4,120 "throws" (throwing a Pokéball is an attempt to capture the Pokémon).

**Analysis of scientizing practices.** In this project, several aspects of scientizing and inquiry practices exist. First, the group attempted to figure out how anomalous data could have occurred. The hypothesis test was generated because *The Silph Road* could not get direct answers from the developers. The first null hypothesis generated from *The Silph Road* occurred from the initial suspicion of the glitches occurring in the Individual Value of *Pokémon* monsters. The null stated that Pokédex Number is not related to IV. However, the group collected enough data and rejected the null hypothesis (Figure 1 – left). A second hypothesis test was done to compare the results with the first test. The data after Sept 22<sup>nd</sup> (Figure 1 – right) demonstrated no correlation between IV and Attack; the group here assumed the glitch was fixed in a new update. In both cases, two sets of data and interpretation provided very stark results. Chinn and Malhotra's (2002) notes that simple inquiry tasks are too straightforward, and that responses to anomalous data are to change the hypothesis. However, in authentic scientific research, there are different responses to anomalous data. In this case, *The Silph Road* group determined that the anomalous data could explain that a glitch was involved on July 23, 2016, but not on September 22, 2016. The group did not change their hypothesis, but rather assumed there was a more plausible explanation of a glitch in the game. Being able to compare data from two different dates shows an epistemology that considers results as dynamic. Here, like scientists, the group had to coordinate results from multiple studies, particularly with data that conflicted with each other. However, this group only coordinated results on one level of analysis (IV vs. Pokémon identification number). The group did construct their own theoretical explanations, starting with anecdotal observations that were confirmed by evidence between multiple studies.

A second form of scientizing practices that exists is transforming observations. In authentic scientific practices, these observations are often repeatedly transformed into multiple data formats. In simple inquiry tasks, observations are seldom transformed into other data formats, except perhaps straightforward graphs. In this case study, we observed simple XY graphs developed to compare the two hypotheses. Figure 1 are graphs of Individual Value of a Pokémon and their species identification number. While the figures are the same straightforward XY graphs, the case study shows evidence of the group highlighting important aspects of the graphs. The group used large samples of data (Figure 1,  $N = 250$  left;  $N = 341$  right) to reduce the amount of sampling error in the data. As well, the shapes of the distribution (Figure 1, curvilinear distribution in left and randomized distribution on right) demonstrate the use of visualization of the large amount of data gathered by the players for interpretation.



**Figure 1.** The left figure shows one set of results during the “High IV Glitch Phase”, from *The Silph Road*<sup>(3)</sup>. The right figure shows “post-glitch” testing of Pokémon attack IV values<sup>(3)</sup>.

## Case 2: Crowdsourcing to combat misinformation

Another major data collection activity coordinated by *The Silph Road* was the “Egg Hatching” research project. Shortly after the game launched, there was a tremendous amount of confusion and misinformation surrounding the nature of a game feature called *egg hatching*. In *Pokémon GO*, players can find eggs by visiting PokéStops. The eggs found at PokéStops can be incubated and hatched if a player walks far enough in the real world, and a Pokémon will emerge from the egg. A key point of confusion about the game centered on the type of Pokémon that could be hatched from these eggs. Players were uncertain as to whether “regionally locked” Pokémon could hatch from eggs, and intentional misinformation about the subject were rampant. There are four “regionally locked” Pokémon that exist in the game, meaning they can each be caught only on one specific continent: Kangaskhan can only be caught in Australia or New Zealand, Mr. Mime in Europe, Tauros in North America, and Farfetch’d in Asia. The burning question that the *Pokémon GO* community had was “can regionally locked Pokémon hatch from eggs outside of the region in which they appear”? Until September 22, 2016 Niantic had not made a formal statement about this matter.

This particular question became very contentious as “spoofers” accounts began to appear across the world. In ARGs, a “spoofers” is a person that uses software tools to falsify their physical location in the game, and are thus able to make it appear as if they are in another location. Not long after the launch of the game, regionally locked Pokémon began appearing in Gyms and play areas associated with accounts suspected of spoofing. In online communities, some people publicly admitted to spoofing to acquire regionally locked Pokémon from other continents without having to travel there. The area of uncertainty and contention that caused most of the controversy centered on the fact many players claimed they acquired regionally locked Pokémon from eggs that they hatched. In some cases, it appeared that players making the claims believed it was possible, and simply used this belief to obfuscate their cheating. Other players seemed to be intentionally spreading misinformation and “trolling” the *Pokémon GO* community. Between July 15th to September 30th, 2016, our research team monitored over 200 *Pokémon GO* community Facebook pages, and claims of hatching regionally locked Pokémon generated long and contentious arguments in at least half of them, with some players simply believing the poster and congratulating them for hatching the region-locked Pokémon outside of the specified region, and others accusing them of spoofing. The social conflict surrounding this issue was high, with many players feeling angry that members of their community were cheating to “get ahead”. This prompted *The Silph Road* to conduct a large crowdsourced research project to determine whether or not Pokémon from other regions could in fact be hatched from eggs<sup>(4)</sup>. The effort was organized on *The Silph Road Reddit* and was a contributory data collection project, where users provided their “total number of eggs hatched and total

number of foreign region locked Pokémon<sup>(4)</sup>. The data were provided in a structured way in the forum thread. Thousands of contributors provided data on their egg hatches, and after data collection, players had hatched in aggregate 40,507 Pokémon eggs. Zero eggs produced a region locked Pokémon from a different continent. Thus, the project organizers declared the myth “busted”.

**Analysis of scientizing practices.** This case study demonstrates many themes of scientizing practices to counter against misinformation. Unlike WoW studies from Steinkuehler and Duncan (2008), we observed that players needed to navigate complex misinformation. To find the answer to their question of region specific Pokémon hatching from eggs, the group needed to guard against bias. Authentic science practices make an attempt to limit experimenter bias. To reduce errors and sampling problems, the group had to collect data characterized as 1) large in amount (40,507 Pokémon eggs); 2) varied from thousands of random players; and 3) varied in regions around the world (all playable continents). Here, the players worked together to make sure threats against validity were mitigated. Second, the players had to check data against misinformation and anomalous data. Because misinformation may have persisted from spoofer, the players in *The Silph Road* needed to constantly check their data against possible spurious claims. Here, the players’ epistemology focused on theory-data coordination. Chinn and Malhotra (2002) note that in authentic inquiry, scientists coordinate theoretical models with multiple sets of complex, partially conflicting data. In simple inquiry, people coordinate one set of observable results with conclusion about those observable results. In this case, we believe the theory-data coordination leans more toward authentic than simple. Because *The Silph Road* collected thousands of data points from all across the world, they were comparing multiple sets of data. In this case, the players were seeking global consistency, rather than local consistency. Because the players were committed to seeking confirmation of their theory, they had to respond to anomalous data by rationally discounting data. By having a large global dataset, this helped to eliminate biased data and misinformation.

### Case 3: Crowdsourcing to create knowledge together

The largest class of crowdsourced and community sourced data collection that has occurred and continues to occur among the *Pokémon GO* player base is the development of “nest maps”. In *Pokémon GO*, certain locations in the world tend to produce a specific type of uncommon Pokémon at a relatively high frequency, referred to as nests<sup>(6)</sup>. These nest areas are particularly useful for players because they produce a specific Pokémon in relative abundance (approximately 3 to 15 Pokémon per hour), and players will travel to that location for the certainty of acquiring the resident Pokémon. Mapping the locations of these nests and the Pokémon that inhabit them, has become a large and persistent project for *Pokémon GO* players that want to contribute data to the community. Adding to the complexity of the project and the need for continuous community sourced data collection is the fact that nests change over time. Nests appear to persist for approximately one month and then they change, in what the *Pokémon GO* player community has come to call “nest migration”. As such, player generated nest maps are in constant need of revision. *The Silph Road* manages the largest nest mapping project, though an interesting phenomenon occurring is that many local communities are building regional projects to map and communicate nests in their own area. Typically, an organizer for a community will create a document or spreadsheet, and solicit suspected nest locations from the local players, who then update the list directly, or post reports into an online forum thread. Examples include the San Antonio nest map<sup>(7)</sup>, which is updated by the organizer monthly using the local Facebook group<sup>6</sup>, and the Las Vegas Pokémon Nest spreadsheet, which collects data on nest from community members, and distributes a link to a Google Docs™ spreadsheet reporting 293 nests in and near Las Vegas. The largest effort to map global nests remains with *The Silph Road*, which has developed the contributor capacity to very quickly collect a lot of data after each migration<sup>(8)</sup>. For instance, the third Pokémon migration occurred on September 26th, 2016. Within 24 hours of the migration, *The Silph Road* reported 70,374 individual nest reports submitted by players worldwide, mapping 57,606 unique nest locations and corresponding Pokémon<sup>(9)</sup>.

**Analysis of scientizing practices.** This case study demonstrates aspects of the *social construction of knowledge* epistemology. Chinn and Malhotra (2002) explain that scientists construct knowledge in collaborative groups, often building on prior research from other scientists. To do this, scientists need institutional norms, such as a peer-review process or models of research. In the simple inquiry perspective, students can construct knowledge together too, but they seldom build on prior research or have experience with institutional norms. In this case, *The Silph Road* created a way for the community to contribute towards building this knowledge that is somewhere between authentic and simple. The creation of Google Docs, social media, and mapping tools provided a common way for everyone in *The Silph Road* to contribute. Here, there is a grounding of the methods which creates a guideline for the group to report the data for the nesting project. However, there are no institutional norms in the group for patterns of argument or a model for research in this case. For instance, the group does not (yet) have a way to review arguments about the nest migration.



## Discussion and implications

Steinkuehler and Duncan (2008) provided clear empirical evidence to show that players engaged in WoW online communities not only engaged in informal science literacy, but that an overwhelming majority of the conversation was dedicated to productive forms of scientific practice. Similarly, our findings demonstrate that players engaging in *Pokémon GO* online communities are not only engaging in scientific practice, but aspects of these practices fall within aspects of authentic scientific practice. Our findings extend the work of Steinkuehler and Duncan by showing specifically how these scientizing practices emerge to solve very complex problems. The nature of the problems in *Pokémon GO* require distributed collaboration across time and space through ICT (Hutchins, 1995). Determining “glitches”, filtering through misinformation, and creating a knowledge base for geo-mapping is not an effort that only a small handful of experts could solve. Our findings show that crowdsourcing information and data together to solve problems not only encourages scientizing practices, but *requires those practices to be collaborative*. Players had to collect large amounts of data in multiple forms, organize that data collection effort across multiple continents and locations, generate models from that data, create community practices around how that data is to be managed, create evidence-based arguments, and determine the unknowns within their arguments. We believe that even though *Pokémon GO* touts itself as a simple online location-based ARG, there are deeper scientizing practices that can occur within everyday gameplay. In addition to the mass appeal, abundant opportunities for participation in data collection and sharing, as well as the need for communicating and working together as a group make *Pokémon GO* an attractive venue for engaging people in scientizing practices.

Furthermore, we believe that the implications of our research for informal science learning are meaningful for families and children — *Pokémon GO* can be used to motivate scientizing practices in families and children. *Pokémon GO* is a mainstream game meant for players of all ages, with very close youth and family connections. Sobel et al. (2017) found that many parents had a much more positive attitude towards the screen time their children were having while playing a game like *Pokémon GO* than other video games (such as *Minecraft*) since they are playing outside and together as a family. With the added benefit of strong appeal to parents, we believe a game like *Pokémon GO* presents an ideal opportunity for engaging children to scientize together with their friends and family. Takeuchi and Stevens (2011) note that children’s engagement with parents about digital media has shown great potential for learning. As well, children scientizing together with other adults has shown great potential for engagement in science learning (e.g., Bell, Bricker, Reeve, Zimmerman, & Tzou, 2013). In our future work, we plan to look at how families and children may be engaging in scientizing practices in other online gaming communities, similar to *Pokémon GO*. Our findings are limited at this point because we were unable to dig deeper at the ages and relationships of the players in these online venues. Future work in this area needs to uncover more specifics about whether youth and families are currently engaging in these crowdsourcing scientizing efforts. We believe that if youth and families are not engaged in these practices around their commercial games, online communities can potentially be developed for this purpose. Online communities like Reddit and Facebook are not designed yet for family, teacher, and children crowdsourcing projects for scientizing. However, families, children, and other learning stakeholders could either participate or run these kinds of large-scale investigations themselves. We believe that bridging third spaces (Steinkuehler, 2008) by creating online communities for families, children, and teachers around *Pokémon GO* and other simple location-based games could potentially be useful in demonstrating how everyday practices and engagements can contribute towards authentic science inquiry practices.

## Endnotes

- (1) <https://pokeassistant.com/main/ivcalculator>
- (2) <http://www.pokemon.com/us/pokedex/>
- (3) <https://thesilphroad.com/science/attack-iv-pokedex-number-correlation>
- (4) [https://www.reddit.com/r/TheSilphRoad/comments/4w9d00/have\\_you\\_hatched\\_eggs\\_great\\_i\\_want\\_you\\_to\\_tell\\_me/](https://www.reddit.com/r/TheSilphRoad/comments/4w9d00/have_you_hatched_eggs_great_i_want_you_to_tell_me/)
- (5) <http://heavy.com/games/2016/09/pokemon-go-can-you-hatch-region-specific-pokemon-from-egg-eggs-outside-of-location-tauros-kangaskhan-mime-farfetchd-north-america-europe-asia-australia-john-hanke/>
- (6) <https://pokemongo.gamepress.gg/pokemon-nests>
- (7) <https://docs.google.com/document/d/1a7kzD3ewFCS68yEEV55gUWtsE011S3VtUZIBqBZokw/edit>
- (8) [https://www.reddit.com/r/TheSilphRoad/comments/54ndqm/megathread\\_the\\_3rd\\_great\\_migration\\_appears\\_to/](https://www.reddit.com/r/TheSilphRoad/comments/54ndqm/megathread_the_3rd_great_migration_appears_to/)
- (9) <https://thesilphroad.com/atlas>

## References

- Asbell-Clarke, J., Edwards, T., Rowe, E., Larsen, J., Sylvan, E., & Hewitt, J. (2012). Martian Boneyards: Scientific inquiry in an MMO game. *International Journal of Game-Based Learning*, 2(1), 52–76.

- Azevedo, F. S. (2013). The tailored practice of hobbies and its implication for the design of interest-driven learning environments. *Journal of the Learning Sciences*, 22(3), 462–510.
- Barab, S., Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development*, 53(1), 86–107.
- Bell, P., Bricker, L.A., Reeve, S., Zimmerman, H.T., & Tzou, C. (2012). Discovering and supporting successful learning pathways of youth in and out of school: Accounting for the development of everyday expertise across settings. In B. Bevan, P. Bell, R. Stevens, & A. Razfar (Eds.), *LOST opportunities: Learning in out of school time* (pp. 119–140). London: Springer.
- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218.
- Clegg, T. L., & Kolodner, J. (2014). Scientizing and cooking: Helping middle-school learners develop scientific dispositions. *Science Education*, 98(1), 36–63. <https://doi.org/10.1002/sce.21083>
- Edelson, D. C., & Joseph, D. M. (2004). The interest-driven learning design framework: motivating learning through usefulness. In *Proceedings of the 6th International Conference of the Learning Sciences* (pp. 166–173). International Society of the Learning Sciences.
- Falk, J., & Davenport, G. (2004). Live role-playing games: Implications for pervasive gaming. In *International Conference on Entertainment Computing* (pp. 127–138). Springer.
- Gee, J. P., & Hayes, E. (2011). Nurturing affinity spaces and game-based learning. In C. Steinkuehler, K. Squire, & S. Barab (Eds.), *Games, learning, and society: Learning and meaning in the digital age* (pp. 129–153). New York, NY: Cambridge University Press.
- Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In *The SAGE handbook of qualitative research* (3rd ed., pp. 191–215). Thousand Oaks, CA: SAGE.
- Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, 19(3), 265–288.
- Ito, M., Gutiérrez, K., Livingstone, S., Penuel, B., Rhodes, J., Salen, K., ... Watkins, S. C. (2013). *Connected learning: An agenda for research and design*. Digital Media and Learning Research Hub.
- Ketelhut, D. J. (2007). The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in River City, a multi-user virtual environment. *Journal of Science Education and Technology*, 16(1), 99–111.
- Lee, O., & Fradd, S. H. (1998). Science for all, including students from non-English language backgrounds. *Educational Researcher*, 27(4), 12–21.
- Kuhn, D. (1992) Thinking as argument. *Harvard Education Review*, 62(2), 155–178
- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: John Wiley and Sons.
- Rashid, O., Mullins, I., Coulton, P., & Edwards, R. (2006). Extending cyberspace: Location based games using cellular phones. *Computers in Entertainment (CIE)*, 4(1), 4.
- Rutherford, F. J., & Ahlgren, A. (1990). *Science for all Americans*. Oxford University Press.
- Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, 3(3), 265–283.
- Sobel, K., Bhattacharya, A., Hiniker, A., Lee, J.H., Kientz, J., & Yip, J.C. (2017). "It wasn't really about the Pokémon": Understanding families' experiences with a location-based mobile game. In *Proceedings of SIGCHI Human Factors in Computing Systems (CHI 2017)*. New York, NY: ACM.
- Steinkuehler, C.A. (2008) Cognition and literacy in massively multiplayer online games. In J., Coiro, M. Knobel, C. Lankshear, D. Leu (eds.) *Handbook of research on new literacies*. (pp. 611-634). Erlbaum, Mahwah NJ.
- Steinkuehler, C., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology*, 17(6), 530–543.
- Takeuchi, L., & Stevens, R. (2011). *The new coviewing: Designing for learning through joint media engagement*. New York: Joan Ganz Cooney Center at Sesame Workshop. Retrieved from <http://www.joanganzcooneycenter.org/publication/the-new-coviewing-designing-for-learning-through-joint-media-engagement/>
- Vygotsky, L. S. (1978). Interaction between learning and development. In *Mind in society: The development of higher psychological processes* (pp. 79–91). USA.

## Acknowledgments

We thank the numerous members of the various *Pokémon GO* online communities that allowed us to interact and inquire about their gameplay.

# Making the Invisible Visible: A New Method for Capturing Student Development in Makerspaces

Richard Lee Davis, Stanford University, rldavis@stanford.edu  
Bertrand Schneider, Harvard University, bertrand\_schneider@gse.harvard.edu  
Paulo Blikstein, Stanford University, paulob@stanford.edu

**Abstract:** The contribution of this paper is twofold: we introduce a new kind of assessment developed to capture students' learning in makerspaces, and we present a new perspective on how participating in a maker workshop impacts students. As opposed to traditional pen and paper tests, we designed a series of hands-on tasks that participants complete before and after a maker workshop. In this paper, we contrast high-school students' performance with experts (graduate students in mechanical engineering) and found evidence that the students' behavior became more similar to experts' after participating in a maker workshop. For the scope of this paper, we focus on a single task and describe in detail our coding scheme and analyses. Additionally, we show how a combination of qualitative and computational analysis helped us develop metrics to compare novices' and experts' performances. We conclude by discussing the potential of this type of assessment for supporting students' learning in makerspaces.

## Introduction

Informal communities of tinkerers, inventors, hackers, and builders have existed across the world for decades. The most famous of these are practically household names—the Homebrew Computer Club in Silicon Valley, the Chaos Computer Club in Berlin—but the vast majority are small communities meeting in garages, classrooms, basements, and libraries. In the past decade, a common identity has emerged for the members of these disparate groups: the maker. Since the emergence of *Make* magazine in 2005, the word “make” has become the dominant signifier of this new community. The maker movement is made up of a community of makers, who gather in makerspaces to take part in a common activity: making. As the maker movement spreads widely across the US and Europe, we are beginning to see a growing interest in bringing making into K-12 schools (Halverson & Sheridan, 2014). And rightfully so; many aspects of the maker movement have the potential to disrupt traditional schooling and positively impact K-12 students. The communities of practice that take root in makerspaces are powerful, authentic learning environments (Blikstein, 2013; Halverson & Sheridan, 2014; Pepler & Bender, 2013). The focus on creating complex, personally meaningful artifacts has known learning benefits (Blikstein & Krannich, 2013; Papert, 1980; Piaget, 1973). However, one notable aspect of the maker movement with important educational potential has escaped investigation so far: the maker mindset.

From an educational perspective, the maker mindset is one of the most compelling aspects of the maker movement. Its existence implies that becoming a maker involves more than learning how to create products; it involves a change in one's view of the world. If we wanted to identify the maker mindset, what would we look for? Because the maker mindset has many definitions and descriptions, it is difficult to know where to start. Dale Doherty, the founder and CEO of Maker Media, Inc., defines the maker mindset as a “can-do attitude... an invitation to take ideas and turn them into various kinds of reality... It is a chance to participate in communities of makers of all ages by sharing your work and experience. Making can be a compelling social experience, built around relationships” (Dougherty, 2013). Martin writes that it is “playful, asset- and growth-oriented, failure-positive, and collaborative” (Martin, 2015).

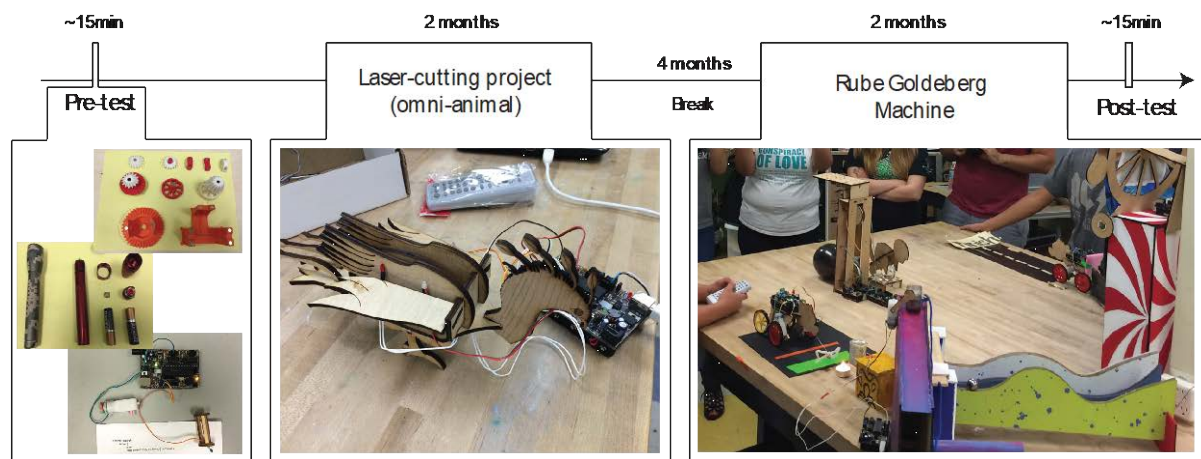
Here, we are interested in a more focused definition of the maker mindset. In previous work, it has been observed that students seem to be more capable of reasoning about and debugging complicated mechanisms after participating in a making workshop (Blikstein, 2013). Fields et al. developed an assessment called a Debuggem to capture this change (Fields, Searle, Kafai, & Min, 2012), finding that after a four-week electronic textiles workshop, students were more able to fix faulty designs like short circuits, poor crafting, and incorrect code. We view this change as part of the shift towards a maker mindset. More specifically, our interpretation of these findings is that after taking part in a maker workshop, some students have learned to think more like engineers. Although this type of thinking is difficult to find in common descriptions of the maker mindset, we feel it is an important outcome with special relevance for K-12 education.

The current study was designed to answer three questions about students' participation in a maker workshop. First, do students think more like engineers as a result of taking part in a long-term maker workshop? Second, is there a way to reliably and efficiently capture this change? And third, if this change in thinking does

occur, can we look more closely at the data and begin to understand the specific ways in which the students' thinking changes?

## General description of the study

A class of high-school seniors took part in a workshop in our FabLab for several months (Fig. 1). Since the workshop targeted students' understanding of complex mechanical systems, we designed assessments that would capture their ability to build, fix or debug them. Before the workshop, we administered 3 tasks to the high-schoolers and they completed 3 similar tasks after the workshop. Once the workshop was over, we recruited 18 experts (graduate students in mechanical engineering) to complete all six tasks from the pre-test and post-test.



**Figure 1.** Timeline of the study. On the left, three examples of tasks used before and after the workshop to assess its effect on students' mindset (described in details in the methods section). In the middle, an example of an Omni-Animal laser cutting project. On the right, the final Rube Goldberg machine.

**Hypotheses:** We expect to see three trends in our data based on participants' performances at the pre- and post-test: 1) novices should improve from pre to post-test; 2) experts should perform significantly better than novices; and 3) novices' behavior should become more similar to experts at the post-test.

## Methods

### Subjects

20 high-school seniors (4 males, 16 females) from a low-SES school took part in a workshop organized in our FabLab. Three female students had to be excluded from our analyses because they dropped out of the workshop. 18 graduate students (9 females and 9 males, mean age=24.67, SD=2.13) in mechanical engineering from an R1 university were asked to complete the study tasks as a comparison group. They received a \$20 gift card for their participation. We will refer to high-schoolers as "novices" and mechanical engineers as "experts" henceforth.

### Intervention

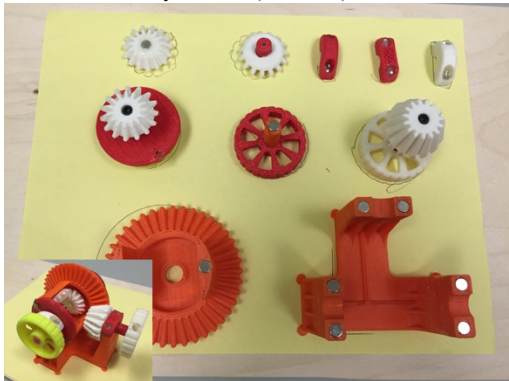
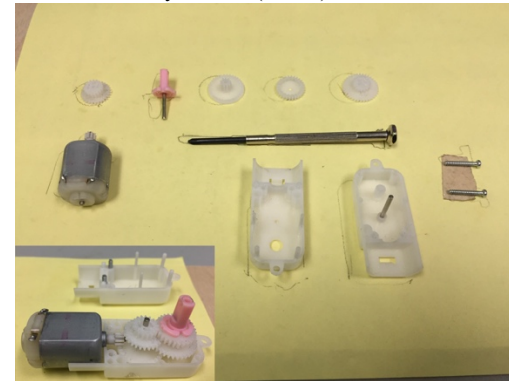




The students took part in a year-long workshop in our FabLab that stretched across two two-month blocks (Figure 1). During each block, the students worked in the lab for 1.5 hours per day, two times a week. In the first two months, the students worked in pairs on a project called the Omni-Animal. During this project students learned how to design their own three-dimensional creature using computer-aided drawing (CAD) software (Figure 1). Each creature was constructed from 6-14 individual pieces, which the students cut out of a piece of plywood using the laser cutter. Due to the difficulty of translating between two and three dimensions, none of the students' initial designs were successful. The students encountered problems like making slots too wide, failing to take the thickness of the material into account, and making pieces too big or too small. All of the students iterated on their initial designs multiple times and ultimately completed a successful Omni-Animal. In the second two-month period, the students worked in groups to create an electro-mechanical Rube Goldberg machine (Figure 1). The students worked in groups of two to three with each group working on one part of the machine. Each group created a laser-cut mechanism paired with a microcontroller platform called the GoGo board (Sipitakiat et al., 2004) to add motion, react to the environment, or create an effect.

## Materials

We used three tasks as a pre-test and three tasks as a post-test. The first set (A) is shown in the first column of Table 1, and the second set (B) is shown in the right column. Task 1 is about rebuilding a gear system: 1a is a differential, and 1b a motor. Task 2 is about fixing a broken system: 2a is a flashlight where the bulb is broken and the batteries are inverted; 2b is a remote control where the batteries and spring are inverted, the fan is switched off. Task 3 is about reconfiguring a micro-controller: the sensor and actuator are plugged into the wrong ports.

Those tasks went through several iterations before we were confident that they could capture students' learning. More specifically, since the laser cutting project was about creating three-dimensional creatures from flat pieces of wood, we expected that students would increase their ability to assemble complex mechanical objects from their individual pieces. They also learned to use the GoGo board, which is an easy to use micro-controller with plug-and-play sensors and actuators and block-based programming environment. Since many everyday devices include a similar input-output structure (e.g., a button / light sensor / motion sensor detects an event, a few lines of code analyze the data, and the system triggers a response with a motor or a speaker), we hypothesized that the students would be better equipped to reason about, troubleshoot, and repair everyday devices. Finally, the Rube Goldberg project was designed to convey the idea that complex systems are more than the sum of their parts: during the workshop, each group built one step of the machine using the laser cutter and a GoGo board. All six tasks described in Table 1 tried to capture those three aspects to various extent.

Table 1: the six tasks given to our participants (a smaller image on the bottom left shows the final product for task 1 or a step toward fixing the system for task 2). For the scope of this paper, we focused on task 1a.

<p>Task 1a – Assembly Puzzle (Gearbox)</p> 	<p>Task 1b – Assembly Puzzle (Motor)</p> 
<p>Task 2a (left)– Fixing an everyday object (Flashlight)</p>  <p>Task 2a (right)– Fixing an everyday object (Flashlight)</p> 	<p>Task 2b (left) – Fixing an everyday object (remote control)</p>  <p>Task 2b (right) – Fixing a remote control</p> 
<p>Task 3a (right)– Reconfiguring a Micro-Controller</p>	<p>Task 3b (right) - Reconfiguring a Micro-Controller</p>

## Design

Half of the novices completed set A as a pre-test and set B as a post-test. We counter-balanced those tasks for the other half of the participants. In other words, half of the novices completed tasks 1a, 2a and 3a for the pre-test and tasks 1b, 2b and 3b for the post-test. The other half of the novices completed tasks 1b, 2b and 3b for the pre-test and tasks 1a, 2a and 3a for the post-test. The experts completed all six tasks in a single session, and we also counter-balanced set A and B for them.

## Procedure

Novices were asked to leave the workshop for a short period of time to complete the study. Because of time constraints, they were run in pairs sitting side by side at a 3' by 6' table. Each participant was given their own puzzle to work on. They were instructed not to work together on the puzzles or to look at what the other person

was working on. First, an *Assembly Puzzle* was placed in front of each participant on a wooden board (task 1). The participants were told that the object in front of them had been disassembled, that it was their job to try and put it back together in five minutes, and that they should try their hardest and not be frustrated if they were unable to solve the puzzle. They were not given any further information about the object (i.e., no instructions on how to assemble the object). Participants were instructed not to touch the puzzle until the timer was started. Once the timer was started they had five minutes to try and solve the assembly puzzle. When the time expired, the puzzle was removed from the table. Next, the repair puzzle was placed in front of the participant (task 2). The participants were given four minutes to work on this puzzle. When the timer expired, the repair puzzle was removed from the table, and the reconfiguration puzzle was placed on the table (task 3). They had three minutes to work on the reconfiguration puzzle. When the time expired, the participants were thanked for their time and left the room.

Since there was no time constraint associated with participating to a workshop, experts were run one at a time and worked through all six tasks in a single session. Half of the experts worked on the tasks from set A first, followed by the tasks from set B (1a-2a-3a-1b-2b-3b), while the other half did the tasks from set B followed by set A (1b-2b-3b-1a-2a-3a). The time-per-task was the same for experts and novices. Video, audio, and body position (using a Kinect sensor) were recorded for the duration of the study.

## Video coding

We had two major goals which resulted in the development of two distinct video coding schemes. First, we needed a way to determine how close participants came to the correct solution. The 11-point assessment scale was developed for this purpose. Second, we needed a way to model the participants' full sequence of meaningful actions as they worked through each puzzle. We designed the time-based coding scheme for this purpose.

### The 11-point assessment scale

In order to meaningfully compare pre-workshop students, post-workshop students, and experts, it was important to develop a metric that accurately measured how close each participant came to the correct solution. However, most of the participants—including experts—made significant progress on the problem but failed to fully complete the gearbox problem in such a short amount of time (5 minutes), so we needed a more nuanced way of measuring progress than percentage of participants who completed the puzzle. Through an iterative coding process we created an 11-point time-agnostic scale for this purpose.

Coding videos with this metric occurred as follows. While watching the video, if the participant performed an action that matched one of the codes, we assigned a 1 to that code. If the participant's action matched a code partially, we assigned a 0.5 to that code. No time information was recorded (hence time-agnostic); that is, at the end of the video, two participants who carried out the same actions in different orders would receive the same score. At the end of the video, we summed up the scores across codes and assigned this to the participant. A score of 0 meant the participant made no progress on the problem, while a score of 11 meant the participant completely solved the problem. The higher the score, the closer to finishing successfully.

### The time-based coding scheme

Since the 11-point assessment scale did not capture any temporal information, we designed a second time-based video coding scheme to categorize the different types of actions participants carried out while attempting to solve the gearbox problem. This coding scheme allowed us to analyze how sequences of actions differed between groups. The final coding scheme contained 14 codes that belonged to 4 categories: planning, evaluation, context, and action. The coding scheme complements the 11-point assessment scale and was designed so that it could be translated into other coding schemes used in similar analyses (Tschan, 2002; Worsley & Blikstein, 2013).

This time-based coding scheme went through a number of iterations. Initially, we coded the exact state of the gearbox as participants worked on the problem. Each code consisted of a set of pieces with an optional sub-code indicating if any pieces were added or taken away in that moment. A major flaw of this coding scheme was the inability to distinguish between a correctly-assembled set of parts and an incorrectly-assembled set of parts. Ultimately, this coding scheme proved to be too noisy for any useful analysis. The next iteration of our coding scheme treated the participants' actions in a more general way. Instead of hundreds of possible codes, we narrowed down the meaningful actions to 9 codes: exploring, looking, rotating, plastic connection, magnetic connection (correct), magnetic connection (incorrect), meshing gears, disassembling, and placing an axle in a hole or bracket.

Finally, in order to make our coding scheme cross-compatible with other coding schemes of interest (Tschan, 2002; Worsley & Blikstein, 2013), we added new codes and further refined the existing ones. The final coding scheme contained 14 codes in 4 categories. We designed custom software to streamline the process of coding the videos. Each time the participant carried out an action, the appropriate code was entered and linked to the video using a timestamp (Figure 2). After coding a participant's video, we were left with a full sequence of

the participant's actions during the problem. In other words, we transformed video of the participants' actions into a time-stamped sequence of codes.



Figure 2: an example of a sequence of actions at different time points coded with the Time-Based Coding Scheme. From left to right: **axle**, **rot**(ate), **plas**(tic connection), **mesh**(ing gears), **axle**, **mag**(netic connection).

## Results

### Hypotheses 1 (novices' improvement from pre to post) and 2 (experts vs novices)

For hypotheses 1 and 2, we used the 11-Point Assessment Scale described in the coding section to assign each participant a score between 0 (no progress) and 11 (finished solution) on the Gearbox task (1a). We first visually explored our dataset using boxplots (Fig. 3, left side). Half of the novices completed the gearbox task before the workshop, and the other half completed it afterwards. All experts completed the task. The descriptive statistics are as follows: for the novices on the pre-test,  $N = 9$ , mean = 2.5,  $SD = 2$ , for the novices on the post-test,  $N = 8$ , mean = 3.69,  $SD = 1.85$ , and for the experts  $N = 18$ , mean = 6.97,  $SD = 2.45$ . The boxplots also revealed that there was an outlier among the novices in the pre-test. One student scored a 7 while the second best student scored a 3. This particular participant had worked in the lab full-time as an assistant for 3 months preceding the workshop and was excluded from the following analyses. To test our first hypothesis (whether our novices improved from pre to post-test), we used an ANOVA to compare the novices' performance before and after the workshop. We found that a significant improvement from pre to post:  $F(1,14) = 5.17, p < 0.05$ , Cohen's  $d = 1.14$ . To test our second hypothesis (whether experts performed better than novices), we grouped pre- and post-tests for novices and used an ANOVA to compare them to the experts. We found that experts did significantly better at the gearbox task than novices:  $F(1,34) = 27.00, p < 0.001$ , Cohen's  $d = 1.82$ .

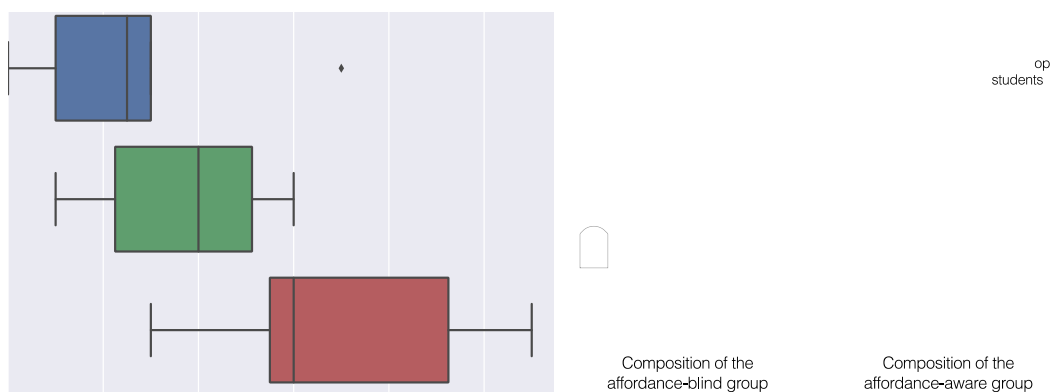


Figure 3. Left: Boxplot of the novices' scores on the 11-point scale for the pre and post-test compared to experts' scores on the gearbox task. Right: Composition of the novice and expert groups found by clustering on problem-solving sequences.

### Hypothesis 3: novices' similarity to experts after the workshop

Since one of the goals of the study was to see if novices became more like experts after taking part in a long-term workshop in the FabLab, we needed a way to capture the similarity between any two problem-solving sequences. We used two different techniques to test this hypothesis: first, we translated sequences into an existing coding scheme (Tschan, 2002) to estimate how many "ideal cycles of cognition" (described below) novices and experts went through. Second, we used computational techniques to compute a similarity score between all participants and used an unsupervised clustering algorithm to separate participants into groups based on their problem-solving sequences. The main idea behind both analyses is to see whether novices at the post-test start to display more "expert-like" behaviors.

### Ideal cycles of cognition

Tschan (2002) found that individuals and groups of students performed better on a problem-solving task when they completed a higher number of “ideal cycles of cognition”. An ideal cycle is composed of three steps: planning, acting, and evaluating. We collapsed the 14 codes of the time-based coding scheme as follows: planning (fdis, look, org), acting (axle, axlex, dis, disx, mag, magx, plas) and evaluating (mesh, meshc, rot, test). This distinction is obviously not ideal, because the planning and evaluating phases overlap. But as a preliminary result, we found a trend similar to the left plot of Figure 2 (see boxplots on Fig. 5).

An ANOVA revealed a significant difference between novices and experts:  $F(1,34) = 14.45, p < 0.001$ , Cohen's  $I = 1.43$  (novices mean=1.07, SD=0.77; experts mean=2.81, SD=1.55). For novices, there was a trend in the same direction (although non-significant):  $F(1,14) = 3.10, p = 0.10$ , Cohen's  $d = 0.99$  (pre-test mean=0.75, SD=0.83; post-test mean=1.43, SD=0.49).

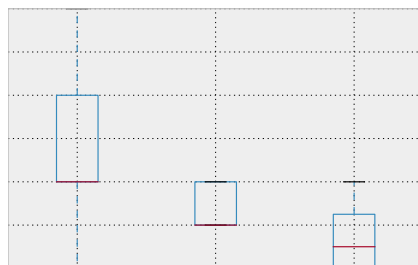


Figure 5. boxplots showing the number of ideal cycles of cognition for novices and experts.

### Clustering on raw actions

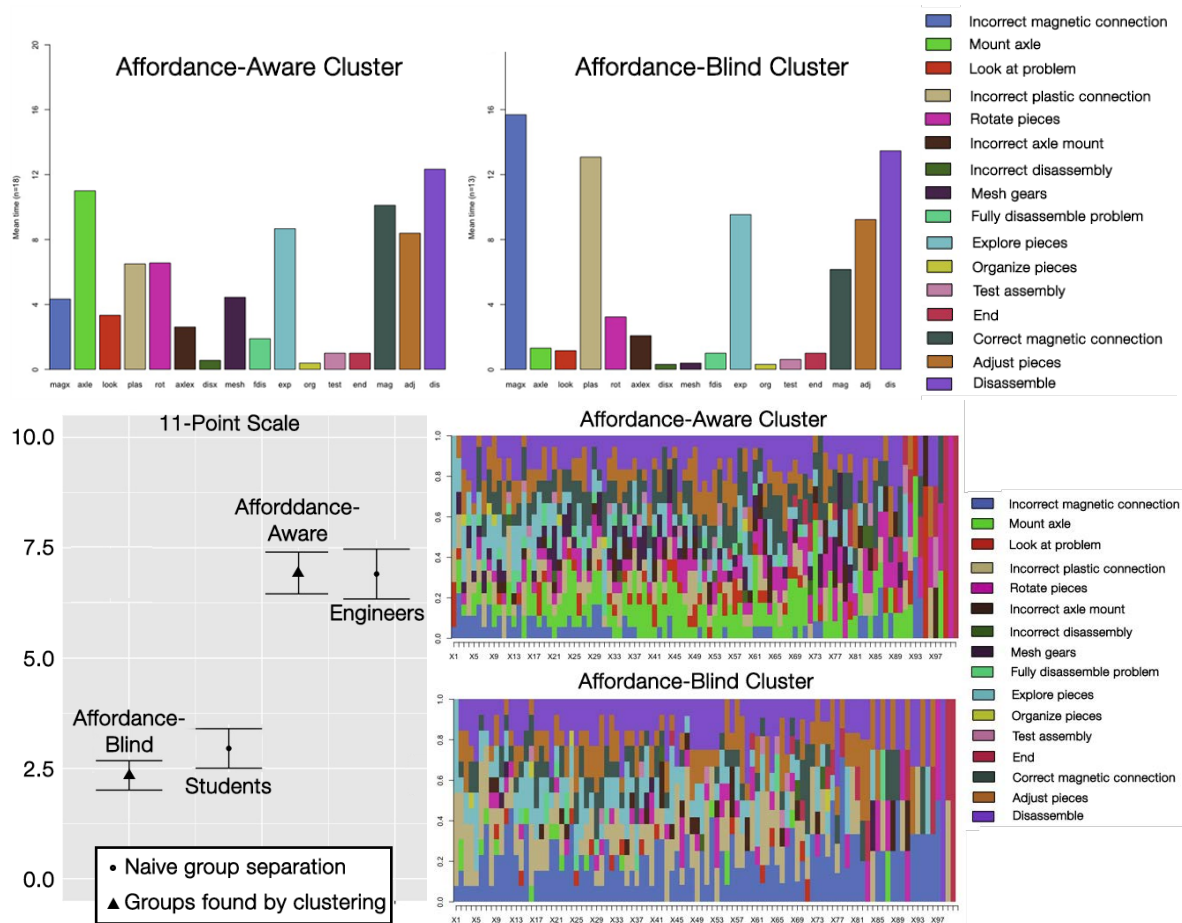
By clustering participants according to their entire set of actions on the gearbox problem, we were able to identify two groups of participants. One group contained 13 participants, composed of all of the pre-workshop high-school students, 62.5% of the post-workshop high-school students, and 6% of the experts. The other group contained 19 participants, composed of zero pre-workshop high-school students, 37.5% of the post-workshops students, and 94% of the experts (Fig. 3, right side).

We used the R package TraMineR for clustering participants on their raw actions (Gabadinho, Ritschard, Mueller, & Studer, 2011). By computing the edit distance between all pairs of sequences using TraMineR's optimal matching algorithm, we were able to construct a symmetric distance matrix that captured the similarity between all pairs of participants. That is, for all pairs of participants, we computed a single index that captured the similarity of their problem-solving trajectories. Finally, we used agglomerative hierarchical clustering (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2016) to separate out the two groups of participants who were most similar to each other.

Based on the compositions of these groups, we labeled the first group affordance-aware and the second group affordance-blind. We chose these labels based on the most prevalent actions within each group. The participants in the affordance-aware group carried out a higher proportion of axle-related actions, gear-meshing actions, correct magnetic connections, and rotations (Figure 4). In comparison, the affordance-blind group carried out a higher proportion of incorrect magnetic connections and incorrect plastic connections (Figure 4). More importantly, the affordance-blind group carried out almost zero axle, gear, or rotation actions. The affordance-blind group seemed unable to perceive the important affordances of the pieces, while the affordance-aware group took these into account when solving the gearbox problem.

The only sub-group split between the two clusters were the post-workshop high-school students. Our interpretation for this split is simple: the workshop had a positive effect. After participating in the workshop, nearly half of the students were more similar in their problem-solving behavior to experts than to pre-workshop high-school students. To validate the integrity of the two clusters, we compared each cluster's scores on the time-agnostic 11-point assessment scale using a two-tailed t-test. Not only was there a strongly-significant difference between the affordance-blind group (mean=2.35, sd=1.20) and the expert-like group (mean=6.89, sd=2.00) ( $t(28.23)=7.89, p < 0.001$ ), but the difference in means was even-larger than the previous tested difference (effect size of previous test is  $d=1.97$ , while effect size of current test is  $d=2.65$ ) (Figure 4).





**Figure 4.** Top: Frequency of actions by cluster. Bottom: Proportion of actions at each time step for each cluster. Note the higher proportion of axle-related actions (green), meshing gears (dark purple), and rotation (fushia) in the Affordance-Aware cluster, and the higher proportion of incorrect magnetic connections (sky blue) and incorrect plastic connections (beige) in the Affordance-blind group.

## Discussion

The motivation for this work was to develop a new methodology for capturing students' learning in makerspaces. We have designed a new kind of task-based assessment where students had to rebuild a gear system, fix everyday devices, and debug a microcontroller. These tasks were designed to assess skills that participants learned during the workshop, such as constructing a 3D object from 2D parts, building an input-output system, and understanding that complex systems are made of small interconnected parts.

For our preliminary analyses, we have focused on the first task of this assessment (the gearbox). We developed an extensive coding scheme to estimate students' performance and found that participants significantly improved from pre- to post-test. This suggests that the workshop had a positive effect on their ability to reason about and assemble complex mechanisms. Additionally, we asked experts (graduate students in mechanical engineering) to complete the same tasks and found that novices became more like experts in their perception of the salient features of the problem and in their problem-solving approach. This finding suggests that novices' improvement actually reflects one (or several) skills that experts have gained through many hours of studying and interacting with mechanical systems. It makes us more confident that we are capturing an important aspect of what constitutes an engineer.

These results are promising, but they are also preliminary: we have only analyzed one task out of six. The main reason for this narrow focus is that the coding schemes took a significant amount of time to develop and apply. Next, we plan to analyze the 5 other tasks and attempt to replicate the findings of this paper. Another limitation of this work is the fact that by limiting our analysis to a single task we were forced to use between subject analyses, which considerably reduced our statistical power. Finally, we acknowledge that our tasks only capture a small portion of what constitutes an expert. The job of mechanical engineers is vastly more complex than rebuilding gear systems and fixing everyday devices.

Even with those limitations, we consider our contribution to be a significant advance in capturing learning in makerspaces. Previous work was mostly limited to qualitative accounts of students' experiences in those spaces or traditional pen and paper questionnaires. We designed semi-authentic engineering tasks, and found preliminary evidence that participating in a maker workshop had a positive impact on participants. Not only did they improve from pre to post, but they also became more similar to mechanical engineers in their actions.

While this approach can be viewed as a new type of assessment, we also view it as a way to gain a more nuanced understanding of the types of cognitive change that are fostered in maker spaces (and potentially other unstructured learning environments). In its current form, our approach provides a starting point for other types of task-based assessments and analyses. We are currently working on a more general framework that provides guidelines for the design of additional tasks and coding schemes. Finally, it is worth mentioning that we also collected all of students' gestures and body postures using a Kinect sensor. Future work includes the analysis of this dataset to extract indicators of expertise that would act as a proxy for the coding schemes we developed in this paper. In the long run, our hope is to automatically collect and process task-independent measures of students' performance using sensors and machine learning algorithms.

## Conclusion

Makerspaces are inherently messy learning environments. In them, students learn a variety of skills: they come up with original project ideas; they address problems in their communities; they learn to overcome failure; they learn to communicate and collaborate with their peers; and finally, they learn to think like engineers. These skills are central to many current (and upcoming) career paths, but it can be difficult to teach them in formal learning environments. Furthermore, it has proven difficult to capture these changes using traditional methods. This paper introduces a new method for capturing these changes through a combination of task-based assessments and qualitative/computational analysis techniques. Using this method, we found that students who took part in a maker workshop became more like engineers in their ability to reason about and solve complex problems. More specifically, the students learned to recognize the functional affordances of complex mechanisms—that wheels are for rotating and gears are for meshing. This shift in perspective is an important, empowering educational outcome, and provides new motivation for studying the educational impact of fostering a maker mindset in youth.

## References

- Blikstein, P. (2013). Digital fabrication and “making” in education: The democratization of invention. *FabLabs: Of Machines, Makers and Inventors*, 1–21.
- Blikstein, P., & Krannich, D. (2013). The makers' movement and FabLabs in education: experiences, technologies, and research. In *Proceedings of the 12th international conference on interaction design and children* (pp. 613–616). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=2485884>
- Dougherty, D. (2013). The maker mindset. *Design, Make, Play: Growing the next Generation of STEM Innovators*, 7–11.
- Fields, D. A., Searle, K. A., Kafai, Y. B., & Min, H. S. (2012). Debuggems to Assess Student Learning in e-Textiles (Abstract Only). In *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education* (pp. 699–699). New York, NY, USA: ACM. <https://doi.org/10.1145/2157136.2157367>
- Gabadinho, A., Ritschard, G., Mueller, N. S., & Studer, M. (2011). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1–37.
- Halverson, E. R., & Sheridan, K. M. (2014). The Maker Movement in Education. *Harvard Educational Review*, 84(4), 495–504.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2016). *Cluster: Cluster Analysis Basics and Extensions*.
- Martin, L. (2015). The Promise of the Maker Movement for Education. *Journal of Pre-College Engineering Education Research (J-PEER)*, 5(1). <https://doi.org/10.7771/2157-9288.1099>
- Papert, S. (1980). *Mindstorms: children, computers, and powerful ideas*. New York: Basic Books.
- Peppler, K., & Bender, S. (2013). Maker movement spreads innovation one project at a time. *Phi Delta Kappan*, 95(3), 22–27.
- Piaget, J. (1973). To understand is to invent: The future of education. Retrieved <http://philpapers.org/rec/PIATUI>
- Tschan, F. (2002). Ideal Cycles of Communication (or Cognitions) in Triads, Dyads, and Individuals. *Small Group Research*, 33(6), 615–643. <https://doi.org/10.1177/1046496402238618>
- Worsley, M., & Blikstein, P. (2013). Towards the development of multimodal action based assessment. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 94–101). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=2460315>

# Dual Gaze as a Proxy for Collaboration in Informal Learning

Kshitij Sharma, Faculty of Business and Economics, University of Lausanne; Computer Human Interaction in Learning and Instruction, EPFL, kshitij.kshitij@unil.ch  
Ioannis Leftheriotis, NTNU, Trondheim, Norway, iolef@acm.org  
Jama Noor, NTNU, Trondheim, Norway, jamawadi@gmail.com  
Michail Giannakos, NTNU, Trondheim, Norway, michailg@idi.ntnu.no

**Abstract:** Interactive displays are increasingly employed in informal learning environments as a technology for enhancing students' learning and engagement. Interactive displays allow students to collaborate and interact with the content in a natural and engaging manner. Despite the increased prevalence of interactive displays for learning, we know very little about how students collaborate in such settings and how this collaboration influences their performance. In this dual eye-tracking study, with 36 participants, a two-staged within-group experiment was conducted to investigate students' collaboration and learning gains in an interactive display. The results show that collaboratively, pairs who have high gaze similarity have high learning outcomes. Individually, participants spending high proportions of time in acquiring the complementary information from images and textual parts of the learning material attain high learning outcomes. We show that the gaze is an effective proxy to cognitive mechanisms underlying collaboration not only in formal settings but also in informal learning scenarios.

**Keywords:** interactive displays, informal learning, collaborative learning, dual eye-tracking

## Introduction

There is growing interest in investigating the use of interactive displays in a plethora of domains, due to their decreasing cost and increasing commodity/availability. Interactive displays have been used for supporting informal learning activities (Klopfer et al., 2005; Dillenbourg & Evans, 2010), among other purposes. However, little evidence exists on how to create and put into practice engaging, efficient, and highly collaborative activities on interactive displays. There have been numerous studies that point out that utilization of interactive displays proved to be an excellent tool to promote collaboration and cooperation while learning (e.g. Schäfer et al., 2013); and since work and interactive displays, such as a tabletop surface, can be considered as a ubiquitous feature of learning, the software that accompanies such displays could be augmented to adapt to various scenarios (Dillenbourg & Evans, 2010).

There is an emerging literature on interactive displays for collaborative interaction and their use in educational settings. In their review, Higgins et al. (2012) propose a typology of features of this technology and offer an analysis of the pedagogic potential of these features. Researchers have to pay attention to two common mistakes that have been repeated each time a new technology is introduced in education: over-generalization and over-expectation (Dillenbourg & Evans, 2010). For instance, in Zaharias et al. (2013) empirical study, researchers assessed the learning performance and user experience between a group that followed the traditional learning procedure in a museum and a group in which students interacted with a multi-touch application dedicated to the museum. Although their results show statistically non-significant differences in the learning performance, the second group reported significantly higher levels of fun and engagement than the first group. Most studies are focused on the experience, fun, enjoyment and engagement of the users (e.g. Schneider et al. 2012; Schäfer et al. 2013). It seems that in order to better understand the way that users learn and collaborate on an interactive displays, further tools and studies are needed.

Based on recent studies regarding collaboration and learning, one important tool that could be used to unveil the cognitive mechanisms underlying collaboration is the dual eye-tracking (DUET). There are studies explaining the expertise (Jermann et. al. 2010), collaboration quality (Sharma et. al 2015), learning outcome (Jermann and Nuesli, 2012), and the task-based performance (Nuessli et. al., 2009) using dual eye-tracking data. However, to the best of our knowledge, DUET has not yet been applied to investigate collaborative gaze patterns in a combination of physical and digital collaborative spaces. This was our main motivation behind this study. In this contribution, we combine the two aforementioned research areas: interactive displays in informal educational settings and DUET for collaborative learning. We designed an experiment where participants were asked to go through a set of posters and play a game (both in collaborative and competitive ways) using content from the area of Neuroscience. We recorded the gaze of the participants while they were watching the posters and while they were playing the game. In this contribution, we focus on how to explain the relation between the

learning gains in an informal learning setting and collaboration, using DUET. Precisely, we address the following research questions:

1. How can the individual and collaborative gaze patterns explain the learning gains?
2. What is the relation between the collaborative gaze patterns in two different contexts of the experiment (physical versus digital)?

## Related work

In this section, we will briefly report on the studies using the interactive displays in education, and the dual eye-tracking studies in the collaborative settings. This section is not exhaustive in terms of the studies reported, but it contains the grounding necessary for this paper.

### Interactive displays in informal educational settings

One of the first studies on touch technology applied to education was “Read-it”, a game-based application, designed to support the development of reading skills in children aged 5-7 years old (Sluis et al., 2004). The results of the pilot experiment showed that children enjoyed playing the game and that the technology was not an obstacle to learning. Different design practices like gamification elements (badges, achievements, points and levels) (Lo et al., 2013) were used in interactive display applications that allow students to engage and collaborate with the application. Schafer et al. (2013) developed a multi-touch learning environment and designed a game that consists of multiple learning and playing modes in which teams of students can collaborate or compete against each other. This multiplayer approach of supporting competition, collaboration and cooperation is perceived as motivating and “fun”. Greater playfulness and enjoyment have been indicated while students were working with the multi-touch display. In a more recent study, Ardito et al., (2013) proposed a new educational format, inspired by the Discovery Learning Technique, which integrates educational games, designed to be played on large multi-touch displays, with other types of formal and informal learning. Ardito et al., (2013) showed that their proposed educational format is effective and that games on these novel multi-touch systems engage users, stimulate collaboration and help consolidating knowledge.

Antle et al., (2011) developed the tabletop game Futura (reported effective and enjoyable by the majority of the general public), with a focus to identify and understand key design factors of importance in creating opportunities for learning. In this study, some special affordances of a multi-touch display are depicted, for instance, the fact that the interface allows all the players to see how and what their co-players are doing. Besides, Watson et al. (2013) suggests that there may be something inherently more appealing about the direct nature of multi-touch interaction, particularly when applied to a game. Kirriemuir and McFarlane (2004) claim that before games can take on a meaningful role in formal or informal education, the education sector and the wider public and media need to better understand the potential and diversity of such ‘tools’. In this study we investigate these ‘tools’ with the use of dual eye-tracking.

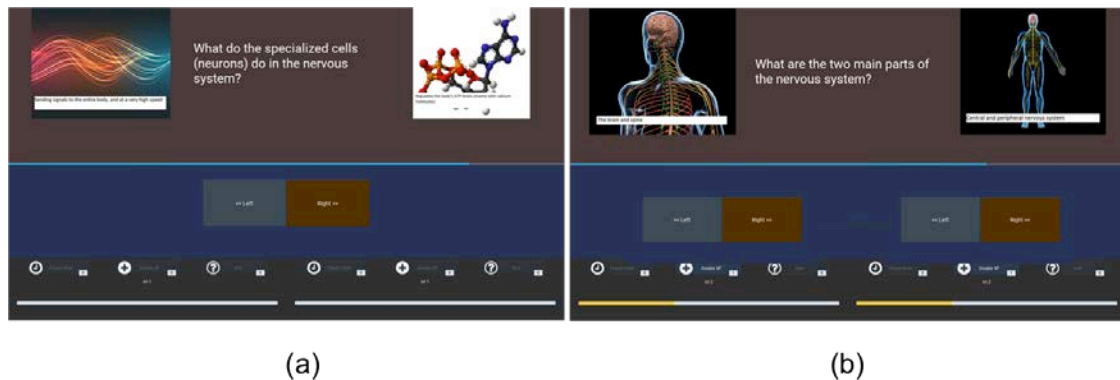
### Dual eye-tracking (DUET)

Previous research has shown the importance of DUET to unveil the cognitive mechanisms underlying collaboration. In a dual eye-tracking study concerning listening comprehension, Richardson and Dale (2007) showed that the pairs having high cross-recurrence (probability to look at the same object at the same time) have high comprehension results. Jermann and Nuessli (2012) showed similar results in a pair program comprehension tasks. The pairs which had high cross-recurrence also had high understanding levels (Jerman & Nuessli, 2012). Nuessli et al. (2009) used DUET data to predict the task-based success in Raven matrices and Bongard problems. The authors used gaze density and fixation dispersion of the pair to predict the success of the pair with an accuracy of 78%. In a collaborative version of Tetris, Jermann et al. (2010) predicted the pair configuration (expert-novice, novice-novice, expert-expert) using the DUET data with an accuracy of 75%. Sangin et al. (2008) used a Knowledge Awareness Tool (KAT) to inform the peers about their partner’s knowledge in a collaborative concept map task. The results show that the gaze on the KAT was correlated to the relative learning gain of the pairs. In a pair-programming task, Sharma et al. (2011) showed that the pairs with a high level of understanding look at the data flow of the program while the pairs with a low level of understanding read the program as if it was English text. In another DUET experiment with collaborative concept map, Sharma et al (2015) showed that the gaze similarity (the probability to look at the same set of objects in the same time-window) is higher for the pairs with a high collaboration quality than that for the pairs with a low collaboration quality.

All the aforementioned studies show that using dual eye-tracking data, one can explain the expertise, collaboration quality and the task-based performance. In the present study, we utilize dual eye-tracking to explain pairs’ learning gains in an informal learning context.

## Experiment

In this section we will present the details of the experiment and the variables involved in the analysis. In the experiment, the participants were asked to fill in a pretest about the content they were going to learn. Then they went through five posters about the structure of neurons, different areas of the brain and their functions, three neurological disorders, and the limbic system (two examples are shown in Figures 3.a and 3.b). The next phase was the first individual posttest. The next phase was a gamified quiz application played in an interactive display and focusing on the same content as the posters. The game had two modalities: in one modality the team played collaboratively while in the second, the members of a team played against each other (the interfaces for the collaborative and the competitive versions of the games are shown in Figures 1.a and 1.b respectively). The order of the game modalities was balanced among the teams. Finally, the participants individually took a second posttest. All the tests and the quizzes in the games were multiple-choice questionnaires. The poster phase was 12-15 minutes long and each modality of the game (collaborative/competitive) was 6-7 minutes long. The gaze of the participants was recorded during the poster phase and the game phase using SMI and TOBII eye-tracking glasses at 60 Hz.



**Figure 1.** Interfaces for the (a) collaborative, and (b) competitive versions of the game. The elements for the two versions were identical except in the collaborative version there was only one set of buttons to select the correct answer, while in the competitive version there were two different sets of buttons for each player.



**Figure 2.** Experiment screenshots, (a) Teams are watching the posters as they would do in a museum, and (b) Teams are playing gamified quizzes (collaborative/competitive) on interactive display.

## Participants and procedure

There were 36 university students (18 randomly formed dyads), who participated in the experiment; there were 13 females among the participants. The average age was 24.4 years ( $SD = 5$  years). Upon their arrival in the lab, they filled in a pretest about the poster content; afterwards, they watched the five posters. The simple instruction for the poster phase was “go through all the posters as if you were visiting a museum with your friend”. The participants were allowed to discuss the content of the posters with their partners (figure 2.a). The dyads were not mandated to stick to each other, however, most of them went through the posters together. Once they finished watching and discussing the posters, the participants individually filled the first posttest. Further, they played the gamified quiz (collaborative/competitive) where they received one of the three power ups for each correct answer: “double xp”, “pause time”, and “hint” (figure 2.b). During the game phase, the participants had

a maximum of 30 seconds to answer each question; they were allowed to go back to the posters and look for the answers. Once they finished both the modalities of the game, they filled in a final second posttest. All the participants were rewarded an equivalent of \$10.

### Dependent variables

We normalized test scores to be between 0 and 1. The two dependent variables are the scores in the first and the second posttests. We do not consider the learning gain in this experiment, as we observed a floor effect on the pretest scores (Mean = 0.2, Median = 0.1).

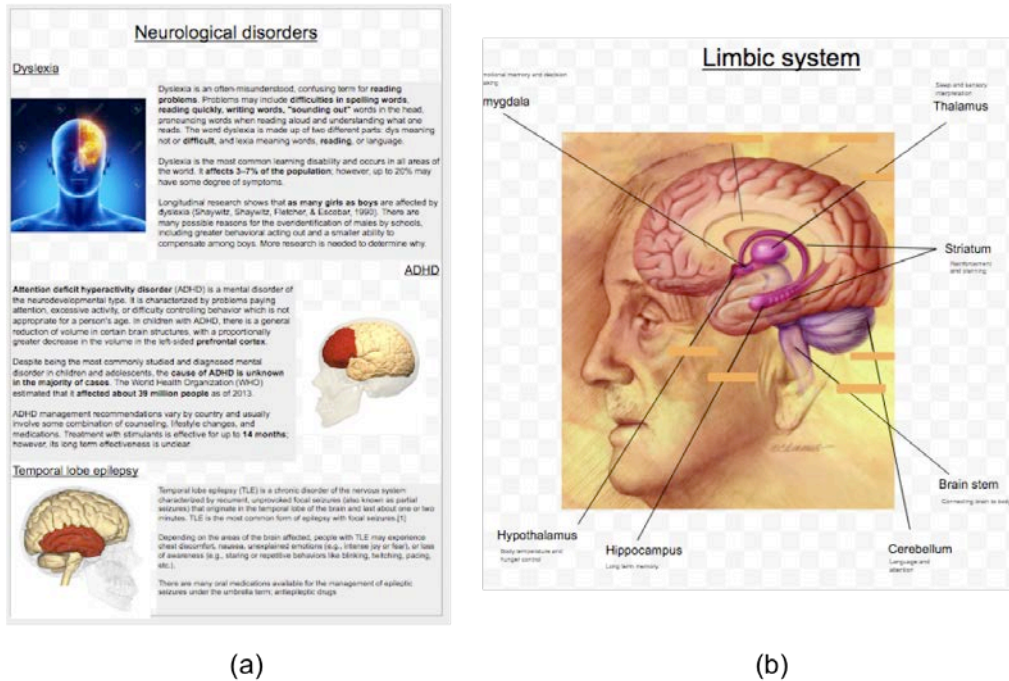


Figure 3. Examples of the posters, (a) Neurological Disorders, and (b) The Limbic System.

### Process Variables

**Performance in the game:** The participants received individual experience points (xp) during the game (both collaborative and competitive) when they replied correctly to a question. We consider the xp value as their game performance index.

**Individual Gaze - AOI Transitions:** We divided the posters into different Areas of Interests (AOIs), for example, text blocks and image blocks. Next, we computed the proportions of the gaze transitions from the images in the poster to the corresponding text and also the proportions of the gaze transitions from the text to the corresponding images. For example, in the Figure 3(a), an image to text transition would be a shift of gaze from the top left image to any of the first three paragraphs and the opposite for the text to image transitions. Any transition from the top-left image to any paragraph other than the first three ones would not be counted as a valid image to text transition. The opposite is also true for the text to image transitions.

**Collaborative Gaze - Gaze Similarity:** To compute the metric for the collaborative gaze patterns, we used the same measure as used by Sharma et al., (2012 and 2015). This measure is called the gaze-similarity and is computed as the cosine similarity of the proportionality gaze vector. The proportionality gaze vector is the vector denoting the proportion of the time spent by each participant looking at the different elements of the visual stimulus for a small window of time (in our case 10 seconds). A gaze similarity value of 1 will depict that the two peers spent exactly the same proportions of 10 seconds on different AOIs. Whereas, a gaze similarity value of 0 will depict that the two peers were looking at completely different elements during a given time window of 10 seconds.

### Results

In this section, we will present the various relations we find among the process and dependent variables. We observe the following relations:

**Improvement in the knowledge from pretest to the first posttest:** we observe a significant improvement from the pretest score to the first posttest score for all the participants ( $t(69.96) = -7.91, p < .0001$ ). The scores in the first posttest are significantly higher than the pretest scores (Figure 4).

**No improvement in the knowledge from the first to second posttest:** however, we do not observe significant improvement from the first to second posttest ( $t(69.88) = 0.39, p > .05$ ). The scores in the two posttests were similar for almost all the participants (Figure 4).

**Correlation between the game score, the first and the second posttest:** we observe three significant correlations: 1. The scores in the first and the second posttests are correlated ( $r(34) = 0.69, p < .0001$ ). The participants who score high in the first posttest also score high in the second posttest (Figure 5.a). 2. The score in the game is correlated to the score in the first posttest ( $r(34) = 0.42, p = .01$ ). The participants who score high in the first posttest also perform well in the game (Figure 5.b). 3. The score in the game is correlated to the score in the second posttest ( $r(34) = 0.34, p = .04$ ). The participants who perform well in the game also score high in the second posttest.

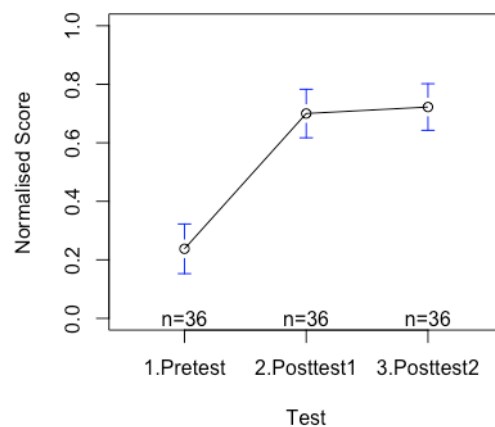


Figure 4. Comparison of scores from the pretest, the first and the second posttest. All values are normalized between 0 and 1. The points show the mean values among all the participants and the blue bars show the 95% confidence intervals.

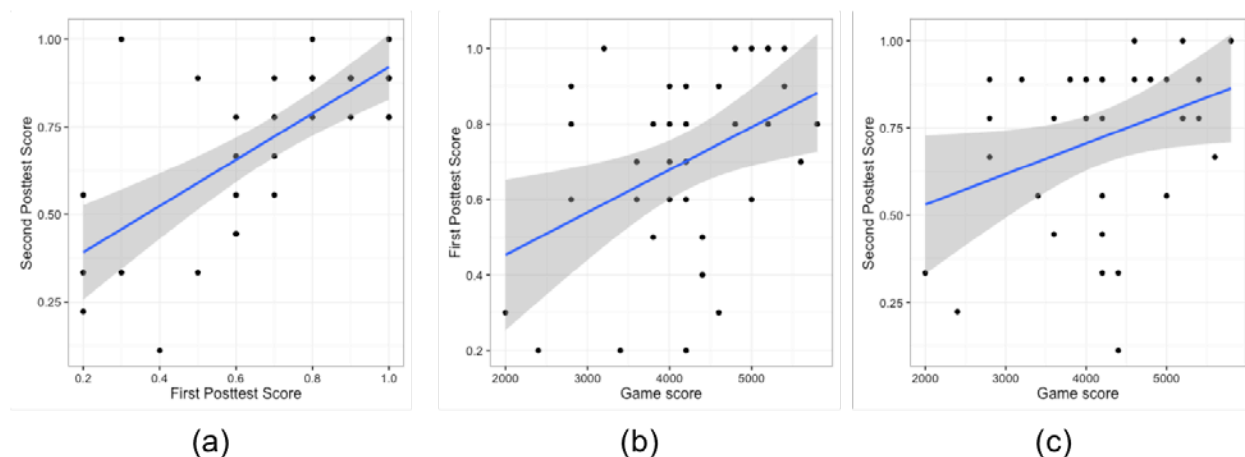
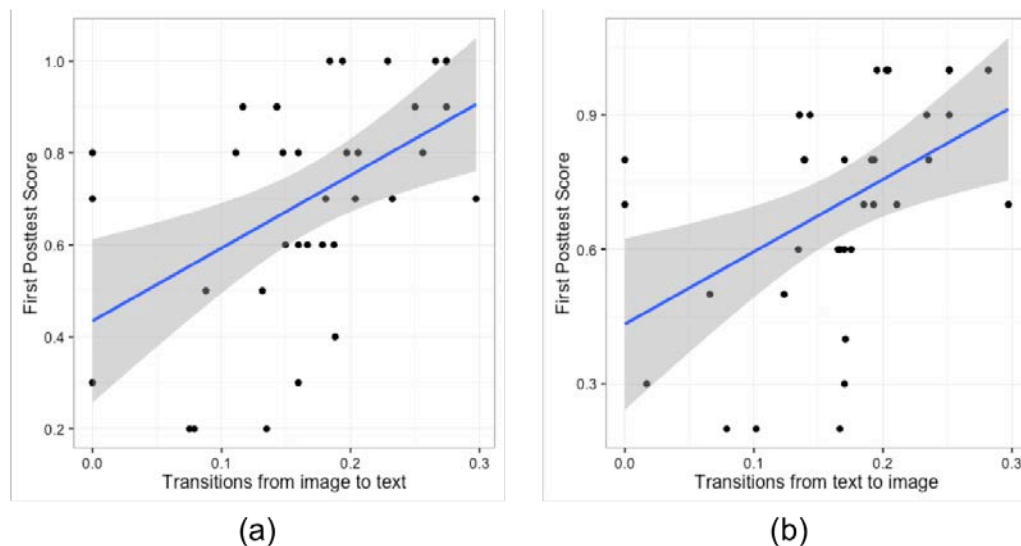


Figure 5. Scatter Plots for (a) first and second posttest score, (b) game score and the first posttest score, and (c) game score and the second posttest score. In all the plots the blue line shows the linear model for the variable on y-axis given the variable on the x-axis. The grey area shows the 95% confidence interval.

**AOI transitions and the first pretest score:** next, we consider the individual gaze patterns during the poster phase. We observe a significant correlation between the transitions from image to text and the first posttest score ( $r(34) = 0.47, p = .003$ ). The participant having high proportion of the image to text transitions, score high in the first posttest (Figure 6.a). Moreover, we also observe a significant correlation between the transitions from the text to image and the first pretest score ( $r(34) = 0.48, p < .002$ ). Participants that have high proportion of the text to image transitions score high in the first posttest (Figure 6.b).

**Gaze similarity and the first pretest score:** further, concerning the collaborative gaze patterns, we observe a significant correlation between the gaze similarity during the poster phase and the first posttest score ( $r(16) = 0.51, p = .03$ ). The pairs having high gaze similarity have high average first posttest score (Figure 7.a).

**Gaze similarity during posters and during the game:** moreover, we also observe a significant correlation between the gaze similarity during the poster phase and the gaze similarity during the game phase ( $r(16) = 0.49, p < .04$ ). The pairs having high gaze similarity during the poster session also have high gaze similarity during the game phase (Figure 7.b).



**Figure 6.** Scatter Plots for the individual transitions (a) image to text, and (b) text to image; and the score in the first posttest. In all the plots the blue line shows the linear model for the variable on y-axis given the variable on the x-axis. The grey area shows the 95% confidence interval.

## Discussion and conclusions

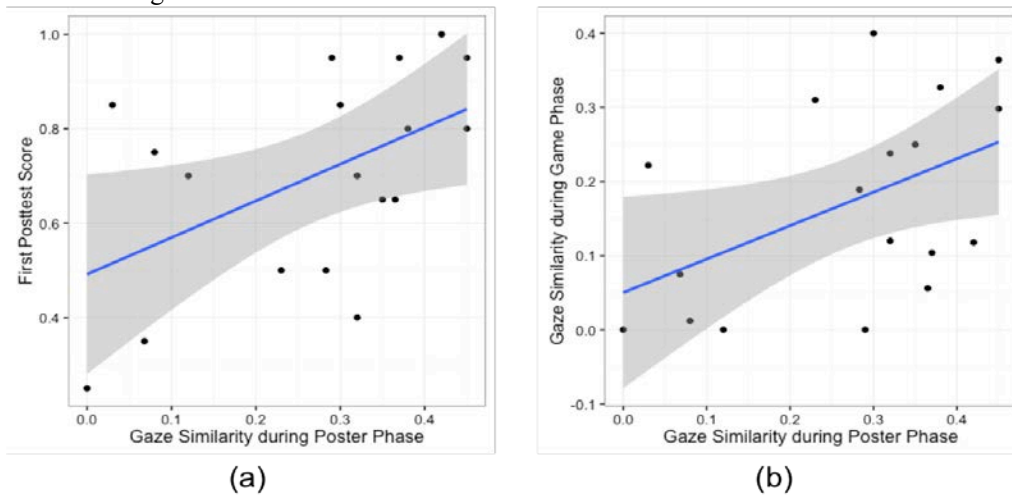
The results presented in the previous section represent two behavioural gaze patterns (individual AOI transitions and collaborative gaze similarity), both of which are correlated to participants' learning outcomes (Question 1).

The first behavioural gaze patterns we report on are the individual transitions to and from the images and the text chunks. The results show that the participants having more such transitions have a higher score in the first posttest than those who have fewer images to text and text to image transitions. One plausible explanation for this relation is that the image to text and text to image transitions translate to the behaviour of combining the information present in the images and the text. The complementarity of the two information sources is necessary to understand the content. Those participants who understood the relation between the two information sources got high first posttest scores. This result is inline with an eye-tracking study by Shrama et al. (2015) where the authors showed that the understanding the complementarity of the graphical and textual elements in a video lecture was a key process to attain high learning gain.

The second behavioural gaze patterns we report on are the collaborative gaze patterns. The gaze similarity denotes the proportion of time the peers spent looking at the same set of objects within the given time window. While looking at the same areas, the peers reflect together on the content of the posters and thus build upon a mutual understanding about the learning material; hence attain higher learning outcomes than the pairs who have lower gaze similarity. This result was also found by Richardson and Dale (2005), Jermann and Nuessli (2012) and Sharma et al. (2013 and 2015). In different contexts, these contributions have shown that the



gaze similarity (or gaze cross-recurrence) is correlated to the task-based performance and/or learning outcomes in collaborative settings.



**Figure 7.** Scatter Plots for (a) collaborative gaze similarity during the poster phase and the first posttest score, and (b) collaborative gaze similarity during the poster phase and the game phase. In all the plots the blue line shows the linear model for the variable on y-axis given the variable on the x-axis. The grey area shows the 95% confidence interval.

The fact that there is no significant improvement from the first posttest scores to the second posttest scores has its roots in the level of scores the participants attained in the first posttest (Mean = 0.70, SD = 0.24). This leaves a small room for improvement in the second posttest. However, we see a slight (statistically non-significant) improvement in the second posttest (Mean = 0.72, SD = 0.23). We also see a correlation between first and second posttest scores, suggesting that the interactive application did not hinder the learning process. This is inline with the results found by Sluis et. al (2004) and Zaharias et al. (2013) who also found that interactive displays were not an obstacle for the learning processes. We also found that the learning outcome was correlated with task-based performance (game xp). This is also inline with the findings of Sangin (2009) who found that task-based performance was correlated to learning gains in a collaborative concept-map task.

Finally, considering the relation between collaborative gaze patterns during the two experimental phases (Question 2), we find a positively significant correlation (Figure 7.b) between the gaze similarity during the poster phase (physical) and the gaze similarity between the game phase (digital). The two phases were quite different from each other: in the poster phase the participants collaborated voluntarily while in the game phase they were told to collaborate. Despite this fact, pairs with high gaze similarity during the poster phase, also have high gaze similarity during the game phase. This supports the interaction style hypothesis of Sharma et. al. (2015), which states that there are two different interaction styles in collaborative settings: “Looking AT” and “Looking THROUGH”. The former appears when the peers are interacting with the content only, while the later appears when the peers are using the content as a medium to interact with their partners. In the present study, we find the same two profiles: the pairs having low gaze similarity during both the poster and game phases appear to interact with the content only (looking AT), while the pairs having high gaze similarity during both the phases appear to use the content as the medium to interact with their partners (looking THROUGH).

In this contribution, we showed that there are individual and collaborative gaze patterns, which can explain the learning outcomes of the participants in a collaborative informal educational setting. These explanations are coherent with studies conducted in more formal educational settings. This motivates us to interlace these findings with dialogues and actions on the touch technology and interactions to understand more about pairs’ collaborative dynamics. Moreover, these results will also lead us to design more hands-on activities with interactive displays within the informal settings to study their influence on the learning outcomes.

## References

- Antle, A. N., Bevans, A., Tanenbaum, J., Seaborn, K., Wang, S. (2011). Futura: design for collaborative learning and game play on a multi-touch digital tabletop. In *Procs. of the fifth international conference on Tangible, embedded, and embodied interaction*, ACM, 93-100.
- Ardito, C., Lanzilotti, R., Costabile, M. F., & Desolda, G. (2013). Integrating traditional learning and games on large displays: an experimental study. *Journal of Educational Technology & Society*, 16(1), 44-56.

- Dillenbourg, P., & Evans, M. (2011). Interactive tabletops in education. *International Journal of Computer-Supported Collaborative Learning*, 6(4), 491-514.
- Higgins, S., Mercier, E., Burd, L., & Joyce-Gibbons, A. (2012). Multi-touch tables and collaborative learning. *British Journal of Educational Technology*, 43(6), 1041-1054.
- Jermann, P., Nüssli, M.-A. and Li, W. (2010) Using dual eye-tracking to unveil coordination and expertise in collaborative tetris. In *Procs. of the 24th BCS Interaction Specialist Group Conference*, pages 36–44. British Computer Society.
- Jermann, P., and Nüssli, M.-A. (2012) Effects of sharing text selections on gaze cross- recurrence and interaction quality in a pair programming task. In *Procs. of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 1125–1134. ACM.
- Kirriemuir, J., & McFarlane, A. (2004). Literature review in games and learning. A NESTA Futurelab Research report – report 8.
- Klopfer, E., Perry, J., Squire, K., Jan, M. F., & Steinkuehler, C. (2005). Mystery at the museum: a collaborative game for museum education. In *Procs. of the conference on Computer support for collaborative learning*. 316-320.
- Lo L.-J., Chiang C.-D., & Liang R.-H. (2013) HexDeck: Gamification of Tangibles for Brainstorming. Consilience and Innovation in Design – In *Procs. of the 5th IASDR 2013 Tokyo*. 3165-3175.
- Nüssli, M.-A., Jermann, P., Sangin, M., and Dillenbourg, P. (2009) Collaboration and abstract representations: towards predictive models based on raw speech and eye- tracking data. In *Procs. of the 9th international conference on Computer supported collaborative learning-Volume 1*, pages 78–82. International Society of the Learning Sciences.
- Richardson, D. C., and Dale, R (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science*, 29(6):1045–1060.
- Sangin, M., Molinari, G., Nüssli, M.-A., and Dillenbourg, P. (2008) How learners use awareness cues about their peer's knowledge: insights from synchronized eye-tracking data. In *Procs. of the 8th international conference on International conference for the learning sciences-Volume 2*, pages 287–294. International Society of the Learning Sciences.
- Schäfer, A., Holz, J., Leonhardt, T., Schroeder, U., Brauner, P., & Ziefle, M. (2013). From boring to scoring—a collaborative serious game for learning and practicing mathematical logic for computer science education. *Computer Science Education*, 23(2), 87-111.
- Schneider, B., Strait, M., Muller, L., Elfenbein, S., Shaer, O., & Shen, C. (2012). Phylo-Genie: engaging students in collaborative 'tree-thinking' through tabletop techniques. In *Procs. of the SIGCHI Conference on Human Factors in Computing Systems*. 3071-3080.
- Sharma, K., Jermann, P., Nüssli, M. A., & Dillenbourg, P. (2012). Gaze Evidence for different activities in program understanding. In *24th Annual conference of Psychology of Programming Interest Group*
- Sharma, K., Jermann, P., Nüssli, M. A., & Dillenbourg, P. (2013). Understanding collaborative program comprehension: Interlacing gaze and dialogues. In *Computer Supported Collaborative Learning*.
- Sharma, K., Caballero, D., Verma, H., Jermann, P., & Dillenbourg, P. (2015). Looking AT versus Looking THROUGH: A Dual Eye-Tracking Study in MOOC Context. In *Procs. of 11th International Conference of Computer Supported Collaborative Learning, Gothenburg, Sweden, CSCL*.
- Sharma, K., Caballero, D., Verma, H., Jermann, P., & Dillenbourg, P. (2015). Shaping learners' attention in Massive Open Online Courses. *International Journal of Technology in Higher Education*. 12(1-2).
- Sluis, R.J.W., Weevers, I., Schijndel, C.H.G.J.V., Kolos-Mazuryk, L., Fitriani, S., & Mar-tens, J. B. O. S., (2004). Read-It: Five-to-seven-year-old Children Learn to Read in a Tabletop Environment, In *Procs. of International Conference on Interaction Design with Children, ACM*, 73-80.
- Watson, D., Hancock, M., Mandryk, R. L., & Birk, M. (2013). Deconstructing the touch experience. In *Procs. of the ACM international conference on Interactive tabletops and surfaces*. 199-208.
- Zaharias, P., Michael, D., & Chrysanthou, Y. (2013). Learning through multi-touch interfaces in museum exhibits: An empirical investigation. *Journal of Educational Technology & Society*, 16(3), 374-384.

## Acknowledgements

The project has been recommended by the Data Protection Official for Research, Norwegian Social Science Data Services (NSD), following all the regulations and recommendations for research with children. This work was funded by the Norwegian Research Council under the projects FUTURE LEARNING (number: 255129/H20), the Centre for Excellent IT Education (ExcITeD <http://www.ntnu.edu/excited>) and by the GEMINI Center in Technology Enhanced Lifelong Learning (<http://www.tellgemini.org/>).

# A Mixed-Methods Approach for Studying Collaborative Learning Processes at Individual and Group Levels

Catherine Dornfeld, Naxin Zhao, and Sadhana Puntambekar  
cldornfeld@wisc.edu; nzhao23@wisc.edu; puntambekar@education.wisc.edu  
University of Wisconsin–Madison

**Abstract:** Learning processes that unfold during small-group collaboration may impact conceptual outcomes for individual students. To study how learning processes unfolded for eighth grade students collaborating in an e-textbook research activity, we analyzed data sources at individual and group levels using multiple methods, including nonparametric tests, text mining, Markov modeling, and quantitative discourse analysis. Individual measures revealed learning gains on content tests and documentation of shared ideas during collaboration. Group measures revealed increased conceptual discourse over time and streamlining of the research process. Measures at each level indicated distinct paths of inquiry for students and groups; however, these differences were not associated with negative conceptual outcomes. These findings have implications for how we understand how collaborative learning processes unfold, especially between the individual and the group, and how we design support for collaboration and knowledge sharing.

## Introduction

CSCL researchers aim to understand both products and processes of learning (Stahl, Koschmann, & Suthers, 2006; Reimann, 2007). Products reveal the relative success of an implementation, but processes reveal the actual learning mechanisms that occur in collaborative settings (Dillenbourg, 1999). To understand collaborative learning processes, we must examine learning at both the individual and group level (Stahl et al., 2006; Dillenbourg, 1999). Individual students contribute unique experiences and prior knowledge, while the group co-constructs knowledge through negotiating and revising shared understandings (Wertsch, 1984).

Understanding learning processes at multiple levels may require mixed-methods approaches, such as quantitative discourse analysis and computational methods (Strijbos & Fischer, 2007; Puntambekar, 2013; Li, Wang, Liao, Zhao, & Huang, 2007). Mixed-methods approaches allow us to combine complementary perspectives and data sources that may triangulate findings about collaborative learning, such as how individuals in groups move toward conceptual convergence (Roschelle, 1992; Kapur, Voiklis, & Kinzer, 2011). We also capture variance in groups' learning processes, which may impact individual students' conceptual outcomes (Barron, 2003). However, with any mixed-methods approach for multiple levels, there are a few major issues. First, in triangulating findings, we must reconcile multiple data sources and identify how each source reveals new understandings of collaborative learning processes (Suthers & Medina, 2011). Also, we must carefully interpret patterns and relationships between data sources (Lajoie, 2011). Finally, we must consider the quality of our data, as poor quality of data can lead to spurious findings (Reimann, Yacef, & Kay, 2011).

In this study, we used a mixed-methods approach incorporating methods from quantitative discourse analysis and learning analytics to understand how learning processes unfolded at both the individual and group levels as students used various forms of support integrated into a curriculum. We analyzed multiple data sources that reflected learning at the beginning and end of an eight-week biology curriculum in order to understand variance between groups and over time. Our research question was: *How do multiple data sources, in combination, reveal how learning processes unfold at both the individual and group levels?* This question has implications for how we understand learning processes using mixed methods for multiple data sources and granularities.

## Methods

Here we describe our mixed-methods approach to understanding how different data sources and analyses revealed learning processes at individual and group levels. First, we describe participants and context, then describe the data and connected analyses.

### Participants and context

This study focused on three groups of eight-grade students (Groups A, B, and C) working in groups of four ( $N = 12$ ) in the same classroom at a semi-rural public school in the Midwestern U.S. The majority of students

attending this school were Caucasian, and over half of students were eligible for free or reduced-price lunch. Students in the three groups demonstrated similar prior knowledge as assessed on a pre-test.

Students participated in a *Make Your Own Compost* unit. In this eight-week design-based unit, students addressed a challenge about compost and ecosystems. The unit incorporated embedded distributed scaffolds to support students' inquiry, such as physical and virtual experiments; small group collaboration; teacher-led whole class discussions; e-textbook (VidyaMap) research sessions; and scientists' journals for tracking design decisions and relevant content. Physical and virtual experiments supported students' modeling of an authentic problem. Small group collaboration elicited students' current understanding and supported co-construction of knowledge. Whole class discussions facilitated idea sharing between groups and revealed opportunities for teacher support. VidyaMap helped students explore design-relevant content. Lastly, journals included prompts for different aspects of design, such as documenting relevant content.

Several scaffolds intersected during VidyaMap research sessions. These sessions were designed to inform students' design decisions by providing relevant content. In these sessions, students engaged in small-group collaboration to brainstorm research topics; worked in pairs to research topics in VidyaMap using Chromebooks; and recorded individual notes about findings and whole-class ideas in their journals. Here, we emphasize the close interplay between the journal prompts and VidyaMap as scaffolds for learning.

We investigated how each group of students used VidyaMap to research topics at the beginning and end of the unit using journal responses and VidyaMap log data. We also investigated how individual journal responses differed, based on prompts. These prompts guided students to i) brainstorm questions and topics in groups, ii) take notes about research on VidyaMap, and iii) share and record new ideas. We also analyzed group discourse during VidyaMap sessions and content test scores as assessments of conceptual outcomes.

## Data and analyses

Here we describe the data sources involved in VidyaMap research sessions and conceptual measures, along with analyses for each data source. Table 1 shows analyses for each unit of analysis. Findings for each analysis are reported in the Results section.

Table 1: Approaches for each unit of analysis

Unit of Analysis	Analytical Approach for Data Source
Individual	<ul style="list-style-type: none"> <li>• Nonparametric tests of content test scores</li> <li>• Topic modeling of journals</li> </ul>
Group	<ul style="list-style-type: none"> <li>• Markov models of VidyaMap log data</li> <li>• Small-group discourse analysis</li> </ul>

### Nonparametric tests of content test scores

To assess learning products at the individual level, we used Wilcoxon Signed-Rank tests to compare students' scores on pre- and post-unit content tests (see Results). Content test focused on concepts related to ecosystems (e.g., biotic and abiotic factors, human impacts, roles and relationships, and cycling of energy and matter). The maximum possible score was 38.5 points, with questions being worth 0.5-3 points depending on question complexity. We also used independent-sample Kruskal-Wallis tests to test for differences between groups at both times. While test scores can demonstrate conceptual outcomes, we considered that the scores do not reveal learning *processes* during the unit. As such, we explored additional analyses to better understand how learning processes unfolded for each group.

### Topic modeling of students' journal responses

To assess learning processes at the individual level, we analyzed journal responses associated with the first and last VidyaMap research sessions. The first session (2 days) focused on decomposition factors, while the last session (1 day) focused on ecosystems. Students collaboratively decided on topics for research, but each student recorded questions and notes in their own journal. The journal acted as an artifact of individual learning in that students chose what to record, but it also reflected collaborative discussion of topics – thus showing learning at the *intersection* of the individual and the group.

We used topic modeling to understand how individual students' responses overlapped during group collaboration with VidyaMap. We transcribed and categorized responses within each research session based on the journal prompts. Prompts guided students to i) brainstorm questions and topics, ii) research and take notes

on topics, and iii) share and record new ideas during whole class discussions. We selected responses from the first session (4 responses) and the last session (3 responses) for a total of seven responses per student. We transcribed responses from 159 journals for a total data set of 1113 responses.

We programmed the analysis using Python 2.7 and the *NLTK*, *pickle*, and *gensim* packages. We chose the *term frequency-inverse document frequency* (tf-idf) algorithm because it identifies words that uniquely characterize each individual response relative to the whole data set (Witten, Frank, & Hall, 2011). These characteristic words can be considered relatively rare in that they discriminate that response from others (Witten et al., 2011). We manually identified co-occurrences of these relatively rare words within the three groups to examine how individual responses reflected shared topics of discussion (and potentially knowledge co-construction). This topic modeling procedure revealed overlap in how individual students recorded ideas from their group discussions. However, to understand *how* these ideas were discussed, we need to analyze log data from VidyaMap and collaborative discourse during small-group work.

### Markov models of log data

Students' journals indicated how individual students recorded research during group collaboration with VidyaMap. We further investigated students' use of VidyaMap at the group level by analyzing groups' log data from the first and last research sessions. These log data reveal how groups coordinated research activities and how pairs of students within each group navigated through VidyaMap. To clean data, we removed records that involved superficial reading (<10 seconds) unless these records were the first topic in a session or acted as the only connection between prior or subsequent topics. We also removed records that were the only instance for that session. Lastly, we combined data for individual logs with the same name and method of access.

We manually identified concepts that were unique to each group for each session and calculated the number of concepts and time per session. We used Markov models to visualize patterns in how students navigated through VidyaMap. Markov models quantify navigation by showing the probability of moving from one concept to the next, such as from "Compost" to "Ecosystem," within a series of records (Witten et al., 2011). Markov models can also reveal snapshots of groups' activity during each session. For each group, we generated Markov models of their VidyaMap activity for the first and last sessions. We programmed the models using Python 2.7 and *NetworkX* and *matplotlib* packages. While the Markov models revealed how groups researched concepts in VidyaMap, they did not reveal how students discussed these concepts within each session. Thus, we examined students' group discourse during collaborative research.

### Discourse analysis of small-group talk

To examine how groups discussed concepts in VidyaMap, we investigated group discourse during the first and last VidyaMap sessions. We coded students' turns of talk for i) conceptual talk, which identified concepts and relationships; ii) procedural talk, which indicated collaborative decisions without explaining concepts or relationships; iii) off-task talk, or (iv) N/A for unclear talk (Cohen's  $\kappa = 0.904$ ; see Dornfeld & Puntambekar, 2016). After coding, we calculated the frequency and proportion of each code within each group's discourse. We used z-score tests of homogeneity to identify significant differences in proportions of talk.

## **Results**

### Nonparametric tests of content test scores

Table 2 shows summary statistics for groups' scores. Wilcoxon Signed Rank tests indicated that all students did significantly better on the post-test ( $W\text{-value} = 1 < 13, p < 0.05$ ). To check for group differences, we used Kruskal-Wallis tests to compare mean scores. These tests indicated that groups' scores were not significantly different for the pre-test ( $H = 2.202 < 5.692, p > 0.05$ ) or post-test ( $H = 2.375 < 5.692, p > 0.05$ ). Students appeared to have similar prior knowledge and learning gains. While this indicates the unit supported learning for all students, we added analyses to examine students' *learning processes*.

Table 2: Summary statistics for pre- and post-tests

Group	N	Pre-test			Post-test		
		Mean	SD	Variance	Mean	SD	Variance
Group A	4	29.44	4.49	20.18	32.31	3.40	11.56
Group B	4	30.06	3.24	10.52	32.69	2.64	6.97
Group C	4	26.56	2.17	4.72	30.31	1.21	1.47
Overall	12	28.69	3.49	12.19	31.77	2.58	6.64

### Topic modeling of students' journals

Topic modeling revealed relatively rare words within each student's response that discriminated that response against others in the data. Some relatively rare words overlapped for individuals within groups, indicating shared topics of discussion. In Table 3, we list co-occurrences of relatively rare words and their frequency. The most frequent co-occurrences across all groups were *effect/affect* (8 responses), *light* (7 responses), *decomposer* (5 responses), *worms* (4 responses), *temperature* (4 responses), and *helps* (4 responses). Group A showed the least overlap (14 co-occurrences), while Groups B and C showed twice as much overlap (35 and 31 co-occurrences, respectively). Overlap was more frequent during the first session (60 co-occurrences) than the last (20 co-occurrences). Figure 1 also shows that overlap was also more frequent during brainstorming sessions (53 co-occurrences) than research sessions (17 co-occurrences) or whole class discussions (10 co-occurrences).

To summarize, topic modeling revealed how relatively rare words that characterized individual responses (per the tf-idf algorithm) overlapped within each group as shared topics of discussion. We found evidence of overlap within each group; however, Groups B and C demonstrated greater overlap than Group A. While we also found overlap between groups, we found that each group investigated unique topics that other groups did not. Lastly, we see that all groups demonstrated less overlap in the last session compared to the first. To triangulate these patterns at the group level, we next examined the VidyAMap log data for each group.

Table 3: Relatively rare word co-occurrences for each group (frequencies in parentheses)

Group	First Session			Last Session		
	Brainstorm	Research	Share	Brainstorm	Research	Share
Group A	Fruit (2) Light (2) Helps (2) Makes (2) Affect/effect (2)	Begin (2)	Affect/ effect (2)	---	---	---
Group B	Moisture (2) Warmer (2) Light (3) Environment (2) Temperature (4) Lower (2) Higher (2) Worms (4)	---	Need (2) Degrees (2)	Flow (3) Together (2)	Plants (2)	Work (2)
Group C	Decomposer (5) Affect/effect (4) Helps (2) Light (2) Factors (2)	Slow (2) Levels (2)	Everything (2)	Life (2)	Depend (2) Components (2) Characteristics (2) Released (2)	---

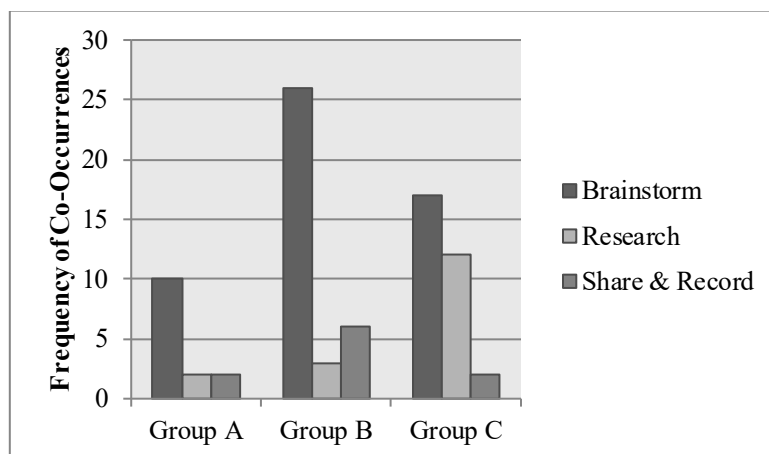


Figure 1. Frequency of word co-occurrences within each group.

### Markov models of log data

Analysis of students' journals revealed that each group focused on particular topics of discussion. To triangulate this, we investigated log data from the first and last VidyaMap sessions. The log data revealed that groups read about similar concepts during each session, such as *compost*, *temperature*, and *decomposer*. This makes sense given that groups received the same prompts about decomposition and ecosystems. However, each group's log data also revealed concepts unique to that group. Table 4 lists these concepts.

We found that Group B investigated more unique concepts (12 topics) than Groups A and C (3 and 4 concepts, respectively). On average, Group B investigated more concepts per session (6.7 concepts) than Groups A or C (4.8 and 5.4 concepts, respectively). Group B also spent more time researching (11.9 minutes per session) than Groups A or C (11.2 and 10.7 minutes, respectively). We found that all groups spent less time researching during the last session compared to the first, with a mean session time of 6.2 minutes for the last session and 13.5 minutes for the first.

Table 4: Unique VidyaMap topics in the log data

<b>Session</b>	<b>Group A</b>	<b>Group B</b>	<b>Group C</b>
First Session, Part 1 (Decomposition)	---	Soil Carbon Cycle Water Ecosystems	Nitrogen Consumers Biotic Factors
First Session, Part 2 (Decomposition)	Food Web Biodiversity	Energy Transformation Producers Leaves Stomata Roots Chloroplast	Nitrogen Cycle
Last Session (Ecosystems)	Soil	Food Web Abiotic Factors	---

In Table 5 (next page), we present Markov models that show how students navigated between concepts during each session. These models serve as snapshots of groups' VidyaMap activity that show the probabilities of moving from one concept to the next. Markov models can reveal if students engage in similar or different inquiry, both in terms of content and navigation. We found that the models for the first session were relatively complex compared to the second session. For example, Group C transitions from researching many concepts in the first research session to researching a focused trajectory of concepts during the second session. This decrease in model complexity aligns with the decrease in average session time for all groups. However, these models do not reveal if less model complexity and time spent researching with VidyaMap imply less conceptual discussion of key ideas. Therefore, to understand the focus of students' collaborative discussions with VidyaMap, we used discourse analysis to examine conceptual, procedural, and off-task talk.

### Discourse analysis of small-group talk

Discourse analysis revealed that conceptual discourse significantly increased for Groups B and C from the first to last session ( $z = 3.35, p < 0.001$ ;  $z = 6.89, p < 0.001$ , respectively). In contrast, procedural discourse significantly decreased over time for all groups ( $z = 3.09, p = 0.002$ ;  $z = 4.34, p < 0.001$ ;  $z = 8.59, p < 0.001$ ). Off-task talk did not significantly change over time. Figure 2 (next page) shows these differences. Groups A and B engaged in mostly off-task talk (40.3% and 48.9%, respectively), followed by conceptual talk (35.5% and 32.1%) and procedural talk (24.2% and 19.0%). Group C engaged in mostly conceptual talk (47.3%), followed by procedural talk (35.3%) and off-task talk (17.3%).

### Summary of results

Nonparametric tests revealed that students showed conceptual gains on the content post-test. Topic modeling revealed that students working in groups showed overlap in their recorded ideas during VidyaMap research sessions. Markov models of VidyaMap log data also showed overlap in concepts between groups for some concepts, such as *compost* and *decomposer*, but also showed that groups investigated different concepts, such as *biodiversity* and *producers*. Log data also showed that students spent less time researching with VidyaMap during the last session compared to the first, which also aligns with the decrease in model complexity for the last sessions. This decrease in time and model complexity was not concerning, though, as students actually demonstrated more conceptual talk and less procedural talk during the last session compared to the first.

Table 5: Markov models showing topic probabilities for Group B's first and last sessions

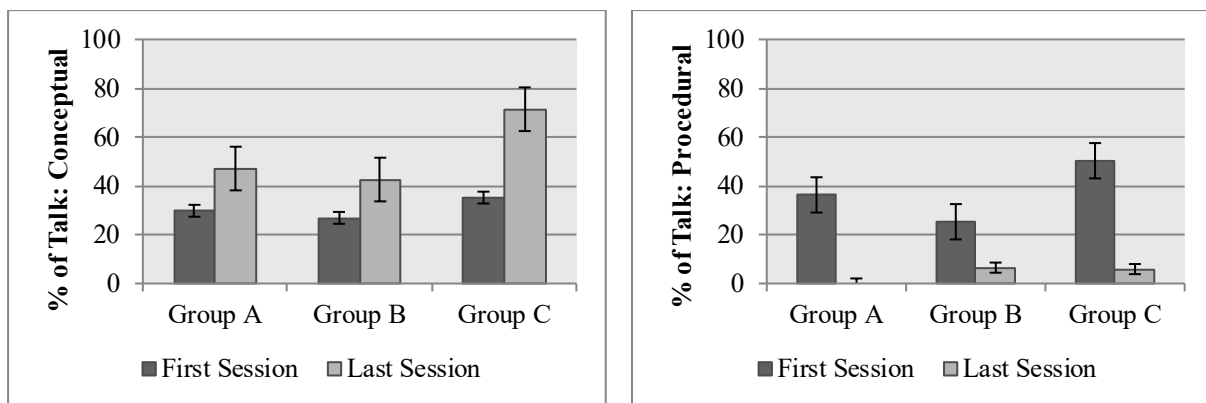
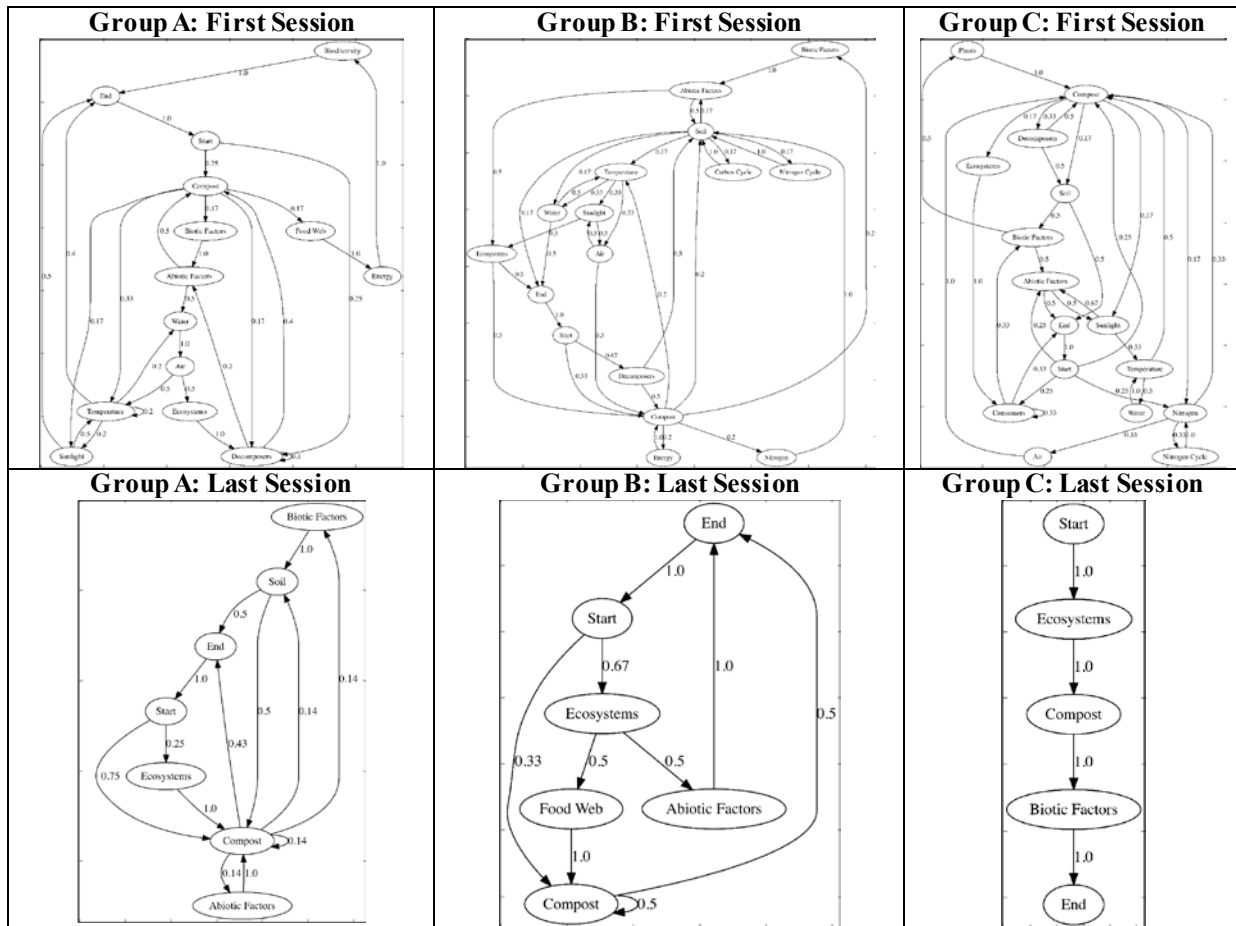


Figure 2. Changes in group discourse over first and last VidyaMap sessions.

## Discussion

In this study, we used a mixed-methods approach that incorporated content assessments, topic modeling, Markov models, and quantitative discourse analysis in order to understand the following question: *How do multiple data sources, in combination, reveal how learning processes unfold at both the individual and group levels?* Each analysis addressed a piece of this question, which we summarize and discuss here.

A comparison of learning outcomes showed significantly better performance on the post-test than pre-test along with no differences between groups in pre- or post-test scores, indicating that all students



demonstrated learning gains. However, test scores only give a limited understanding of conceptual outcomes. To understand learning *processes* at the individual and group levels, we investigated student's journal responses to see how individual students working in groups overlapped in their documentation of collaborative research within VidyaMap. Interestingly, when identifying relatively rare words within individual responses, we found overlap among members of the same group, indicating that students discussed and documented shared ideas within their groups. We also found that each group investigated unique topics, based on overlap in journal responses and records from VidyaMap log data. This indicates that each group engaged in collaborative inquiry through distinct paths when using VidyaMap. Even when we detected decreases in average session time and number of topics researched over the unit, we found that this decrease might not be problematic. Students spent less time researching in VidyaMap, but they engaged in more conceptual discourse and less procedural discourse over time. One interpretation of this is that groups used the e-textbook more efficiently to streamline their research, rather than reduce the quality of their research. Groups might have focused more on meanings and applications of concepts instead than procedural decisions about VidyaMap.

Stahl and colleagues (2006), Dillenbourg (1999), and Reimann (2007) have emphasized the importance of studying collaborative learning at both the individual and group levels. By using mixed-methods to study learning processes at both of these levels, we have a better understanding how different groups in the same classroom took different paths to learning, yet arrived at similar conceptual outcomes (Author, 2013; Kapur, Voiklis, & Kinzer, 2011). Using multiple analyses for different data sources allowed us to triangulate how individual students engaged in collaborative discussion of key topics and increased participation in conceptual discourse while keeping track of their own ideas and conclusions (Suthers & Medina, 2011). Examining the relationship between two data sources—the journal and e-textbook—and how they were used in conjunction with each other revealed reciprocal learning processes between the individual and the group; students co-constructed knowledge through brainstorming, researching, and sharing ideas together while individually documenting their ideas. Interestingly, while we found variance in the paths groups took while learning with VidyaMap, students still achieved similar conceptual outcomes. In this study, the variance in content exploration may not have negatively impacted collaboration dynamics (Barron, 2003). Also, the variance may represent a level of tolerance for differences between students in how they individually and collaboratively developed solutions for an open-ended design challenge. Even with different paths toward learning, whole class discussions may have reinforced key ideas between groups. However, for this exploratory study, we cannot be certain of these interpretations without further analysis of whole class discourse (Lajoie, 2011).

To further understand variance between groups and impacts on conceptual outcomes, we plan to investigate embedded opportunities for knowledge sharing during the unit, such as whole class discussions. These discussions facilitate sharing of ideas between groups, which may explain how groups researching different topics demonstrated similar learning gains. These discussions may explain how divergent paths for inquiry are not only tolerable but maybe even helpful if these paths result in greater knowledge sharing, such as with jigsaw activities for knowledge co-construction. As curriculum designers, we may find opportunities for knowledge building between individuals and groups in our design of journal prompts and scaffolding strategies for teachers. We also plan to further investigate how students learn to use VidyaMap as a resource, including streamlining of their research process, by examining their log data and discourse over all sessions in the unit.

The implications of this study involve how we understand collaborative learning processes through combinations of methods that study both the individual and the group. By using multiple methods for each unit of analysis, we better understand how ideas are shared between individuals collaborating in groups, which may impact conceptual outcomes (Barron, 2003). Understanding how conceptual understanding is interwoven between the individual and the group—and also between groups—is essential to our understanding of collaborative learning processes and the design of embedded supports for them.

## Conclusion

In this study, we used an exploratory mixed-methods approach to understand how learning processes unfolded at the individual and the group levels during small-group collaboration with an e-textbook. We found that groups engaged in divergent paths of inquiry but still demonstrated similar conceptual outcomes across groups. We plan to further investigate how individuals and groups shared ideas in order to track how groups with divergent paths of inquiry co-constructed shared understandings together. Understanding the progression of knowledge co-construction across the levels of the individual and the group helps us to support collaboration and to track and assess learning outcomes.

## References

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, 12(3), 307-359.

- Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and Computational Approaches* (pp. 1-19). Oxford: Elsevier.
- Dornfeld, C. & Puntambekar, S. (2016, June). Negotiation towards intersubjectivity and impacts on conceptual outcomes. In C. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016, Volume 1* (pp. 562-569). Singapore: The International Society of the Learning Sciences.
- Hutchins, E. (1993). Learning to navigate. In S. Chaiklin & J. Lave (Eds.), *Understanding practice: Perspectives on activity and context* (pp. 35-63). New York: Cambridge University Press.
- Kapur, M., Voiklis, J., & Kinzer, C. K. (2011). A complexity-grounded model for the emergence of convergence in CSCL groups. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL* (pp. 3-23). New York, NY: Springer.
- Lajoie, S. P. (2011). Is the whole greater than the sum of its parts? Explaining the role of individual learning and group processes in CSCL. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL* (pp. 235-246). New York, NY: Springer.
- Li, Y., Wang, J., Liao, J., Zhao, D., & Huang, R. (2007, July). Assessing collaborative process in CSCL with an intelligent content analysis toolkit. In *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)* (pp. 257-261). IEEE.
- Puntambekar, S. (2013). Mixed methods for analyzing collaborative learning. In C. E. Hmelo-Silver, C. A. Chinn, C. Chan, & A. M. O'Donnell (Eds.), *The International Handbook of Collaborative Learning* (pp. 220-232). New York, NY: Routledge.
- Reimann, P. (2007). Time is precious: Why process analysis is essential for CSCL (and also can help to bridge between experimental and descriptive methods). In C. Chinn, G. Erkens & S. Puntambekar (Eds.), *Minds, Minds, and Society. Proceedings of the Computer-Supported Collaborative Learning Conference (CSCL 2007)* (pp. 598-607). New Brunswick, NJ: International Society of the Learning Sciences.
- Reimann, P., Yacef, K., & Kay, J. (2011). Analyzing collaborative interactions with data mining methods for the benefit of learning. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL* (pp. 161-185). New York, NY: Springer.
- Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *Journal of the Learning Sciences*, 2(3), 235-276.
- Stahl, G., Koschmann, T., & Suthers, D. D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 409-426). Cambridge: Cambridge University Press.
- Strijbos, J. W., & Fischer, F. (2007). Methodological challenges for collaborative learning research. *Learning and Instruction*, 17(4), 389-393.
- Suthers, D. & Medina, R. (2011) Tracing interaction in distributed collaborative learning. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL* (pp. 341-366). New York, NY: Springer.
- Wertsch, J. V. (1984). The zone of proximal development: Some conceptual issues. *New Directions for Child and Adolescent Development*, 1984(23), 7-18.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Algorithms: The Basic Methods Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed., pp. 85-146). Burlington, MA: Morgan Kaufmann.

## Acknowledgments

We would like to thank Matthew Berland for his assistance with Python. This project received funding from the National Science Foundation (award #1418044).

# Learning Biology Coherently Through Complex Systems, Scientific Practices, and Agent-Based Simulations

Miyoung Park, Emma Anderson, and Susan A. Yoon  
parkmi@gse.upenn.edu, ejanderso@gmail.com, yoonsa@upenn.edu  
University of Pennsylvania

**Abstract:** The Next Generation Science Standards (NGSS) calls for greater coherence in how science is taught and learned in K-12 classrooms. Research on biology classrooms has shown that different units are often taught in a disconnected way, with little focus on unifying themes (e.g., systems) that connect various concepts together in the study of biology. In this study, we hypothesized that a curricular model based on scientific practices and agent-based simulations to teach biology through a complex systems lens would support students' coherent understanding of biology. We investigated the extent to which and ways our curriculum supported students' biology coherence. Units covered topics of diffusion, ecology, enzymes, evolution, genetics, and modeling. Fifty-four students were randomly selected for focus group interviews from a larger study of 463 students. Findings provide promising evidence that students developed a coherent understanding of biology.

**Keywords:** Biology Teaching, Complex Systems, Modeling, Simulations, Scientific Practices, Integrated Knowledge, Curricular Coherence, NGSS

## Introduction

The Next Generation Science Standards (NGSS) in the United States has required a shift in how science is taught and learned in K-12 classrooms. There exists a greater focus on cross-cutting themes such as systems thinking and modeling in order for students to have a deep and more connected understanding between various concepts that are taught (NRC, 2012). Too often, long lists of disconnected facts are taught to students with a focus on breadth, rather than on the overall coherence of how students understand science. This type of approach is alienating to students and also leaves them with fragments of knowledge that provides no sense of the creative achievements of science, logic and consistency in science, and the universality of science (NRC, 2012). Some science education researchers have suggested that, in order to be a scientifically literate adult, knowledge of relationships among ideas is key—understanding the ways important ideas fit together (Roseman, Stern, & Koppal, 2010).

A coherent understanding of biology is defined by an integrated understanding of the various units that comprise the study of biology. This means that students are able to connect separate topics with one another in a way that helps them to better understand biological phenomena (Fortus & Krajcik, 2012). Additionally, an understanding of the relationships and patterns across units enables learners to explain and predict phenomena as well as solve problems (Fortus & Krajcik, 2012). Yet, there are challenges that students face in developing a coherent understanding of biology. First, there exists a lack of integration across topics in science (Chiu & Linn, 2011; Chiu & Linn, 2014; Gilbert & Boulter, 2000; Klymkowsky & Cooper, 2012; NRC, 2012). Second, static images and the ways processes are presented in textbooks make it difficult for students to see the dynamic nature of various phenomena, which make it hard for students to learn biology coherently (Hoffler & Leutner, 2007; Plass et al., 2009; Roseman et al., 2010). Third, science is often taught in a didactic manner, requiring students to learn concepts through rote memorization, which adds to the issue that students often learn long lists of disconnected facts (Anderson & Schonbom, 2008; Osborne, 2014).

We developed a curriculum intervention, which was designed to address these challenges that students face in learning biology coherently. This curriculum supported students' connected understanding of biology through complex systems as an integrated theme, use of dynamic visualizations, and student investigations of scientific practices related to inquiry and argumentation. The research questions that guide our study are: (1) To what extent did the curriculum help students to learn biology coherently?; and (2) How did the curricular model support student understanding?

In the following section we discuss the curricular design choices and provide evidence from the literature that demonstrates how, in combination, these choices may address the curricular coherence problem.

## Conceptual framework

The study's conceptual framework is underpinned by three current research areas in science education that include learning about complex systems, instructional use of agent-based simulations, and scientific practices that more closely represent how science is done in the real world. Each unit in our curriculum was taught through a complex systems lens in order to respond to the lack of integration across topics in science (Chiu & Linn, 2011; Chiu & Linn, 2014; Gilbert & Boulter, 2000; Klymkowsky & Cooper, 2012; NRC, 2012). Agent-based simulations are an integral part of our curriculum because they address the disparate and static manner in which textbooks present phenomena (Hoffler & Leutner, 2007; Plass et al., 2009; Roseman et al., 2010). To address the prevalence of didactic instruction and rote memorization strategies in science class, we integrated key scientific practices that encouraged students to actively construct knowledge (Anderson & Schonbom, 2008; Osborne, 2014). We expand on each of these literature bases in the following section.

### Complex systems

Systems and system models have value for integrating and unifying concepts (Pratt, 2012). Yet, students often have misconceptions about systems, believing, among other things, that they are controlled by a central agent and intentionally designed with certain functions (Taber & Garcia-Franco, 2010). Learning about complex systems is important as students develop understandings about the variability and unpredictability of systems (Osborne, 2014). Complex systems are characterized by multiple interrelated parts that form non-linear relationships, which exhibit emergent properties. Because of this non-linearity, small changes can have large consequences (Yoon, 2008; Yoon, 2011). Curriculum developed through a complex systems lens can cut through various domains and concepts in science (Yoon, 2011; Grotzer et al., 2015; Wilensky & Rand, 2015). Although other studies have examined aspects and challenges of students' complex systems understanding (Ben-Zvi Assaraf & Orion, 2010; Ben-Zvi Assaraf & Orpaz, 2010; Chi et al., 2012; Grotzer et al., 2015), we do not know of any studies that looked purposely at how computational thinking through complex systems can contribute to developing a coherent understanding of biology.

### Agent-based simulations

The second aspect of our conceptual framework and curriculum involves the use of agent-based simulations. Learning through simulations and modeling can lead to greater understanding of scientific phenomena through scaffolding student meaning-making (Smetana & Bell, 2012). Visualizing patterns is better accomplished through computer simulations than through static images or descriptions found in textbooks (Yoon et al., 2013). Chiu and Linn (2014) demonstrated that dynamic visualizations helped increase connections among students' ideas about chemical reactions compared to typical instruction. Beyond the simulations themselves, agent-based modeling allows the student to connect micro and macro aspects of scientific phenomena. By tinkering with the programming, students can explore questions, which reveal the implications of their ideas, while simulating new ideas (Wilensky et al., 2014). Agent-based simulations enable students to understand how processes work together in emergent ways. In this study, we are interested in what ways this support enables students to learn about biological complex systems in a dynamic way for coherent understanding.

### Scientific practices

As students engage in scientific practices, they are involved in the very practices that are essential for a deeper, more nuanced understanding of science (NRC, 2012). The NGSS identifies eight scientific practices for K-12 classrooms: asking questions, developing and using models, planning and carrying out investigations, analyzing and interpreting data, using mathematics and computational thinking, constructing explanations, engaging in argument from evidence, and obtaining, evaluating and communicating information (NRC, 2012). In one study, the importance of students' designing, conducting, and critiquing experiments was highlighted to promote a coherent understanding of science (Chang & Linn, 2013). Students are being pushed to move beyond rote demonstration of scientific content to developing, using, and engaging in constructing knowledge to make sense of the world (Berland et al., 2015).

## Methods

### Context

This study is part of a larger project in which a series of units were developed to support improved understanding of biology through a complex systems approach in the following topics: diffusion, ecology, enzymes, evolution, genetics, and modeling. Each unit takes 2-3 days of instruction to implement in a

classroom. The units can be implemented in any order the teacher believes will best suit the curriculum in their classroom. Along with the curricular units, teachers were also provided with off computer tasks that could be used to introduce or reinforce complex systems ideas.

Each unit consisted of a simulation and student packet, which scaffolded students' learning about complex systems, scientific practices, along with biology content knowledge. Each of the curricular units was intentionally constructed with complex systems components to enable students to understand multiple different biological phenomena through a complex systems perspective. For example, in the unit on evolution, complex systems was first emphasized in the packet introduction. It highlighted how genetic drift is due to random chance survival. Additionally, the units on diffusion, ecology, enzymes, and genetics also all emphasized randomness as a key component in understanding biological phenomena.

In terms of scientific practices, students were asked to make hypotheses, collect data, create and interpret graphs, compare results, answer argumentation questions, etc. In addition, several units required students to read or manipulate the simulation's code. For example, students were asked to go 'under the hood' to explore how fish move. Figure 1 provides an example of instruction and questions students are given in order to interpret the code.

Figure 1. Instructions for viewing the code and questions that ask students to interpret the code for the simulations.

All of the units ask students to respond to argumentation prompts that require students to state a claim, and provide evidence and reasoning to support their claim. For example, in one argumentation prompt, through group discussion, students needed to figure out if the simulation has shown them genetic drift or natural selection and why, with the following sentence starters: "Our claim is...", "Our evidence for this is...", and "Our reasons are that..."

## Participants

The larger study involved 463 students in grades 9 through 12 from seven different schools in the northeastern United States during the academic year 2013-2014. We collected demographic information about this larger group. For school-level data, the seven schools ranged from having 11.4% to 83% of students on free or reduced-price lunch. The schools also ranged from 3.4% to 79.1% non-white students, and ranged from 54% to 89% of students above proficient in the state standardized exam. For this smaller study, we randomly selected 54 students in grades 9-11 to conduct 12 focus group interviews at the end of the academic year to understand in more detail how and what students learned.

## Data sources

We conducted 12 focus group interviews with 4-5 students in each. The combined interview time was 3 hours and 15 minutes. Students were asked the following questions: (1) What do you think biology is? (2) Recall all the units you did using the simulations, which units did you cover? Was there anything that these units had in common? What were these common characteristics? (3) How do complex systems fit into biology? (4) Can you please define what science is?

## Data analysis

The interview transcripts were mined for the three different conceptual framework components. For complex systems, a coding scheme emerged through the data analysis, which included any student response that showed an overarching theme of complex systems thinking which could include nested levels, interdependence, and complex systems mechanisms such as randomness, feedback, cascading or nonlinear actions etc. The other two aspects were coded using previously vetted and validated coding schemes. For agent-based simulations, a coding scheme was used from Yoon and Wang (2014), which included affordances of a phenomenon being visible, dynamic, details, interactive, and scaffolding. For instances of students engaging in scientific practices,

a coding scheme was used from the NGSS scientific practices (Pratt, 2012). Table 1 shows the coding scheme for complex systems, agent-based simulations, and scientific practices, with a description of each code and an example and explanation for each of the codes.

If the researchers disagreed on a code, the researchers discussed until they came to consensus on a single code for that particular response. Each student was only coded once for each category. For example, if a student made three different responses that could be coded as understanding the umbrella theme of complex systems, that student was only coded once for that code.

Table 1: Coding Scheme

	Code Description	Exemplar Coded Response
Complex Systems	<b>Umbrella theme of complex systems</b> Student articulates an overarching theme of complex systems thinking.	<b>Example:</b> “I feel like the complex systems govern kind of the overarching patterns that we see from stuff that’s really, really tiny like the organelles in your cell. Like ribosomes and enzymes functioning and in each of those cells go by another and form organs, each of those organs form complex systems, to form your body. Each individual body forms complex systems within a population and it just builds, and builds, and builds.” <b>Explanation:</b> Here the student shows how multiple concepts in biology can be understood from a complex systems lens.
	<b>Visible</b> Allows users to see things that are normally invisible.	<b>Example:</b> “I think it was especially good for visualizing the randomness aspect of a lot of this. You kind of hear that it moves randomly, but you don’t quite register it until you see all these things bouncing all over the place. Then you are like, oh that’s how they ended up over there. They weren’t just making their way for the gap, they just sort of bounced.” <b>Explanation:</b> Here the student is articulating that it wasn’t until she saw the visualization that she could truly understand that the agents were moving randomly, revealing normally invisible information.
Agent-based Simulations	<b>Dynamic</b> Displays the phenomenon in motion, showing changes over time	<b>Example:</b> “So you got to see how over time they changed.” <b>Explanation:</b> Here the student articulates the importance of seeing the dynamic nature of the phenomenon.
	<b>Details</b> Provides scientific details of the phenomenon.	<b>Example:</b> “Yeah, it gave you like a visual of what was actually happening.” <b>Explanation:</b> Student is articulating how actually seeing a phenomenon helped him to better understand biology.
	<b>Interactive</b> Enables the user to interact with the device.	<b>Example:</b> “I think that was definitely the most helpful part; being able to change something in the situation.” <b>Explanation:</b> Here the student articulates how changing a variable in the phenomenon, helped her learn biology better.
	<b>Scaffolding</b> Provides structure that focuses the users’ attention on relevant information.	<b>Example:</b> “I think the fact that everything was going on at once made it a lot clearer that a single action has more than one consequence. It’s not just a chain reaction of events. It’s all of this stuff is more or less happening at once, in various stages, and intersecting and bumping into each other. It’s harder to convey that on paper. You have to use a ton of arrows. You usually just simplify it to one example of each aspect and that’s really not how stuff worked in real life.” <b>Explanation:</b> By being able to show multiple things happening at once the simulation was able to focus the students’ attention to the fact that a single action has more than one reaction—in a way that a textbook diagram cannot show.
Scientific Practices	<b>Asking questions and defining problems</b> Students should be able to ask questions of each other about the texts they read, the features of the phenomena they observe, and the conclusions they draw from their models or scientific investigations. Asking questions and defining problems progresses to formulating, refining, and evaluating empirically testable questions and designing problems using models and simulations.	<b>Example:</b> “I may not directly find out what I want but I feel like I’m finding out new things I didn’t know before and answering problems that I would have never [gotten] to.” <b>Explanation:</b> Here the student articulates how he answered problems by asking questions that he may have otherwise never considered, which helped him understand a phenomenon better.
	<b>Developing and using models</b>	<b>Example:</b> “I’ve seen a flock of birds outside before but when

<p>Modeling can begin with students' models progressing from concrete "pictures" and/or physical scale models (e.g., a toy car) to more abstract representations of relevant relationships in later grades, such as a diagram representing forces on a particular object in a system. It includes using, synthesizing, and developing models to predict and show relationships among variables between systems and their components in the natural and designed worlds.</p>	<p><i>you look at the programming specifically, you can see the rules that they're following whereas if you looked at it outside, you wouldn't see those rules really showing."</i>  <b>Explanation:</b> Here the student articulates how seeing the rules in the code that modeled a real-life phenomenon enabled her to better understand what was happening.</p>
<p><b>Planning and carrying out investigations</b>  Students should have opportunities to plan and carry out several different kinds of investigations. At all levels, they should engage in investigations that range from those structured by the teacher - in order to expose an issue or question that they would be unlikely to explore on their own (e.g., measuring specific properties of materials) - to those that emerge from students' own questions. Planning and carrying out investigations include investigations that provide evidence for and test conceptual, mathematical, physical, and empirical models.</p>	<p><b>Example:</b> "We just made changes in the simulation but it gave us the basis of what it means to make your own hypothesis."  <b>Explanation:</b> Here the student shows how developing a hypothesis was helpful.</p>
<p><b>Analyzing and interpreting data</b>  Because raw data as such have little meaning, a major practice of scientists is to organize and interpret data through tabulating, graphing, or statistical analysis. Such analysis can bring out the meaning of data—and their relevance—so that they may be used as evidence. Analyzing data includes introducing more detailed statistical analysis, the comparison of data sets for consistency, and the use of models to generate and analyze data.</p>	<p><b>Example:</b> "Sometimes we did [compare] data with other groups so we got to changes to see if all complex systems are the same [or] if all were different. From mostly what we did I can remember for the most part, most group[s] kind of got the same results; they weren't the same exact results."  <b>Explanation:</b> Here the student shows how, through data collection and analysis of the data, they were able to see the larger patterns across complex systems in biology, and the non-static nature of science, that science does not have one set answer.</p>
<p><b>Using mathematics and computational thinking</b>  Mathematical and computational thinking includes using algebraic thinking and analysis, a range of linear and nonlinear functions including trigonometric functions, exponentials and logarithms, and computational tools for statistical analysis to analyze, represent, and model data. Simple computational simulations are created and used based on mathematical models of basic assumptions.</p>	<p><b>Example:</b> "I've seen a flock of birds outside before but when you look at the programming specifically, you can see the rules that they're following whereas if you looked at it outside, you wouldn't see those rules really showing."  <b>Explanation:</b> Here the student shows how using a computational simulation with specific rules enabled them to see something they would otherwise not have been able to see and understand.</p>
<p><b>Constructing explanations and designing solutions</b>  The goal of science is the construction of theories that provide explanatory accounts of the world. A theory becomes accepted when it has multiple lines of empirical evidence and greater explanatory power than previous theories. Constructing explanations and designing solutions includes explanations and designs that are supported by multiple and independent student-generated sources of evidence consistent with scientific ideas, principles, and theories.</p>	<p><b>Example:</b> "Yeah because we [did] observations, charts and graphs. And we also had to do that summarizing thing."  <b>Explanation:</b> Here the student is expressing how, in the unit, she had to summarize her findings—therefore she was constructing an explanation.</p>
<p><b>Engaging in argument from evidence</b>  The study of science and engineering should produce a sense of the process of argument necessary for advancing and defending a new idea or an explanation of a phenomenon and the norms for conducting such arguments. In that spirit, students should argue for the explanations they construct, defend their interpretations of the associated data, and advocate for the designs they propose. Engaging in argument from evidence includes using appropriate and sufficient evidence and scientific reasoning to defend and critique claims and explanations about the natural and designed world(s). Arguments may also come from current scientific or historical episodes in science.</p>	<p><b>Example:</b> "...yeah and like the evidence, reasoning, claim thing."  <b>Explanation:</b> The student is explaining how she had to answer questions using evidence, reason, and claims, the scaffolding design in the project helped students answer argumentation questions.</p>
<p><b>Obtaining, evaluating, and communicating information</b>  Any education in science and engineering needs to develop students' ability to read and produce domain-specific text.</p>	<p><b>Example:</b> "And we had to like kind of hypothesize a lot and like explain why this happens and why everything comes on."  <b>Explanation:</b> The student is expressing how her group had to</p>

<p>As such, every science or engineering lesson is in part a language lesson, particularly reading and producing the genres of texts that are intrinsic to science and engineering. This includes evaluating the validity and reliability of the claims, methods, and designs.</p>	<p>communicate why her group had gotten the results to their experiments.</p>
--	---

## Results

To investigate our research questions, we looked at the total number of responses per coded category, which is shown in Figure 2. In total, there were 114 unique codable responses. The most frequent categories included detail (21 responses) in agent-based simulations, umbrella theme of complex systems (18 responses), and planning and carrying out investigations (14 responses) in scientific practices.

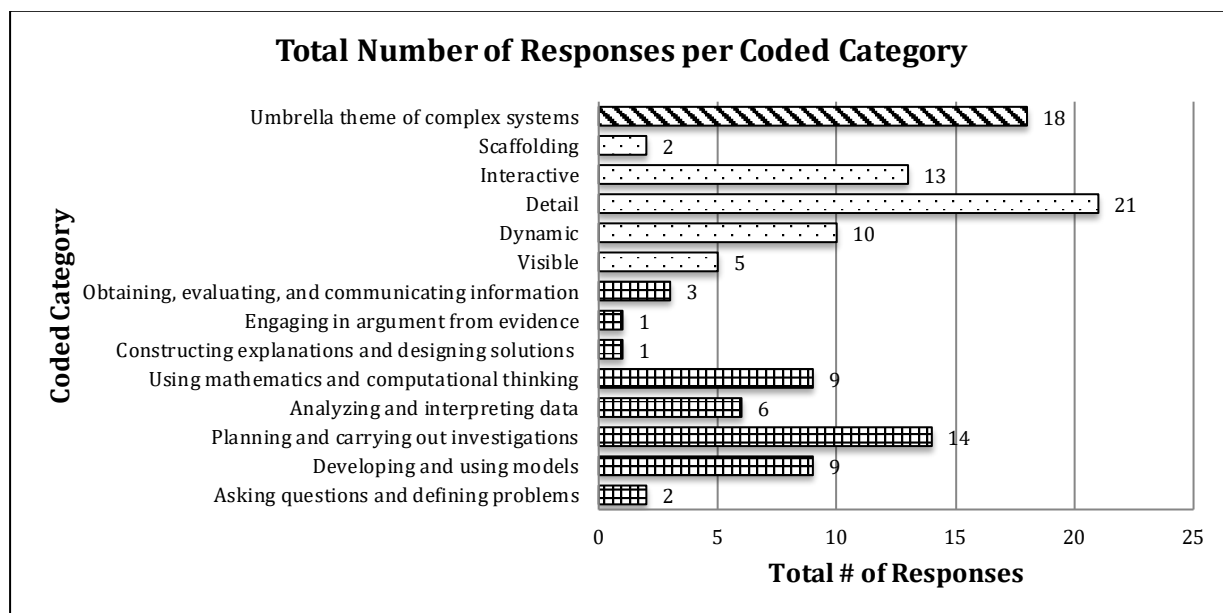


Figure 2. A bar graph representing the total number of responses per coded category.

To answer our first research question, we found that 33% of students in our sample (18 students) articulated an umbrella theme of complex systems in understanding biology. For example, one student stated:

*I mean all [of the units] just had like -- It wasn't just sun hits plant, plant goes, yay. It was like the protein goes over here. Then the RNA reacts like this, and this hooks onto here, but if it hits here, then it does this. If it goes over there, then it does that. There were multiple factors all running around doing their own things and depending on how they interacted, when they bumped into each other mostly, the step would interact differently. Stuff would happen. They were all like that. (Focus Group ID 6, May 2014)*

In the above quote, the student explains that all of the units showed how systems have multiple intersecting agents, who randomly bump into each other, and depending on the ways in which they interact, different outcomes would occur in the system. The student shows a sophisticated understanding of complex systems and how this is a tying theme across the units. Another student simply states, “Everything is a complex system; if you think about it.” (Focus Group ID 6, May 2014). This statement reveals this student sees complex systems everywhere—understanding that complex systems are pervasive.

To investigate the second research question, we analyzed students’ statements to understand the ways in which the most frequently identified supports helped students learn biology in a coherent way. Amongst the student responses related to the simulations, 41% identified detail as an important affordance of the simulations. For example, a student states, “The biggest thing that helps me understand biology was how everything in the simulation has a set of rules that it follows and how things move about randomly in complex systems. It's hard to get that from a diagram that your teacher might draw on the board or something like that.” (Focus Group ID 9, May 2014). Here, we observe that seeing the detail in the simulation enabled the student to understand



randomness in complex systems. This was important because his understanding of a complex system came through an affordance of the simulation, which ultimately contributed to his coherent understanding of biology. Amongst the student responses related to scientific practices, 31% identified planning and carrying out investigations as important. A student articulates that playing with the code of the simulations itself helped the student understand the simulation model, “I like using the coding; when you use the coding to change the program... Because I could control what everything was doing and I saw like how when you took the tumble blocks in and out, I saw like [how] things worked. Like I could just know what they were suppose [sic] to do.” (Focus Group ID 5, May 2014). The student points out that being able to manipulate the code allowed her to understand how the agents function within the model (planning and carrying out an investigation), giving her a greater understanding of the complex system.

## Discussion and significance of the study

We conducted this study in response to the need for students to have a coherent understanding of biology (NRC, 2012; Roseman et al., 2010). Our curriculum was designed with complex systems, agent-based simulations, and scientific practices to address the challenges that students face in developing this coherent understanding. In the results, we identified that there were particular aspects of the agent-based simulations and scientific practices that had been designed into the curriculum that enabled students to learn biology through complex systems, which in turn helped them learn biology in a coherent manner. Developing a coherent understanding of biology using standard curriculum is challenging to do, and here we found that a third of our students were able to very clearly articulate a systems understanding that brought multiple units of biology together. This study was completed in five units that took about ten days of instruction. It was a small portion of the curriculum, and yet we see promising evidence that a third of the students had a clear understanding of complex systems unifying the various topics in biology. From earlier studies we know that students do understand complex systems (Yoon et al., 2015)—what we see in this study is that understanding complex systems, which was enabled through the details in simulations and students’ planning and carrying out their own investigations, may have contributed to their coherent understanding of biology for at least a third of the students.

Agent-based simulations let students see details in processes. Scientific practices enabled students to understand how models function and what is actually happening in the phenomena. These scaffolds work together to help students learn biology coherently. Moreover, this study extends the literature that suggests ways in which supports may help students to better understand scientific phenomena (Berland et al., 2015; Osborne, 2014; Wilensky et al., 2014). This is important as we consider the design of future biology curriculum and the ways we can incorporate complex systems as a unifying theme for various units, with the supports of simulations and scientific practices. The results of this study are encouraging and give us reason to believe that follow-up curricula that demonstrate how biology is interconnected through a systems lens would support pattern recognition across content domains. In the future, an experimental randomized controlled study of this curriculum would validate these findings, since a limitation of the current study is the lack of a control group. Additionally, further study of classroom observations and teacher interviews may reveal additional mechanisms through which students developed a coherent understanding of biology.

## References

- Anderson, T. R., & Schonbom K. J. (2008). Bridging the educational research-teaching practice gap. *Biochemistry and Molecular Biology Education*, 36, 309-315.
- Ben-Zvi Assaraf, O., & Orion, N. (2010). System thinking skills at the elementary school level. *Journal of Research in Science Teaching*, 47(5), 540–563.
- Ben-Zvi Assaraf, O., & Orpaz, I. (2010). The “Life at the Poles” study unit: Developing junior high school students’ ability to recognize the relations between earth systems. *Research in Science Education*, 40, 525–549.
- Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo., A. S., & Reiser, B. (2015). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*, 1-32.
- Chang, H., & Linn, M.C. (2013). Scaffolding learning from molecular visualizations. *Journal of Research in Science Teaching*, 50(7), 858-886.
- Chi, M. T. H., Roscoe, R., Slotta, J., Roy, M., & Chase, M. (2012). Misconceived causal explanations for “emergent” processes. *Cognitive Science*, 36, 1–61.
- Chiu, J. L., & Linn, M. C. (2011). Knowledge integration and wise engineering. *Journal of Pre-College Engineering Education Research*, 1, 1-14.
- Chiu, J. L., & Linn, M. C. (2014). Supporting knowledge integration in chemistry with a visualization-enhanced

- inquiry unit. *Journal of Science Education Technology*, 23, 37-58.
- Fortus, D., & Krajcik, J. (2012). Curriculum coherence and learning progressions. In B.J. Fraser, K.G. Tobin, & C.J. McRobbie (Eds.), *Second international handbook of science education* (pp. 783-798). New York, NY: Springer.
- Gilbert, J. K., & Boulter, C. J. (2000). *Developing models in science education*. Dordrecht, Netherlands: Kluwer.
- Grotzer, T. A., Power, M. M., Derbiszewska, K. M., Courter, C. J., Kamarainen, A. M., Metcalf, S. J., & Dede, C. (2015). Turning transfer inside out: The affordances of virtual worlds and mobile devices in real world contexts to teaching about causality across time and distance in ecosystems. *Technology, Knowledge and Learning*, 20, 43–69.
- Hoffler, T.N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17, 722-738.
- Klymkowsky, M. W., & Cooper, M. M. (2012). Now for the hard part: the path to coherent curricular design. *Biochemistry and Molecular Biology Education*, 40, 271-272.
- National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting concepts, and Core Ideas*. Washington, D.C.: National Academies Press.
- Osborne, J.F. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328, 463-466.
- Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education* 25, 177-196.
- Plass, J.L., Homer, B.D., and Hayward, E.O. (2009). Design factors for educationally effective animations and simulations. *Journal of Computing in Higher Education*, 21, 31-61.
- Pratt, H. (2012). *NSTA's Reader's Guide to a Framework for K-12 Science Education*. Expanded Edition. NSTA Press: Arlington, VA.
- Roseman, J. E., Stern, L., & Koppal, M. (2010). A method for analyzing the coherence of high school biology textbooks. *Journal of Research in Science Teaching*, 47, 47-70.
- Smetana, L. K., & Bell, R. L. (2012). Computer simulations to support science instruction and learning: A critical review of the literature. *International Journal of Science Education*, 34(9), 1337–1370.
- Taber, K. S., & García Franco, A. (2010). Learning processes in chemistry: Drawing upon cognitive resources to learn about the particulate structure of matter. *Journal of the Learning Sciences*, 19(1), 99-142.
- Wilensky, U., Brady, C., and Horn, M.S. (2014). Fostering computational literacy in science classrooms. *Communications of the ACM*, 57(8), 17-21.
- Wilensky, U., & Rand, W. (2015). *An introduction to agent-based modeling. Modeling natural, social and engineered complex systems in NetLogo*. Cambridge MA: MIT Press.
- Yoon, S. (2008). An evolutionary approach to harnessing complex systems thinking in the science and technology classroom. *International Journal of Science Education*, 30(1), 1-32.
- Yoon, S. (2011). Using social network graphs as visualization tools to influence peer selection decision-making strategies to access information about socioscientific issues. *Journal of the Learning Sciences*, 20(4), 549-588.
- Yoon, S., Anderson, E., Koehler-Yom, J., Sheldon, J., Schoenfeld, I., Wendel, D., Scheintaub, H., Klopfer, E., Oztok, M., & Evans, C. (2015). Design features for computer-supported complex systems learning and teaching in high school science classrooms. *Journal of Research in STEM Education*, 1(1), 17-30.
- Yoon, S., Klopfer, E., Wang, J., Sheldon, J., Wendel, D., Schoenfeld, I., Scheintaub, H., & Reider, D. (2013). Designing to improve biology understanding through complex systems in high school classrooms: No simple matter! In the proceedings of the Computer Supported Collaborative Learning, Madison, Wisconsin.
- Yoon, S., & Wang, J. (2014) Making the invisible visible in science museums through augmented reality devices. *Tech Trends*. 58(1), 49-55, 2014.

## Acknowledgements

We thank Eric Klopfer, Josh Sheldon, Ilana Schoenfeld, Daniel Wendel, Hal Scheintaub, and Jessica Koehler for their collaboration on the project. This work was funded by the U.S. National Science Foundation Discovery Research K-12 (DRL 1019228).

# Expressing and Addressing Uncertainty: A Study of Collaborative Problem-Solving Dialogues

Fernando J. Rodríguez, Kimberly Michelle Price, and Kristy Elizabeth Boyer  
fjrodriguez@ufl.edu, kimberlymprice@ufl.edu, keboyer@ufl.edu  
University of Florida, Gainesville, Florida, USA

**Abstract:** To support learners during collaborative problem solving, developing a deeper understanding of collaborative dialogue is essential. This paper focuses on one important aspect of collaborative dialogue: expressions of uncertainty. In a study of undergraduate novice computer science students working in pairs, we observed that the students who produced the lowest quality solutions expressed uncertainty more often than those who produced middle-quality solutions. Perhaps surprisingly, pairs with the highest quality solutions also expressed more uncertainty than the middle performers. Examining the ways in which students expressed and then followed up on uncertainty revealed that higher-performing pairs utilized emerging learning opportunities when uncertainty was expressed, and remained focused on one task at a time. In contrast, the lower-performing pairs often did not resolve their uncertainty before moving on, attempting to work with multiple incomplete pieces of the solution simultaneously. These findings provide insight into how best to support collaborative learning during uncertainty.

## Introduction

Collaborative dialogue is a complex process through which learners express their perspectives and catalyze learning (Gee, 2014; Howley, Mayfield, & Rosé, 2011; Rosé et al., 2008; Vygotsky, 1978). Through dialogue, *uncertainty* often arises as students self-explain (Chi, De Leeuw, Chiu, & LaVancher, 1994), implicitly inviting their collaborators to elaborate (Webb, 1982). Learners also express uncertainty as a form of politeness or hedging, allowing a less knowledgeable collaborator or oneself to avoid embarrassment (Brown & Levinson, 1978; Markkanen & Schröder, 1997). Uncertainty during collaboration can provide opportunities for learning by inciting curiosity and exploration (Berlyne, 1978). However, if collaborators repeatedly (or for prolonged periods of time) do not address their own uncertainty or that expressed by others, frustration and missed learning opportunities can ensue (D'Mello & Graesser, 2012).

There is evidence that adapting to students' uncertainty as expressed through dialogue can have significant benefit. In a study of undergraduate students learning physics through spoken dialogue with an intelligent tutoring system, the students learned significantly better when the system adapted to the presence of uncertainty (Litman & Forbes-Riley, 2014). Unlike intelligent tutoring systems, humans naturally adapt to each other's uncertainty. For collaborative problem solving in particular, in which students collaborate to produce a shared solution (Nelson, 1998), our recent work has shown that the frequency of several types of dialogue utterances, including expressions of uncertainty, are associated with quality of the shared solution (Rodríguez, Price, & Boyer, 2017). This paper takes a deeper look at expressions of uncertainty and how collaborative pairs address them during the problem-solving process.

This paper examines collaborative problem solving in the domain of computer science. Specifically, the learners in this study solve programming problems in pairs within a structured collaborative paradigm known as *pair programming* (Nagappan et al., 2003). There are two collaborator roles in pair programming: the *driver* writes the program, while the *navigator* provides feedback and instructions. Together, driver and navigator produce a single shared solution (Falkner, Falkner, & Vivian, 2013; Porter & Simon, 2013). We collected dialogue and problem-solving data from pairs of students who interacted remotely through textual dialogue. We found that, perhaps not surprisingly, pairs who produced the lowest quality solutions showed more expressions of uncertainty (both in terms of absolute frequency and relative frequency) than pairs with middle-quality solutions. However, pairs who produced high-quality solutions also made significantly more uncertainty expressions than the middle performing pairs. The results show that higher-task-quality pairs were in a position to take advantage of the learning opportunities that uncertainty affords: they often addressed uncertainty by experimenting in their programming code until they resolved the uncertainty, then moved on to the next subtask. In contrast, the lower performing pairs often did not focus their efforts in the same way, leaving uncertainty unresolved and moving on to the next subtask. By understanding these processes, we hope to inform the design of adaptive systems that support student pairs during collaborative problem solving.

## Related work

Prior work has considered numerous types of dialogue moves that express uncertainty. Some uncertainty utterances express confusion (Keltner & Shiota, 2003), e.g., “Why did that happen?” or “I don’t understand.” Closely related to confusion is the notion of *cognitive dissonance* (Festinger, 1962) or *cognitive disequilibrium* (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005), when the observed state of the world does not match what a learner expected. These cognitive states are different from the phenomena underlying a type of politeness-uncertainty (Brown & Levinson, 1978), in which a conversational partner attempts to avoid “face-threatening” moves that could lead to embarrassment on the part of the other partner. Similarly, people often *hedge* their dialogue moves, making those utterances more fuzzy and less certain, in order to save themselves from embarrassment (Markkanen & Schröder, 1997), e.g., “I think...”, a phrasing seen regularly in dialogue during learning (Forbes-Riley & Litman, 2011).

Berlyne (1978) emphasizes the importance of uncertainty because it leads to curiosity and exploration, with substantial potential for students to learn as the uncertainty is resolved. More specifically, if uncertainty related to confusion is not successfully addressed, it can lead to negative outcomes such as interrupted flow, frustration, and boredom (D’Mello & Graesser, 2012). For example, Litman & Forbes-Riley (2014) evaluated adaptive support to student uncertainty within an ITS for physics problems. The intelligent tutoring system provided different levels of adaptation based on student uncertainty and correctness. The system that provided adaptive support based on presence of uncertainty did improve students’ learning gains. The tutor that specifically adapted to the different levels of uncertainty and correctness, however, did not provide further benefit. Our goal is to investigate uncertainty during collaborative problem solving, paying close attention to how students expressed and addressed uncertainty differently, and how those differences relate to the quality of the solution that the pair constructed.

A study by Sharma et al. (2013) investigated pair programming from the perspective of pair program comprehension. In that study, students collaboratively evaluated Java programs and researchers analyzed students’ dialogue and gaze. They found that successful pairs tended to focus together on the same program elements and their dialogue was centered around describing the program, as opposed to less successful pairs whose dialogue focused on managing the collaboration. The study presented in this paper provides new findings along these lines: we investigate how expressions of uncertainty are associated with the quality of the shared solution that pairs produce.

## Study description

Study participants were recruited from an introductory programming course in Java for computer science majors at a university in the southeastern United States. Out of the approximately 450 students enrolled in the course, 54 voluntarily participated in the programming study. The volunteers were 40 male and 14 female students; 25 White, 16 Latino, 11 Asian, 1 Black, and 1 Pacific Islander; with ages between 18 and 31 ( $M=19.6$ ,  $SD=2.21$ ). The students were assigned to pairs based on their mutual scheduling availability, for a total of 27 pairs.

Students were asked to create a program that acts as a math tutor to help young children practice addition, subtraction, and multiplication. They used the *Snap!* block-based programming language to implement this program (Figure 1). In *Snap!*, programmers create programs by dragging blocks and snapping them together to create the necessary logical structure. Block-based programming languages are increasingly common for introductory computer science both in K-12 and at the postsecondary level for programming novices. We chose this programming language because one of the broader goals of our work is to understand the affordances of block-based versus traditional textual programming languages for fostering collaboration in computer science problem solving. Although students were partway through a course in the Java programming language, the block-based programming task presented substantial challenge to them because they were addressing a new problem for which they needed to utilize an appropriate algorithm and reuse previously constructed code modules. Students had one hour to work on the activity, including implementation and software testing, with no requirement that they complete the full activity before the time ended. Only one pair completed all implementation and testing activities within one hour. The students’ math tutor program needed to display an equation with the operator blanked out and prompt the user to select an operator that solves the equation. If no operator satisfied the equation, the user would select “None”. The learning task was to implement the code for the tutor to 1) display the equation, 2) evaluate the user’s choice, and 3) let her know if her answer was correct or incorrect. The modules for selecting the operators and “None” were implemented ahead of time and provided as scaffolding so that the students could focus on implementing the remaining functionality of the program.

When they arrived for the study, students were seated in separate rooms. They collaborated through an interface (Figure 1) that provided a synchronized view of the problem-solving area and textual dialogue through Google Hangouts. This collaboration modality is common in students’ everyday practice: they often share screens

remotely and interact via text messages or instant messaging while solving problems together. Researchers randomly assigned one student to the *driver* role and the other student to the *navigator* role. Due to technical limitations of the screensharing interface, the collaborative roles remained fixed throughout the one-hour session. The driver actively engaged in programming actions, while the navigator viewed the instructions and communicated with the driver.

Students completed an assessment activity before and after the collaborative programming task. They worked individually on the pre- and post-assessments in which they were given three minutes to implement a short program to display the larger of two randomly-generated integers. We used a 10-point rubric to assign each student an assessment score. The average pre-assessment score was 4.6 ( $SD=1.9$ ,  $max=10$ ,  $min=1$ ), and the average post-assessment score was 7.9 ( $SD=1.7$ ,  $max=10$ ,  $min=4$ ). The average learning gain of 3.3/10 (post-assessment minus pre-assessment) is significantly nonzero ( $p<0.0001$ ; paired  $t$ -test). The collaborative role is significantly associated with learning gain: drivers' average learning gain was 4.2/10, while for navigators the average was 2.3/10. This difference is statistically significant ( $p=0.0008$ ; two-sample  $t$ -test).

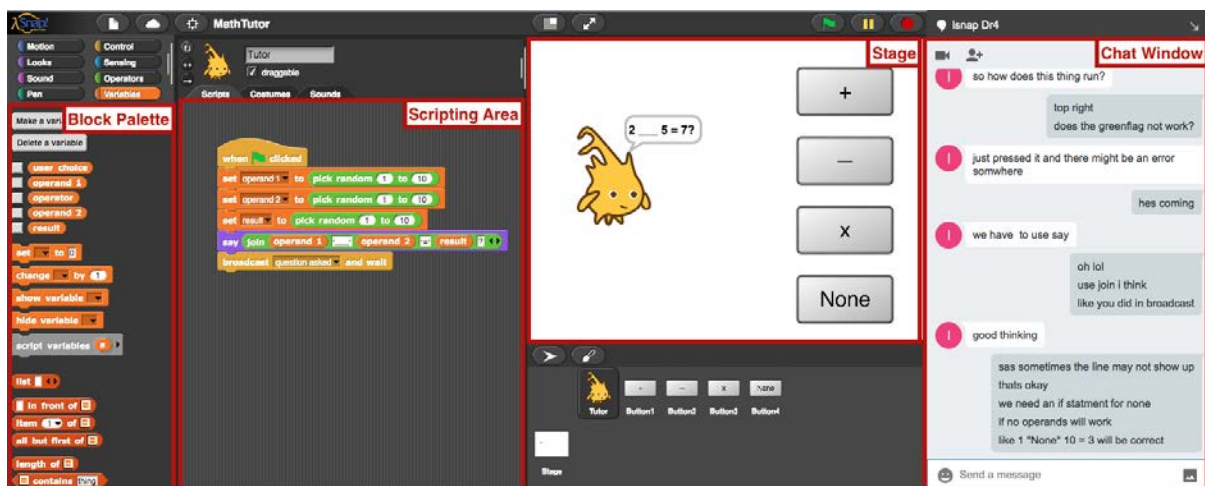


Figure 1. Problem-solving and dialogue interface.

The version of *Snap!* used in this study was instrumented with database logging of programming actions. Whenever a student performed an action in the interface (adding or removing a programming block, connecting or separating two blocks, moving blocks within the interface, editing the parameters of a block, switching between block categories, or running the program), an entry for the action was added to a database. Each event entry included the timestamp, action type, and the current state of the program. The dialogue history was extracted from Google Chat and combined with the action logs based on the relative timestamp, yielding a dataset of programming action and chat sequences. There were a total of 9335 programming actions ( $M=346$  per session,  $SD=112.3$ ,  $max=654$ ,  $min=111$ ) and 3438 chat messages ( $M=127$  per session,  $SD=61.8$ ,  $max=233$ ,  $min=47$ ). Drivers sent a total of 1089 messages ( $M=40$  per session,  $SD=18.7$ ,  $max=73$ ,  $min=10$ ) while navigators sent a total of 2349 messages ( $M=87$  per session,  $SD=53.4$ ,  $max=200$ ,  $min=28$ ).

## Data analysis

We were interested in examining how dialogue unfolded within each pair and comparing these dialogues based on the pairs' performance on the given task. To evaluate the quality of the pairs' solutions, we collected the final versions and graded them with a purpose-built 11-point rubric, which accounted for presence of all necessary code blocks and test results for functionality of the program. The average solution score was 7.1 ( $SD=2.1$ ,  $max=11$ ,  $min=3$ ). For further analysis, student pairs were split into three groups based on their solution quality: pairs with a score of 9, 10, or 11 were classified as *High* ( $N=8$ ); pairs with a score of 6, 7, or 8 were classified as *Medium* ( $N=11$ ); and the remaining pairs, who scored 3, 4, or 5, were classified as *Low* ( $N=8$ ). Given these sample sizes, we used the nonparametric Wilcoxon rank-sum test to evaluate significant differences between the groups.

We extracted the textual dialogue messages and split them into utterances based on punctuation marks (periods, question marks, exclamation marks). Uncertainty utterances were tagged as part of a broader dialogue labeling study that included thirteen distinct tags such as direct and indirect instructions, questions and answers, and partner feedback. Table 1 describes the dialogue act tagging scheme applied to the chat messages. More details about the tagging scheme can be found in our previous work (Rodríguez, Price, & Boyer, 2017). The

dialogue act labeling was reliable, with Cohen's Kappa of 0.73. For the work presented in this paper, we take a closer look at messages tagged as expressions of uncertainty (total=133,  $M=4.9$  per session,  $SD=4.0$ , max=14, min=1), which include explicit statements of confusion (e.g., "Huh?") as well as hedged suggestions (e.g., "Maybe we should...") and uncertain explanations (e.g., "I think it's because..."). With regards to driver uncertainty specifically, we found that these events were more frequent in both high-performing (total=16,  $M=2$  per session,  $SD=0.76$ , max=3, min=1) and low-performing pairs (total=22,  $M=2.75$  per session,  $SD=1.49$ , max=5, min=1) when compared to medium-performing pairs (total=10,  $M=0.91$  per session,  $SD=0.94$ , max=3, min=0) with Wilcoxon rank-sum test  $p$ -values of 0.0180 and 0.0082, respectively.

Table 1: Dialogue act tagging scheme. This paper focuses on the Uncertainty tags

Tag	Name	Description	Example
S	Statement	Statement of information or an explanation	<i>we just need it to have an else</i>
U	Uncertainty	Statement of uncertainty, suggestion, or indication of confusion	<i>unsure how to add strings together</i>
D	Directive	Explicit instruction to the partner (includes references to specific interface elements)	<i>wait put the if back</i>
SU	Suggestion	Polite or indirect instruction to the partner	<i>we could do a "not"</i>
ACK	Acknowledgement	Acknowledging a partner's previous message	<i>oh ok gotcha</i>
M	Meta-comment	Reflection on the problem-solving process (what the student is thinking or doing)	<i>hmmm</i>
QYN	Yes/No Question	Task-related question requesting yes or no	<i>can the answer be negative?</i>
QWH	Wh- Question	Task-related question requesting information (who, what, where, when, why, and how)	<i>how do I take in their input?</i>
AYN	Yes/No Answer	Response to task-related yes/no question	<i>yea</i>
AWH	Wh- Answer	Response to task-related information question	<i>The numbers must be random</i>
FP	Positive Fdbk.	Distinctly positive response to partner actions	<i>oh nice</i>
FNON	Nonpositive Fdbk.	Nonpositive response to partner actions	<i>thats weird</i>
O	Off-task	Unrelated to the task	<i>wow its sweet in this room</i>

## Examining uncertainty in collaborative dialogue

When examining driver uncertainty, our hypothesis was that high- and low-performing pairs dealt with moments of uncertainty differently; in particular, we hypothesized that high-performing pairs addressed and resolved uncertainty while low-performing pairs were unable to or took longer to do so. In our dataset, driver uncertainty manifested itself one of two primary ways: suggestions/hedges, and explicit confusion. In this section, we describe examples of each kind of uncertainty and compare how it was managed by high- and low-performing pairs.

Students' incoming knowledge level is likely an important influencing factor in how students expressed, and addressed, uncertainty during collaboration. Indeed, students in the *High* group for task solution quality were more knowledgeable at the outset: they scored significantly higher on the pre-assessment than students in the *Low* solution quality group. The average pre-assessment score for students in the *High* collaborative solution quality group was 5.2/10, while the average pre-score for students who generated a *Low* quality collaborative solution was 3.8/10 ( $p=0.0078$ ; Wilcoxon rank-sum test). This higher knowledge at the outset likely enabled collaborators to more effectively identify their own confusion or uncertainty, and helped them more successfully address it. The following section examines driver uncertainty events and how they appear to relate to the group's success in the collaborative problem-solving activity. Each excerpt presented in the following subsections contains the original student messages, some of which contain typos. The gender of each student is indicated at the start of the excerpt (*M* for male, *F* for female). Driver uncertainty messages appear in bold text for emphasis.

### Case 1: Expressions of uncertainty as suggestions

Students often communicated with their partners in a hedged manner that can indicate politeness or face-saving, but which manifests as uncertainty. For example, when the pair identified an error, the driver often suggested a reason for the error or proposed a solution. These kinds of utterances typically began with "I think" or "I could." Excerpt 1 parts *a* and *b* show examples of uncertainty events for a high-performing pair and a low-performing pair. For the high-performing pair, the driver and navigator were having trouble completing a subtask. The navigator proposed a solution while the driver hypothesized about the reason for the program error. The driver then made changes to the program, tested it, and proved his intuition was right. In the low-performing pair,

students attempted a different task. Both the driver and the navigator were unsure of how to solve the task. The driver suggested an approach, and the navigator approved by giving positive feedback but then expressed uncertainty by following his feedback with “I think.” The pair did not return to discuss the driver’s uncertainty further. It was left unresolved as the pair moved on to another subtask.

Excerpt 1: Suggestion dialogue excerpts

a) High-Performing Pair (driver: M, nav.: F)	b) Low-Performing Pair (driver: M, nav.: M)
<p><b>Navigator (15:31:12):</b> hmmm  <b>Navigator (15:31:33):</b> maybe they all need to be in one say block  <b>Driver (15:41:34):</b> I think it just says result  <i>(Driver edits program parameters and tests it)</i>  <b>Driver (15:42:06):</b> yup, it skips the first two “say”s</p>	<p><i>(Pair discuss how to assign values to equation parameters for 2 minutes)</i>  <b>Driver (11:49:31):</b> I think we can with “se”t  <i>(Driver experiments with and figures out how to set the value of a variable)</i>  <b>Navigator (11:51:21):</b> yes its gonna be something like what you’re doing  <b>Navigator (11:51:31):</b> i think</p>

This kind of uncertainty is related to the concept of *subjective uncertainty* in that it leads to specific exploration, behavior prompted by events of uncertainty that focuses on eliminating it (Berlyne, 1978). Both instances of uncertainty shown in the excerpts represent an opportunity for learning. In the high-performing pair’s session (excerpt 1a), both the driver and navigator were attempting to solve the same problem, providing suggestions to each other for consideration. The driver then tested and confirmed his thinking, resolving the uncertainty. He even explicitly stated his findings in the dialogue, letting the navigator know that they could move on to the next subtask. They were able to take advantage of the learning opportunity by exploring potential solutions and resolving the uncertainty. The low-performing pair found themselves in a similar situation, but with a different outcome. Both members were trying to solve the same problem, and the driver was able to figure out the solution on his own. However, the driver did not explicitly let the navigator know that he had found a solution. Instead, the navigator acknowledged the driver’s attempt at a solution in the dialogue but expressed his own uncertainty about the approach. They did not leverage the opportunity in part because the driver did not explicitly state that he had arrived at a solution, and the navigator appears to have missed the opportunity to add an understanding of the approach to the pair’s common ground.

**Case 2: Expressions of uncertainty as confusion**

Many uncertainty events are utterances indicating that the speaker is confused or does not understand something. These utterances often begin with “I don’t know” or “I’m not sure”. Excerpt 2 parts *a* and *b* show two examples of confusion-related uncertainty, one from a high-performing pair and one from a low-performing pair. In Excerpt 2a, the driver in a high-performing pair expressed to the navigator that he did not know how to implement a component of the task. The navigator pointed out to the driver where he could find the programming blocks he needed, and the driver added the block to the program. In the low-performing pair (Excerpt 2b), the driver also told the navigator that she did not know how to implement a component of the task. The navigator provided quick guidance, but then the conversation shifted to another part of the task for a few minutes. Afterward, the navigator revisited the previous task component with the driver, and they both mentioned that they did not know how to complete it. At this point, the conversation shifted once more to a separate task component for a longer period of time. After testing the program, the driver brought back the unresolved issue, but the session ended without resolving this issue.

Research on cognitive load theory suggests that having too many simultaneous workflows puts a strain on working memory, limiting the amount of information that can be processed (Renkl & Atkinson, 2010). In the high-performing pair session, the driver expressed uncertainty and the pair worked together to resolve it before moving on to the next steps. Since they focused on a single task at a time, they were able to utilize the full potential of their working memory to quickly overcome the uncertainty. Conversely, the low-performing pair switched between several subtasks, leaving any established uncertainty unresolved. By attempting to complete multiple subtasks simultaneously, the low-performing pair may have taxed their working memory and inhibited their ability to address the uncertainty.

Excerpt 2: Confusion dialogue excerpts

a) High-Performing Pair (driver: M, nav.: F)	b) Low-Performing Pair (driver: F, nav.: M)
<p><b>Driver (13:55:43):</b> unsure how to add strings together <b>Navigator (13:55:47):</b> can you use an operator? <b>Navigator (13:55:31):</b> go to the ops <i>(Driver switches to the Operator block category)</i> <b>Driver (13:56:07):</b> oh shoot nice <i>(Driver adds "join" block to the program)</i> <b>Driver (13:56:21):</b> this one?</p>	<p><b>Driver (14:11:23):</b> im not sure how to display the operands <b>Navigator (14:12:17):</b> try locating the "say" button <i>(Pair discuss a different task)</i> <b>Navigator (14:21:13):</b> Is there a way we can use the "say" button and then put the operators in it <b>Navigator (14:21:26):</b> then use user choice button to store the users result <b>Driver (14:22:30):</b> Yeah I havent figured out yet how to display more than one variable at a time <i>(Pair discuss a different task)</i> <b>Driver (14:45:25):</b> im not sure how to make it display the operands and the result together</p>

**Case 3: Addressing uncertainty around a similar task**

In our third example (Excerpt 3 parts *a* and *b*), the high-performing and low-performing pairs expressed uncertainty toward the same task: selecting a random operator on which to base their equation. Both excerpts occur near the end of the one-hour collaborative session. In the high-performing pair, the driver expressed uncertainty at how to complete the task, and then the navigator asked a clarification question. This prompted the driver to explain his point of view on the task at hand. The navigator understood the question and turned to the task instructions to find an answer, warning the driver of the time remaining a few minutes later. During this time, the driver edited the program for a few minutes and found the solution ten minutes before the end of the session. For the low-performing pair, the driver similarly expressed uncertainty regarding the given task. The navigator provided feedback, but the driver did not explicitly acknowledge this feedback. The driver experimented with the program for a few minutes, and the navigator asked a question about a separate task. The driver answered the navigator's question and proceeded to implement his program, running out of time for the session in the process.

Excerpt 3: "Selecting a random operator" excerpts

a) High-Performing Pair (driver: M, nav.: F)	b) Low-Performing Pair (driver: F, nav.: M)
<p><b>Driver (16:11:12):</b> i dont know how to let it pick a random one to show now <b>Navigator (16:12:10):</b> You mean a random operation symbol? <b>Driver (16:12:21):</b> no, <b>Driver (16:12:26):</b> like you see how i have 4 blocks? <b>Driver (16:12:32):</b> one for each option? <b>Driver (16:12:48):</b> you know how to let the program randomly picks one? <b>Navigator (16:12:52):</b> oh i see... <b>Navigator (16:13:08):</b> i don't know let me check the instructions again <i>(Driver experiments with program)</i> <b>Navigator (16:15:11):</b> &lt;copy/pasted instructions and answer&gt; <i>(Driver experiments with program)</i> <b>Navigator (16:18:34):</b> btw the guy said we have 10 min left <i>(Driver experiments with program)</i> <b>Driver (16:23:38):</b> DONE</p>	<p><b>Driver (14:51:00):</b> I still dont understand how to choose a random operator though <b>Navigator (14:51:02):</b> yeah, we just need to get the question to display and then say correct or incorrect when given the use input <i>(Driver experiments with program)</i> <b>Navigator (14:54:47):</b> Is there like one big "say" button to display the equation all at once? <b>Driver (14:55:27):</b> Hmm.. <b>Driver (14:55:27):</b> not that I see <i>(Driver experiments with program)</i> <i>(Session runs out of time)</i> <b>Navigator (14:59:11):</b> Nice job we were close!</p>

Excerpts 1 and 2 provided evidence that student collaboration in high-performing pairs encouraged specific exploration and properly managed the pair's cognitive load. Excerpt 3 is consistent with the previous two.



In Excerpt 3 part a, the driver made sure that the navigator understood his source of uncertainty, and both partners were able to engage in specific exploration with respect to their given roles: the driver experimented with the program, and the navigator reviewed the task instructions. Additionally, by only focusing on one task, the pair had enough processing power in their working memory to address the current task. In contrast, Excerpt 3 part b suggests that the students were unable to surpass their confusion, and they may have suffered the effects in terms of cognitive load from not addressing one source of uncertainty before moving on to the next task. In the excerpt, the driver expressed uncertainty, the navigator gave feedback, but the driver did not acknowledge this feedback and continued to work on the task. The navigator also did not provide more feedback on the driver's actions. In contrast, the driver from the high-performing pair let the navigator know that he was able to solve the task and that they could move on; by not explicitly stating her process, the driver from the low-performing pair may have left them unable to address their uncertainty.

## Recommendations and limitations

The results described above suggest some recommendations for supporting students during uncertainty in collaborative problem solving. In Excerpt 1a we saw that the driver from the high-performing pair notified the navigator that he had solved the current subtask, while in Excerpt 1b the driver from the low-performing pair did not mention this in his dialogue. Adaptive scaffolding such as that provided by real-time intelligent learning environments could detect the expression of uncertainty during a subtask and prompt the collaborator to tell his partner when he believes the subtask is solved, addressing both students' uncertainty and providing an opportunity for learning. In Excerpt 2, the high-performing pair focused on one subtask while the low-performing pair switched between several subtasks and did not resolve their uncertainty, possibly due to an increased cognitive load. A real-time collaboration scaffolding system could encourage students to resolve uncertainty in one subtask before moving on to the next. Finally, Excerpt 3 shows how communication between the driver and navigator differed in high- and low-performing groups. The navigator in the high-performing pair asked the driver a follow-up question and maintained open communication regarding what she was doing; the navigator from the low-performing pair steered the focus away from the main task and towards another task. A potential suggestion to assist this pair would have been to encourage the navigator to converse with the driver more by asking her about her thought process and providing feedback. Through this interaction, the partners can achieve a common understanding of each other's process and have a clearer picture of what remains to be completed.

The results discussed in this paper must be interpreted in light of its limitations. First, the sample of student participants was based on volunteers who received a small amount of course credit in exchange for completing an alternate assignment. The nature of this homework credit may have introduced bias in the sample. Another limitation involves the implementation of the pair programming roles. Usually, students within a pair alternate between the driver and navigator roles. Due to the technical limitations of the screensharing software, students in our study were not able to switch roles during the activity. Finally, whether these results will generalize to other populations of students or other collaborative paradigms remains to be seen.

## Conclusion

The relationship between uncertainty and task performance during collaboration is complex. We have observed that a larger number of uncertainty utterances were expressed in dialogues of high-performing pairs and low-performing pairs when compared to medium-performing pairs. These expressions of uncertainty are clearly important, and the ways they are dealt with is different between pairs that do well and those who do not. We found that low-performing pairs missed some opportunities for learning and may have pushed their collective working memory to the limit when attempting to multitask. The results suggest that the ways in which collaborators express and address uncertainty could be highly influential in their success in a learning activity, and highlight the importance of supporting this aspect of collaboration.

There are several important directions for future work. Continuing to investigate practices for resolving uncertainty in collaborative problem solving is an important step toward more effectively supporting learners. If uncertainty during collaborative problem-solving activities can be identified, adaptive scaffolds may be able to promote and support specific exploration, reducing the negative effects of unresolved uncertainty and improving learning. Future work should investigate how this adaptive support can manifest itself effectively within collaborative problem solving. Designing and evaluating different forms of support for collaborative problem solving can lead to the next generation of adaptive scaffolding that holds the potential to significantly improve the learning experience.

## References

Berlyne, D. E. (1978). Curiosity and learning. *Motivation and Emotion*, 2(2), 97–175.

- Brown, P., & Levinson, S. C. (1978). Universals in language usage: Politeness phenomena. *E. N. Goody (Ed.), Questions and politeness: strategies in social interaction*, 56–311.
- Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & LaVanher, C. (1994). Eliciting Self-Explanations Improves Understanding. *Cognitive Science*, 18(3), 439–477.
- D’Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145–157.
- Falkner, K., Falkner, N. J. G., & Vivian, R. (2013). Collaborative Learning and Anxiety: A phenomenographic study of collaborative learning activities. *Proceedings of the 44th ACM Technical Symposium on Computer Science Education (SIGCSE)*, 227–232.
- Festinger, L. (1962). *A Theory of Cognitive Dissonance*. Stanford University Press.
- Forbes-Riley, K., & Litman, D. (2011). Benefits and Challenges of Real-Time Uncertainty Detection and Adaptation in a Spoken Dialogue Computer Tutor. *Speech Communication*, 53(9–10), 1115–1136.
- Gee, J. P. (2014). *An Introduction to Discourse Analysis: Theory and Method*. Routledge.
- Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question Asking and Eye Tracking during Cognitive Disequilibrium: Comprehending Illustrated Texts on Devices When the Devices Break Down. *Memory & Cognition*, 33(7), 1235–1247.
- Howley, I., Mayfield, E., & Rosé, C. P. (2011). Missing Something? Authority in Collaborative Learning. *Proceedings of the 9th International Conference on Computer-Supported Collaborative Learning (CSCL)*, 366–373.
- Keltner, D., & Shiota, M. N. (2003). New Displays and New Emotions: A Commentary on Rozin and Cohen (2003). *Emotion*, 3(1), 86–91.
- Litman, D., & Forbes-Riley, K. (2014). Evaluating a Spoken Dialogue System that Detects and Adapts to User Affective States. *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, 181–185.
- Markkanen, R., & Schröder, H. (1997). Hedging: A Challenge for Pragmatics and Discourse Analysis. *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts*, 3–20.
- Nagappan, N., Williams, L., Ferzli, M., Wiebe, E., Miller, C., Balik, S., & Yang, K. (2003). Improving the CS1 Experience with Pair Programming. *Proceedings of the 34th ACM Technical Symposium on Computer Science Education (SIGCSE)*, 359–362.
- Nelson, L. M. (1998). *Collaborative Problem Solving: An Instructional Theory for Learning through Small Group Interaction*. Indiana University.
- Porter, L., & Simon, B. (2013). Retaining Nearly One-Third more Majors with a Trio of Instructional Best Practices in CS1. *Proceeding of the 44th ACM Technical Symposium on Computer Science Education (SIGCSE)*, 165–170.
- Renkl, A., & Atkinson, R. K. (2010). Structuring the Transition From Example Study to Problem Solving in Cognitive Skill Acquisition: A Cognitive Load Perspective. *Educational Psychologist*, 38(1), 15–22.
- Rodríguez, F. J., Price, K. M., & Boyer, K. E. (2017). Exploring the Pair Programming Process: Characteristics of Effective Collaboration. *Proceedings of the 48th ACM Technical Symposium on Computer Science Education (SIGCSE)*, To appear.
- Rosé, C. P., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning. *International Journal of Computer-Supported Collaborative Learning*, 3, 237–271.
- Sharma, K., Jermann, P., Nüssli, M.-A., & Dillenbourg, P. (2013). Understanding Collaborative Program Comprehension: Interlacing Gaze and Dialogues. *Proceedings of the 10th International Conference on Computer-Supported Collaborative Learning (CSCL)*, 430–437.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.
- Webb, N. M. (1982). Student Interaction and Learning in Small Groups. *Review of Educational Research*, 52(3), 421–445.

## Acknowledgments

We would like to thank the members of the LearnDialogue Group for their helpful input. This material is based upon work supported by the National Science Foundation under grants CNS-1622438, DUE-1625908, and a Graduate Research Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

# Collaborative and Individual Scientific Reasoning of Pre-Service Teachers: New Insights Through Epistemic Network Analysis (ENA)

Andras Csanadi, Ludwig Maximilian University of Munich, andras.csanadi@psy.lmu.de  
Brendan Eagan, University of Wisconsin-Madison, beagan@wisc.edu  
David Shaffer, University of Wisconsin-Madison, dws@education.wisc.edu  
Ingo Kollar, University of Augsburg, ingo.kollar@phil.uni-augsburg.de  
Frank Fischer, Ludwig Maximilian University of Munich, frank.fischer@psy.lmu.de

**Abstract:** When assessing scientific reasoning both (1) modeling connections in the discourse and (2) doing so at an appropriate grain size can be challenging for researchers. Our study suggests combining a novel theoretical (Fischer et al., 2014) and a novel methodological (Shaffer et al., 2006) framework to respond to these challenges by detecting epistemic networks of scientific reasoning processes in the context of collaborative vs individual problem solving of pre-service teachers. We investigated (1) whether the combination of these frameworks can be fruitfully applied to model scientific reasoning processes and (2) what unit of analysis researchers or instructors should choose to answer questions of interest. One novel aspect of our study is that we compared epistemic networks in case of collaborative vs individual reasoning processes. Our results show that (1) epistemic networks of scientific reasoning can reliably capture reasoning processes when comparing collaborative vs individual reasoning; and (2) propositional and potentially larger units might be considered as “optimal” units of analysis to detect such differences.

**Keywords:** collaborative problem solving, epistemic network analysis, scientific reasoning

## Introduction

Assessment of scientific reasoning in process data is a critical for the development of appropriate learning support. Although many fruitful approaches have been developed for the evaluation of reasoning and argumentation (Brown, Furtak, Timms, Nagashima & Wilson, 2010); general theoretical and methodological frameworks that allow analysis of scientific reasoning patterns on multiple layers (e.g., Chi, 1997) are scarce. Consequently, the selection of grain size at an early stage of the analysis and a resulting dilemma surrounding creation of larger units that allow further interpretation of the data (e.g., Weinberger & Fischer, 2006) often limit the generalizability of findings (Chi, 1997; Stegmann & Fischer, 2011). Also, using a pre-defined selection of a unit of analysis might cause difficulties when a researcher or a tutor would like to be more conclusive about the reasoning processes: simultaneously making qualitative and quantitative assessments. For example, a researcher (or tutor) may be interested in ideas, or codes, at a very fine grained (e.g., propositional) level in order to detect “elementary” units of reasoning processes. Meanwhile, she might be also interested in the connections, or relationships, between these ideas or codes captured at that fine-grained level, in order to assess the quality of reasoning processes (Chi, 1997; Weinberger & Fischer, 2006). Moreover, when aggregating data into larger chunks, what would be an optimal choice? Would combining multiple propositions or defining a larger, e.g. sentence units, lead to better representation of reasoning processes? The present study investigates whether a combination of a novel theoretical framework on scientific reasoning (Fischer et al., 2014) as well as a novel methodological approach on modelling reasoners’ epistemic networks (Shaffer, 2006) can be meaningfully combined 1) to analyze patterns (epistemic networks) of scientific reasoning and 2) to disambiguate the question on grain size selection and data aggregation when assessing patterns (epistemic networks) of scientific reasoning.

## Scientific reasoning and argumentation

There are different theoretical frameworks to conceptualize and analyze scientific reasoning. Many follow a “structural” approach, focusing on the structure of argumentation (see Brown et al., 2010) while others emphasize the role of engagement in scientific reasoning processes (Okada & Simon, 1997). Our work belongs to the latter stream of research understanding scientific reasoning as engagement of individuals or groups in a sequence of epistemic activities (Fischer et al., 2013). According to this model, scientific reasoning involves reasoners identifying an existing problem (Problem identification), articulating questions of how to proceed with their reasoning processes (Questioning), derive possible explanations of the problem (Hypothesis generation), construct artifacts, such as intervention plans, to solve the problem (Generating solutions), generate

and collect information (Evidence generation), evaluate that information (Evidence evaluation), engage others in the reasoning process (Communicating & scrutinizing), and draw conclusions (Drawing conclusions). Earlier studies found that both individual and collaborative reasoning in a professional problem solving context can be reliably coded using this framework (Csanadi, Kollar & Fischer, 2016).

### **Collaborative vs. individual scientific reasoning processes**

Collaborative scientific reasoning has the potential to lead individuals to higher engagement in epistemic processes such as hypothesis generation and evidence evaluation compared to reasoning alone (Okada & Simon, 1997; Teasley, 1995). Similarly, more recent findings (Csanadi et al., 2016) showed that when pre-service teachers solved a problem from their future practice as dyads, they engaged more in hypothesis generation (i.e., trying to find an explanation to the problem) but less in generating solutions than individuals did. Nevertheless, this purely frequency-based approach for analysis to count the occurrence of certain codes has clear constraints. Most importantly, it cannot be conclusive enough regarding the patterns of epistemic processes that can characterize collaborative vs individual reasoning. For example, although dyads were found to be more explanatory, indicated by a higher engagement in hypothesizing, whether they did this in a more evidence-based manner (i.e. if they made more connections between hypothesizing and evaluating evidence) remained unclear. Being able to identify such connections or patterns in the data is, therefore, important for assessing quality aspects of scientific reasoning.

### **Selection of grain size and data aggregation to capture patterns of reasoning**

To assess and compare reasoners with respect to the patterns of the epistemic activities they engage in, researchers should find answers to two related questions. First, what is an appropriate grain size (i.e., unit of analysis) and second, how should coded data be aggregated in order to gain a deeper understanding of the quality and features of the reasoning processes. Many researchers emphasize that data segmentation should be a separate and preceding step to coding (Chi, 1997; Strijbos, Martens, Prins & Jochems, 2006). This would mean that the division of verbal data into chunks that carry meaningful information for further analysis should precede further analyses. However, this early selection of the unit of analysis has its limitations (e.g., Chi, 1997). Especially the use of smaller grain sizes (e.g., propositional unit) allow for a more fine-grained analysis of reasoning processes (e.g., to interpret the relation between independent clauses of compound sentences) and allow for frequency-based analyses. Indeed, many quantitative approaches to the analysis of scientific reasoning processes (e.g., Okada & Simon, 1997) suggest analyzing frequencies of single categories. However, considering that discourse moves are not unrelated to each other, relying on solely frequency-based information of data can lead to missing meaningful patterns of discourse (Cress & Hesse, 2013). At this point an emerging concern of data aggregation (Stegmann & Fischer, 2011), i.e., how the researcher/tutor can make higher level inferences based on data coded at a lower grain size, often generates uncertainty. When looking for relationships between coded units (e.g., propositions), how far these units can fall from each other? Can we meaningfully detect relationships between two neighboring units or does allowing for slightly “longer distance” connections increase explanatory power? A method that allows more adaptable choice of grain size (Siebert-Evenstone et al., 2016), such as considering multiple units of analysis instead of relying on a pre-defined selection in order to model scientific reasoning could help to answer such questions.

Another issue associated with coding-independent segmentation may arise if some codes turn out to be highly frequent ones while others occur relatively rarely. “Uneven” frequency distributions can bias further analyses of the dataset (e.g., Csanadi, Daxenberger, Ghanem, Kollar, Fischer & Gurevych, 2016). For example, high frequency codes might generate many connections with each other while also being related to many other codes. On the other hand, low frequency codes may lack enough connections with other codes to demonstrate the power to discriminate between epistemic networks of different groups (e.g., dyads vs individuals). Thus, in case of modeling reasoning processes, this can mean that some reasoning patterns may emerge as mere artifacts while other connections in the data may remain undetected, and therefore, models of scientific reasoning should account for such limitations.

To summarize, using a hierarchical segmentation procedure and reliance on solely frequency-related information when analyzing scientific reasoning processes and comparing reasoners, leaves open the questions of (1) how to aggregate and identify meaningful larger patterns in the data that can (2) help more validly capture the reasoning performance beyond simply counting the occurrences of single codes.

### **Epistemic Network Analysis: A method to analyze (multiple scopes of) scientific reasoning**

One solution of the abovementioned problems can be to code on multiple levels of granularity (Stegmann & Fischer, 2011). As Chi (1997) notes, this approach has the advantage of leading to more reliable results and interpretations at different levels. Generally speaking, segmentation might be a matter of the researchers' focus of interest (Chi, 1997), the theoretical framework they apply (Clara & Mauri, 2010), the nature of data (e.g. synchronous vs asynchronous discussions) and more. Still, selecting multiple levels of analysis can contribute to more valid interpretations about the data (Chi, 1997; Weinberger & Fischer, 2006) as different lenses may capture different aspects of collaborative learning and reasoning processes.

Epistemic Network Analysis (ENA; Shaffer, 2006) is a method to identify meaningful and quantifiable patterns in discourse/reasoning. It can provide an alternative to the widespread "code and count" approach. ENA moves beyond the traditional frequency-based assessments by examining the structure of the co-occurrence, or connections in coded data. Moreover, compared to other methodological approaches, e.g., sequential analysis (see in Cress & Hesse, 2013), ENA has the novelty of (1) modeling whole networks of connections and (2) it affords both quantitative and qualitative comparisons between different network models.

A main theoretical assumption of ENA is that repeated co-occurrences of two or more codes in the discourse can reveal epistemic networks which characterize an underlying Discourse (Gee, 1999; Collier et al., 2016), e.g., to collaborative (vs. individual) scientific reasoning. To identify a unit of analysis for calculating such co-occurrences, ENA provides an adaptable feature: the *moving stanza window size* (MSWS; Siebert-Evenstone et al., 2016). The term stanza window refers a window or scope within which ENA is searching for connections. This means that a MSWS=1 allows search for connections only between a proposition of reference and its preceding proposition. Therefore, a MSWS=1 results in connections only between neighboring propositions. A MSWS=2, however, allows one further step: it allows connection between a proposition of reference and the two preceding propositions. By changing MSWS from smaller values to larger it is possible to open the "search window" from very narrow context to wider ones. As a result, the researcher or tutor can look for connections not only within propositions (as in case of "coding and counting" approaches) or between neighboring propositions, but even between propositions that are two, three or more steps further from each other in the discourse. In short, it offers the advantage of multiple scopes for analysis. Here we aim to investigate if ENA can reveal some characteristics of collaborative (compared to individual) scientific reasoning processes as well as to articulate what grain sizes should be considered when using ENA for that analysis.

Furthermore, ENA provides the opportunity to quantitatively and qualitatively compare different epistemic network models with each other. Quantitative comparison is possible by using calculated centroids for every epistemic networks generated by ENA. Such centroid values are determined by the strength of connections between nodes in the epistemic network. Nodes are the codes (such as epistemic activities, see below) while the strength of connections between them are generated based on their local co-occurrences (within each stanza window: see above). These centroid values can be used for quantitative analyses. Furthermore, qualitative comparison of epistemic networks is possible using various options for visualization. One option is "Subtracting networks" which means contrasting two network models by subtracting their nodes and connections weights from each other. A resulting "subtracted network" represents the difference between two reasoning networks and therefore, can illustrate what makes dyadic reasoning different from individual reasoning.

## Research questions

RQ1: Do collaborative and individual reasoners exhibit different epistemic networks of scientific reasoning while solving a professional problem?

While earlier studies demonstrated differences between collaborative and individual reasoning in terms of their engagement in different epistemic activities (Csanadi et al., 2016; Okada & Simon, 1997), these results were mainly frequency-based. E.g., the researchers compared proportions as well as raw frequencies of engagement in different epistemic activities, such as evaluating evidence or hypothesizing. Thus, an open question is whether dyads also differ from individuals in the patterns of epistemic activities they engage in during scientific reasoning. In this study we address this question using ENA (Shaffer et. al. 2009) to capture meaningful patterns of co-occurrences between epistemic activities (i.e., epistemic networks of scientific reasoning), and to compare dyads with individual reasoners.

Epistemic networks can, however, also be defined based on larger speech units (e.g., across multiple propositions) and we can also implement larger grain sizes beyond analyzing neighboring propositions or within sentences. To fully answer RQ1, therefore, we investigated whether some grain sizes can provide potentially better explanation of patterns in the data than others.

RQ2: Do the epistemic networks we detect investigating RQ1 differ from epistemic networks based on the same data set that has been randomly resorted (i.e. with the same frequency information)?

ENA models co-occurrences of codes, since some codes occur more frequently than others, it is more likely that these highly frequent codes make connections (co-occur) with other codes more often than lower frequency codes. Consequently, ENA may “overestimate” some connections. Therefore, to answer our second research question, we compared ENA results from RQ1 to ENA results obtained from a dataset that contained only frequency information of the original discourse (see below). If the epistemic networks identified in relation to RQ1 cannot be explained merely by the frequency distribution of epistemic activities, the epistemic networks detected in relation to RQ1 should differ from the epistemic networks of the randomly resorted dataset.

## Method

The data analyzed in this study is a re-analysis of process data from another study (Csanadi et al., 2016). In the original study  $N=76$  preservice teachers (59 female,  $M_{Age}=21.22$ ,  $SD_{Age}=3.98$ ) solved a problem case from their future profession in one of two between-subject conditions: either as individuals ( $N=16$ ) or as dyads ( $N=30$  dyads). Think aloud and discourse data of their problem solving were first manually segmented into propositional units and then coded for further analysis. The coding scheme of that study was developed based on the framework of scientific reasoning by Fischer et al. (2014). Epistemic activities identified by the framework (see above) were applied (Table 1): Problem identification for an initial attempt to build an understanding of the problem; Questioning for statements or questions triggering further inquiry; Hypothesis generation for developing explanations of the problem; Evidence generation for reference to information or lack of information that could support a claim; Evidence Evaluation to evaluate a claim; Communicating and scrutinizing for planned discussions with others (e.g., in order to find out further information); Drawing conclusions for concluding outcomes of reasoning. Finally, the epistemic activity of “Constructing artefacts” (in Fischer et al., 2014) was operationalized as developing interventions or solution plans, and such propositions were labelled as Generating solutions. Moreover, the codes for Evidence generation and Evidence evaluation were merged into Evidence evaluation. Both segmentation (79.73% of agreement by Coder 1 and 85.09% of agreement by Coder 2) and coding ( $\kappa = 0.68$ ) proved to be reliable. We used this dataset (original dataset) to analyze further in our present study.

We used the abovementioned original dataset to answer RQ1. To be able to answer RQ2 we created a randomized dataset in the following way. Using the original dataset within each dyad and individual participants we created a random sequence of the pre-segmented propositions (Csanadi et al., 2016). That meant, the original sequence of propositions were randomized while the relative frequency of propositions was preserved (no propositions were deleted). This new randomized dataset preserved the information of the occurrence of epistemic activities, yet, in a randomized order; containing the information to which individual or dyad the epistemic activities belong to, how frequently they occur, but without any information regarding their sequence in the original dataset.

We used ENA to identify epistemic networks of scientific reasoning in order to answer both RQ1 and RQ2. We built epistemic network models using ENA in four steps. First, we calculated co-occurrences between epistemic activities (MSWS=1, means rotation was applied) for dyads and for individuals. At the same time ENA automatically generated a centroid value for each dyad or individual that served as a numeric representation of their epistemic network and it was included in further analysis to compare dyadic and individual epistemic networks of scientific reasoning. Second, mean, or “average,” networks were defined for both the dyadic and the individual reasoning conditions, respectively. Each of these networks visually represented all the connections that participants (dyads or individuals) generated in the given condition. Third, we quantitatively compared epistemic networks for dyads with epistemic networks for individuals by comparing the mean centroid values (calculated in step 1) in the two conditions. Fourth, we subtracted the mean dyadic and mean individual networks from each other (by using the “Subtracting networks” option in ENA). The resulting subtracted networks visualized what connections contributed to the difference between the two reasoning conditions (dyadic vs individual, calculated in step 3).

To be able to fully answer RQ1 regarding grain size, we sequentially set MSWS from 1 to 7, step-by-step, performing the same analysis for each stanza window size. The resulting epistemic network models at each MSWS level allowed us quantitative as well as qualitative (visual) comparisons.

To answer RQ2, we used the randomized dataset selecting the same parameters and performing the same analysis as in case of RQ1. We compared the outcomes of this analysis with the ENA results from RQ1.

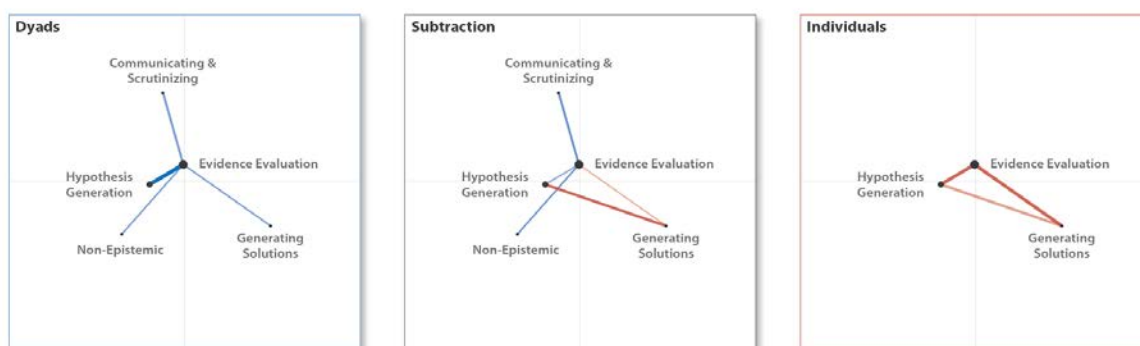
Table 1: Coding scheme for epistemic activities

Code	Short Description	Example
------	-------------------	---------

Problem identification	An attempt to understand the problem.	"So it is about a student, // who has low grades"
Questioning	A question orienting inquiry.	"Ok, so what is the reason for that?"
Hypothesis generation	Explanation of the problem.	"...the reason is her learning method"
Evidence generation	Referring to any information/ lack of inf. relevant for the inquiry	"She studies diligently at home"
Evidence evaluation	Evaluation information.	"..you can even exclude the problem of exam nerves"
Generating solutions	Planning an intervention/ solution to the problem.	"You should discourage her from using surface strategies"
Communicating & scrutinizing	Planning to engage others.	"You can also talk to the parents"
Drawing conclusions	Concluding the outcomes of the earlier steps of inquiry.	"For me these would be the most important points..."
Non-epistemic	Everything else, e.g. coordination.	"Ok, have you read it through?"

## Results

RQ 1: To answer RQ1, as a first step, we compared dyadic and individual networks at the grain size of MSWS=1 which lead to the following results. The mean centroid value for individuals' epistemic networks ( $M=.21, SD=.32$ ) was significantly different from the mean centroid value for dyads' epistemic networks ( $M=-.11, SD=.21$ ),  $t(44)=3.65, p<.01, d=1.32$ . Plotting epistemic networks (Figure 1) further revealed that the central epistemic activity accounting for most of the connections was evidence evaluation. Moreover, in case of dyads evidence evaluation showed more complex network than in case of individuals: for dyads it was connected to hypothesis generation, communicating and scrutinizing, generating solutions and non-epistemic propositions; while in the case of individuals it was only connected to hypothesis generation and generating solutions. Finally, subtracting individual from dyadic networks revealed that in case of individual networks it was solution generation rather than evidence evaluation that played a central role in contrast to dyadic networks where only evidence evaluation showed multiple connections after subtraction.



**Figure 1.** Epistemic networks of dyads (blue, left), individuals (red, right) and the difference between their networks (center) using the original dataset.

To completely answer RQ1 and in order to see whether there is an optimal grain size that can best capture the differences between epistemic networks of dyads and individuals, we compared epistemic networks at  $1 < MSWS \leq 7$  levels which led to the following results. All comparisons were statistically significant at least under  $p<.01$ . Although effect size showed a small increase at every MSWS level, these differences were small: the explained variance increased only by 5.35% ( $\Delta R^2=.05$ ) from MSWS=1 ( $R^2=.30$ ) to MSWS=7 ( $R^2=.36$ ).

Finally, a visual inspection of the epistemic networks conducted at  $1 \leq \text{MSWS} \leq 7$  levels suggested highly similar patterns at every MSWS levels (see Figure 1).

RQ 2: Similar to the outcomes of RQ1, when using the randomized dataset, the mean centroid value for individuals' epistemic networks ( $M=.17, SD=.26$ ) was significantly different from the mean centroid value for dyads' epistemic networks ( $M=-.09, SD=.20$ ),  $t(44)=3.35, p<.01, 95\%, d=1.15$ . Plotting epistemic networks (Figure 2), however, revealed no visible difference between dyadic and individual networks. Dyadic and individual networks showed identical patterns regarding complexity: connections occurred among the three most frequent epistemic activities: hypothesis generation, solution generation and evidence evaluation. This was in clear contrast with the results of RQ1 where epistemic networks were different for collaborative vs individual reasoning (Figure 1). A further important difference is that Figure 2 does not indicate any central epistemic activity, neither for dyadic and individual nor for the subtracted pattern. Moreover, Figure 2 shows very low level of network complexity for dyads (connections among the highest-frequency activities) compared to Figure 1. Finally, the subtracted network model on Figure 2 consists of only blue lines, indicating that dyads made more connections among the highly frequent codes than individuals.

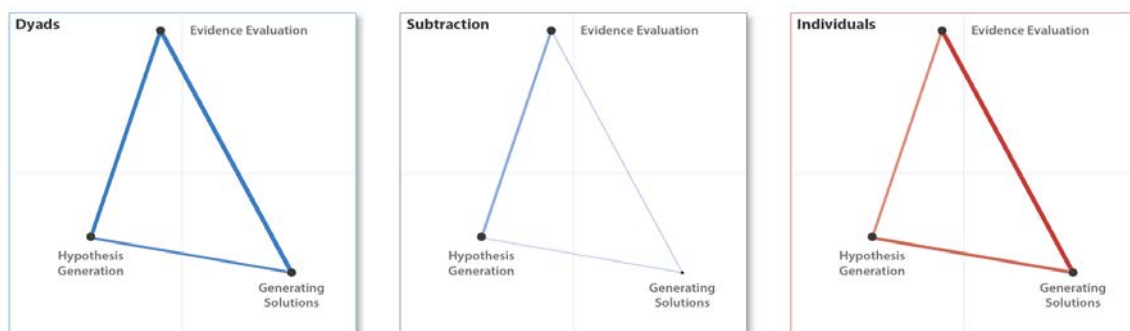


Figure 2. Epistemic networks of dyads (blue, left), individuals (red, right) and the difference between their networks (center) using the randomized dataset.

## Discussion

The two main aims of our study were (1) to see whether we can aggregate data to capture meaningful patterns (epistemic networks) of scientific reasoning processes regarding collaborative and individual reasoning (RQ1 & RQ2) and (2) to search for an optimal grain size, or unit of analysis, for such aggregation (RQ1). We sought to answer these questions by the application of a novel theoretical framework on scientific reasoning (Fischer et al., 2014) and a novel methodological approach on modelling epistemic networks (Shaffer, 2006).

The outcomes for RQ1 suggest that epistemic networks of scientific reasoning can meaningfully differentiate between collaborative and individual reasoning processes. More specifically, dyads seemed to engage in a more complex manner in scientific reasoning compared to individuals: they made more connections between epistemic activities (specifically, with evidence evaluation). Moreover, while individual reasoning was rather solution-focused; dyadic reasoning seemed to be more evidence-focused. These results are also in accordance with previous frequency-based findings (Csanadi et al., 2016; Okada & Simon, 1997).

To be able to fully answer RQ1 we ran further analyses at different stanza window sizes that resulted in patterns quite similar to those in Figure 1. On the one hand, this suggests the robustness of our findings, on the other, a question of the optimal grain size to detect meaningful patterns of scientific reasoning cannot be conclusively answered. A partial answer is, however, that choosing larger speech unit (e.g., sentences) at a first step may represent reasoning patterns in the data at least closely as well as propositions do. Yet, further empirical research could test (1) whether this is true and if (2) varying stanza window sizes on sentence units would lead to different results. Based on the results of this study and considering the exhaustiveness of hand-coding procedure, however, choosing larger units of analysis that still carry the information needed to model scientific reasoning may be an efficient choice for the researcher/tutor.

The outcomes on RQ2 show that epistemic networks extracted on discourse data (original dataset) are likely to be valid models for the evaluation of reasoning patterns in the data as they are not reducible to the frequency distribution of codes. Furthermore, it is clear that merely frequency-information in the data resulted in only "poor" network models: networks represented solely the most frequent codes and their connections. Additionally, after subtracting those networks the results suggested that dyads made more connections everywhere. These results did not add much explanatory value to the frequency-based outcomes of the earlier



findings (Authors, 2016a), which underlines the assumption that ENA conducted on real discourse data can detect meaningful patterns of scientific reasoning.

Finally, the results imply that identifying epistemic processes on the propositional level and aggregating data by conducting epistemic network analysis can offer a powerful way to meaningfully assess scientific reasoning in discourse.

## Final conclusions

Our results have further important consequences.

First, the theoretical (Fischer et al., 2014) and the methodological (Shaffer, 2009) frameworks could be fruitfully combined to result in a series of robust analyses of identifying epistemic networks of scientific reasoning.

Second, dyadic vs. individual reasoning networks can be valid models of scientific reasoning in discourse. Yet, we need more empirical research to see if this result holds as well as see the predictive validity of our findings. For example, the extent to which dyads' more extensive connections could potentially predict learning outcomes and whether some connections might play a stronger moderating role in that process, are questions for future research.

Finally, additional analyses that can more directly address the impact of frequency distribution of codes on epistemic networks could also contribute to conclusions regarding the validity of the findings. For example, alternative measures provided by ENA could account for "imbalanced" frequency distribution in the data. Those measures could apply, for example, some weighting method for assigning less weight to higher frequency codes or to connections among higher frequency codes, in order to reduce the chance of detecting artefactual connections due to higher probability of co-occurrence between high-frequency codes. Similarly, if ENA could generate a simple frequency-based epistemic network model (similar to the outcomes on RQ2) and would allow its subtraction from the epistemic network model on the real dataset; that would afford the visualization of reasoning patterns beyond highly frequent connections. Yet, such measures should be implemented with caution: connections captured in the discourse should always represent connections in the Discourse (Gee, 1999).

## References

- Brown, N. J., Furtak, E. M., Timms, M., Nagashima, S. O., & Wilson, M. (2010). The evidence-based reasoning framework: Assessing scientific reasoning. *Educational Assessment, 15*(3-4), 123-141.
- Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences, 6*(3), 271-315.
- Clarà, M., & Mauri, T. (2010). Toward a dialectic relation between the results in CSCL: Three critical methodological aspects of content analysis schemes. *International Journal of Computer-Supported Collaborative Learning, 5*(1), 117-136.
- Cress, U., & Hesse, F. W. (2013). Quantitative methods for studying small groups. In C. E. Hmelo-Silver, C. A. Chinn, C. K. Chan, & A. O'Donnell (Eds.), *The international handbook of collaborative learning*, 85-111. New York: Routledge, Taylor & Francis.
- Csanadi, A., Kollar, I., Fischer, F. (2016). Scientific reasoning and problem solving in a practical domain: Are two heads better than one? In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The International Conference of the Learning Sciences (ICLS) 2016*, Vol. 1 (pp. 50-57). Singapore: International Society of the Learning Sciences.
- Csanadi, A., Daxenberger, J., Ghanem, C., Kollar, I., Fischer, F., & Gurevych, I. (2016a). *Automated text classification to capture scientific reasoning and argumentation processes in different professional problem solving contexts*. Paper presented at the 26th Annual Meeting of the Society for Text & Discourse, Kassel, Germany.
- Collier, W., Ruis, A. R., & Shaffer, D. W. (2016). Local versus global connection making in discourse. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) Vol. 1* (pp. 426-433). Singapore: International Society of the Learning Sciences.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., . . . Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research, 5*, 28-45.
- Gee, J. P. (1999). *An introduction to discourse analysis*. New York: Routledge.
- Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *The Journal of the Learning Sciences, 14*(2), 201-241.

- Shaffer, D.W. (2006). Epistemic frames for epistemic games. *Computers and Education* 46(3): 223-234.
- Shaffer, D. W., Hatfield, D. L., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E. A., ... Frank, K. (2009). Epistemic Network Analysis: A prototype for 21st century assessment of learning. *International Journal of Learning and Media*, 1(1), 1–21.
- Siebert-Evenstone, A. L., Arastoopour, G., Collier, W., Swiecki, Z., Ruis, A. R., & Shaffer, D.W. (2016). *In search of conversational grain size: Modeling semantic structure using moving stanza windows*. Paper presented at the 12th International Conference of the Learning Sciences, Singapore.
- Stegmann, K., & Fischer, F. (2011). Quantifying qualities in collaborative knowledge construction: the analysis of online discussions. In S. Putambekar, G. Erkens, C. Hmelo-Silver (Eds.). *Analyzing Interactions in CSCL: Methods, approaches and issues* (pp. 247-268). Springer US.
- Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. (2006). Content analysis: What are they talking about? *Computers & Education*, 46(1), 29-48.
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & education*, 46(1), 71-95.

## Acknowledgments

This research was supported by the Elite Network of Bavaria under Grant K-GS-2012-209. This work was funded in part by the National Science Foundation (DRL-0918409, DRL-0946372, DRL-1247262, DRL-1418288, DUE-0919347, DUE-1225885, EEC-1232656, EEC-1340402, REC-0347000), the MacArthur Foundation, the Spencer Foundation, the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

# Time and Semantic Similarity – What is the Best Alternative to Capture Implicit Links in CSCL Conversations?

Gabriel Gutu, Mihai Dascalu, Traian Rebedea, and Stefan Trausan-Matu,  
gabriel.gutu@cs.pub.ro, mihai.dascalu@cs.pub.ro, traian.rebedea@cs.pub.ro, stefan.trausan@cs.pub.ro  
University Politehnica of Bucharest, Computer Science Department

**Abstract:** The goal of our research is to compare novel semantic techniques for identifying implicit links between utterances in multi-participant CSCL chat conversations. Cohesion, reflected by the strength of the semantic relations behind the automatically identified links, is assessed using WordNet-based semantic distances, as well as unsupervised semantic models, i.e. Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA). The analysis is built on top of the *ReaderBench* framework and multiple identification heuristics were compared, including: semantic cohesion metrics, normalized cohesion measures and Mihalcea’s formula. A corpus of 55 conversations in which participants used explicit links between utterances where they considered necessary for clarity was used for validation. Our study represents an in-depth analysis of multiple methods used to identify implicit links and reveals the accuracy of each technique in terms of capturing the explicit references made by users. Statistical similarity measures ensured the best overall identification accuracy when using Mihalcea’s formula, while WordNet-based techniques provided best results for un-normalized similarity scores applied on a window of 5 utterances and a time frame of 1 minute.

## Introduction

Chat represents a commonly used collaboration tool nowadays that can also be successfully employed in learning processes, such as Computer Supported Collaborative Learning (CSCL) (Stahl, 2006). Creativity fostering is a key element in multi-participant chat conversations (Trausan-Matu, 2010), where multiple changes of perspectives and of points of interest are frequently encountered, which are helpful for social knowledge building in educational settings. However, these mixed discussion topics and threads may be difficult to follow and understand and thus, participants add explicit links to previous utterances when they have this facility (Holmer, Kienle, & Wessner, 2006) in order to ensure threading and, consequently, coherence between utterances. As this is a cumbersome task for many users, which tend to introduce few explicit links (if any), the need of automation has dramatically increased when analyzing chats of hundreds of utterances and with more than 3 participants. This process of linking related utterances is referred to as *implicit links detection* (Trausan-Matu & Rebedea, 2010) and represents an important step that allows the integration of additional operations on texts such as topic mining, sentiment analysis, detection of lexical chains, and evaluating the degree of collaboration in problem solving and CSCL.

Natural Language Processing (NLP) techniques (Manning & Schütze, 1999) are more and more used nowadays since they provide efficient analyses of written texts. In contrast to other types of web collaboration tools such as forums or social networks, most chat systems do not provide any “*reply-to*” option. This lack makes difficult to follow the threads of discussions in chats with more than two participants, generating discourse segmentation. Therefore, due to this lack of a referencing facility in the clear majority of online chats, the usage of NLP tools for the detection of implicit links between utterances represents an important research topic.

The purpose of this paper is to determine which state of the art semantic similarity measure performs best for the detection of implicit links in multi-party chat conversations and what is the optimum distance in terms of utterances and time frames to look for them. The corpus for this comparative analysis consists of a collection of 55 conversations lasting up to two hours, performed by computer science students from our faculty using the *ConcertChat* environment (Holmer, Kienle, & Wessner, 2006), which enables users to explicitly reference previous utterances. Within these conversations, participants had to discuss about the benefits and disadvantages of each several web collaboration technologies (i.e., wiki, blog, forum chat) and identify the most suitable tool to be used by an enterprise (Trausan-Matu & Rebedea, 2010). This collection represents a refined version of an initial set of 200 conversations used in previous studies (Gutu, Rebedea, & Trausan-Matu, 2015) based on the following criteria: multiple conversation sessions with the same participants were discarded, as well as discussions with a limited timeframe (less than 30 minutes), too few utterances (less than 50 utterances) or too few explicit links added by users (less than 10 explicit links per conversation).

In terms of the structure of the paper, the following section provides general information about semantic similarity and introduces the five measures available in the *ReaderBench* framework (Dascalu, Dessus, Bianco, Trausan-Matu, & Nardy, 2014; Dascalu et al., 2015a; Dascalu et al., 2015b): *Latent Semantic Analysis* (LSA)

(Landauer & Dumais, 1997), *Latent Dirichlet Allocation* (LDA) (Blei, Ng, & Jordan, 2003) and three *WordNet*-based distance functions: Leacock Chodorow (Leacock & Chodorow, 1998), Wu Palmer (Wu & Palmer, 1994) and path length (Budanitsky & Hirst, 2006). The third section presents the results of our analysis alongside statistical information about the chat corpus. In the end, conclusions and future work are presented.

## Related work on semantic models

Semantic cohesion reflects the degree to which two text fragments are related one to another in terms of meaning (Bestgen, 2012) and can be automatically evaluated using several approaches. In previous studies in the Natural Language Processing field, several techniques gained high popularity. The first one consists of applying different semantic distance functions on ontologies (Budanitsky & Hirst, 2006), such as the WordNet lexical database (Miller, 1995). Second, Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997) is the most frequently used method to compute semantic similarity by relying on vector spaces of keywords (terms). Third, probabilistic topic modeling has gained an increasing attention lately, Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003) being the most frequently used method of this kind.

LSA (Landauer & Dumais, 1997) is a NLP technique based on a vector space model highlighting term co-occurrences within documents. LSA is frequently used to compute the similarity between documents and between terms (Manning & Schütze, 1999). LSA is based on a “bag of words” approach in which word order is disregarded and words’ occurrences are normalized through Term frequency – Inverse Document frequency or log entropy. A singular-value decomposition (SVD) (Golub & Reinsch, 1970), followed by a reduction of the matrices’ dimensionality through a projection on  $k$  dimensions is performed in order to determine indirect links induced between groups of terms and underlying documents. The optimal empiric range for  $k$  is  $300 \pm 50$  (Landauer, McNamara, Dennis, & Kintsch, 2007; Lemaire, 2009). LSA can be perceived as a mathematical optimization for representing the meaning of words and group of words in a vector space by adopting an unsupervised learning technique applied on a corpus of texts. Our LSA model was trained on a pre-processed version of a custom corpus obtained by concatenating the TASA corpus (Touchstone Applied Science Associates, Inc., <http://lsa.colorado.edu/spaces.html>) that contains general texts, novels and newspaper articles and a corpus of more than 500 CSCL-related scientific papers. Stop-words and non-dictionary words were disregarded, inflectional word forms were reduced to their lemmas and only paragraphs with more than 20 content words were considered.

LDA (Blei, Ng, & Jordan, 2003) is a generative probabilistic process built on top of the assumption that documents integrate multiple topics and can be therefore considered a mixture of corpus-wide topics. Each topic represents a Dirichlet distribution (Kotz, Balakrishnan, & Johnson, 2000) over the vocabulary where related concepts have similar probabilities based on co-occurrence patterns from the training corpora. Although each topic contains all the words from the vocabulary, a clear differentiation in terms of corresponding probabilities can be observed between salient versus dominant concepts. Similar to LSA, LDA relies on the “bag of words” approach and classifies new texts in terms of the latent topics inferred from the model trained on a text collection. Documents and words alike become topics distributions drawn from Dirichlet distributions, while semantic similarities between textual fragments are determined using the Jensen-Shannon dissimilarity (JSH) (Manning & Schütze, 1999), a symmetric smoothed alternative of the KL divergence (Kullback & Leibler, 1951). A drawback of the traditional LDA model is that it uses an imposed number of topics. (Teh, Jordan, Beal, & Blei, 2006) have introduced an alternative, Hierarchical Dirichlet Process (HDP), a nonparametric Bayesian approach used to cluster grouped data. The HDP mixture model can be considered as a generalization of LDA in which the number of topics is inferred from the training text corpora (Teh, Jordan, Beal, & Blei, 2006). HDP applied on the custom-built TASA–CSCL corpus inferred 118 topics as the optimal number of topics.

WordNet (Miller, 1995) represents a lexical ontology having words organized in four different trees based on their corresponding parts of speech: nouns, verbs, adverbs or adjectives. The hierarchical representation of words using synsets (sets of synonyms) describes relations between them. Thus, a word which is a descendant of another concept in WordNet’s internal representation is a more specific term, while the parent represents a more general concept. Given WordNet’s internal representation, several distance measures were developed, out of which the most notable three, implemented in the *ReaderBench* framework, were used in this study: Leacock Chodorow (Leacock & Chodorow, 1998), Wu Palmer (Wu & Palmer, 1994) and path length (Budanitsky & Hirst, 2006), all presented in detail in Table 1.

Table 1: Semantic distances based on WordNet that are used in the study

Semantic distance	Formula	Description
Leacock Chodorow	$sim_{LC}(c_1, c_2) = -\log \frac{l(c_1, c_2)}{2D}$	The path length normalized by the depth of the ontology, where depth represents the path length from the current concept to the global root.
Wu Palmer	$sim_{WP}(c_1, c_2) = \frac{2 \times d(lso(c_1, c_2))}{l(c_1, lso(c_1, c_2)) + l(c_2, lso(c_1, c_2)) + 2 \times d(lso(c_1, c_2))}$	Similarity is computed considering the depths of the two concepts, as well as their least common ancestor.
Path length	$l(c_1, c_2)$	The length of the shortest path between the two concepts.

## Method

Two accuracy measures were considered within this study: a) *exact* implicit links detection, when the computed reference is the same as the explicit reference attribute set by the chat participant, and b) *in-turn* implicit links detection, when the computed reference belongs to the same turn (i.e., a collection of adjacent utterances belonging to the same participant, including the utterance mentioned within the explicit link). Our corpus of 55 chat conversations was initially cleaned using several NLP refinements (Manning & Schütze, 1999): *stop-words* (words with no semantic relevance and no contextual information) were eliminated, duplicate words frequently encountered in chat conversations were removed and the remaining words were lemmatized using the CoreNLP library (Manning et al., 2014).

Table 2 introduces two examples of identified implicit links extracted from the same conversation. The examples show the differences between in-turn matching, when implicit links must belong to the same participant in a continuous block of utterances, while in case of exact-matching implicit links must overlap perfectly with the explicit links defined by the user. The first example shows utterance with id 74 having an explicit reference to utterance 65 which was manually added by a participant (explicit links are in the third column – *Ref. ID*). However, when imposing a window of maximum 5 utterances and 1 minute time frame, the detected implicit link is utterance 72 (emphasized) using the Path Length similarity measure. As turn 72 is enclosed in a continuous series of utterances belonging to the same user (i.e., Monica), we consider this to be a correct *in-turn matching*, but an incorrect *exact matching*. In the second excerpt, the identified implicit link for utterance 140 is turn 138 using the same parameters for time frame, distance and semantic similarity. Turn 138 is also the explicit reference for utterance 140, as can be easily observed from the *Ref. ID* column, and this is a correct *exact matching*.

Table 2: Excerpt from a chat conversation

Measure	Utt. ID	Ref. ID	Speaker	Time	Content
In-turn matching	65		Monica	09:08:27	features to add RSS feeds, file sharing and so on
	... (several utterances of the same participant, Monica) ...				
	<b>72</b>		<b>Monica</b>	<b>09:09:57</b>	<b>and they embed only what you need</b>
	73		Monica	09:10:16	users tend to be scared away by a multitude of features that they need to figure out
	74	65	Razvan	09:10:22	The thing that I think would be a problem with wikis is that they will not allow a person to keep track of the latest information added. Ok RSS are good but not everybody wants to use an RSS feed reader.
...					
Exact matching	135		Monica	09:21:18	actually apart from the user setup
	136		Monica	09:21:30	that personally I find quite an effort
	137		Monica	09:21:41	blogs are a good solution
	<b>138</b>		<b>Stefan</b>	<b>09:22:01</b>	<b>you know the biggest disadvantage of wikis? that anybody can input and that makes wikis a not-so-reliable source of info</b>
	139	138	Alex	09:22:20	that depends on the configuration
	140	138	Razvan	09:22:24	you could have admins that check the information

## Results

Several measurements centered on per-chat analyses and distance-related or time-related statistical tests were performed on the corpus of chat conversations before running the implicit links detection process. The conversation corpus totals 17,612 utterances ( $M = 320.22$ ,  $SD = 136.04$ ,  $min = 79$ ,  $max = 817$ ) with an average of 4.35 participants per conversation ( $SD = 0.97$ ,  $min = 3$ ,  $max = 8$ ) and 4,463 explicit references in total ( $M = 81.15$ ,  $SD = 45.29$ ,  $min = 15$ ,  $max = 226$ , per chat). The average coverage (i.e., the percentage of referred utterances from the total number of utterances from a chat) was 28.62% ( $SD = 16.62\%$ ,  $min = 5\%$ ,  $max = 72.9\%$ ). The average time duration of a conversation is about 2 hours.

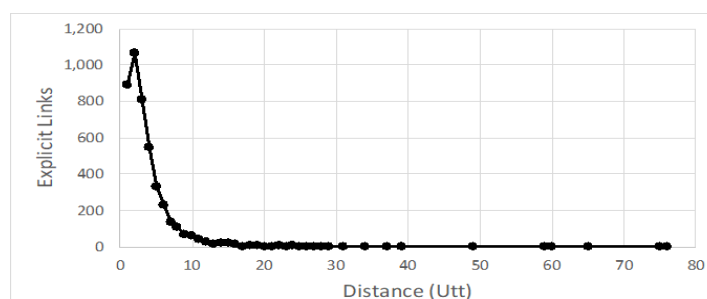


Figure 1. Distribution of explicit links per distance between turns.

Table 2 presents the statistical data from which it can be easily observed that the number of references decreases exponentially with distance (see Figure 1 for a visual representation of the distances' distribution for the entire corpus of conversations). Moreover, distances (D) of 1 to 10 cover more than 95% of the references, while a window of 20 utterances covers more than 99% of potential explicit links. In terms of authors of the original and referred utterance, as expected, there is a higher percentage of explicitly linked utterances belonging to different speakers ( $M = 88.51\%$ ) than to the same chat participant for a window size of 20 adjacent utterances.

Figure 2 presents the graphical evolution of the coverage of explicit links as a function of distance and time. We can observe in a visual manner that a window size of 20 utterances ensures the coverage of most (99%) explicit links, while a distance of 10 utterances enables a sufficient level of certainty (covering more than 95% of the total links). Given these distributions, we decided to compute the semantic similarity between each utterance and the previous ones considering window sizes of 20, 10 and 5 utterances respectively.

Table 3: Explicit links occurrences and coverage in terms of distance

Distance	Explicit links	M (SD)	Min / Max	Percentage		Same speaker		Different speaker	
				Local	Cumulative	#	%	#	%
1	890	16.18 (13.10)	0 / 54	19.94%	19.94%	213	23.93%	677	76.07%
2	1065	19.36 (11.92)	1 / 49	23.86%	43.80%	126	11.83%	939	88.17%
3	810	14.73 (9.84)	2 / 42	18.15%	61.95%	79	9.75%	731	90.25%
4	548	9.96 (6.56)	1 / 31	12.28%	74.23%	84	15.33%	464	84.67%
5	332	6.04 (4.61)	0 / 24	7.44%	81.67%	44	13.25%	288	86.75%
6	230	4.18 (3.36)	0 / 16	5.15%	86.83%	26	11.30%	204	88.70%
7	134	2.44 (2.39)	0 / 11	3.00%	89.83%	10	7.46%	124	92.54%
8	106	1.93 (1.69)	0 / 7	2.38%	92.20%	6	5.66%	100	94.34%
9	68	1.24 (1.39)	0 / 6	1.52%	93.73%	5	7.35%	63	92.65%
10	62	1.13 (1.33)	0 / 6	1.39%	95.12%	7	11.29%	55	88.71%
11	43	0.78 (1.03)	0 / 4	0.96%	96.08%	6	13.95%	37	86.05%
12	30	0.55 (0.83)	0 / 3	0.67%	96.75%	6	20.00%	24	80.00%
13	16	0.29 (0.60)	0 / 2	0.36%	97.11%	2	12.50%	14	87.50%
14	20	0.36 (0.36)	0 / 2	0.45%	97.56%	3	15.00%	17	85.00%
15	21	0.38 (0.71)	0 / 3	0.47%	98.03%	1	4.76%	20	95.24%
16	17	0.31 (0.57)	0 / 2	0.38%	98.41%	1	5.88%	16	94.12%
17	4	0.07 (0.33)	0 / 2	0.09%	98.50%	0	0.00%	4	100.00%
18	9	0.16 (0.46)	0 / 2	0.20%	98.70%	3	33.33%	6	66.67%
19	6	0.11 (0.37)	0 / 2	0.13%	98.83%	1	16.67%	5	83.33%
20	4	0.07 (0.26)	0 / 1	0.09%	98.92%	1	25.00%	3	75.00%

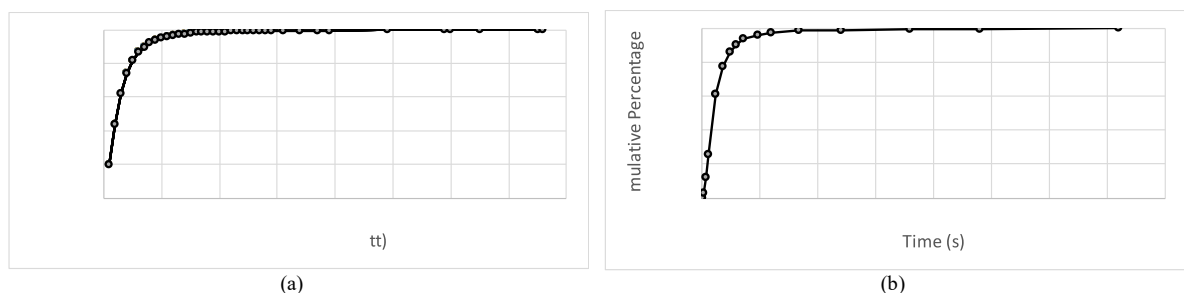


Figure 2. Cumulative coverage of explicit links (a) per distance and (b) per time.

Regarding the time difference between utterances, several time frames were set and explicit links' coverage within these time frames were computed. Table 3 shows the coverage of the links within the selected time frames. Significant changes in terms of cumulative percentage were desired, which made us to select 5 time frames for the study: 30 seconds, 1, 2, 3 and 5 minutes. As it was observed, within 5 minutes 97% of the links are covered, while the time frame of 1 minute covers 61% of them.

### Implicit links identification

Three metrics were used in order to identify the best semantic match between the current utterance and all previous ones within the imposed window sizes: semantic similarity (SIM), normalized similarity by inverse distance between current utterance and referred utterance (NSIM), and Mihalcea's similarity formula (MSIM) (Mihalcea, Corley, & Strapparava, 2006). SIM represents the baseline similarity metric for each semantic model; for example, when talking about LSA, we use SIM to refer to the standard formula for computing LSA cosine similarity. NSIM is used to refer to a normalized value of the previously introduced similarity. The specific formula developed by Mihalcea was introduced as a third metric named here MSIM. In a nutshell, SIM, NSIM and MSIM refer to specific optimization formulas which rely on the underlying semantic models.

The implicit link reflects the highest score within the window in terms of distance or time frame, namely the pair of utterances having the highest semantic similarity between them. Accuracy was assessed in terms of the overlap between the automatically identified implicit links and the explicit ones added by users on two criteria: *exact matching* and *in-turn matching*, presented in pairs delimited by slash in all subsequent tables. Table 5 presents the percentage of detected explicit links using both exact and in-turn accuracy measures per window size. Table 6 presents the same percentages computed per time frame. Emphasized values represent the semantic similarity formula that provided the best accuracy for each technique and for each windows size.

Table 4: Explicit links occurrences and coverage in terms of time frame

Time	Explicit links	Cumulative Percentage	Same speaker		Different speaker	
			#	%	#	%
1 sec	1	0.03%	0	0.00%	1	100.00%
2 secs	1	0.03%	0	0.00%	1	100.00%
3 secs	2	0.05%	0	0.00%	2	100.00%
5 secs	4	0.11%	1	25.00%	3	75.00%
10 secs	100	2.71%	39	39.00%	61	61.00%
20 secs	459	12.42%	126	27.45%	333	72.55%
30 secs	957	25.89%	222	23.20%	735	76.80%
1 min	2246	60.77%	364	16.21%	1882	83.79%
1.5 mins	2850	77.11%	423	14.84%	2427	85.16%
2 mins	3176	85.93%	453	14.26%	2723	85.74%
2.5 mins	3342	90.42%	467	13.97%	2875	86.03%
3 mins	3451	93.37%	477	13.82%	2974	86.18%
4 mins	3553	96.13%	487	13.71%	3066	86.29%
5 mins	3593	97.21%	493	13.72%	3100	86.28%
7 mins	3634	98.32%	496	13.65%	3138	86.35%
10 mins	3658	98.97%	499	13.64%	3159	86.36%
15 mins	3673	99.38%	503	13.69%	3170	86.31%
20 mins	3682	99.62%	505	13.72%	3177	86.28%
30 mins	3685	99.70%	505	13.70%	3180	86.30%
1 hour	3688	99.78%	507	13.75%	3181	86.25%
1.5 hours	3689	99.81%	507	13.74%	3182	86.26%
2 hours	3690	99.84%	507	13.74%	3183	86.26%
> 2 hours	3696	100.00%	508	13.74%	3188	86.26%

**Table 5: Percentage of correctly matched explicit links per window size (Exact matching / In-turn matching)**

Window size	Measure	LSA	LDA	Leacock	Wu-Palmer	Path length
20 (chance = 5%)	SIM	17.47% / 22.99%	14.84% / 20.48 %	15.46% / 21.55%	17.99% / 24.11%	20.27% / 26.6 %
	NSIM	<b>29.27% / 37.98%</b>	<b>26.48% / 35.88%</b>	<b>26.40% / 35.82%</b>	<b>27.48% / 36.72%</b>	<b>31.00% / 39.98%</b>
	MSIM	17.97% / 23.47%	19.02% / 24.35%	14.03% / 19.40%	16.67% / 22.10%	20.05% / 26.54%
10 (chance = 10%)	SIM	23.41% / 31.19%	20.18% / 28.30%	21.67% / 30.03%	24.69% / 32.81%	26.62% / 34.99%
	NSIM	<b>29.82% / 38.65%</b>	<b>26.53% / 35.88%</b>	<b>27.03% / 36.44%</b>	<b>28.08% / 37.31%</b>	<b>31.45% / 40.48%</b>
	MSIM	24.03% / 31.36%	24.93% / 32.76%	19.22% / 27.31%	21.86% / 29.67%	25.00% / 32.80%
5 (chance = 20%)	SIM	29.23% / 38.79%	27.31% / 37.84%	<b>27.65% / 37.49%</b>	<b>30.19% / 39.87%</b>	31.66% / 41.03%
	NSIM	<b>30.50% / 39.49%</b>	26.35% / 35.72%	27.29% / 36.89%	28.31% / 37.62%	<b>31.99% / 41.29%</b>
	MSIM	30.16% / <b>39.51%</b>	<b>30.08% / 39.82%</b>	25.76% / 35.97%	27.88% / 38.08%	30.05% / 39.43%

**Table 6: Percentage of correctly matched explicit links per time frame (Exact matching / In-turn matching)**

Time frame	Measure	LSA	LDA	Leacock	Wu-Palmer	Path length
5 mins	SIM	20.31% / 27.28%	16.85% / 23.70%	18.21% / 25.46%	20.53% / 27.40%	22.51% / 29.62%
	NSIM	<b>29.22% / 37.91%</b>	<b>26.28% / 35.58%</b>	<b>26.19% / 35.46%</b>	<b>27.27% / 36.39%</b>	<b>30.85% / 39.71%</b>
	MSIM	20.50% / 27.03%	21.85% / 28.57%	16.40% / 23.59%	18.92% / 25.85%	22.03% / 29.06%
3 mins	SIM	25.22% / 33.50%	22.42% / 30.69%	24.15% / 32.86%	26.41% / 34.81%	28.12% / 36.49%
	NSIM	<b>29.98% / 38.74%</b>	26.29% / <b>35.55%</b>	<b>26.95% / 36.41%</b>	<b>27.94% / 37.12%</b>	<b>31.71% / 40.82%</b>
	MSIM	25.81% / 33.60%	<b>26.78% / 34.76%</b>	21.79% / 30.63%	24.15% / 32.74%	27.18% / 35.19%
2 mins	SIM	29.03% / 37.38%	26.75% / 35.95%	<b>27.52% / 36.95%</b>	<b>30.10% / 39.45%</b>	31.49% / 40.41%
	NSIM	<b>30.56% / 39.23%</b>	25.55% / 34.83%	26.85% / 36.24%	27.86% / 36.94%	<b>31.70% / 40.79%</b>
	MSIM	29.82% / 37.91%	<b>29.52% / 38.13%</b>	25.79% / 35.11%	27.69% / 36.61%	30.27% / 38.82%
1 min	SIM	28.74% / 37.20%	27.11% / 35.98%	<b>27.97% / 36.95%</b>	<b>29.12% / 37.96%</b>	<b>29.62% / 38.46%</b>
	NSIM	27.17% / 35.73%	22.87% / 31.75%	24.72% / 33.51%	25.41% / 34.07%	28.20% / 36.72%
	MSIM	<b>28.87% / 37.25%</b>	<b>28.78% / 37.15%</b>	26.89% / 35.71%	27.68% / 36.55%	29.04% / 37.73%
30 secs	SIM	<b>15.13% / 22.97%</b>	14.98% / 22.94%	<b>14.44% / 22.59%</b>	<b>14.48% / 22.62%</b>	<b>14.58% / 22.64%</b>
	NSIM	14.53% / 22.31%	13.20% / <b>20.78%</b>	13.67% / 21.58%	13.73% / 21.62%	14.11% / 22.12%
	MSIM	14.71% / 22.56%	<b>14.98% / 22.93%</b>	14.23% / 22.26%	14.25% / 22.36%	14.39% / 22.51%

All semantic measures per window size exceeded random chance, i.e. the percentage of matches by randomly selecting a reference from the imposed window. Normalized semantic similarity provided the best accuracy for most of the used techniques and for all the three selected window sizes: about 30% for perfect match and about 40% for in-turn match. As presented in Table 3, the window size of 10 utterances covers more than 95% of the explicit links; corroborated with a random chance of only 10%, NSIM produced significant results with almost 30% exact accuracy and 40% in-turn accuracy. Accuracies are just a little higher (less than 1%) for the window size of 20 which covers 99% of the explicit links, but with the disadvantage of doubling the computational effort. Thus, we conclude that the window size of 10 using the NSIM technique is the best alternative to be used for determining implicit links in chat conversations in terms of distance. The window size of 5 utterances provided close results, too. Both window sizes were considered for the further experiment.

In terms of time frames, the best results were obtained for the window size of 2 minutes, both for exact matching and in-turn matching. The 1-minute window provided close values to the 2-minutes window, hence both of the time frames were considered. Results for the subsequent analysis with window sizes of 5 and 10 utterances, and time frames of 1, respectively 2 minutes, are presented in Table 7. With regards to the optimal window to search for implicit links, for our corpus, a distance of 5 utterances and a time window of 1 minute provided the leading scores for both exact matching and in-turn matching. Furthermore, compared to previous results in Table 5 and Table 6, better similarity scores were obtained by narrowing both the distance and the time frame.

By analyzing the semantic similarity scores obtained for the chosen pairs of window sizes and time frames, the results show specificity based on the employed measure. Hence, using LSA, for a window of 10 utterances, the NSIM formula detects most links, while for a window of 5 utterances the MSIM formula is the best. When it comes to LDA, the MSIM formula detects most explicit links. The WordNet-based techniques provide the best results with the SIM formula. In terms of formula to be used, we can observe that WordNet-based techniques brought better results using the un-normalized semantic similarity, while for the statistical methods, namely LSA and LDA, Mihalcea's formula can be considered the best for most scenarios.



**Table 7: Percentage of correctly matched explicit links per chosen (window size, time frame) pairs (Exact matching / In-turn matching)**

Window size, time frame	Measure	LSA	LDA	Leacock	Wu-Palmer	Path length
5 utt, 1 min	SIM	30.68% / 39.72%	28.77% / 38.61%	<b>28.55% / 38.18%</b>	<b>30.95% / 40.44%</b>	<b>32.44% / 41.49%</b>
	NSIM	30.01% / 38.74%	25.43% / 34.73%	26.78% / 36.19%	27.73% / 36.81%	31.35% / 40.51%
	MSIM	<b>31.45% / 40.41%</b>	<b>30.98% / 40.09%</b>	27.06% / 36.75%	28.89% / 38.45%	31.04% / 39.97%
10 utt, 1 min	SIM	29.25% / 37.80%	27.12% / 36.27%	<b>27.82% / 37.30%</b>	<b>30.27% / 39.62%</b>	<b>31.88% / 40.78%</b>
	NSIM	<b>30.28% / 38.94%</b>	25.55% / 34.83%	26.91% / 36.29%	27.88% / 36.94%	31.57% / 40.65%
	MSIM	30.28% / 38.48%	<b>30.08% / 38.67%</b>	26.16% / 35.59%	27.99% / 36.95%	30.60% / 39.16%
5 utt, 2 mins	SIM	30.68% / 39.72%	28.77% / 38.61%	<b>28.55% / 38.18%</b>	<b>30.95% / 40.44%</b>	<b>32.44% / 41.49%</b>
	NSIM	30.01% / 38.74%	25.43% / 34.73%	26.78% / 36.19%	27.73% / 36.81%	31.35% / 40.51%
	MSIM	<b>31.45% / 40.41%</b>	<b>30.98% / 40.09%</b>	27.06% / 36.75%	28.89% / 38.45%	31.04% / 39.97%
10 utt, 2 mins	SIM	29.25% / 37.80%	27.12% / 36.27%	<b>27.82% / 37.30%</b>	<b>30.27% / 39.62%</b>	<b>31.88% / 40.78%</b>
	NSIM	30.28% / <b>38.94%</b>	25.55% / 34.83%	26.91% / 36.29%	27.88% / 36.94%	31.57% / 40.65%
	MSIM	<b>30.28%</b> / 38.48%	<b>30.08%</b> / <b>38.67%</b>	26.16% / 35.59%	27.99% / 36.95%	30.60% / 39.16%

## Conclusions and future work

In this research, our aim was to compare various methods for the identification of implicit links by computing semantic similarity between every utterance and the previous utterances, within imposed distance and time windows. The performed comparative analysis provides evidence that using a combined window of the previous 5 utterances and 1-minute pair provides the best trade-off in terms of both exact and in-turn accuracies for the implicit links detection process. Plus, WordNet-based techniques provide the best overall results if un-normalized semantic distances are used, while statistical techniques such as LSA and LDA provide best results when Mihalcea's formula is employed. Current results surpass previous studies that used only un-normalized semantic similarity measures (Rebedea, Chiru, & Gutu, 2014; Rebedea & Gutu, 2013) and provide a deeper analysis on how to achieve the best accuracy for identifying implicit links in chat conversations. Chat conversations represent a modern tool used for collaborative working and are widely used in educational environments and learning tasks as they allow automatic analyses of natural language such as topic mining, sentiment analysis, lexical chains and others. Detection of implicit links established between utterances represents an initial step in processes that refer to text cohesion in general. The results presented in this paper are intended to facilitate the development of a more advanced chat tool that can automatically extract such NLP-related information.

Practical outcomes of implicit links identification include the detection of students' interaction patterns in ongoing conversations, as well as the extraction of lexical chains and of discussion threads. Integrated within a chat client or any CSCL environment housing multi-participant conversations, these features may provide better awareness to users in terms of understanding and following the current discussion threads. Moreover, the identified implicit links facilitate tutor assessment while highlighting members' active involvement, as well as their collaboration with other participants.

However, several adjustments are considered for increasing the accuracy of the implicit links identification process. First, machine learning techniques could be used to create an aggregated similarity score relying on multiple semantic measures. Second, dynamic sliding windows could be enforced by considering cut-offs induced by topic changes or long pauses within the discourse. Third, certain patterns extracted using speech acts (e.g., continuations, question answering) (Searle, 1969) and discourse connectors may be indicative of implicit links within the discourse. For example, the presence of a contrast connective ("but") or of a question answering sequence may indicate an implicit link between two utterances that might have a low in-between semantic similarity score.

## References

- Bestgen, Y. (2012). Évaluation automatique de textes et cohésion lexicale. *Discours*, 11. doi: 10.4000/discours.8724
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(4-5), 993-1022.
- Budanitsky, A., & Hirst, G. (2006). Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics*, 32(1), 13-47.
- Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S., & Nardy, A. (2014). Mining texts, learner productions and strategies with ReaderBench. In A. Peña-Ayala (Ed.), *Educational Data Mining: Applications and Trends* (pp. 345-377). Cham, Switzerland: Springer.

- Dascalu, M., Stavarache, L. L., Dessus, P., Trausan-Matu, S., McNamara, D. S., & Bianco, M. (2015a). ReaderBench: An Integrated Cohesion-Centered Framework. In G. Conole, T. Klobucar, C. Rensing, J. Konert & É. Lavoué (Eds.), *EC-TEL 2015* (pp. 505–508). Toledo, Spain: Springer.
- Dascalu, M., Stavarache, L. L., Trausan-Matu, S., Dessus, P., Bianco, M., & McNamara, D. S. (2015b). ReaderBench: An Integrated Tool Supporting both Individual and Collaborative Learning. In *5th Int. Learning Analytics & Knowledge Conf. (LAK'15)* (pp. 436–437). Poughkeepsie, NY: ACM.
- Golub, G. H., & Reinsch, C. (1970). Singular value decomposition and least squares solutions. *Numerische Mathematik*, *14*(5), 403–420.
- Gutu, G., Rebedea, T., & Trausan-Matu, S. (2015). A comparison of semantic similarity techniques for a corpus of CSCL chats. In *RoEduNet 2015* (pp. 178–183): IEEE.
- Holmer, T., Kienle, A., & Wessner, M. (2006). Explicit Referencing in Learning Chats: Needs and Acceptance. In W. Nejdl & K. Tochtermann (Eds.), *EC-TEL 2006* (pp. 170–184). Crete, Greece: Springer.
- Kotz, S., Balakrishnan, N., & Johnson, N. L. (2000). Dirichlet and Inverted Dirichlet Distributions *Continuous Multivariate Distributions* (Vol. 1: Models and Applications, pp. 485–527). New York, NY: Wiley.
- Kullback, S., & Leibler, R.A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics*, *22*(1), 79–86.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*(2), 211–240.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Erlbaum.
- Leacock, C., & Chodorow, M. (1998). Combining local context and WordNet similarity for wordsense identification. In C. Fellbaum (Ed.), *WordNet: An electronic lexical database* (pp. 265–283). Cambridge, MA: MIT Press.
- Lemaire, B. (2009). Limites de la lemmatisation pour l'extraction de significations. In *9es Journées Internationales d'Analyse Statistique des Données Textuelles (JADT2009)* (pp. 725–732). Lyon, France: Presses Universitaires de Lyon.
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In *52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 55–60). Baltimore, MA: ACL.
- Mihalcea, R., Corley, C., & Strapparava, C. (2006). *Corpus-based and knowledge-based measures of text semantic similarity*. Paper presented at the 21st Int. Conf. AAAI, Boston, Massachusetts.
- Miller, G.A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*(11), 39–41.
- Rebedea, T., Chiru, C.-G., & Gutu, G. (2014). How useful are semantic links for the detection of implicit references in CSCL chats? In *RoEduNet Conference 2014* (pp. 1–6): IEEE.
- Rebedea, T., & Gutu, G. (2013). Detecting Implicit References in Chats Using Semantics *Scaling up Learning for Sustained Impact* (pp. 627–628): Springer.
- Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, UK: Cambridge University Press.
- Stahl, G. (2006). *Group cognition. Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.
- Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). Hierarchical Dirichlet Processes. *Journal of the American Statistical Association*, *101*, 1566–1581.
- Trausan-Matu, S. (2010). Computer support for creativity in small groups using chats. *Annals of the Academy of Romanian Scientists, Series on Science and Technology of Information*, *3*(2), 81–90.
- Trausan-Matu, S., & Rebedea, T. (2010). A polyphonic model and system for inter-animation analysis in chat conversations with multiple participants. In A. F. Gelbukh (Ed.), *11th Int. Conf. Computational Linguistics and Intelligent Text Processing (CICLing 2010)* (pp. 354–363). New York: Springer.
- Wu, Z., & Palmer, M. (1994). Verb semantics and lexical selection. In *32nd Annual Meeting of the Association for Computational Linguistics, ACL '94* (pp. 133–138). New Mexico, USA: ACL.

## Acknowledgments

This research was partially supported by the FP7 2008-212578 LTfLL project, by the 644187 ECH2020 *Realising an Applied Gaming Eco-system* (RAGE) project, as well as by University Politehnica of Bucharest through the “Excellence Research Grants” Programs UPB–GEX 12/26.09.2016 and 13/30.09.2016.

# The Dangers of Assuming Before Analysis: Three Case Studies of Argumentation and Cognition

Kristine Lund and Matthieu Quignard

kristine.lund@ens-lyon.fr, matthieu.quignard@ens-lyon.fr  
ICAR, CNRS, Ecole Normale Supérieure Lyon, University of Lyon

**Abstract:** In this article, we argue that researchers in the hypothetico-deductive tradition expose themselves to dangers when they first make assumptions about theoretical constructs and second, when they gather data specifically in order to test the predictions arising from hypotheses about these constructs. We review our own research in computer supported collaborative learning and in computer supported collaborative work on argumentation and cognition to show ways to partially surmount these dangers while raising new methodological concerns. We conclude by underlining the importance of the role of theory and the importance of reflecting on these issues, especially in a multidisciplinary community such as the learning sciences.

## Our vision of theory may expose us to dangers

The ways in which a researcher leverages theory in the human sciences directs the types of questions asked, the type of participants who will be studied, the way data will be collected, how the data will be analyzed and how results will be presented (Kawulich, 2009). In the hypothetico-deductive framework, researchers make hypotheses about theoretical constructs and data is gathered specifically in order to test the predictions arising from hypotheses. In what follows, we review three of our own research projects, mainly situated within the hypothetico-deductive approach. Our goal is to illustrate a set of dangers that a researcher could encounter when making assumptions about theoretical constructs before analysis and to show how each danger can be partially surmounted. We hope to encourage reflection on the possible consequences of the epistemological foundations of chosen research methods.

## Making theoretical assumptions about conceptual constructs prior to collecting data

The hypothetico-deductive approach in Computer Supported Collaborative Learning is illustrated by what can be called the Munich group (e.g. Stegmann & Fischer, 2011; Weinberger, Ertl, Fischer & Mandl, 2005). Theirs is a brand of research where theoretical assumptions are made before gathering the data that will be used to test them. This does not mean that theory is not also elaborated from previous empirical studies; it just means that a question is formulated before collecting the data for a particular study aimed at answering that question. In what follows, we note examples of these researchers' questions, the type of participants they study, their data collection, analysis and presentation of results and show how they are all influenced by their adopted theoretical position. The Munich group builds studies in order to test theoretical assumptions of relations between specific features of collaborative processes and successful knowledge construction. A theory used in this sense of the term is a general proposition, or a logically connected system of general propositions, which establishes a relationship between two or more variables, in general – independently of time and place (Abend, 2008).

## Research questions and the notion of learning

The hypothetico-deductive method is used where hypotheses are proposed, predictions are generated and then tested in an experimental setting in order to find out to what extent the hypothesis is founded. For example, a typical question would be: "To what extent are specific epistemic activities (e.g. relating conceptual and problem space) associated with improved domain knowledge among participating individuals?" (Stegmann & Fischer, 2011). The prediction corresponding to this question is that completeness of single arguments during discussions is positively related to depth of cognitive processing of the individual and that longer argumentation sequences (those that include for example counter-arguments) are positively related to improved domain knowledge among participating individuals. Another example research question is: "What are the effects of an epistemic script and a social script and their combination on the individual acquisition of knowledge as the outcome of collaborative learning in a text-based computer-supported peer discussion environment?" (Weinberger, et al. 2005). Predictions here involve both types of scripts (alone and in combination) enhancing individual knowledge acquisition in comparison with a non-scripted environment. Although these research questions are "confirmative", research questions can also be more "exploratory" in order to generate hypotheses (Stegmann & Fischer, 2011). Regarding both of the research questions mentioned above, learning is defined in

terms of individual acquisition of knowledge. Generally, such acquisition is measured by differences between a pre-test and a post-test, in between which an experiment with different conditions takes place (see more below).

#### Participants, data, analysis and presentation of results

When researchers studying learning use an experimental method and have predictions to test or questions to explore, they often choose their participants from students already taking a class and insert their pedagogical experiment into the lesson plan. The studies referred to in Stegmann & Fischer (op. cit.) were carried out at the University level and the example data shown were collected from an online discussion (threaded chat) that used a collaboration script to prompt students to elaborate grounds and qualifications for claims they make about a particular situation and so was specifically set up to test the aforementioned predictions arising from the theory and previous empirical studies.

Data were analyzed by applying measures that evaluated the quality of single arguments, the quality of argumentation sequences, and the quality of individual domain knowledge. The categories used in analyzing the quality of the collaborative process can be developed in a top down or bottom up manner. In the former, the effects and relations between the variables being examined have been well conceptualized in theory whereas in the latter, these effects and relations are to be explored during analysis. Often, there is a mix of both (Stegmann & Fischer, op. cit.). In such studies, results are presented in terms of the extent to which a particular experimental condition (analyzed in terms of the quality of their process) gave rise to a statistically significant result in terms of individual domain knowledge gains. In the case where no pre-test is possible (e.g. due to a lack of initial knowledge on the part of the participants or because the goal is to trace knowledge expressed individually during a post-test to knowledge expressed during collaboration), the post-test is evaluated alone for its expression of knowledge, using scales adapted to content.

Results are often expressed in terms of ANOVA calculations showing significant effects of experimental conditions on outcome measures. Treatment checks are sometimes also performed in order to ensure that participants reacted to the intervention as was intended by the experiment designer and these measures are also presented as results.

#### Summary of the hypothetico-deductive approach

This discussion of the typical experimental psychology approach to studying learning (some would say quasi-experimental as these studies are often not carried out in strict laboratory conditions, but rather in a classroom) shows that the consequences of a researcher's chosen theoretical framework are felt throughout the whole research process. The definition of learning is specific to the approach and the measurement of learning is highly dependent upon assumptions about it. In addition, this experimental psychology approach determines the way participants are chosen, data collected, analyzed and presented. The hypothetico-deductive method theorizes about the relations between conceptual constructs and their effect on outcome measures. Experimental research is accused of being less ecologically valid than, say, ethnomethodological research in addition to being "a product of the researcher's or informant's manipulation, selection, or reconstruction of preconceived notions of what is probable or important" (ten Have, p. 2), but ethnomethodological approaches are accused of not obtaining results that are as generalizable. That said, there are numerous studies that give reasons for experimental research failing to generalize, so this method is not immune to this criticism either. For example different results may be obtained when a variable that was not taken into consideration originally is paid attention to, thus illustrating the importance of context. As a case in point, McGarrigle & Donaldson (1974) showed how the experimenter crucially influenced children for Piaget's traditional conservation task. These authors described how children were taking into account the actions of the experimenter and not just the experiment, per se, when formulating their answers to the questions the experimenter asked them. In brief, if the experimenter was performing an action and then asking the children if something changed, they responded that something did change because why would the experimenter ask them if something had not changed? This taking into account theory of mind (Antaki, 2004) as an important element in interpreting experimental results was illustrated when the children responded correctly when a "naughty teddy bear" intervened, manipulated by the experimenter, and who, according to the experimenter, had the habit of "messing up toys" or "spoiling the game" (McGarrigle & Donaldson (1974). Bringing this alternative explanation of results to light gave a potential explanation for why it was difficult to replicate the age at which children were supposed to understand conservation of matter.

Clearly, experimental methods are designed with a mind to controlling conditions so that relations between constructs and outcomes can be clearly established, but it's also clear that researchers focus on particular aspects of the situation that they deem important, in light of their world view (e.g. what they consider is important to pay attention to in regards to learning), thereby allowing them to miss other influential factors. In a more basic sense, there can be problems with experimental protocols (e.g. small effect sizes, etc.), but that is

not our argument here. Instead, we argue that viewing methods and results that are derived from them with different perspectives is an effective way to eliminate alternative explanations for results and to thus have more confidence in the causal link between condition/treatment and outcome. Schooler (2014) argues for Metascience — the science of science — where the goal of researchers is to examine how scientific practices may influence the validity of scientific conclusions. It is an argument both for careful methodologies and recognition of underlying assumptions and biases.

In the remainder of this paper, we take a closer look at one particular type of danger in the hypothetico-deductive approach that researchers expose themselves to: formulating theoretical assumptions about conceptual constructs prior to collecting data. In what follows, we first recount an anecdote to illustrate what pre-theorizing before analysis can lead to. We then review our own work on cognition and argumentation and show how we surmounted a set of dangers due to pre-theorization. In doing so, we are giving the backstory to the studies we refer to, something that is rarely done. Granted, analytical methods courses address some of the issues we raise, but nevertheless, some researchers do not question how their underlying epistemologies influence their work (Lund, Rosé, Suthers, & Baker, 2013). Although we do not have space to fully compare the hypothetico-deductive method with ethnomethodological approaches — two important approaches in the CSCL community — it is our hope that building a narrative regarding some of our experiences with the first approach will enable young researchers to anchor such reflections in their own practice. We conclude by underlining the importance of the role of theory and the importance of reflecting on these issues, especially in a multidisciplinary community such as the learning sciences.

### **Changing your perception of data to fit your theory**

The following anecdote illustrates quite clearly one of the dangers of having assumptions about theoretical concepts before analysis. A researcher in particle physics was invited to the first author's home; let's call him Pierre. Upon seeing the pet cat, Pierre said: "That cat must be a female, it's a calico three-color cat". Indeed as genetic theory specifies (Kaelin & Barsh, 2013), the gene that determines how the orange color is displayed is on the X chromosome. Female cats have two X chromosomes whereas male cats have an X and a Y chromosome. A cat can only be calico if it has two X chromosomes and so in the majority of instances, the cat will be a female. However, a calico cat can be male if the cat has three sex chromosomes — two X and one Y, although this is extremely rare. When the first author told Pierre that the cat was male, he did a surprising thing. He altered his perception of the data. Instead of recognizing the three colors he originally saw, he said he instead now saw only two colors. It seems that Pierre was capable of changing what he saw in the data so that the data fit his theory; it was apparently too risky for him from a statistical point of view to imagine that he was looking at the rare case of a calico male.

This anecdote is similar to confirmation bias (Nickerson, 1998). This bias occurs when researchers actively seek evidence that confirms their hypothesis while ignoring evidence that could disconfirm it and is not limited to researchers trained in the exact sciences (e.g. particle physics). In the case of Pierre, even though he saw one thing initially, he changed his perception to be more compatible with what he should have been seeing, according to the theory. In fact, he distorted his view of the data (albeit presumably not intentionally) so that it would be in alignment with the expectations of the pre-chosen theory. Although such behavior can be considered to be a danger to good scientific inquiry, many researchers risk such behavior as choosing a theory that makes predictions about data before gathering data is a popular way of doing research — in the exact sciences, but also in the human and social sciences. In the next sections, we describe three of our own studies where we were exposed to the dangers of making assumptions about theoretical constructs before data analysis. Each study illustrates a particular danger and describes how we surmounted it.

### **The study of cognition and argumentation**

As with any domain of study, researchers interested in cognition and argumentation take different theoretical approaches. In this section, we concentrate on research falling mainly into the hypothetico-deductive approach, where conceptual constructs of argumentation are defined according to theory and then studied in computer-supported interaction. These conceptual constructs were assumed to predict learning gains in the first example, built into patterns that illustrated procedures of decision-making in the second example and used to predict the amount of conflict in Wikipedia in the third example.

### **Widen the context of the phenomenon of interest**

The danger of making assumptions about theoretical constructs illustrated by our first example is potentially not as serious as confirmation bias, but it is one where the researcher's focus may be considered to be too narrow and therefore has the consequence of excluding other phenomena that may also have repercussions for the

outcome measures. As Greeno (1998) points out, if we investigate cognitive subsystems, some general activity structures have to be arranged in which these subsystems function. It is currently the case that we don't understand the relations between the different subsystems, yet when we run an experiment, we are still assuming that the particular subsystem we are investigating does not significantly depend on how the other subsystems behave. This is most likely not the case and so we risk generalizing in an unjustified manner.

This research originated within the European project SCALE where we elaborated a coding scheme called Rainbow, for analyzing on-line pedagogical debates during which students could both chat and draw argument diagrams (Baker, Andriessen, Lund, Amesvoort & Quignard, 2007). The scheme was inspired both by theory (e.g. Toulmin, 1958; Barth & Krabbe, 1982; van Eemeren & Grootendorst, 1984; Plantin, 1990) and other coding schemes (e.g. Meier et al. 2007; Suthers, 2006; Andriessen, Erkens, van de Laak, Peters, & Coirier, 2003; van Bruggen and Kirschner, 2003 and Veerman, 2003; de Vries, Lund, & Baker, 2003) and further elaborated in confrontation with part of our own corpus. It was then applied to the rest of our corpus and validated with inter-coder reliability. We obtained results relating specific on-line distance pedagogical situations (e.g. using an argumentation diagram as a means of debate or using it as a means for representing a debate) and quality of an outcome measure of argumentation (e.g. Lund, Molinari, Séjourné & Baker, 2007), but we were also victims of our own pre-conceived view of what was important, according to our dominant theoretical framework. For example, we attributed the most importance to categories 5-7 (cf. Table 1) because that was where the conceptual notions under debate were being delineated, disentangled and deepened and we were interested in how such processes related to learning.

Table 1: The Rainbow coding scheme initially elaborated for analyzing pedagogical debates (cf. Baker, Andriessen, Lund, Amesvoort & Quignard, 2007)

1) Outside Activity	Any interaction that is not concerned with interacting in order to carry out the teacher/ researcher-defined task, including socio-relational interaction that does not relate to interacting in order to achieve the task, e.g., talk about last night's party.
2) Social Relation	Interaction that is concerned with managing the students' social relations with respect to the task (debating about X), e.g. greeting, leave-taking, politeness, expressions of frustration with the way the partner is interacting, etc.
3) Interaction management	Interaction concerned with managing the interaction itself: who will speak or not and when (coordination), establishing contact, perception, understanding, attitudes (communication management), topic shifting, time management, ...
4) Task management	Management of the progression of the task itself: planning what is to be discussed, establishing whether problem solved or not, ...
5) Opinions	Interaction concerned with expressing (stating, requesting) opinions (beliefs, acceptances, ...) with respect to the topic debated, especially (but not only) at opening and closing of sequences of argumentative discussion (dialectical outcomes).
6) Argumentation	Expression of (counter-)arguments directly related to a thesis (e.g. GMOs increase famine because farmers become dependent on seed companies), theses themselves, requests for justification
7) Explore and deepen	Interaction concerned with (counter-)arguments linked to (counter-)arguments, argumentative relations, and meaning of arguments themselves (elaboration of them, definition, extension, contraction, i.e. any discursive or conceptual operation performed on content of arguments themselves).

However, other types of interaction on which we were not immediately focused turned out to also be important for the way argumentation unfolded. For example, the dynamics of power and influence between group members can explain to what extent some members admit publicly to being influenced by other members in how they change their opinions (Molinari & Lund, 2012). In Baker, Andriessen & Lund, (2009), we explore extending the socio-cognitive paradigm to include how the dynamic interplay of emotions relate to processes of knowledge co-elaboration (see also Baker, Andriessen & Järvelä, 2013 and Polo, Lund, Plantin, & Niccolai, 2016). Finally, even a phenomenon specifically judged as initially uninteresting by the theoretical framework can be revealed to be interesting later on. For example, although spatially arranging elements of an argumentation graph might have been thought to be a waste of time, as opposed to arranging them logically, or thematically, it can also allow students to review the arguments discussed during the interaction and therefore increase the quality of their argumentative texts written afterwards (Baker, Quignard, Lund & Séjourné, 2003).

These examples illustrate that although we initially focused on aspects that restricted our understanding of the phenomena that interested us, we were subsequently able to explore other aspects that allowed us to more fully explain the phenomena; the question now focuses on combining these insights. This is not an illustration of confirmation bias, per se, where researchers only see evidence that support the conclusions they want to make.

But it still shows how suppositions about theoretical constructs orient the gaze of researchers. The first studies may be represented by the old adage of looking for your keys where the light is good (e.g. first we looked at conceptual aspects of argumentation) whereas during the further analyses, we turned on other lights (e.g. highlighting social relations and emotion during argumentation and spatial organization of arguments) and looked under *those*. The idea is to link up all the different light sources and build a coherent narrative.

### Code not to count, but to render phenomena observable

Our second example comes from a long term project focused on studying decision making during collaborative design in industrial engineering contexts (e.g. Lund, Prudhomme & Cassier, 2013; Prudhomme, Pourroy, Lund, 2007). We began this study in a similar way to the initial studies carried out within the SCALE project. Inspired by theory (Plantin, 2016; Simon, 1969; Gero, 2002; Vera, 2003; Hutchins, 2000) and previous empirical results (e.g. Baker et al. 2007, Détienne, Boujut & Hohmann, 2004), we developed an initial coding scheme (cf. Figure 1) and further elaborated it in confrontation with part of our corpus, using the rest of the corpus to perform inter-coder reliability. We did yet not have specific predictions to test by a coding and counting procedure. We were using the hypothetico-deductive approach but in an initial more exploratory and hypothesis generating stage. That said, our theoretical bases and empirical experience directly underpinned the nature of our coding scheme but the scheme's initial set-up was unable to show us the *criteria* that designers use during decision-making, when they argue in favor or against solutions.

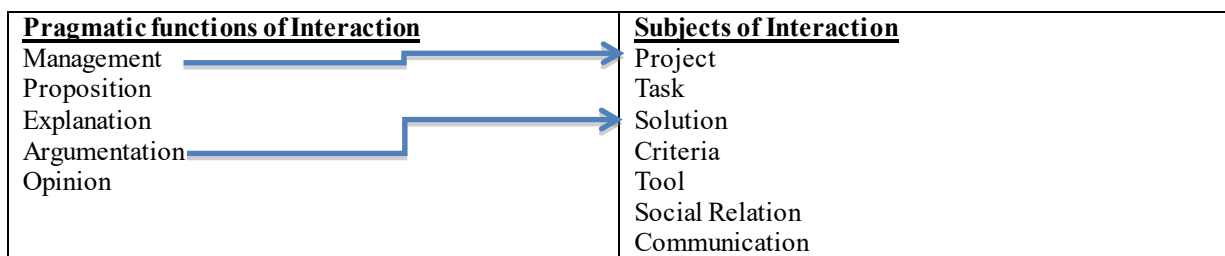


Figure 1. The Design Interaction Framework (DIF).

Our coding system enabled us to see how designers employed the pragmatic functions of their utterances (e.g. for managing, or for arguing) on particular subjects of interaction (the project, the solution). The arrows show two possible ways – a designer can manage a project and argue about a solution and argue about a solution. But he or she can also manage a task, a tool, a social relation, or communication. And he or she can argue about the project, per se, about a task, solution, criteria, tool, social relation or communication. We did not find all possible combinations of pragmatic functions of the interaction applied to subjects of the interaction in our data — for example, we did not find designers who managed solutions or criteria.

Although our coding system enabled us to make sense of how the design process evolved, it did not allow us to pinpoint exactly how designers used criteria when they argued. In other words, when argumentation was carried out on proposed solutions, the criteria each designer used to evaluate the solution were mobilized within the arguments and were thus not displayed with the way the coding scheme was set up. We could see arguments *about* criteria, but not arguments about solutions that *mobilized* criteria. So we went back to the corpus and specifically looked at how criteria were mobilized during argumentation. This work enabled us to build static and dynamic visualizations of how criteria are mobilized in collaborative design and illustrate two patterns of decision-making we found multiple times, in two different design contexts. The first pattern is when the criteria mobilized allow designers to choose between solutions already on the table and the second is when the criteria mobilized force the designers to propose and choose a new solution.

We argue that this example illustrates a way to potentially escape some of the limits of coding, often criticized in the literature: coding and counting cannot adequately capture important aspects of human interaction such as emotion (Peräkylä, 2004) and counting removes utterances from their meaning-making context (Schegloff, 1993). Here, we used coding as an exploratory aid in understanding the detailed articulation between argumentation and decision-making. In particular, systematically studying how criteria specific to a particular profession (e.g. mechanical engineering, assembly line work, project managing) were mobilized during arguments for and against solutions that were proposed during collaborative design allowed us to bring to light two patterns of decision procedures, without any counting of categories. In other words, coding can be seen as a filter that brings to light a phenomenon that could not have been noticed had the coding not been done, thus enabling a detailed understanding of the interactive mechanisms.

## Admit that even simple indicators can have explanatory power

In our last example, we show how comparing methods for predicting the existence of conflict in Wikipedia forums enabled us to 1) become aware of the validity of a model based on theory we would not have initially considered and 2) to evaluate the relative contributions of a model built from data versus one built on theory (Denis, Quignard, Fréard, Détienne, Baker, & Barcellini, 2012).

We began by elaborating an automated tool, based on natural language processing (NLP) techniques that could categorize Wikipedia threads as containing (or not) conflicts. Our tool was elaborated on theory stemming from Barth & Krabbe (1982) and Mackenzie, (1985) where a dispute can be modeled as a dialogue game whose goal is to attack or defend a controversial statement and eventually solve the conflict. According to Barth & Krabbe (op. cit.) a conflict is formally defined when participants have committed to at least one attack and one defense move with respect to a statement. By those moves, participants overtly take position for and against a point of view and commit to their respective positions as long as they can. Such a dialectical model gives us a practical method for identifying conflicts: find two utterances with opposite argumentative orientations (for or against) in relation to a preceding third utterance. However, implementing this method with NLP techniques is not trivial and involves both checking for markers of first and second person (e.g. me, you, your, etc. to distinguish viewpoints) and for the global connotation of the utterance (to obtain positive or negative nature). According to our method, there was a conflict if there were at least two negatively connoted contributions, one with a first person marker and the other with a second person marker. For a corpus of 320 discussions (122 were in conflict, or 38%), the method succeeds 77,8% of the time (0.86 of f-measure, kappa of 0.7), as compared to an expert analysis.

We compared this result to another method applied to the same corpus in order to appreciate its efficiency and to test whether or not the same markers used with simpler decision rules gave better results. The statistical induction method (Quinlan, 1993) calculates the global rates for each marker across the corpus and finds the most pertinent thresholds. Contrary to our own decision rule, this one is very simple to implement: if the discussion is at least 5 messages and if there are more than 8 2<sup>nd</sup> person markers per 1000 words, then there is a good chance that the text contains conflicts. This simpler rule has a success rate of 50% (0.64 of f-measure, kappa of 0.5), compared to an analysis done by an expert. Although this is a better result than by chance (which would be 38%) and significantly worse than our own, it still is surprisingly accurate, for its simplicity.

It is impossible to predict beforehand what type of rule will result from this type of induction method. What is surprising is that this rule does not predict conflicts as much as it predicts the co-existence of multiple voices that are interacting with each other. And since there is an overlap between the co-existence of two voices and expressed conflicts, this rule also predicts conflicts, but less efficiently. We can therefore draw two main lessons from our experience. First, the second rule made us aware of a global characteristic of the forums (i.e. presence of voices that interact); this could partially explain a local characteristic (i.e. presence of conflict). Remember that forums do not necessarily have multiple voices. There can be only one voice in the “discussion” area that never interacts with anyone else. Focusing on the interacting multiple voices forced us to consider the importance of this second rule, something we would never have done as the assumptions we made about the theoretical constructs involved oriented us to thinking that this rule would be unimportant. Second, this result also forced us reconsider the contribution of statistical approaches (driven by data) in relation to symbolic approaches (driven by theoretically based models). The former allows us to test the explanatory power of the markers without presuming how they should be interpreted whereas the latter furnishes the way in which the markers should be interpreted. This result argues for alternating data driven and theoretically motivated approaches.

## Conclusions

In this paper, we have described the hypothetico-deductive approach for carrying out research oriented to cognition and argumentation in computer supported collaborative learning and in computer supported work and problem solving situations. We have discussed some of the theoretical assumptions mobilized in this approach, how theory is used and its influence when it comes to defining research questions, conceptualizing learning, gathering and analyzing data, and representing results in relation to cognition and argumentation. We chose to examine one of the weaknesses of the hypothetico-deductive method—that of the different dangers researchers expose themselves to, due to formulating theoretical assumptions about data constructs prior to collecting data. In examining three of our own research projects, we illustrated three specific dangers due to making such assumptions and how we were able to partially surmount each of them, while also posing new methodological questions. First, the design of coding and counting schemes elaborated to test predictions of relations between conceptual constructs concerning collaborative processes and individual learning gains necessarily orient the gaze of the researcher toward particular phenomena. The trick is to be able to switch foci and break free of one’s



first assumptions. If a researcher can focus her gaze on other phenomena, this can lead to the discovery of other significant relations. But in this case, one must also build a narrative that coherently combines all of these results. The question then arises as to how inclusive a narrative should be before it is deemed sufficient. Second, elaborating a coding and counting scheme for fine grained exploratory analysis of collaborative processes also orients a researcher's gaze, but avoiding specific initial hypothesis testing and avoiding counting codes opens up the analyst to attuning to what can emerge from the corpus, while keeping the interactional context intact. But in this case, how should generalization be addressed? We answered this question by seeking and finding the same patterns of decision-making in two very different design contexts (Prudhomme, Pourroy, & Lund, 2007); Lund, Prudhomme, & Cassier, 2013), thus adding strength to an argument for generalization. Finally, comparing results in a benchmarking context where each result stemmed from a different method forces the researcher to question the assumptions underlying her method and to recognize the intangibility of her theoretical framework. Here, we wonder how we may place theoretically motivated and data driven approaches in dialogue to one another.

Our three examples are particularly targeted toward young researchers who may not have begun to reflect on the epistemological foundations of their research methods, or on how their practices may influence both the focus of their work and the validity of their results. This is more common than may be supposed. Finally, in a community such as the learning sciences where researchers collaborate in multi-disciplinary contexts, it's important to consider the theoretical assumptions that underlie our research. The epistemological encounters that researchers from different traditions may experience may be leveraged in order to explore the extent to which approaches can be integrated (Lund, Rosé, Suthers, & Baker, 2013).

## References

- Abend, G. (2008). The Meaning of 'Theory'. *Sociological Theory*, 26(2), 173-199
- Antaki, C. (2004). Reading Minds or Dealing with Interactional Implications? *Theory & Psychology*, Sage Publications, 14(5): 667-683.
- Baker, M.J., Andriessen, J., Lund, K., van Amelsvoort, M., & Quignard, M. (2007). Rainbow: a framework for analysing computer-mediated pedagogical debates. *International Journal of Computer-Supported Collaborative Learning*, 2:315-357.
- Baker, M.J., Quignard, M., Lund, K. & Séjourné, A. (2003). Computer-supported collaborative learning in the space of debate. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.) *Designing for Change in Networked Learning Environments : Proceedings of the International Conference on Computer Support for Collaborative Learning 2003*, pp. 11-20. Dordrecht : Kluwer Academic Publishers.
- Baker, M., Andriessen, J., Lund, K. (2009). Socio-relational, affective and cognitive dimensions of CSCL interactions: integrating theoretical-methodological perspectives. Symposium in the proceedings of the *International Conference of Computer Supported Collaborative Learning (CSCL09) on CD-ROM*. June 13-18, University of the Aegean: Rhodes, Greece.
- Baker, M.J., Andriessen, J. & Järvelä, S. (2013). *Affective Learning Together, Social and emotional dimensions of collaborative learning*. New York: Routledge.
- Barth, E. M., & Krabbe, E. C. W. (1982). *From axiom to dialogue: A philosophical study of logics and argumentation*. Berlin: Walter de Gruyter.
- Denis, A., Quignard, M. Fréard, D., Détienne, F., Baker, M & Barcellini, F. (2012). Détection de conflits dans les communautés épistémiques en ligne. Dans Conférence Traitement Automatique des Langues Naturelles, Grenoble, 2012.
- Détienne, F., Boujut, J.-F., & Hohmann, B., (2004). Characterization of collaborative design and interaction management activities in a distant engineering design situation. In (Eds.) F. Darses, R. Dieng, C. Simone & M. Zacklad, *Cooperative systems design*, (pp. 83- 98). Amsterdam, The Netherlands: IOS Press.
- de Vries, E., Lund, K. & Baker, M.J. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. *The Journal of the Learning Sciences*, 11(1), 63-103
- van Eemeren, F. H., & Grootendorst, R. (1984). *Speech acts in argumentative discussions*. Dordrecht: Foris.
- Gero, JS (2002) Computational models of creative designing based on situated cognition, in T Hewett and T Kavanagh (eds), *Creativity and Cognition 2002*, ACM Press, New York, NY, pp. 3-10.
- Greeno, J. G. (1998). "The situativity of knowing, learning, and research". *American Psychologist* 53 (1): 5-26.
- Hutchins, E. (2000). *Cognition in the Wild*. MIT Press, London.
- Kaelin, C.B. & Barsh, G. S. (2013) Genetics of Pigmentation in Dogs and Cats. *Annual Review of Animal Biosciences*. Vol. 1, 125-156

- Kawulich, B. (2009). The Role of Theory in Research. In Garner, M., Wagner, C., & Kawulich, B. (Eds.). *Teaching research methods in the social sciences* (pp. 37-82). Burlington, VT: Ashgate.
- Lund, K., Molinari, G., Séjourné, A. & Baker, M.J. (2007). How do argumentation diagrams compare when student pairs use them as a means for debate or as a tool for representing debate? *International Journal of Computer-Supported Collaborative Learning*. 2:273–295.
- Lund, K., Rosé, C. P., Suthers, D. D., & Baker, M. (2013). Epistemological encounters in multivocal settings. In D. D. Suthers, K. Lund, C. P. Rosé, C. Teplovs & N. Law (Eds.), *Productive Multivocality in the Analysis of Group Interactions*. In C. Hoadley & N. Miyake (Series Eds.), *Computer Supported Collaborative Learning Series: Vol. 15* (pp. 659-682). New York: Springer.
- Lund, K., Prudhomme, G., & Cassier, J.L. (2013). Pivotal moments for decision making in collaborative design: are they teachable? In (Eds.) S. Goggins, I. Jahnke & V. Wulf. *CSCL@Work: Case Studies of Collaborative Learning at Work (Computer-Supported Collaborative Learning Series)* (pp 243-268). New York: Springer.
- Lund, K., Rosé, C. P., Suthers, D. D., & Baker, M. (2013). Epistemological encounters in multivocal settings. In D. D. Suthers, K. Lund, C. P. Rosé, C. Teplovs & N. Law (Eds.), *Productive Multivocality in the Analysis of Group Interactions*. In C. Hoadley & N. Miyake (Series Eds.), *Computer Supported Collaborative Learning Series: Vol. 15* (pp. 659-682). New York: Springer.
- Mackenzie, J. D. (1985). No Logic before Friday. *Synthese*, 63:329–341.
- McGarrigle & Donaldson (1974). Conservation Accidents. *Cognition*. 3(4), 341-350.
- Molinari, G. & Lund, K. (2012). How a power game shapes expressing opinions in a chat and in an argument graph during a debate: A case study. In J. van Aalst, B.J. Reiser, C. Hmelo-Silver, K. Thompson (Eds.), *The Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012), (Vol. II, pp. 232-236)*. July 2-6, Sydney: International Society of the Learning Sciences.
- Nickerson, R. S. (1998). Confirmation Bias; A Ubiquitous Phenomenon in Many Guises, *Review of General Psychology* (Educational Publishing Foundation) 2(2), 175–220.
- Peräkylä, A. (2004). Two traditions of interaction research. *British Journal of Social Psychology*, 43, 1-20.
- Plantin, C. (1990) *Essais sur l'argumentation*. Paris : Kimé.
- Plantin, C. (2016). *Dictionnaire de l'argumentation – une introduction notionnelle aux études d'argumentation*. Lyon: ENS Editions.
- Polo, C., Lund, K., Plantin, C. & Niccolai, G. (2016). Group Emotions: The Social and Cognitive Functions of Emotions in Argumentation. *International Journal of Computer Supported Collaborative Learning*. 11, 123–156.
- Prudhomme, G., Pourroy, F., & Lund, K. (2007). An empirical study of engineering knowledge dynamics in a design situation. *Journal of Design Research*. 6(3), 333-358.
- Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers.
- Schegloff, E.A. (1993). Reflections on Quantification in the Study of Conversation. *Research on Language and Social Interaction*, 26(1), 99-128.
- Schooler, J. W. (2014). Metascience could rescue the 'replication crisis'. *Nature*. 515 (7525): 9
- Simon, H. (1069). *The Sciences of the Artificial*. MIT Press, Cambridge, Mass, 1st edition
- Stegmann, K., Fischer, F. (2011). Quantifying Qualities in Collaborative Knowledge Construction: The Analysis of Online Discussions. In Puntambekar, Erkens and Hmelo-Silver (Eds.) *Analyzing Collaborative Interactions in CSCL: Methods, Approaches and Issues* (pp. 247–268). NY: Springer.
- Suthers, D. (2006). A qualitative analysis of collaborative knowledge construction through shared representations. *Research and Practice in Technology Enhanced Learning*, 1(2), 1–28.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Vera, A.H. (2003). By the Seat of Our Pants: The Evolution of Research on Cognition and Action. *The Journal of the Learning Sciences* 12(2), 279–284.
- Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science*, 33(1), 1–30.

## Acknowledgments

Research discussed here was funded by the EU (IST-1999-10664-SCALE project), the Rhône Alpes region (PhD stipend), and the French ministry (ANR-08-CORD-004 CCCP-Prosodie project), The authors are also grateful to the ASLAN project (ANR-10-LABX-0081) of Université de Lyon, for its financial support within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR). We would also like to acknowledge the colleagues who worked with us on the projects we review here. In particular we owe a great debt to Michael Baker.

# Teaching Accessibility in a Technology Design Course

Kristen Shinohara, Cynthia L. Bennett, Jacob O. Wobbrock, and Wanda Pratt  
kshino@uw.edu, benec3@uw.edu, wobbrock@uw.edu, wpratt@uw.edu  
University of Washington

**Abstract:** The goal of college computer science and informatics design curricula is to prepare the next generation of technology designers. While accessibility is considered important for disabled users, it is treated as a niche or stigmatized subdomain of technology design. As a result, traditional computer science and informatics curricula do not expose students to diverse user needs, and students transition to the workforce unaware of the importance of accessibility. We incorporated accessibility in a design thinking course and observed student learning. We sought to challenge accessibility as niche, learn methods and barriers for including accessibility in the curricula, and identify impact on student learning. We found that students grasped general design concepts while designing for disabled and nondisabled users, and they faced challenges designing for both populations in concert. Contributions include insights for future accessibility and design courses and challenges facilitating accessible user-centered design process.

**Keywords:** Design Thinking, Accessible Design, Teaching Accessibility.

## Introduction

People with disabilities use accessible technologies to engage with others (Cook and Hussey 2002; Scherer 1993), yet most personal technologies are not designed to be usable by people with impairments (Scherer 1993), indicating that accessibility is not part of the main function of technical development. Although the goal of college computer science and informatics education is to prepare students for careers as technology innovators, typical curricula cover accessibility under legal issues (Section 508<sup>1</sup>, ADA<sup>2</sup>), and accessible pedagogy (teaching students with disabilities) (Rosmaita 2006). Therefore, despite research emphasizing the need to create accessible technologies (Stephanidis et al. 1998), computer science and informatics students are not exposed to the needs of diverse users as more than an “edge-case” (Rosmaita 2006) in learning design and development.



Figure 1. Students in our course receive feedback from a user with a hearing impairment.

Including accessibility as a key topic early in HCI and computer science curricula benefits technology design, society at large, and technological pedagogy; it is instrumental to ensuring that technology is usable for people with disabilities *and* that instruction is accessible for students with impairments. Including accessibility in computer science curricula positively impacted student learning (Poor et al. 2012; Rosmaita 2006) and resulted in accessible designs (Bigelow 2012; Ludi 2007; Waller et al. 2009). Yet, few computer science courses include accessibility as a *main* theme, and those that do remain disability-specific (Bigelow 2012; Ludi 2007; Poor et al. 2012; Waller et al. 2009), perpetuating the separation of users with disabilities from those without. Not including accessibility in computer science and informatics education risks omitting important elements of diversity, not just in technical domains, but in technology education overall. The consequences are that technologies, like laptops, mobile phones, etc., continue to be unusable by those with disabilities.

The present technical landscape includes virtual and augmented reality (Microsoft’s HoloLens, Oculus Rift), intelligent speech recognition (Apple’s Siri, Amazon’s Alexa, Microsoft’s Cortana, Google Now), with self-

driving cars on the horizon (Tesla, Google), with potential to benefit disabled<sup>3</sup> and nondisabled users. These opportunities for accessible design behoove designers and developers to engage in accessibility-focused practices.

To investigate how we, as technology educators, can effectively infuse accessibility as a main theme in technology design, we conducted a design course study assessing how students incorporate accessibility into an introductory design thinking course (Figure 1). Our study investigated tensions between concepts in accessibility and general design and we found that including accessibility as a core thread—as part of the practice of learning design (Wenger and Lave 1991)—did not introduce barriers to learning design thinking overall. In fact, it broadened student thinking about diverse approaches to design. We identified challenges faced by students and instructors, with respect to pedagogical and aspirational goals. Our contributions include a course outline for teaching accessibility as part of the main design theme, and insights on how to incorporate teaching accessibility in technology design courses. To allay instructors' fears that including accessibility in design thinking courses is challenging enough to preclude teaching accessibility as well, we designed our study to identify *how* we can effectively incorporate teaching technical accessibility. We asked: do students learn enough about design and accessibility to be able to create a prototype evaluated as usable by users with and without disabilities? How does including accessibility impact student learning about design thinking? We present our results and experiences.

## Background and related work

Teaching undergraduates accessibility is one way to increase accessible technologies. In teaching design thinking, it is common to acculturate students to industry practices and strategies through hands-on learning objectives and projects. Engaging legitimate peripheral participation (Wenger and Lave 1991) and communities of practice (Wenger 1998) manifest in exposing students to disabled experiences while addressing design problems. We briefly discuss teaching accessibility and design thinking, two key elements that served as objects of study.

### Teaching accessibility

Researchers have looked at how accessibility is addressed in teaching computer science, engineering, and design students. Rosmaita incorporated accessibility as a main theme in introductory web design courses (Rosmaita 2006). Bigelow included Universal Design principles in introductory engineering courses, and found that students emphasized accessibility in their engineering design projects (Bigelow 2012). Rosmaita and Bigelow's courses did not involve working with people with disabilities, which we have done in our study. Waller et al., investigated a multi-year program on accessibility, focusing on integrating accessibility throughout the learning experiences (Waller et al. 2009). Waller et al.'s multi-year approach frames our efforts here at a high level; we focused on specific elements in a single course. Ludi included stakeholders in requirements gathering in an engineering course to increase accessibility awareness, but designs were not produced (Ludi 2007).

Despite research showing the benefits of a range of teaching accessibility practices (Putnam et al. 2015), students rarely consider disabled users without provocation, and are not taught to include accessibility as part of the “main event” of design. For example, we conducted a brief overview of technology design courses from three different undergraduate technical computing programs from two separate institutions. We assessed project information as posted on public websites and evaluated project descriptions to determine target audience and accessible design. We found twelve out of 179 (6.7%) student projects targeted users with disabilities, or might be accessible (e.g., some projects targeted caregivers, disabled people). Project descriptions showed high level prompts, and that student groups self-defined design problems and selected target users for their projects. We elevate the concern that most computer science and informatics students do not consider the role of disability in technology design without prompting, relegating accessibility as a niche subdomain to mainstream design.

### Design thinking

Not all computer science curricula include design thinking principles and practice, but the number is growing. Far from the traditional waterfall method, computer science education research highlights the benefits of abductive reasoning in software design and engineering problem solving, including design thinking's close relationship with “computational thinking” (Hu 2011, 2016). Alongside technical skills, design skills, by way of learning design thinking, has become sought after. Using abductive reasoning to create new forms, design thinking emphasizes rational (Simon 1969) and reflective (Schön 1987) envisioning of new artifacts. In contrast with deductive approaches in functional requirements gathering, reflective and iterative design thinking involves a cyclical create-evaluate-revise approach, not to achieve a perfect solution, but toward an idealized particular (Brown 2008; Stolterman and Nelson 2012): (1) Understand the user and develop empathy; (2) Define the problem space; (3) Explore multiple ideas through ideation; (4) Prototype quickly and often, eliciting feedback; (5) Test and iterate.

Design thinking principles in HCI are commonly applied via user-centered design (UCD) (Gould and Lewis 1985), encapsulating the above process in a way that centers on users' needs and preferences. In teaching

design in computer science, Hu examined metrics used to evaluate software design effort, and emphasized “the ways [students] approach design decision making” (Hu 2016) as more important than the artifacts themselves when students are learning. In the spirit of Hu’s argument, we assessed how incorporating accessibility impacted learning about design, and if artifacts designed from an accessibility-driven approach met disabled and nondisabled user needs. Students in our course learned design thinking concepts and principles, applied through typical user-centered methods and tools. We show that teaching accessibility in a design thinking course did not impede students’ “approach to design decision making” (Hu 2016).

## Accessibility in design

Like Bigelow, Ludi, and Waller’s approaches to teaching accessibility, we view the key element of design and accessibility as working directly with disabled people in at least one point in the design process. It is hard to design for disability without receiving input directly from people with disabilities. Thus, researchers defined a variety of ways to design for disabled users, including: Design for User Empowerment (Ladner 2015), emphasizing increasing the number of people with disabilities in technology design disciplines; User Sensitive Inclusive Design (Newell et al. 2011), encouraging designers to get to know users with disabilities as part of their design work; Universal Design (Bigelow 2012), employing an access-for-all approach in design, and in curricula for students with disabilities; and, Ability-Based Design (Wobbrock et al. 2011), focusing on abilities users do have. We distinguish our approach by requiring student designers to include users with and without disabilities, and structuring the course with multiple face-to-face sessions.

## Method

We conducted an IRB-approved design course study investigating how students learned design when we incorporated accessibility in overall design requirements. The introductory course was part of an informatics-based curriculum, focusing “on computer systems from a user-centered perspective and study[ing] the structure, behavior and interactions of natural and artificial systems that store, process and communicate information<sup>4</sup>.” Learning objectives were that students demonstrate an ability to: (1) create a technical design to a specific prompt, (2) complete the stages of the UCD process while working with a user, (3) incorporate components into a final design concept, (4) produce a usable prototype exemplifying that concept, and (5) adequately communicate that concept for development. The 10-week course introduced design thinking concepts through readings (course texts included (Buxton 2007; Norman 1988)) and lectures, and students applied these concepts through techniques and tools based UCD (Gould and Lewis 1985). Design thinking concepts and UCD methods included needs assessment, user interviews, brainstorming, ideating, synthesizing, low-fidelity prototyping, high fidelity prototyping and user testing. Students were instructed to design for users with and without disabilities. At the beginning of the course, a blind guest speaker covered general etiquette tips in a question and answer forum to prepare students to work with people with disabilities. Every week, students applied concepts from class in a lab section. In every other lab, students worked directly with users with disabilities to test ideas and elicit feedback. At the end of the course, students presented their work, evaluated by users with disabilities.

## Participants

The course had 42 students. Twelve were female. Students did not have disabilities or a design background, and few interacted with disabled people previously. Introductory programming courses were a prerequisite and students had a basic understanding of coding techniques, data structures, and simple algorithms. Users with visual and hearing impairments were recruited from listservs, and referred to as “expert users” emphasizing their expertise in assistive technology use. Each student group was assigned to work with one expert user for the course. Students met with expert users 4 times; expert users evaluated student designs during final presentations.

## Course work and projects

Students worked in randomly assigned groups; each group was assigned one of two projects to work on throughout the term. Groups 1-6 worked with visually impaired users and were tasked to design a real-time augmented reality navigation application. Groups 7-11 worked with hearing impaired users and were tasked to design a real-time live captioning application (see Tables 1 and 2).

Classes consisted of a traditional lecture format, followed by in-class activities where students used the techniques they learned. Lab sections were held once a week, during which students applied the week’s concepts and techniques toward their project. Concepts were introduced successively such that students added to their projects as the course progressed. Expert users attended lab sections, allowing students to practice their design skills with and without expert users. Expert users assessed student work at the ideation, synthesis, prototyping and

testing stages. Assignments followed the design process: interview protocols and summaries showed that students asked appropriate questions and could brainstorm.

Table 1. “Project A” Real-Time Augmented Reality Navigation, groups and expert users.

Group	Student Designers	Expert User
G1	S12 (M), S22 (M), S41 (M), S31 (F)	E1 (M), Blind
G2	S1 (F), S26 (M), S28 (F), S36 (M)	E2 (M), Blind
G3	S19 (M), S21 (M), S33 (M), S35 (F)	E3 (F), Low Vision
G4	S6 (F), S8 (M), S23 (M), S34 (M)	E4 (F), Low Vision
G5	S2 (F), S9 (M), S15 (F)	E5 (F), Blind
G6	S11 (M), S13 (M), S25 (F), S42 (M)	E6 (F), Blind

Table 2. “Project B” Real-Time Live Captioning, groups and expert users.

Group	Student Designers	Expert User
G7	S14 (M), S30 (M), S38 (M), S39 (M)	E7 (F), Deaf
G8	S3 (M), S5 (F), S20 (M), S32 (M)	E8 (F), Hard of Hearing
G9	S10 (F), S16 (F), S29 (M), S37 (M)	E9 (M), Deaf
G10	S4 (M), S7 (M), S27 (M)	E10 (F), Hard of Hearing
G11	S17 (M), S18 (M), S24 (F), S40 (M)	E11 (M), Deaf

## Data and analysis

Study data comprised student assignments: interview protocols and summaries, conceptual models, brainstorm ideas, sketches, low and high fidelity prototypes, usability heuristics, and user test results. Each group created a design specification detailing the form and operation of the final design, and each student completed a process book describing their own experience and individual contribution. Students completed 1-page reflective journal writing assignments each week. Journal prompts asked students to reflect on each week’s topic, technique, lab activity or the student’s overall experience. For example, after spending time on the concept of ideation, including brainstorming sessions with and without expert users, students were asked: “How did your feedback session with your expert user go? What did you learn from the session? Do you feel that you received helpful feedback? Why or why not? What could you have done better? How will your group use this feedback?” A summative survey at the end of the term captured student thoughts of the course overall.

Qualitative analysis was conducted on the data described above, via openly coding student assignments, with focus on journal entries. Following open coding analysis methods (Miles and Huberman 1994) in the spirit of grounded theory (Glaser and Strauss 1967), two researchers independently coded 10% of the journals, then discussed and refined the codes before one researcher coded the rest. Once categories and themes were discussed, researchers focused on student learning and outcomes: where did students demonstrate knowledge of key concepts and how were designs exemplars of student learning? Criteria for assessing designs included how well students identified and addressed expert user needs (evaluated by expert users and instructors), prototyped ideas, and usability evaluation (i.e., with usability heuristics?). Analysis prioritized how students applied UCD techniques and design thinking concepts, worked with expert users, and the quality of final designs.

## Findings

We found that incorporating accessibility in a design thinking course can be accomplished via a holistic approach requiring accessibility as a key element in overall design. We intentionally set expectations that design thinking concepts and UCD techniques ought to sufficiently address accessibility and usability requirements for users *with and without* disabilities. We assessed student understanding of design thinking concepts, and how well students used UCD techniques and tools to create designs. Although some students encountered challenges using some techniques, they were able to adapt concepts and tools to create complete designs. We identify challenges and discuss implications for implementation of future courses.

## Assessing student learning

All groups completed all phases of the design process as indicated by their assignment completion rate and successful presentation of final designs. Students demonstrated skill in presenting ideas, eliciting feedback from expert users, in completing assignments, and creating projects that met course expectations. They sometimes had mixed success in producing quality assignments (e.g., some students found that some interview questions yield

more useful responses than others), but journals gave insight into if and how they rebounded from less than stellar design activities. We discuss how we assessed students' grasp of key concepts, and how assignments were judged.

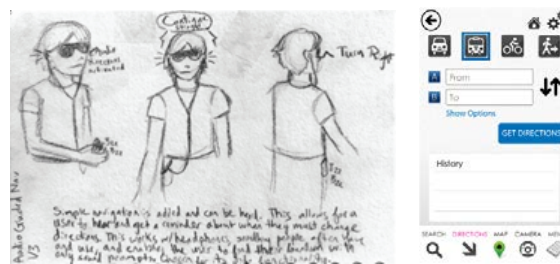
### Learning design thinking

We looked to the quality of completed assignments to gauge understanding at each stage of the design process (Figure 2) and to assess students' ability to keep pace with the course. The sequential aspect of UCD (i.e., brainstorming before prototyping) and the schedule of meetings with expert users meant students had to keep up. If students failed to keep up with pre-requisite concepts, they would have difficulty completing subsequent assignments. For example, students had to whittle down 90 brainstormed ideas into a single suitable prototype, as required of the ideation assignment. A prototype not driven by a common vision would be less likely to succeed in obtaining feedback needed to move to the next stage. Students needed to understand design constraints and develop a conceptual model before a prototype could take shape. Expert user meetings motivated students to be prepared for each session, as S23 explained:

We would spend the previous evenings before meeting with E4 preparing prototypes [so] that once she saw them, she was instantly able to help us improve. (S23, Process Book)

In-class observations showed students engaging design ideas, debating pros and cons of each choice or attribute, and demonstrating an understanding of concepts such as constraints and reflection. Journals corroborated students' grasp the process toward an understanding of the overall course learning objectives. For example, S10 wrote:

I feel that my sense for design has improved especially from the readings that include examples of studies, design approaches, and design techniques. The examples that Buxton and the other authors present give me a solid idea of how the design process works and how to make the process successful. I have come to really appreciate the different steps individually and together. Feedback is just as important as the brainstorming and sketching can be quick and easy while extremely valuable. (S10, Journal6)



**Figure 2.** Completed assignments demonstrated students' grasp of concepts. (L) G4's sketch shows a smartphone with headphones directing a user. (R) A screenshot from G3's high fidelity prototype.

S30's understanding of iteration led him to value the sequence over selecting ideas too early in the process:

Sketches are now no longer "prototypes" or firm ideas like I used to think... but now a way of putting some start of an idea on paper and then having something to build off of. (S30, Journal7)

### Working with expert users

Working with disabled users was initially challenging for students who were not familiar with disabilities, and who did not at first view their role as designers for people with disabilities. However, students quickly adjusted perceptions and addressed design requirements as the course continued, engaging users in a productive way. We observed students interacting professionally and generatively with expert users, asking questions specific to design and use, and conducting feedback and testing sessions focused on improving designs. A challenge in working with users, whether disabled or not, is that designers have to make on-the-fly changes to accommodate unexpected changes; S5's group was unprepared for initial outcomes:

... sketch, the expert user affirmed us that our idea looked great, and then the conversation was pretty much done in about 20 minutes. Because the meeting was so pale in context, we had to pull out our very first draft on paper prototype to prevent the awkwardness of nothing to talk about. (S5, Journal8)

Instead of wasting time, they adapted to get as much information as possible from their expert user. In hindsight, the students should have been better prepared for their user. They reflected on their actions (e.g., should have prepared better questions) and made adjustments to improve their situation (asking pointed questions on on half-

baked prototypes). Thus, students were able to follow the design process and meet the learning goals because they also adapted to accessibility challenges as they arose. Next, we show how student work illustrated how well students learned design, and how they incorporated accessibility.

## Evaluating design artifacts

All student groups produced a high-fidelity prototype that was tested with expert users, demonstrating that students were successful at identifying and translating user needs into designs. User test results were mixed: some expert users had no trouble using student designs, while others found bugs. We recognize the non-trivial effort required in creating a high-fidelity prototype to a level that can be tested and we consider a variation in user testing outcomes typical, especially in an introductory design course. In this section, we break down how design specifications, student reflections, and process books provide evidence of students substantively and reflectively meeting course learning goals, particularly around designing an accessible solution.

### Design specifications

Groups were required to create a specification communicating the technical details of their design in a way that a software developer could implement. Design specifications were assessed on the level of detail necessary for user interface implementation and were not required to have code snippets or algorithms. Given that accessibility was inextricably tied to course requirements, design specifications were expected to incorporate elements of accessible interactions. Although we could not be sure how developers would use the specifications, the instructor had prior experience as a software engineer and in creating design specifications, and could adequately assess the work.

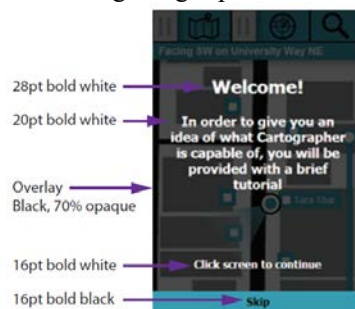


Figure 3. G4's design specification shows descriptions of each element within the interface.

As expected in an introductory design course, specifications varied in quality, but all contained minimal required elements: screenshots, details about specific elements of a user interface (Figure 3), descriptions of key elements of the design (the level of detail varied by group), how it should operate and be used, and a rationale for each major design decision (as required by the assignment). Specifications demonstrated that course learning objectives were met. Students: (1) could create a technical solution to a specific design prompt, (2) complete the various stages of the UCD process while working with a user *with a disability*, (3) incorporate *accessible* components into a final design concept, (4) produce a usable prototype exemplifying that concept, and (5) adequately communicate that concept for development. (Emphasis added to highlight how students met expectations for accessibility incorporated in learning objectives).

### Student reflections and process books

To evaluate design specifications, we referenced student journals and process books to corroborate their knowledge of course concepts. S23 described how his group incorporated feedback to improve their design:

...E4 still was often confused trying to navigate our prototype. Due to [her] feedback, we concluded that it was not clear with the amount of options we had at the top of the screen what view mode she was in and would be going into. Following this meeting, our group sat down and tried to simplify our design, and even considered cutting one or two of the views. But [teammate S34] came up with a solution that even let us add a view while making the system similar for the user. ...[In] a two button toggle system,...One toggle would be the Directional/Global toggle to determine the user was looking all around them or in a specific direction and the other would be List/Map to determine how the user wanted to data displayed. (S23, Process Book)

Toward course conclusion, journals offered detailed reflections what students felt they learned:

I really enjoy the fact that we are learning about the design process by getting the chance to apply every step and practice it... I am getting a lot out of this method, and I will retain this



knowledge far better with memories of my experiences with it. I've been going through a cycle with feeling mildly overwhelmed with what the next step requires me to accomplish (such as all the options and decisions we had to make for our prototype) and once I break it down into the required assignments it suddenly ends up conquered. (S2, Journal7)

## Challenges

Our findings indicated that students grasped general design concepts while tasked to create accessible designs, but they faced challenges specific to designing for disabled and nondisabled users. Groups with hearing impaired users sometimes had difficulty communicating, and receiving feedback on sketches was challenging for groups with visually impaired users. As with any project involving disabled people, it was difficult to find volunteers who could participate at the same time and place with students. With financial support, we mitigated issues with recruiting by generously compensating expert users. Thus, having resources can make a difference in finding people from non-typical user populations to work with. In this section, we identify how challenges impacted learning goals, and highlight methodological and substantive challenges.

### Inaccessible design methods

We changed as little of the design process as possible to expose accessibility issues, and working with disabled expert users revealed where the UCD methods and tools used in this course assume nondisabled users. Visual techniques—such as sketching or paper prototyping—assumed sighted users and students struggled to find ways to make each successive step accessible in a way that showed progress. Inaccessible techniques and tools highlighted where the UCD process overall tends to assume users and designers without disabilities. These challenges emphasize opportunities to develop alternative, accessible design methods.

### Lack of disability-specific knowledge

Students began the class with little knowledge or experience of disability, and though the initial Q&A with the blind guest speaker was helpful, it was their only introduction and was insufficient in contextualizing experiences of using accessible technologies that students would come to rely on. Students spent considerable time becoming acquainted with disability-specific technical knowledge and many took it upon themselves to learn more about disability. Inaccessible design methods made it difficult for students to meet specific learning objectives of eliciting feedback and creating testable prototypes. Students overcame these challenges in creating minimally usable and accessible designs. Yet, we cannot be sure of how much more students could have accomplished if tools at their disposal were accessible. Unfortunately, such challenges revealed barriers to a truly accessible design process, though they did not critically block students' ability to create designs for their expert users.

## Discussion

Students in our study met course objectives and completed a design project while incorporating *accessible design*. Despite challenges, barriers to *teaching* accessibility are low. We add insights to related work (Poor et al. 2012; Putnam et al. 2015; Rosmaita 2006) for future design thinking courses: (a) expectations should include accessibility, (b) students should work with disabled users, (c) students should be required to create designs for disabled *and* nondisabled users, (d) courses should cover disability etiquette and existing accessible technologies.

Establishing a baseline of technical accessibility or awareness of issues around disability would equip students to take on disability and design, improving the design outcomes. For example, students could learn what orientation and mobility skills a blind person might use to navigate. Facilitating accessible methods (paper prototyping for blind users?) and creating accessible tools enables students to thoroughly engage with accessible aspects of design. Accessible design skills and tools would be useful and practical for student designers to learn.

## Conclusion

Students reported an increased awareness of implications for inaccessible design; and changed their perspectives that accessibility is “someone else’s job” to understanding their role as designers in creating an accessible future. In immersing accessibility, students grasped concepts around design thinking and (1) created technical solutions to design prompts, (2) completed the UCD process while working with a user *with a disability*, (3) incorporated *accessible* elements into designs, (4) produced usable prototypes, and (5) adequately communicated their designs for development. Incorporating accessibility in design is yet imperfect due to inaccessible methods and tools, but teaching accessibility in a design thinking course was effective when resources were in place and when expectations included accessible outcomes.

## Endnotes

(1) <https://www.section508.gov/>

- (2) <https://www.ada.gov/>
- (3) It is accepted practice in disability studies to use ‘people with disabilities,’ ‘disabled people,’ and ‘nondisabled’ to center disability (Linton 1998).
- (4) Classification for Instructional Programs: <http://nces.ed.gov/ipeds/cipcode/cipdetail.aspx?y=55&cipid=89325>

## References

- Bigelow, K. E. (2012). Designing for Success: Developing Engineers Who Consider Universal Design Principles. *Journal of Postsecondary Education and Disability*, 25(3), 211–225.
- Brown, T. (2008). Design Thinking. *Harvard business review*, 86(6), 84–92.
- Buxton, W. (2007). *Sketching user experiences: getting the design right and the right design*. Amsterdam; Boston: Elsevier/Morgan Kaufmann.
- Cook, A. M., & Hussey, S. M. (2002). *Assistive technologies: principles and practice*. St. Louis: Mosby.
- Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory*. Chicago, IL: Aldine Publishing Co.
- Gould, J. D., & Lewis, C. (1985). Designing for usability: Key principles and what designers think. *CACM*, 28(3), 300–311.
- Hu, C. (2011). Computational thinking: what it might mean and what we might do about it. In *Proc of the 16th conf on Innovation and technology in computer science education* (pp. 223–227). Darmstadt, Germany: ACM.
- Hu, C. (2016). Can Students Design Software?: The Answer Is More Complex Than You Think. In *Proc SIGCSE* (pp. 199–204). Memphis, Tennessee, USA: ACM.
- Ladner, R. E. (2015). Design for user empowerment. *interactions*, 22(2), 24–29.
- Linton, S. (1998). *Claiming Disability: Knowledge and Identity*. New York: New York University Press.
- Ludi, S. (2007). Introducing Accessibility Requirements through External Stakeholder Utilization in an Undergraduate Requirements Engineering Course. In *Proc Software Engineering '07* (pp. 736–743). IEEE Computer Society.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: an expanded sourcebook*. Thousand Oaks: Sage Publications.
- Newell, A., Gregor, P., Morgan, M., Pullin, G., & Macaulay, C. (2011). User-Sensitive Inclusive Design. *Universal Access in the Information Society*, 10(3), 235–243.
- Norman, D. (1988). *The Design of Everyday Things* (2002 Edition.). New York: Basic Book.
- Poor, G. M., Leventhal, L. M., Barnes, J., Hutchings, D. R., Albee, P., & Campbell, L. (2012). No User Left Behind: Including Accessibility in Student Projects and the Impact on CS Students Attitudes. *ACM Transactions on Computing Education (TOCE)*, 12(2), 1–22.
- Putnam, C., Dahman, M., Rose, E., Cheng, J., & Bradford, G. (2015). Teaching Accessibility, Learning Empathy. In *Proc ASSETS 2015* (pp. 333–334). Lisbon, Portugal: ACM.
- Rosmaita, B. J. (2006). Accessibility first!: a new approach to web design. In *Proc SIGCSE 2006* (pp. 270–274). Houston, Texas, USA: ACM.
- Scherer, M. J. (1993). *Living in the state of stuck: how technologies affect the lives of people with disabilities*. Cambridge, MA: Brookline Books.
- Schön, D. A. (1987). *Educating the reflective practitioner*. San Francisco: San Francisco : Jossey-Bass.
- Simon, H. A. (1969). *The sciences of the artificial*. Cambridge: Cambridge, M.I.T. Press.
- Stephanidis, C., Akoumianakis, D., Sfyraakis, M., & Paramythis, A. (1998). Universal accessibility in HCI: Process-oriented guidelines and tool requirements. In *Proc. User Interfaces for All '98*.
- Stolterman, E., & Nelson, H. G. (2012). *The design way: intentional change in an unpredictable world*. Cambridge, Massachusetts: Cambridge, Massachusetts.
- Waller, A., Hanson, V. L., & Sloan, D. (2009). Including accessibility within and beyond undergraduate computing courses. In *Proc. ASSETS '09* (pp. 155–162). Pittsburgh, PA, USA: ACM.
- Wenger, E. (1998). *Communities of practice : learning, meaning, and identity*. Cambridge, U.K.: Cambridge.
- Wenger, E., & Lave, J. (1991). *Situated learning: legitimate peripheral participation*. Cambridge [England]: Cambridge England.
- Wobbrock, J. O., Kane, S. K., Gajos, K. Z., Harada, S., & Froehlich, J. (2011). Ability-based design: Concept, principles, and examples. *ACM TACCESS*, 3(3), 1–27.

## Acknowledgments

This work was supported in part by the National Science Foundation under grants IIS-1217627, IIS-1230435, and IIS-0952786. Any opinions, findings, conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect those of any supporter listed above.

# Group and Individual Level Effects of Supporting Socio-Cognitive Conflict Awareness and Its Resolution in Large SNS Discussion Groups: A Social Network Analysis

Dimitra Tsovaltzi, Nikita Dutta, Thomas Puhl, and Armin Weinberger  
d.tsovaltzi@edutech.uni-saarland.de, ndutta.uds@gmail.com t.puhl, a.weinberger@edutech.uni-saarland.de  
Saarland University, Germany

**Abstract:** SNS (social networking sites) provide great opportunities for access to knowledge sources and equity in publicly expressing ideas, opinions and attitudes. They define a rich context of socio-cognitive interactions in which conflict can arise. Supports for conflict awareness and argumentation processes to resolve the conflict can foster learning. This article presents a comparative 2×2 field study ( $N=98$ ) of such supports in a university course that included weekly SNS discussions. Group Awareness Tools (GATs) were used to increase attitude conflict awareness, and argumentation scripts as a cognitive guidance to help learners capitalize on this awareness and resolve the conflict productively. We use Social Network Analysis (SNA) to gain a group-level understanding of the effects of argumentation on attitude change relative to peer interactions. We measure number of interactions, information flow control, influence distribution, and attitude similarity. Both GATs and argumentation script influence group processes, but argumentation script shows more substantial influence.

**Keywords:** computer-supported collaborative learning, social network analysis, group awareness tools, argumentation script, social networking sites

## Attitude change through argumentation in SNS using group awareness tools and argumentation scripts

### Socio-cognitive conflict and attitude change in SNS

SNS provide a great possibility for equity and access to sources of knowledge, but also for public expression of ideas, opinions and attitudes that can lead to pluralistic exchange. They define a rich context of socio-cognitive interactions in which conflict can arise, but they do not necessarily promote productive interactions. In particular, discussions in public settings foster the externalization of attitudes and afford chances for learners to scrutinize their own attitudes and those of their peers (Nussbaum, 2008), thus facilitating interaction and elaboration that could lead to socio-cognitive conflict and attitude change. However, research has shown that while on the one hand social media can instigate socio-cognitive conflict beyond the grasp of existing purpose-specific collaborative learning tools by leveraging differences in attitude, on the other hand the public nature of the discussions may rather reinforce private beliefs and attitudes and prohibit change (Lampert, Rittenhouse, & Crumbaugh, 1996). Peers play an important part in shaping new opinions and perspectives during group discussions in public settings. They, thus, influence group decisions by encouraging networked communication and collective decision-making (Chaiken, Wood, & Eagly, 1996; Cialdini & Trost, 1998; Wood 2000). More research is needed to better understand the socio-cognitive processes involved in SNS peer interactions in their own right, but also to identify supports that can enhance interactions and help learners to productively resolve conflicts in SNS and leverage attitude differences and socio-cognitive conflict for attitude change.

### Argumentative knowledge construction to promote communication attitude change

Attitudes influence our ability to learn and acquire new skills, and may often hinder learning (Erber, Hodges, & Wilson, 1995). Although, communication theory seminars are becoming increasingly common in teacher trainee university courses to modify teachers' attitudes and improve their communication skills, negative attitudes towards the need for good communication skills are pronounced among teacher trainees (Ihmeideh & Al-omari, 2010). However, communication is a key skill for teachers and higher education professionals due to their everyday involvement with students, parents, colleagues and school administrative staff, requiring skillful intervention through communication. Additionally, individual attitudes are stable and long-term deep learning and conflict awareness are prerequisites for attitude change (Erber et al, 1995). *Argumentative Knowledge Construction* (AKC), which is the deliberate practice of elaborating learning material by constructing formally and semantically sound arguments (Weinberger & Fischer, 2006) can aid such deep learning. Argumentation

can not only induce increased self-reflection and conflict awareness, but can also help in attitude co-construction (Andriessen, 2006; Asterhan & Schwarz, 2009; Baker, 2003; Felton & Kuhn, 2001, Sassenberg & Boos, 2003, Tsovaltzi, Puhl, Judele, & Weinberger, 2014; Wenger, McDermott, & Snyder, 2002)).

### Computer-Supported Collaborative Learning supports for attitude change in SNS

Computer-Supported Collaborative Learning (CSCL) environments have been systematically enhanced and utilized to induce attitude change and promote knowledge acquisition (Buder & Bodemer, 2008; Dillenbourg & Fischer, 2007; Puhl, Tsovaltzi, & Weinberger, 2015a; 2015b). SNS, like Facebook, provide a modifiable platform through apps that can host argumentative CSCL learning scenarios in which learners can exchange and formulate their elaborate opinions and arguments during meaningful discussions (Greenhow, 2008; Greenhow, Menzer, & Gibbins, 2012; Puhl et al; 2015a; 2015b). Prominent CSCL supports like GAT and argumentation scripts can, thus, be augmented to implement AKC in SNS and make use of group level interactions for learning and attitude change. *GATs* provide learners with information regarding actual group processes. They can, for instance, visualize group process information (Buder & Bodemer, 2008; Gutwin & Greenberg, 2002) to highlight conflicting opinions in discussions and foster learning and collaboration through socio-emotional and motivational processes. The increased awareness of ones' opinion and attitude afforded by *GATs*, especially in comparison to others in a group, can help in making differences salient, a prerequisite of dissonance and attitude change (Festinger, 1957). Similarly, in SNS, *GATs* could foster socio-cognitive conflict and prompt more individuals to actively engage in meaningful dialogue regarding conflicts, especially when quantitative representations are combined with qualitative ones (Erkens, Schlottbom & Bodemer, 2016). *GATs* can, thus, assist in forming new communication channels between learners and increase the number of interactions in an attempt to understand and resolve conflicts. Additionally, increasing the number of learners central to discussions, *GATs* could also enhance individual control on information flow and the subsequent distribution of influence in the network promoting increased idea exchange and knowledge co-construction. In effect, *GATs* can support constructive discussions and conflict resolution, through which learners may form similar attitudes. *Scripts* are socio-cognitive structures that can provide specific guidance and scaffold group discussions during collaborative learning scenarios (Fischer, Kollar, Stegmann & Wecker, 2013). *Argumentation scripts* aim to improve argumentation quality. They may prompt learners to elaborate their arguments and clarify their opinions, and thus foster AKC (Noroozi, Weinberger, Biemans, Mulder, & Chizari, 2012). They may guide learners to analyze the lines of argumentation provided by both their peer group and themselves during discussions, scrutinize their attitudes, and critically reflect on them. Argumentation scripts may also help participants in providing rational epistemic counter-arguments, leading to mutual conflict resolution (Kollar, Fischer & Slotta, 2007) which forges similar knowledge (knowledge convergence; Weinberger, Stegmann, & Fischer, 2007). Argumentation scripts in SNS might also forge similar attitudes overtime. In previous studies, inferential statistics has indicated that over more time (semester long), SNS discussions with incorporated *GATs* and argumentation scripts, and their combination influence learners' attitude towards communication skills by introducing socio-cognitive conflict and improving argumentation quality (Puhl et al, 2015a).

The present work aims to inculcate learners' mutual attitude change and knowledge gain during argumentative SNS discussions in distinctively designed and implemented Facebook apps. The apps provide *GATs* support to foster socio-cognitive conflict and promote interactivity towards its resolution. They also implement argumentation scripts which intent to improve argumentation quality and resolve conflict productively. We aim to capture and understand group dynamics in SNS discussions as part of a teacher communication seminar. We define group dynamics as the unforeseeable patterns of communication flow among learners, and the *flow* of resources: knowledge in the form of ideas and opinions of the peers that may lead to attitude change through discussions. We also want to observe how *GATs* vs. argumentation scripts influence group dynamics in SNS, and whether they provide opportunities or form constraints for interactivity, learning and attitude change. We combine SNA and inferential statistics (Halatchliyski, 2011) to test the effect of these supports on group dynamics. We test if *GATs* and argumentation scripts enhance network measures, support learners and facilitate discussions in SNS. We aim at enhancing our understanding on how CSCL supports may be used to leverage social influence and promote attitude change and learning in SNS.

### Method

We conducted a 2x2 semester long field study (see Table 1) with German teacher trainees (N=98) in two rounds (two consecutive semesters). The data for script conditions were collected in the second round. Facebook, a prominent SNS, was used to integrate online argumentative discussions with face-to-face teacher training university seminars on communication theory over 9 weeks. The participants filled out a case-based questionnaire weekly based on every-day social interaction scenarios in the school, to capture their

communication attitudes. Each question comprised of two communication attitude dimensions: multi-perspective / flexible attitudes vs. goal-oriented / structured attitudes, following Buder & Bodemer (2008) and Jermann & Dillenbourg (1999, 2002), rated on four Likert-scale (0 to 6) answers on how they as teachers would assess the situation. The dimensions were balanced in the cases. Furthermore, the explorative factor analysis of the communication attitude questionnaire resulted in two independent factors similar to our theoretical construct of multi-perspective/flexible for the first factor (Cronbach's  $\alpha = .87$ ) and goal-oriented/structured for the second factor (Cronbach's  $\alpha = .87$ ) (see also, Puhl et al, 2015b). GAT and argumentation script, were implemented in closed Facebook groups, where participants engaged in argumentative discussions on the problem cases and could reference communication theories to support their arguments.

Table 1: 2x2 factorial design

	Argumentation Script	No Argumentation Script
GAT	$N = 26$	$N = 24$
No GAT	$N = 30$	$N = 18$

The GAT provided a graphical visualization of participants' attitude questionnaire results (Figure 1). The Facebook application included the visualization in order to increase socio-cognitive conflict (Jermann & Dillenbourg, 1999) by making attitude differences salient. Students were asked to reflect on these. Browsing the GAT was obligatory, but reflection was not controlled. Participants in *argumentation script* conditions had to "like" the best argument made by peers. Additionally, they received weekly feedback in the form of argument analysis on the epistemic (theoretical concepts and relations) and the formal (reasoning and evidence) argumentation quality for the most "liked", and for the instructors' favorite argument. These analysed arguments should be used as models in subsequent sessions. The control group received no additional guidance and was only required to hold discussions in their Facebook group. Discussion threads were organized based on posting time (last one on the top) and not on the number of "likes".

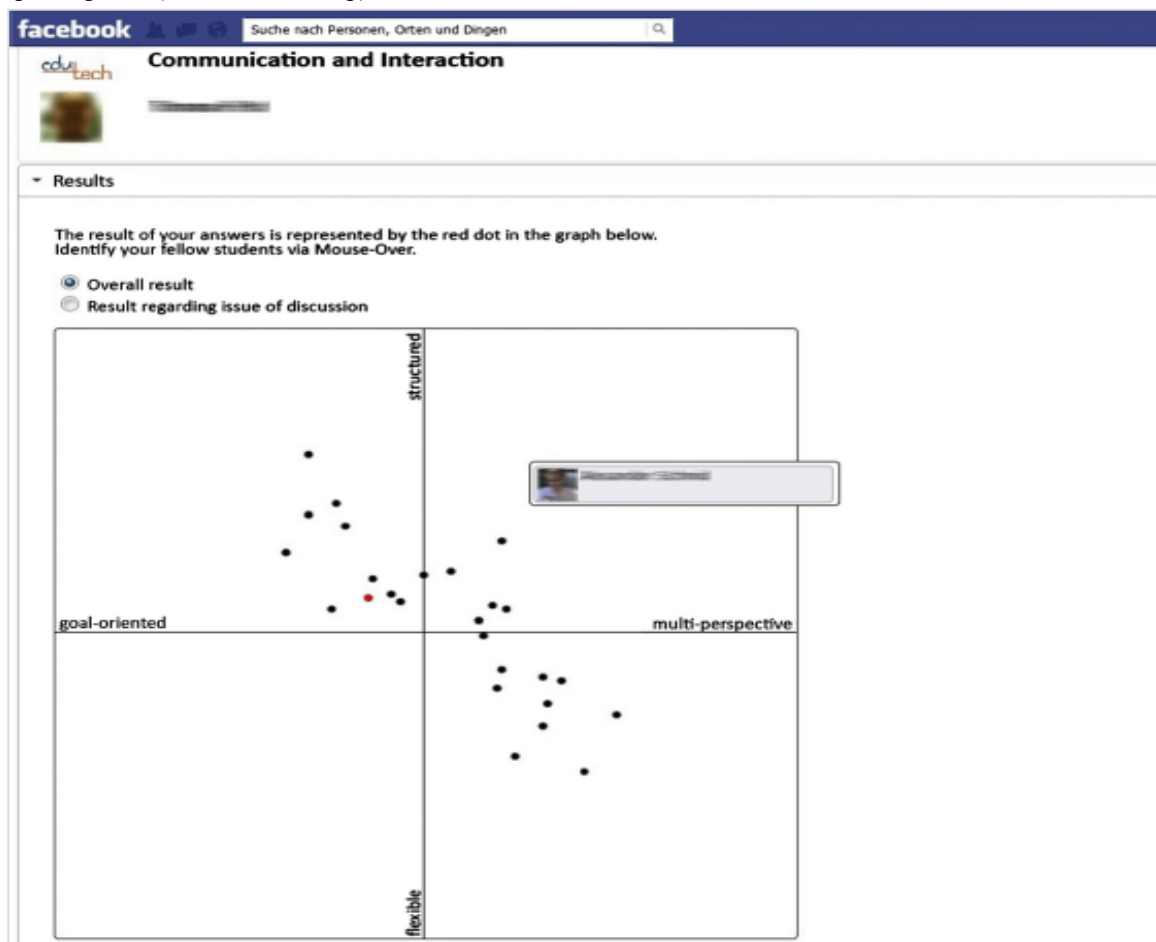


Figure 1. Group Awareness Tool in Facebook.

## Social network analysis and hypotheses

We formed a network for the participants in each SNS discussion group using *Prominence-based placement* to visualize the group as a social network, such that each node (participant) position is a reflection of its centrality inside the network and their interactions are represented using the edges. To measure interactions, the number of outgoing posts and the number of incoming replies to every post by each participant of the group was recorded. The communication attitude questionnaire was used to record attitude of each participant after each discussion. We use SNA techniques to model attitude change and mutual weight balancing in the communication network formed by participants during discussions. Centrality measures obtained through the network are then used to assess intra group characteristics of participants affected by the use of experimental conditions in terms of their prominence based placement regarding attitude change. Thus, centrality is indicative of the importance of a particular participant in the discussion group. That is the participant is important for the proper propagation of information in the whole group and is, thus, a key member during discussions. If a particular group has more key members then that signifies an even distribution of influence, importance and information in the group, which helps in increased idea exchange and knowledge co-construction.

Additionally, the similarity in attitude reached by participants in each group is assessed using *density based clustering (DBC)* (Ester, Kriegel, Sander & Xu, 1996) instead of the more commonly used *clustering coefficient*, as the latter measures simply the tendency of a group's ability to cluster together. However, DBC measures the density of a group and thus, the information about how close the participants are regarding their attitudes after the discussions. We use measures of graph theory and network analysis (Freeman, 1978; Stephenson & Zelen, 1989; Wasserman & Faust, 1994; Ester, Kriegel, Sander & Xu, 1996) to operationalize our dependent variables in SNA (see Table 2), which are generated for each of the participating groups (experimental and control group) using programming and network analysis software (SocNetV1.9). The generated data was then empirically analyzed in SPSS and a 2×2 factorial ANOVA with conditions GAT (GAT, no GAT) vs. argumentation script (ARG, no ARG) as between-subjects factors was conducted to analyze the effects, differences and interactions introduced by the GATs and the argumentation script conditions.

Table 2: Operationalisation of SNA measures

Dependent Variable	SNA Operationalization
<i>Number of interactions</i>	The sum of incoming and outgoing interactions for each member of a group measured across time.
<i>Information flow control (information centrality)</i>	The proportion of total information flow that is controlled by the participants of the network (each group member). It uses the distance between two nodes, traversing the attitude values (Euclidean distance over time) of all nodes that mediate these two nodes.
<i>Influence of individual attitudes on the group's attitude (outdegree centrality)</i>	The effect of one member's interactions on the magnitude of attitude change of other group members over the course of the seminar; where influence is quantified as the weighted outgoing interactions of the participants.*
<i>Influence of the group on individual attitudes (indegree centrality)</i>	The effect of the group's interactions on the magnitude of attitude change of each individual participant throughout the seminar, where influence is quantified as the weighted incoming interactions on each individual participant.*
<i>Attitude similarity (density based clustering)</i>	The quantification of the closeness of group members to the center of the communication network w.r.t. their attitudes, where attitude similarity within a group is measured as their respective cluster density.**

\* For example, if there are three participants in a group discussion, A, B and C, and "A" posts a view on the scenario and both "B" and "C" read it but only "C" replies once, then each participant might display a small change in their attitude by reading the discussion post alone. Therefore, a small weight of 0.1 is assigned. However, "A" and "C" is expected to have a major effect on each other's attitude due to the exchange of posts, so their weights would be represented as the relative attitude change between them, calculated as the Euclidean distance between them w.r.t the two dimension of the communication attitude questionnaire, and distributed in the network using uniform probability distribution. Therefore, if the euclidean distance between the attitude change of "A" and "C" is 4 and 6 in the two dimensions, then their relative weight would be 5.1 and the total weight of "B" would be 0.2 .

\*\* Calculated using the density based clustering algorithm (Ester et al., 1996).

We hypothesize that, the GAT, the argumentation script, and their combination will foster network measures when compared to the control group due to increased awareness and externalization of attitude differences and thus increased engagement shown in higher number of interactions (H1). Additionally, the formation of new communication channels within the groups will lead to higher centralized control per person on the total information flow (H2); and also reflection on conflicts and attitude differences would thereof increase the distribution of individual influence on the magnitude of attitude change of the group (H3); and lead to higher distribution of the group's influence on the magnitude of attitude change of individuals (H4), due to an even distribution of people central to the network. Finally, the experimental groups will foster higher attitude similarity (H5), due to the above changes in group processes, which will lead to attitude co-construction.

## Results

In terms of control variables, we found no significant differences between participants of the different conditions regarding computer use (hours per week;  $F(3, 93) = 1.48, p = .22, \eta_p^2 = .05$ , Facebook use (hours per week);  $F(3, 48.49) = 1.11, p = .35, \eta_p^2 = .03$ , familiarity on computers and SNS ( $F(3, 94) = 1.99, p = .12, \eta_p^2 = .06$ ), knowledge on communication theories ( $F(3, 94) = 1.57, p = .20, \eta_p^2 = .05$ ) and on either factor of attitude change – multi-perspective/flexible:  $F(3, 94) = .30, p = .82, \eta_p^2 = .01$  or goal-oriented/structured:  $F(3, 94) = 1.45, p = .23, \eta_p^2 = .04$ .

To test our hypotheses, we run inferential statistics using the SNA measures. We found significant differences between the groups regarding total number of interactions,  $F(3, 94) = 199.45, p = .00, \eta_p^2 = .86$ . The argumentation script had strong effects on the number of interactions with, alone,  $F(1, 94) = 479.70, p = .00, \eta_p^2 = .84$ , and in combination with the GATs,  $F(1, 94) = 119.96, p = .00, \eta_p^2 = .56$  (interaction effect). However, the GAT alone did not significantly affect the number of interactions between participants,  $F(1, 94) = 0.52, p = .47, \eta_p^2 = .01$  (Figure 2). Also, there were significant differences between groups regarding information flow control,  $F(3, 94) = 395.85, p = .00, \eta_p^2 = .93$ . We found a strong significant main effect of the argumentation script,  $F(1, 94) = 301.93, p = .00, \eta_p^2 = .76$  and the GAT,  $F(1, 94) = 64.39, p = .00, \eta_p^2 = .41$ , and a strong interaction effect,  $F(1, 94) = 934.27, p = .00, \eta_p^2 = .91$ , indicating a higher distribution of information in the network and participants displayed more centralized control over information flow (Figure 2).

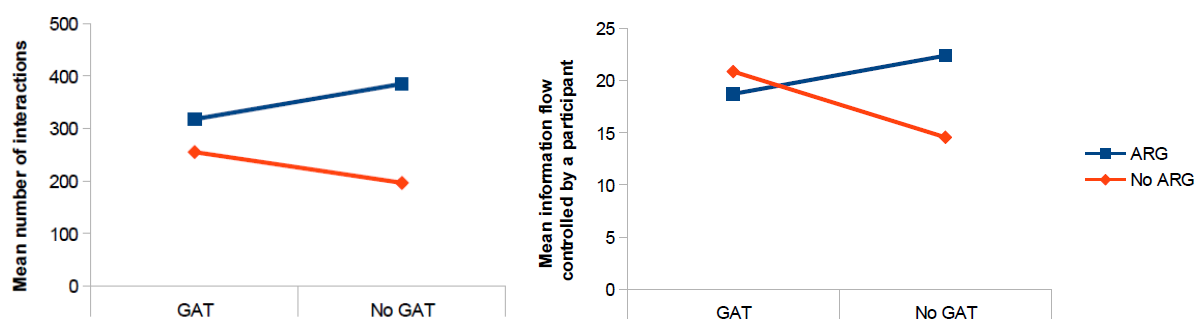


Figure 2. Total number of interactions and Information flow controlled by a participant measured by information centrality.

The analysis of the influence distribution in the network showed significant differences between groups regarding the effect of participant on group attitude change,  $F(3, 94) = 39.31, p = .00, \eta_p^2 = .56$ . There was a strong main effect of the argumentation script,  $F(1, 94) = 106.18, p = .00, \eta_p^2 = .53$ , and an interaction effect,  $F(1, 94) = 15.49, p = .00, \eta_p^2 = .14$ , that is there was an even distribution of individual influence in the network. However, there was no significant effect of GAT,  $F(1, 94) = 1.60, p = .21, \eta_p^2 = .02$  (Figure 3). Additionally, there were significant differences between the groups regarding the centralized effect of group on participant's attitude change,  $F(3, 94) = 275.21, p = .00, \eta_p^2 = .90$ . We found a strong main effect of the argumentation script,  $F(1, 94) = 743.34, p = .00, \eta_p^2 = .89$  and the GAT,  $F(1, 94) = 11.22, p = .001, \eta_p^2 = .11$ , and a strong interaction effect,  $F(1, 94) = 108.44, p = .00, \eta_p^2 = .54$ , that is there was an even distribution of group influence over the attitude change of individual participants in the network (Figure 3).

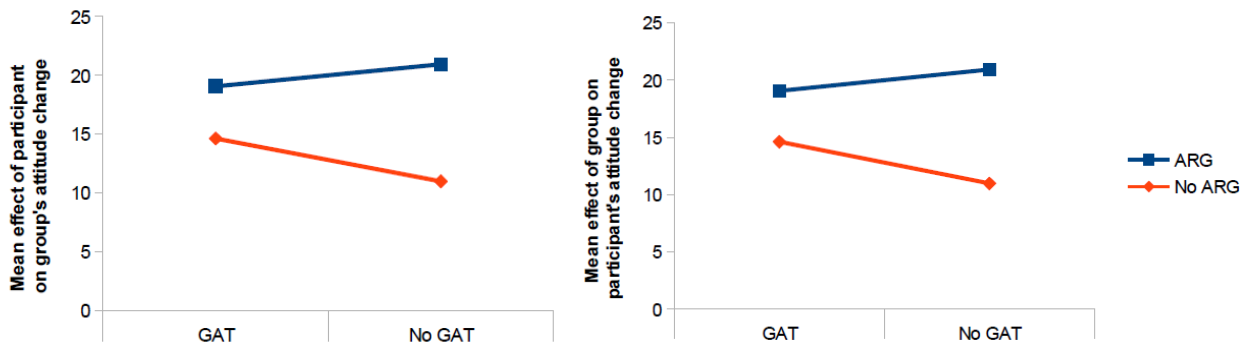


Figure 3. Effect of participant on group's attitude change and Effect of group on participant's attitude change measured by outdegree & indegree centrality respectively.

There were significant differences between the groups regarding attitude similarity – measured as their respective cluster density,  $F(3, 16.58) = 9.14, p = .001, \eta_p^2 = .26$ . We found a significant main effect of the argumentation script,  $F(1, 32) = 4.54, p = .04, \eta_p^2 = .12$ , and an interaction effect,  $F(1, 32) = 5.36, p = .03, \eta_p^2 = .14$ , that is the intra-group attitude similarity was higher with the argumentation script fostering higher similarity among participants. On the contrary, the GAT,  $F(1, 32) = 1.62, p = .21, \eta_p^2 = .05$  did not show significance regarding attitude similarity (Figure 4).

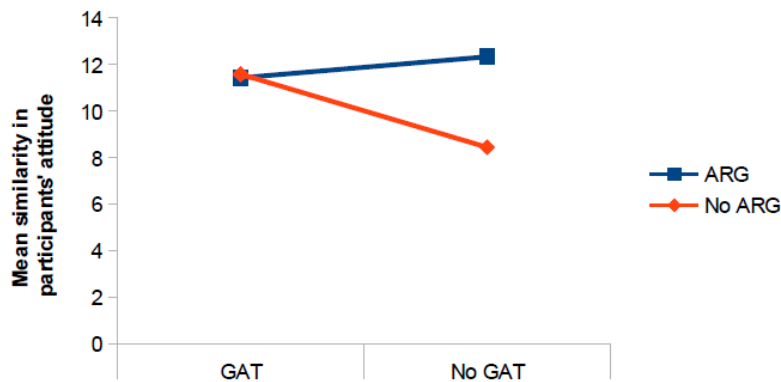


Figure 4. Similarity in participants' attitude measured by DBC.

## Discussion and conclusion

The presented study uses SNA to provide a group-level analysis of SNS group interactions on learners' attitude change. It is one of the few studies to our knowledge (Puhl et al, 2015a; 2015b) that considers attitude change not as a byproduct of discussions but as a learning outcome in itself in monitored educational settings. The study also moves beyond individual-centric considerations of the knowledge co-construction process and provides a group-centric perspective, to do justice to the theoretical origins of knowledge co-construction. We defined and analyzed internal group dynamics during group discussions in SNS and their subsequent effects on the attitude change and its relation to knowledge acquisition of the learners using a combination of SNA measures and inferential statistics.

The results revealed that group level processes – number of interactions, information and influence distribution, and similarity in peer attitude – were all influenced by the argumentation scripts and by its combination with the GAT, and some were also influenced by the GAT alone. These supports cater for a more equal distribution of information flow and influence distribution. Distributed and centralized control per participant was facilitated by the argumentation script. Together with the effects of the argumentation script on attitude change, this result suggests that increased centralized control of learners on group processes may have led to a better representation of individual attitudes and may have thus facilitated information and opinion



exchange. We had hypothesized that facilitating centralized control for more learners in group discussions would cater for more attitudes becoming transparent in the group, strengthening attitude co-construction, and that it would lead to similarity in attitudes. Centralized control of information in the GAT groups facilitated representation and promotion of individual attitudes and increased the interactions compared to the unsupported group. However, centralized control of information (*H2*) did not initiate meaningful resolution of conflicts and therefore attitude similarity due to mutual change was to not observed in the GAT groups (*H5*).

The results also suggest that the argumentation script either alone or combined with the GAT increased interactivity, despite the fact that GAT alone did not (*H1*). This is a possible indication that the GAT may have instead taken over the role of promoting attitude externalization otherwise promoted by argumentation scripts. It also indicates that conflict awareness, does not necessarily lead to conflict resolution. Moreover, the results suggest that the cognitive guidance provided by the argumentation script might have prompted learners to resolve conflicts constructively and thus reach consensus regarding attitude dissimilarity. This cannot be a case of group bias (Sassenberg & Boos, 2003) or quick consensus building (Weinberger et al, 2007), as it is unlikely that these would have led to the observed increased interaction. The GAT alone, however, may have fostered group bias without the additional argumentative processes promoted by the script that further allow reflection on attitudes. The results, thus, proclaim the necessity to combine the conflict awareness provided by GATs visualizations with structuring interactions via argumentation scripts in order to enhance and encourage meaningful exchange of opinions and arguments during SNS discussions and promote conflict resolution.

In conclusion, the interaction network formed during group discussions identifies the potential of GATs and argumentation scripts to utilize the rich context of socio-cognitive interactions provided by popular SNS during argumentative discussions, in order to support pluralistic exchange and attitude change. SNA measures can be helpful in understanding the interdependence of learning objectives on individual's interactions and group communication and, thus, also for understanding the relative effects of learning interventions. However, SNA measures are a combination of several mathematical computations and lack an intuitive interpretation, posing a fundamental dependency on the framework and interpretation of the research subjected to the used measures. Nevertheless, the ability of SNA to incorporate multiple dependencies during analysis can be crucial to the examination of learning supports in the wild in contrast to monitored course settings and can be essential in gaining meaningful insights into the long term societal influences that SNS discussions may trigger.

## References

- Andriessen, J. E. B. (2006). Arguing to learn. In K. Sawyer (Ed.), (Tran.), *Handbook of the Learning Sciences*, 443-459. Cambridge: Cambridge University Press.
- Asterhan, C. S. C. & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialogue. *Cognitive Science*, 33, 374-400.
- Baker, M. (2003). Computer-mediated argumentative interactions for the co-elaboration of scientific notions. In *Arguing to learn*, 47-78. Springer Netherlands.
- Buder, J. & Bodemer, D. (2008). Supporting controversial CSCL discussions with augmented group awareness tools. *International Journal of Computer-Supported Collaborative Learning*, 3(2), 123-139.
- Chaiken, S., Wood, W., & Eagly, A. H. (1996b). Principles of persuasion. In E.T. Higgins and A. Kruglanski (Eds.), *Social psychology: Handbook of basic mechanisms and processes* (Higgins & Kruglanski), 702-42. Guilford Press, New York, US.
- Cialdini, R. B., Trost, M. R., (1998). Social influence: social norms, conformity, and compliance. In D.T. Gilbert and S.T. Fiske and G. Lindzey (Eds.), *The handbook of social psychology*, Vols. 1-2 (pp. 151-92). McGraw-Hill, New York, US.
- Dillenbourg, P. & Fischer, F. (2007). Basics of Computer-Supported Collaborative Learning. *Zeitschrift für Berufs- und Wirtschaftspädagogik*. 21, 111 – 130.
- Erber, M. W., Hodges, S. D., & Wilson, T. D. (1995). Attitude strength, attitude stability, and the effects of analyzing reasons. In Petty, R. E. & J. A. Krosnick (Eds.), *Attitude Strength: Antecedents and consequences* (pp. 433-454).
- Erkens, M., Schlottbom, P., & Bodemer, D. (2016). Qualitative and quantitative information in cognitive group awareness tools: Impact on collaborative learning. In U. Cress (Ed.), *Transforming learning, empowering learners: the international conference of the learning sciences (ICLS) 2016*. International Society of the Learning Sciences: Singapore.
- Ester, M., Kriegl, H. P., Sander, J., Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. AAAI Press. 226-231.
- Felton, M. & Kuhn, D. (2001). The Development of Argumentative Discourse Skill. *Discourse Processes*, 32(2), 135-153.

- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117–140.
- Fischer, F., Kollar, I., Stegmann, K. & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, 48(1), 56-66.
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215-239.
- Greenhow, C. (2008). Connecting informal and formal learning experiences in the age of participatory media: Commentary on Bull et al. *Contemporary Issues in Technology and Teacher Education*, 8(3), 187–194.
- Greenhow, C., Menzer, M., & Gibbins, T. (2012). When friends debate science: Socioscientific argumentation in a Facebook application. Paper accepted for presentation at the annual meeting of the American Association of Educational Researchers. San Francisco, CA.
- Gutwin, C. & Greenberg, S. (2002). A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work*, 11(3), 411–446.
- Halatchliyski, I. (2011). Social network analysis of collaborative knowledge creation in Wikipedia. In H. Spada, G. Stahl, N. Miyake, & N. Law (Eds.), *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings*, Vol. III (pp. 1254-1258). Hong Kong: International Society of the Learning Sciences.
- Ihmeideh, F. M. & Al-omari, A. A. (2010). Attitudes toward Communication Skills among Students' -Teachers' in Jordanian Public Universities, 35(4).
- Janssen, J. & Bodemer, D. (2013). Coordinated Computer-Supported Collaborative Learning: Awareness and Awareness Tools. *Educational Psychologist*, 48(1), 40-55
- Jermann, P. & Dillenbourg, P. (1999). An analysis of learner arguments in a collective learning environment. *Proceedings of the 1999 Conference on Computer Support for Collaborative Learning - CSCL '99*, 33.
- Jermann, P. & Dillenbourg, P. (2002). Elaborating new arguments through a cscl script. In J. Andriessen, M. Baker & D. Suthers (Eds.), *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning environments*, 1–6. Kluwer Academic Publishers. Printed in the Netherlands.
- Lampert, M. L., Rittenhouse, P., & Crumbaugh, C. (1996). Agreeing to disagree: Developing sociable mathematical discourse. In D. R. Olson & N. Torrance (Eds.), *Handbook of human development in education* (pp. 731–764). Cambridge, MA: Blackwell.
- Noroozi, O., Weinberger, A., Biemans, H. J. A., Mulder, M., & Chizari, M. (2012). Argumentation-Based Computer Supported Collaborative Learning (ABCSCCL): A synthesis of 15 years of research. *Educational Research Review*, 7(2), 79-106.
- Nussbaum, E. M. (2008). Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology*, 33(3), 345–359.
- Puhl, T., Tsovaltzi, D., and Weinberger, A (2015a). A long-term view on learning to argue in Facebook: The effects of group awareness tools and argumentation scripts. Lindwall, O., Häkkinen, P., Koschman, T. Tchounikine, P. & Ludvigsen, S. (Eds.) (2015). *Exploring the material conditions of learning: The computer supported collaborative learning (CSCL) Conference 2015, Volume 1*. Gothenburg, Sweden: The International Society of the Learning Sciences, ISBN: 978-0-9903550-6-9 (pp. 110-117).
- Puhl, T., Tsovaltzi, D., & Weinberger, A. (2015b). Blending Facebook into seminars for practicing argumentation. *Computers in Human Behavior*, 53, 605-616.
- Sassenberg, K., & Boos, M. (2003). Attitude Change in Computer-Mediated Communication: Effects of Anonymity and Category Norms. *Group Processes & Intergroup Relations*, 6(4), 405–422.
- Stephenson, K. & Zelen, M. (1989). Rethinking centrality: Methods and examples. *Social Networks*, 11, 1-37.
- Tsovaltzi, D., Puhl, T., Judele, R. & Weinberger, A. (2014). Group awareness support and argumentation scripts for individual preparation of arguments in Facebook. *Computers & Education*, 76, 108-118.
- Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, ENG and New York: Cambridge University Press.
- Wecker, C. & Fischer, F. (2014). Where is the evidence? A meta-analysis on the role of argumentation for the acquisition of domain-specific knowledge in computer-supported collaborative learning. *Computers & Education* 75, 218-228.
- Weinberger, A. & Fischer, F. (2006). A framework to analyse argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46, 71-95.
- Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction*, 17(4), 416-426
- Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Review of Psychology*, 51, 539-570.

# High School Students' Collaboration and Engagement With Scaffolding and Information as Predictors of Argument Quality During Problem-Based Learning

Brian R. Belland, Nam Ju Kim, David M. Weiss, and Jacob Piland  
brian.belland@usu.edu, namju\_1001@gmail.com, dmark.weiss@usu.edu, jacob.piland@aggiemail.usu.edu  
Utah State University

**Abstract:** Strong information literacy, collaboration, and argumentation skill are essential to success in problem-based learning (PBL). Computer-based scaffolding can play a key role in helping students enhance these skills during PBL. We examined how information literacy, collaboration, and time spent in various scaffolding sections combine to predict argument quality, and qualitative analysis from the social cognitive framework perspective to explain why significant variables predicted argumentation score. Quantitative results indicated that information literacy, time spent doing individual work, and time spent on the scaffold stages *define problem* and *link evidence to claims* significantly predicted argument quality. Qualitative results suggest that there was little connection between the content of the written argument and what students wrote in the scaffold when students spent more time in individual work. Results are discussed in light of the literature.

## Conceptual framework

The Next Generation Science Standards and the Common Core invite educators to enhance K-12 students' (USA students aged 5-18) problem solving and argumentation skills (Krajcik, Codere, Dahsah, Bayer, & Mun, 2014; McLaughlin & Overturf, 2012). One way to do this is to have students address authentic, ill-structured problems within the framework of problem-based learning (PBL; Hmelo-Silver, 2004). In PBL, students work in small groups to define the problem, determine, gather and synthesize needed information, make a claim about problem resolution, and link the claim to evidence (Barrows, 1985; Belland, Glazewski, & Richardson, 2008; Hmelo-Silver, 2004). Meta-analyses indicate that PBL is particularly helpful in enhancing principles-level skill - knowledge of rules that dictate the direction and strength of relationships between concepts (Gijbels, Dochy, Van den Bossche, & Segers, 2005; Sugrue, 1995), and application level skill - ability to use concept- and principles-level knowledge to address new problems (Sugrue, 1995; Walker & Leary, 2009).

Two key factors contribute to success in PBL - argumentation ability (Kuhn, 2010) and information literacy (Hakkarainen, 2009). Argumentation ability can be defined as the ability to lead an audience to accept a claim as valid by linking evidence to the claim by way of premises (Perelman & Olbrechts-Tyteca, 1958). Many K-12 students have limited argumentation skills, as evidenced by tendencies to (a) fail to provide evidence to back up claims (Belland et al., 2008) and (b) back up claims with explanations of the claim (Glassner, Weinstock, & Neuman, 2005), irrelevant evidence, and/or evidence of questionable validity (Kuhn, 2010). Argumentation ability is crucial in PBL because students need to be able to (a) provide evidence that their solutions are well reasoned, and (b) weigh other's arguments for their solutions (Belland et al., 2008). Information literacy refers to students' abilities to identify credible/relevant sources related to a topic, and weigh the credibility and relevance of gathered information when synthesizing research (Forte, 2015; Van de Vord, 2010). K-12 students often lack sufficient information literacy, and this can result in them backing claims with evidence of questionable credibility and/or relevance (Forte, 2015; Kuiper, Volman, & Terwel, 2005).

Within PBL, effective collaboration is central to student success (Arts, Gijsselaers, & Segers, 2002). Effective collaboration arises when (a) shared work addresses a shared learning goal, and (b) students offer ideas to the group and engage with each other's ideas in a critical but constructive manner (Rojas-Drummond & Mercer, 2003), and (c) synergy among group members' efforts results in output superior to what would result from adding all group member's individual efforts (Schmidt, Rotgans, & Yew, 2011).

Scaffolding serves an important role in structuring and problematizing problem solving (Reiser, 2004), and can be defined in terms of strategies (e.g., enlisting interest, indicating important problem elements to consider, and questioning) and form (e.g., expert modeling, question prompts, and concept mapping; van de Pol, Volman, & Beishuizen, 2010; Wood, Bruner, & Ross, 1976). Some research suggests that students engage with scaffolding on the basis of their goals and the affordances that they perceive in the scaffolding (Belland & Drake, 2013). This depends on students' ability to be agentic (Bandura, 1986). The allowance for student agency also means that students can engage with scaffolding as a whole, and specific scaffolding elements to various degrees, and this can influence the quality by which they address the problem (Oliver & Hannafin, 2000).

Meta-analyses indicated that scaffolding in STEM education led to an average between-subjects effect of  $g = 0.46$  (Belland, Walker, Kim, & Lefler, In Press) and an average pre-post effect of at least  $g = 0.7$  across concept, principles, and application assessment levels, indicating that scaffolding can produce strong gains in cognitive outcomes ranging from declarative knowledge to problem solving and argumentation (Belland, Walker, & Kim, Under review; Sugrue, 1995).

## Research questions

1. Do the time that students spend in each scaffolding stage, time spent working individually, time spent working collaboratively, and information literacy scores predict argument quality?
2. How and why is argument quality explained by information literacy and student engagement with a computer-based scaffold?
3. What choices do high school students make as they construct arguments in PBL with the support of a computer-based scaffold, and why?

## Method

### Participants and setting

The study took place in one class section of environmental sciences in a medium high school in a rural setting in the Intermountain west. Twenty-four 10th grade and 11th grade students participated. The unit centered on how to improve soil quality within the county. The teacher, with 30 years of experience teaching science, had been previously exposed to computer-assisted problem-based learning in a summer school class for credit recovery students (Belland, Gu, Weiss, & Kim, 2016). Students brought to class soil samples from their houses, which they then analyzed for nitrogen, potassium, phosphorus, calcium, copper, iron, and silt/clay consistency. Working in groups of 3-4, they needed to address what could be done to optimize soil quality for specific beneficial uses (e.g., growing vegetables) and mitigate potential problems (e.g., erosion). They then needed to argue what should be done to optimize soil quality by drawing on research on the soil samples and soil quality elements that are essential to specific beneficial uses, and how to enhance such elements. The teacher provided one-to-one, dynamic scaffolding to complement the computer-based scaffolding.

### Design

We took a sequential, mixed methods approach to data analysis. Quantitative analysis in the form of Bayesian regression came first, and qualitative analysis was used to explain mechanisms underlying the relationships uncovered in Bayesian regression (Onwuegbuzie, Slate, Leech, & Collins, 2009).

### Materials

The *Connection Log* is a database-driven web application designed to scaffold middle and high school students' construction of evidence-based arguments during PBL (Belland, Gu, Armbrust, & Cook, 2015). The *Connection Log* invites students to define the problem, determine needed information, find and organize needed information, make a claim, and link evidence to claim (Belland et al., 2015). In each stage (e.g., make claim), students (a) articulate ideas individually and (b) come to consensus with groupmates to maximize perspectives and enhance collaboration. Consensus entries were entered by a group member designated scribe. The *Connection Log's* support can be classified as strategic and metacognitive (Hannafin, Land, & Oliver, 1999), and served to structure and problematize the problem solving and argumentation process (Reiser, 2004). As with CSCL tools more generally, the *Connection Log* invited and supported students in "engaging in productive processes" and "engaging in co-construction" but also allowed students to monitor and regulate their groupmates' and their own work, and allowed the teacher to do the same (Jeong & Hmelo-Silver, 2016, p. 250).

In studies in 7th grade science, lower-achieving students who used the *Connection Log* gained significantly more from pre to post on an argument evaluation test than matched controls ( $ES = 0.93$ ; Belland et al., 2015), lower-achieving experimental students performed significantly better in argument evaluation ( $ES = 0.61$ ) than matched controls (Belland, Glazewski, & Richardson, 2011) and average-achieving experimental students performed significantly better in argument evaluation ( $ES = 0.62$ ) than matched controls (Belland, 2010), and groups who used the *Connection Log* engaged with data, synthesized evidence from multiple sources, and adhered to stakeholder positions, while control groups largely tried to find out online if the river was polluted and labeled water quality elements dichotomously (Belland, Gu, Kim, & Turner, 2016). Across several studies, students used scaffolds in diverse ways in response to their needs (Belland, 2010; Belland et al., 2011, 2015). In a mixed method study among 6th grade students at risk of academic failure, students had greater science interest when they saw activities as authentic, and their epistemic aims were more sophisticated when

their science interest was higher (Gu, Belland, Weiss, Kim, & Piland, 2015).

## Data collection

### Screencasting data

Everything that students (a) did on their computers and (b) verbalized was recorded using Screencastify. This was used to indicate how students searched for information, how quickly they read and interacted with sources (e.g., scrolled rapidly), and how they interacted with the *Connection Log*. All student discourse was transcribed.

### Interviews

At the end of the unit, students engaged in 30-minute interviews covering how they used information to find solutions to the problem. Sample questions included "How did you judge the accuracy of information?" and "How did you decide on search terms while performing searches?" The interviews were transcribed verbatim.

### Log files

For each student, time spent on individual pages and text written in response to prompts was collected. We summed time spent on each stage, resulting in the following variables: time - define the problem, time - determine needed information, time - find and organize information, time - develop claims, and time - link evidence to claims. We also counted the number of words written in response to each prompt, and summed that across pages within each stage, resulting in the following variables - word count - define the problem, word count - determine needed information, word count - find and organize needed information, word count - develop claims, and word count - link evidence to claims. We also added the amount of time spent on individual and collaborative learning tasks, resulting in the following variables: time - groupwork, and time - individual work.

### Pre and post information literacy assessment

To measure information literacy, we used Tools for Real-Time Assessment of Information Literacy Skills. The 25 items cover (a) development and implementation of search strategies b) evaluation of information, and c) ethics in information use (Kovalik, Yutzey, & Piazza, 2012). Internal consistency was 0.83 (Cronbach's alpha), which is consistent with other studies ( $\alpha = 0.8$  to  $0.82$ ; Arnone, Small, & Reynolds, 2010; Salem, 2014).

### Essay rating

Student argument quality was assessed by rating their essays, in which students needed to make a claim for addressing soil quality, provide evidence to support their claims, and link evidence to claims. Two raters scored students' essays with a rubric designed to rate the structural soundness of the argument and the work that it does in presenting a solution to the stated problem. The rubric has six subcategories to assess argument quality - a) claim, b) relatedness of claim to topic, c) evidence, d) relatedness of evidence to topic, e) articulation of connection of evidence to claim, and f) relatedness of articulation of connection of evidence to claim to the topic. A maximum score was 12 and minimum score was 0. Each rater scored students' essays independently and then came to consensus. The initial inter-rater reliability between two raters with 10 samples was 0.87.

## Data analysis

### RQ1: Do the time that students spend in each scaffolding stage, time spent working individually, time spent working collaboratively, and information literacy scores predict argument quality?

We used Bayesian linear regression analysis to estimate the relationship between predictor variables (pre and post information literacy scores, and time spent (a) in each stage of the *Connection Log*, (b) working individually, and (c) working collaboratively) and essay rating, which represents argument quality. We used uniform distribution as the non-informative prior distribution and generated the posterior distribution by Metropolis-Hastings MCMC algorithms. In contrast to classic linear regression, Bayesian regression does not generate coefficients of determination (i.e., R-squared) and F values to evaluate model results because the observed data was not included in the posterior distribution and each predictor in Bayesian regression model has its own parameter at the population level (Kaweski & Nickeson, 1997). But, t-tests can still be used to identify the significance of coefficients from each predictor included within a Bayesian regression model.

### RQ2: How and why is argument quality explained by information literacy and student engagement with a computer-based scaffold? and RQ3: What choices do high school students make as they construct arguments in PBL with the support of a computer-based scaffold, and why?

*Theoretical framework.* Social cognitive theory suggests that knowledge is acquired when students choose to observe and connect with others as models of learning behavior through social interactions and experiences. Student choices can be organized into the following categories: intentionality, forethought, self-reactiveness, and self-reflectiveness. Intentionality refers to a students' proactive commitment to carry out a course of action.

Forethought occurs when students motivate themselves to guide their own actions toward a contemplated future goal. Self-reactiveness is when students exercise self-monitoring of their choices and actions by evaluating their values and judge their choices against actual and potential outcomes. Agentic perspective is evidenced through the choices a student makes that result in learning experiences contributing to knowledge building.

*Process.* We followed the process of a) data reduction, b) data display, c) data transformation, d) data correlation, e) data consolidation, f) data comparison and g) data integration. An initial set of codes were derived from reading transcripts, post unit interviews, student soil quality essays and viewing screen-capture video to account for patterns in the data. An evolving coding scheme was further applied and themes clarified and strengthened. Coding categories included information literacy topics in the Tool for Real-time Assessment of Information Literacy: develop use and revise search strategies, identify potential sources, and evaluate sources. Additional codes from Bandura’s social cognitive theory included intentionality, forethought, self-reactiveness, and self-reflectiveness.

## Results

### RQ1: Do the time that students spend in each scaffolding stage, time spent working individually, time spent working collaboratively, and information literacy scores predict argumentation quality?

Bayesian regression indicates that students’ post information literacy, individual time, and time spent on ‘Define the problem’ and ‘Link evidence to claims’ significantly predict their argument quality (See Table 1).

Table 1: Bayesian Regression Results

Bayesian normal regression		MCMC iterations = 12,500				
Random-walk Metropolis-Hastings sampling		Burn-in = 2,500				
		MCMC sample size = 10,000				
		Number of obs = 24				
Log marginal likelihood = -153.70089		Acceptance rate = .2071				
Predictor Variables	Mean	Std. Dev.	MCSE	t-value	Equal-tailed (95% Credible Interval)	
Pre information literacy	-1.736	2.486	0.542	-0.698	-6.987	2.782
Post information literacy	4.743	1.819	0.351	2.607*	1.140	8.090
Word count – Define the Problem	-3.386	2.716	0.562	-1.247	-8.781	1.710
Word count– Determine needed info	2.720	2.701	0.556	1.010	-3.115	7.525
Word count – Find and organize info	-0.748	2.034	0.299	-0.368	-4.642	3.011
Word count – Develop claims	2.179	3.227	0.440	0.675	-4.342	8.775
Word count – Link evidence to claims	2.567	1.706	0.284	1.505	-0.654	5.938
Time - Groupwork	-5.263	3.178	0.464	-1.656	-11.268	1.254
Time - Individual work	-16.178	6.043	1.580	-2.677*	-27.167	-3.486
Time – Define the Problem	11.359	3.147	0.635	3.609*	5.508	17.912
Time – Determine needed info	-2.657	6.361	1.591	-0.418	-15.500	9.900
Time - Find and organize info	5.854	4.272	1.011	1.370	-2.399	13.432
Time – Develop claims	3.302	4.147	0.858	0.796	-3.885	12.085
Time – Link evidence to claims	-13.509	3.804	0.894	-3.551*	-19.843	-5.558
Cons	49.549	10.710	2.748	4.626	30.093	69.446

The coefficient for post-information literacy was 4.74, which means that for every additional point in information literacy, students’ argument quality can be expected to increase 4.74 points. Moreover, when students spent one more hour in the section ‘Define Problem’, their argument quality will increase by 11.36 points. On the other hand, Individual time in the *Connection Log* ( $\beta = -16.178$ ) and time spent in the section ‘Link evidence to claim’ ( $\beta = -13.509$ ) negatively affected students’ argument quality.

### RQ2: How and why is student argument quality explained by information literacy and student engagement with a computer-based scaffold? and RQ3: What choices do high school students make as they construct arguments in PBL with the support of a computer-based scaffold, and why?

Due to space constraints, we only present results for one group.

*Quantitative data suggest that when information literacy increases, argumentation quality goes up.* When students recognize a need for information beyond their current level of understanding, they need to locate, evaluate and apply information effectively. The *Connection Log* has steps that encourage students to identify what they know, what they don’t know and plan for acquiring and using information. Trace data indicated that group 1 members identified topics to investigate, and Internet sources of information were listed. As a group,

they decided which of these elements to include in their study, and made assignments for group members to view and report on the sources of information. A final step was to eliminate information that they didn't think they would use. From a time perspective, these steps were steps that students spent the least amount of time accomplishing. However, their word counts were higher than all other steps except linking evidence to claims. Student interviews indicated that this group had almost a nonchalant attitude about information searches

There are three key reasons why an improvement in information literacy had a positive effect on these students' argument quality. First, when conducting web searches, students view not only words but also focus on images displayed on the page. Students not only learn how others articulate thoughts on the subject but also come to understand abstract ideas as they view images and charts. Thus, their confidence in articulating their own ideas increases. Second, web searches explain ideas about the topic that the soil analysis does not. And third, during post unit interviews, students in Group 1 considered information found on the Internet to be authoritative when it was (a) repeated on several sites, or (b) consistent with information the teacher had explained. Student essays created by this group were well organized and often detailed in support of their ideas.

*Quantitative data suggest that when students spend more time in "Define the Problem," argument quality goes up.* The rubric used to evaluate argument quality assessed claims, evidence and linking evidence to claims. The measurement was conducted on two levels: the logic of argument in the essay and the connection of the argument to the topic of the PBL unit. Of these six rubric sections, the two on which students generally received the highest scores were those related to claims. Within the computer-based scaffold two of the five steps relate directly to claim making namely, "Define the Problem" and "Develop Claims". "Define the Problem" is the first activity in the *Connection Log*. Their motivation is high and a novelty effect exists.

For all four Group 1 members, both the topic of problem statement and the topic of the claim are identical. The similarity of both suggests an alignment in the minds of the students that was reinforcing their view of both the problem and the claim. In addition, the claim topics in group 1 essays were aligned with students' problem statements and claims in the *Connection Log*. Among the rubric scores, the claims elements were highest. In this sense, there is an alignment between their problem statement and their claim in the software and their claim in their essay. In addition to essay topics, problem statement topics and claim topics being centered around low nitrogen as the problem, low nitrogen was also a topic covered by the instructor specifically as a problem in their locale. Further, low nitrogen was a determination resulting from soil tests.

It is important to consider that the "Define the Problem" step includes both individual work and group work. The quantitative measurements suggest a negative effect for individual work and an unknown effect for group work. Yet the combination of individual work and group work in this step resulted in a positive effect with respect to argumentation. Perhaps the alignment of activities in this step, subsequent activities, their soil test results and the teacher's instruction combined to create a positive effect on argument quality.

Quantitative data suggest that when students spend more time on "Linking Evidence to Claims," argumentation quality does not improve. The "link evidence to claim" step, the fifth of the five major steps in the *Connection Log*, guides students through four individual sub-steps. Two of the sub-steps are completed individually (Select Evidence and Test Evidence) and two are group activities (Put it all Together and Conclusion). The individual activities request that students enter their work by themselves. When the software presents a group activity, students gather around the scribe's computer, discuss information on the screen and decide together what their entries should be. These steps are perhaps most important as they finalize their claim, assign and validate the evidence they are using to support their claim. However, students spent only 5% of their total time using the *Connection Log* working through this step. Further, individual work by members of Group 1 represented less than 2% of the total word count entered by the group for the entire step. Additionally, two students in Group 1 made no individual entries for this step at all. At the conclusion of this step, students were expected to apply the claims and evidence they had entered into the software to their argumentation essays.

All four essays from group 1 students suggest that low nitrogen was a problem and composting was a possible solution. However, only two of the four students in Group 1 supported their claim by referencing soil analysis tests. Only one student of the four in Group 1 referenced measurements from those tests.

There may be several explanations for the quantitative finding that spending more time on this step would result in a decrease in argument quality. First, as measured by time, word count and the review of their essays, the linking of evidence to claims step was characterized by a lack of individual effort in deference to group work, largely done by the scribe. Further, when the software instructed them to work as a group, only the scribe was tasked with articulating through writing into the software. Screen-capture video and transcripts confirm that Group 1 did discuss briefly what to enter into the software, but the actual recording of their ideas was the scribe's responsibility. Spending more time in this step would likely not contribute to individual argument quality when articulation of ideas in this step was largely the scribe's responsibility.

Second, observations indicated that the teacher (a) made specific reference to nitrogen levels being low

in their locale, and (b) informed students that composting was an important solution to soil quality problems. As students spent only 5% of the total time in the software on this step, and little time individually articulating their own ideas in those steps, it is likely that they depended more on their teacher's lecture points in their essays than the work they did in the software. Essays confirm that students focused on low nitrogen and composting in their claims and evidence. Thus, spending more time on this step would not contribute to better argumentation scores.

Third, fatigue is a consideration. Given that two students made no individual entries at all, and the other student's individual entries were minimal, the desire to "finish" the unit may have overcome the need to complete this last step carefully. Argumentation rubric scores support the struggle this group had with articulating evidence and linking evidence to claims. During the post unit interview, students agreed that they stalled using the software, inducing frustration and confusion. Identifying points of evidence that require investigation into collected data, and connecting data to claims, is a higher order skill that these students seem to have resisted especially at the end of the unit.

*Quantitative data suggest that when students spend more time engaged with individual activities as opposed to group activities in the software, student argumentation quality does not improve.* The *Connection Log* consists of 20 separate steps, 11 of which are completed by individual students and 9 of which are completed as a group. Individual activities include a) define problem individually, b) determine needed information, c) find and organize information, d) generate claim and e) link evidence to claims. Some Group 1 members were methodical entering information step by step. Others spent time on each page but made no entries at all. Students were guided through each step of the PBL process to increase their awareness of what they already knew about the topic, what they needed to know and how they can organize this information as evidence to support a claim. As each student identified what they knew and gathered information needed to understand the problem they identified in the very first step, it was expected that what was entered into the software would make it into the final student essays. However, student essays reflected more of the teacher's lectures than the student's investigative work or interpretation of collected data from internet sources. In some cases, interview data and screen capture software confirm that students completed individual activities using Google searches that led to promising sites on the Internet, as well as visiting the web-based resource page prepared by the research team. Entries were made in the software identifying these sites as sources of evidence as they wrote their arguments. However, successive individual entries referenced the information delivered by their teacher rather than the information they searched for and read on the internet.

There are several explanations why spending more time doing individual work in the computer-based scaffold would not translate into improved argumentation scores. There seems to be a disconnect between their individual work in the scaffold and what they wrote in the essays. High school students often focus on "getting things done". In this way, they may have completed each step in the software without connecting what they did in the software to their essay writing. Spending more time would not be reflected in argumentation scores.

While some students admitted to having a hard time using the software occasionally, they perceived that it invited them to do both individual and group work, citing the pros and cons of both approaches to learning. Students experienced high motivation to write a report that would inform their parents about soil close to their place of residence. However, word counts demonstrated that group activities generated higher interaction with the software than individual activities. Individual activities might have been perceived by students as check-off activities if, in the end, the scribe would articulate answers with greater detail. This is further supported in that for some students in group 1, individual activities had no or little response recorded.

## **Discussion and implications**

One key contribution of this paper was collection of empirical evidence of the prediction of argument quality by information literacy. This implies that the pursuit of enhanced information literacy is not important only in terms of making sure that students have the tools to engage with information effectively, but also so that they can engage with and generate effective arguments.

That time spent in define problem was a significant, positive predictor of argumentation quality, and the one that explained the most variance in argument quality makes sense, as quality of problem definition is often considered to be one of the most important contributors to problem solving ability (Chi, Feltovich, & Glaser, 1981; Jonassen, 2000). And to successfully define a problem, one needs to take time to think about it qualitatively. Problem definition is not always central to argumentation scaffolding (Scheuer, Loll, Pinkwart, & McLaren, 2010), but this result may suggest that it should be.

Time spent working individually was a significant negative predictor of argument quality, which makes sense because students who spent much time working individually and correspondingly less time working collaboratively did not put in the time to critically and constructively engage with their groupmates' ideas (Hmelo-Silver & Barrows, 2008; Rojas-Drummond & Mercer, 2003). At the same time, some students spent



relatively little time working individually because they expected the scribe to answer the questions. This would similarly lead to poor collaboration and poor argument quality. Ultimately, what is needed is the right balance between individual and collaborative work time. Further research is needed.

The finding that time spent in the link evidence to claims stage was a significant negative predictor of argumentation ability was surprising. But evidence indicates that the longer students spent in this stage, the less they actually did individual work, and simply left it up to the scribe to do the work. This, of course, is not effective collaboration, and thus conditions were not set for work that leverages the collective group strengths.

## References

- Arnone, M. P., Small, R. V., & Reynolds, R. (2010). Supporting inquiry by identifying gaps in student confidence: Development of a measure of perceived competence. *School Libraries Worldwide*, *16*, 47.
- Arts, J. A. R., Gijsselaers, W. H., & Segers, M. S. R. (2002). Cognitive effects of an authentic computer-supported, problem-based learning environment. *Instructional Science*, *30*, 465–495.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Barrows, H. S. (1985). *How to design a problem-based curriculum for the preclinical years*. NY: Springer.
- Belland, B. R. (2010). Portraits of middle school students constructing evidence-based arguments during problem-based learning: The impact of computer-based scaffolds. *Educational Technology Research and Development*, *58*(3), 285–309. doi:10.1007/s11423-009-9139-4
- Belland, B. R., & Drake, J. (2013). Toward a framework on how affordances and motives can drive different uses of computer-based scaffolds: Theory, evidence, and design implications. *Educational Technology Research & Development*, *61*, 903–925. doi:10.1007/s11423-013-9313-6
- Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2008). A scaffolding framework to support the construction of evidence-based arguments among middle school students. *Educational Technology Research and Development*, *56*(4), 401–422. doi:10.1007/s11423-007-9074-1
- Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2011). Problem-based learning and argumentation: Testing a scaffolding framework to support middle school students' creation of evidence-based arguments. *Instructional Science*, *39*, 667–694. doi:10.1007/s11251-010-9148-z
- Belland, B. R., Gu, J., Armbrust, S., & Cook, B. (2015). Scaffolding argumentation about water quality: A mixed method study in a rural middle school. *Educational Technology Research & Development*, *63*(3), 325–353. doi:10.1007/s11423-015-9373-x
- Belland, B. R., Gu, J., Kim, N., & Turner, D. J. (2016). An ethnomethodological perspective on how middle school students addressed a water quality problem. *Educational Technology Research & Development*, *64*, 1135–1161. doi:10.1007/s11423-016-9451-8
- Belland, B. R., Gu, J., Weiss, D. M., & Kim, N. J. (2016). An examination of credit recovery students' use of computer-based scaffolding in a problem-based, scientific inquiry unit. Presented at the Annual Meeting of the American Educational Research Association, Washington, DC, USA.
- Belland, B. R., Walker, A. E., & Kim, N. J. (Under review). A Bayesian network meta-analysis to synthesize the influence of contexts of scaffolding use on cognitive outcomes in STEM education.
- Belland, B. R., Walker, A. E., Kim, N., & Lefler, M. (In Press). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research*.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science: A Multidisciplinary Journal*, *5*(2), 121–152.
- Forte, A. (2015). The new information literate: Open collaboration and information production in schools. *International Journal of Computer-Supported Collaborative Learning*, *10*(1), 35–51.
- Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research*, *75*, 27–61.
- Glassner, A., Weinstock, M., & Neuman, Y. (2005). Pupils' evaluation and generation of evidence and explanation in argumentation. *British Journal of Educational Psychology*, *75*, 105–118.
- Gu, J., Belland, B. R., Weiss, D. M., Kim, N. J., & Piland, J. (2015). Middle school students' science interest and epistemic beliefs in a technology-enhanced, problem-based, scientific inquiry unit. Presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, USA.
- Hakkarainen, P. (2009). Designing and implementing a PBL course on educational digital video production: lessons learned from a design-based research. *Educational Technology Research & Development*, *57*(2), 211–228. doi:10.1007/s11423-007-9039-4

- Hannafin, M., Land, S., & Oliver, K. (1999). Open-ended learning environments: Foundations, methods, and models. In C. M. Reigeluth (Ed.), *Instructional design theories and models: Volume II: A new paradigm of instructional theory* (pp. 115–140). Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Hmelo-Silver, C. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, *16*, 235–266. doi:10.1023/B:EDPR.0000034022.16470.f3
- Hmelo-Silver, C., & Barrows, H. (2008). Facilitating collaborative knowledge building. *Cognition and Instruction*, *26*, 48–94. doi:10.1080/07370000701798495
- Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist*, *51*, 247–265.
- Jonassen, D. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, *48*(4), 63–85. doi:10.1007/BF02300500
- Kaweski, D., & Nickeson, M. (1997). *C-SHRP Bayesian modelling: A user's guide*. Ottawa, Canada: Canadian Strategic Highway Research Program. Retrieved from <https://trid.trb.org/view.aspx?id=474116>
- Kovalik, C. L., Yutzey, S. D., & Piazza, L. M. (2012). Assessing change in high school student information literacy using the tool for real-time assessment of information literacy skills. *Contemporary Issues in Education Research (CIER)*, *5*(3), 153–166. doi:10.19030/cier.v5i3.7092
- Krajcik, J., Codere, S., Dahsah, C., Bayer, R., & Mun, K. (2014). Planning instruction to meet the intent of the Next Generation Science Standards. *Journal of Science Teacher Education*, *25*(2), 157–175.
- Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, *94*, 810–824.
- Kuiper, E., Volman, M., & Terwel, J. (2005). The web as an information resource in K–12 education: Strategies for supporting students in searching and processing information. *Review of Educational Research*, *75*, 285–328. doi:10.3102/00346543075003285
- McLaughlin, M., & Overturf, B. J. (2012). The Common Core: Insights into the K–5 standards. *The Reading Teacher*, *66*, 153–164. doi:10.1002/TRTR.01115
- Oliver, K., & Hannafin, M. J. (2000). Student management of web-based hypermedia resources during open-ended problem solving. *Journal of Educational Research*, *94*, 75–92.
- Onwuegbuzie, A. J., Slate, J. R., Leech, N. L., & Collins, K. M. (2009). Mixed data analysis: Advanced integration techniques. *International Journal of Multiple Research Approaches*, *3*(1), 13–33.
- Perelman, C., & Olbrechts-Tyteca, L. (1958). *La nouvelle rhétorique: Traité de l'argumentation [The new rhetoric: Treatise on argumentation]*. Paris, France: Presses Universitaires de France.
- Reiser, B. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*, 273–304. doi:10.1207/s15327809jls1303\_2
- Rojas-Drummond, S., & Mercer, N. (2003). Scaffolding the development of effective collaboration and learning. *International Journal of Educational Research*, *39*(1–2), 99–111.
- Salem, J. A. (2014). *The development and validation of all four TRAILS (tool for real-time assessment of information literacy skills) tests for K-12 students* (PhD Dissertation). Kent State University, Kent, OH.
- Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, *5*, 43–102.
- Schmidt, H. G., Rotgans, J. I., & Yew, E. H. (2011). The process of problem-based learning: what works and why. *Medical Education*, *45*(8), 792–806. doi:10.1111/j.1365-2923.2011.04035.x
- Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem-solving ability. *Educational Measurement: Issues and Practice*, *14*(3), 29–35.
- van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, *22*, 271–296. doi:10.1007/s10648-010-9127-6
- Van de Vord, R. (2010). Distance students and online research: Promoting information literacy through media literacy. *The Internet and Higher Education*, *13*(3), 170–175. doi:10.1016/j.iheduc.2010.03.001
- Walker, A., & Leary, H. (2009). A problem based learning meta analysis: Differences across problem types, implementation types, disciplines, and assessment levels. *Interdisciplinary Journal of Problem-Based Learning*, *3*(1), 12–43. doi:10.7771/1541-5015.1061
- Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*, 89–100. doi:10.1111/j.1469-7610.1976.tb00381.x

## Acknowledgements

This research was supported by Early CAREER Grant 0953046 awarded to the first author by the National Science Foundation. Any opinions, findings, and or conclusions are those of the authors and do not necessarily represent official positions of NSF.

# Using Multimodal Learning Analytics to Identify Aspects of Collaboration in Project-Based Learning

Daniel Spikol, Malmö University, Sweden, daniel.spikol@mah.se  
Emanuele Ruffaldi, Scuola Superiore Sant'Anna, Italy e.ruffaldi@santannapisa.it  
Mutlu Cukurova, University College London, UK, m.cukurova@ucl.ac.uk

**Abstract:** Collaborative learning activities are a key part of education and are part of many common teaching approaches including problem-based learning, inquiry-based learning, and project-based learning. However, in open-ended collaborative small group work where learners make unique solutions to tasks that involve robotics, electronics, programming, and design artefacts evidence on the effectiveness of using these learning activities are hard to find. The paper argues that multimodal learning analytics (MMLA) can offer novel methods that can generate unique information about what happens when students are engaged in collaborative, project-based learning activities. Through the use of multimodal learning analytics platform, we collected various streams of data, processed and extracted multimodal interactions to answer the following question: which features of MMLA are good predictors of collaborative problem-solving in open-ended tasks in project-based learning? Manual entered scores of CPS were regressed using machine-learning methods. The answer to the question provides potential ways to automatically identify aspects of collaboration in project-based learning.

## Introduction

Collaborative learning activities are key part of education and are part of many common teaching approaches including problem-based learning, inquiry-based learning, and project-based learning. Such constructivist teaching approaches have the potential to help foster the 21st-century learning skills we require of young people across subject domains (Banks & Barlex 2014). Particularly within the context of computer science education, most of these activities take place as open-ended, collaborative, small group work where learners make unique solutions to tasks that involve robotics, physical computing, and programming code as well as designing artefacts. However, the evidence on the effectiveness of using such teaching methods to satisfy common learning outcomes is rare (Klahr & Nigam 2004; Kirschner et al. 2006). Blikstein and Worsley (2016) argue that one reason for this be that evaluation in this context is notoriously laborious and requires measurement methods that the current standardised testing strategies and psychometrics cannot provide. Learning analytics research, particularly multimodal learning analytics (MMLA) can offer novel methods that can generate distinctive information about what happens when students are engaged in collaborative project-based learning activities (Worsley & Blikstein 2014). In this paper, we focus on collaborative problem-solving (CPS). We present an empirical study through which we explored CPS in groups of university engineering students (aged 20-22 years) through the use of specially designed workstation and MMLA system that collected diverse multimodal interaction data.

MMLA offers researchers new tools to capture different types of data from these complex learning activities (Ochoa et al. 2013). The ability to collect multimodal data from bodily movements, face tracking, affective sensors, log files from the hardware and software, user and research generated data provide opportunities to obtain useful features for understanding collaborative learning. Through the use of multimodal learning analytics platform, we collected diverse streams of data from learning activities. We processed and extracted multimodal interactions to answer the following question: *which features of MMLA are good predictors of CPS in open-ended tasks in project-based learning?* In particular, we performed a regression task over human evaluated CPS scores by means of machine learning techniques. The answer to the questions provides ways to automatically identify aspects of students CPS practices and provides means for different types of interventions to support and scaffold the students and inform teaching practices.

## Collaborative problem-solving

CPS is a term that is increasingly used to refer to the process of people working together to solve a problem with equivalent roles. It brings together individual problem-solving and the social collaborative process of more than one individual learner working together. Both the subject of problem-solving and the subject of collaborative learning have a substantial research history in their right. However, it is important that we make clear what we mean by the term CPS, because, as learning analytics developers, we rely on effective frameworks to drive the

analysis of our data that answers the research questions we pose. Research questions that are themselves shaped by our theoretical understanding, which enables us to make sense of our data, to identify data sets that indicate the effective implementation of the educational construct under investigation and differentiate them from data sets that evidence a less effective implementation. The OECD Collaborative Problem Solving Framework, for example identifies three dimensions for understanding and assessing problem solving: context, task and process (2015).

Context in CPS is described as the circumstances of the problem being solved. Context consists of the resources that are available to learners to support their collaborative learning activity (Luckin, 2010). A CPS task can be thought of as a set of features that represent a gap or crossroads where the way forward to solve the problem is to an extent unknown and must be generated and/or co-constructed by two or more participants. CPS might be as much about identifying a possible solution as about identifying and producing the solution. The process of CPS requires the combination or the inter-relation of social and cognitive processes. Ideally, interaction and joint problem-solving will centre on a number of parallel cognitive activities, such as understanding the problem situation, clarifying sub goals and reflecting on assumptions. In addition to the OECD approach, which was developed for assessment of individual student capacities, we can also consider CPS as a tuition approach (Cukurova et al. 2016) and at groups and communities' levels (Dillenbourg, & Jermann, 2007). These considerations would also increase the complexity further.

CPS is a complex concept with multiple dimensions. In the literature, many variables have been identified as indicators of successful CPS, or collaborative learning and warrant further investigation. These variables include equality and mutuality (Damon & Phelps 1989), symmetry (Dillenbourg 1999), synchrony of groups' actions and gaze (Schneider & Blikstein, 2015; Schneider & Pea, 2013), individual accountability of participants (Johnson, & Johnson, 2003; Springer, Stanne, & Donovan, 1999), reaction time of participants to the actions of members of the the group (Raca, Tormey, & Dillenbourg, 2014), and reaction of students to the prompts of teachers (Sharma, Jermann, & Dillenbourg, 2014). In this paper, we further explore synchrony and individual accountability as independent variables to identify CPS. It is important to note here that the results on synchrony of groups' body actions are not conclusive. Although, overwhelming majority of existing evidence suggests that synchrony can predict collaboration in Educational Psychology research (cf. Lakens, & Stel, 2011; Wiltermuth, & Heath, 2009), this may not be the case for all data sets and interpretations of synchrony. For instance, although Schneider and Blikstein (2015) find that synchrony in groups' actions may not predict collaborative learning, their gaze synchrony *does* predict it. Therefore, in our interpretation, we used a combination of their hand movements and head direction (where they are looking) to interpret synchrony. In addition to these two variables, we use the amount of physical engagement of students as this is an important aspect of project-based learning environments.

### Multimodal learning analytics

To better accommodate learning in small groups, researchers typically use low-cost sensors and inexpensive computational power for obtaining data from diverse sensors that include computer vision, audio, biometric, and data from the learning objects (like physical computing components or laboratory equipment) to collect insights. The multimodal data from these sensors provides new opportunities for investigating learning activities in the real-world between small groups of learners working on tasks with physical objects (Halverson & Sheridan 2014; Blikstein & Worsley 2016).

There is an emerging body of work with MMLA to capture small group work on project-based learning that has grown out of the work of (Blikstein & Worsley 2016; Chen et al. 2014; Ochoa et al. 2013) explored multimodal techniques for capturing code snapshots to investigate students learning computer programming and video and gesture tracking for engineering tasks. Worsley and Blikstein (2014) presented different approaches for data integration and fusion and how these can have a significant impact on the relation of research and learning theories. These approaches provided the means for other researchers to begin to explore MMLA with small groups of students across different subjects. Ochoa et al. work (2014) used existing multimedia processing technologies to produce a set of features for accurate predictions of experts in groups of students solving math problems. Grover et al. (2016) have explored how to develop test computational models of social in CPS learning environments. Their approach has been to classify the quality of collaboration from body movement and gestures of pair programmers working together. Drawing from the literature we can observe that MMLA has a role to play to support CSCL in project-based learning through looking at what types of multimodal interaction is relevant for understanding CPS. Additionally, opportunities exist for more investigation with MMLA to gain insights into CPS.

### PELARS system and context

The work discussed in this paper is based on the European project Project-based Learning Analytics for Research and Support (PELARS<sup>1</sup>). One of the aims of the project is to develop learning analytics tools for hands-on, open-ended STEM and STEAM learning activities using physical computing. The current system includes customised furniture with an integrated Learning Analytics System (LAS) such as tracking hands, faces and other objects and the Arduino platform with a visual web-based Integrated Development Environment (IDE) that captures interaction information of physical computing. The learners and observers use mobile devices to capture multimedia data (text, images, and video) to self-document the learning activities (Ruffaldi et al. 2016). See the system in action in Figure 1 below.

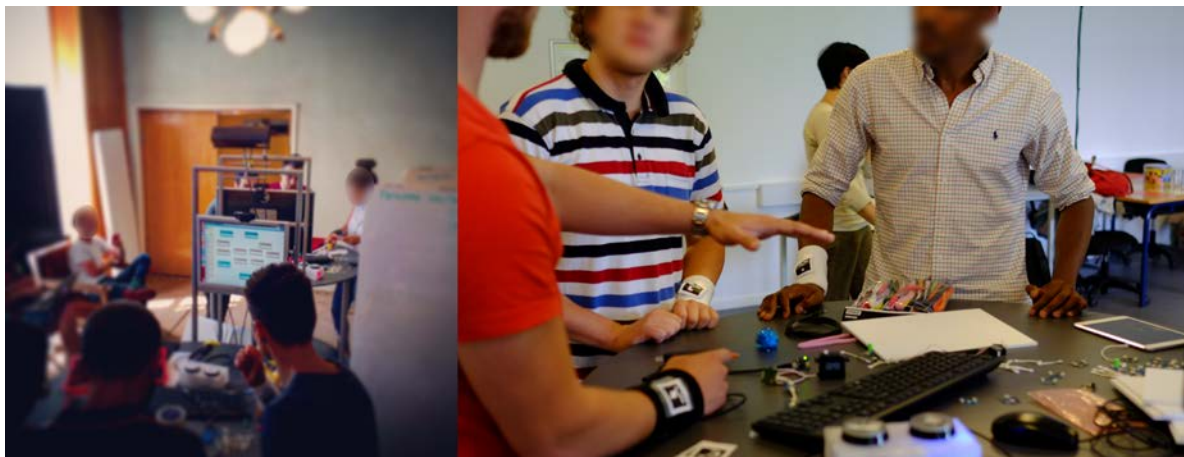


Figure 1. Different views of the PELARS system.

The PELARS LAS captures data streams from face and hand tracking, the ARDUINO IDE that includes hardware and software log-files, and audio levels that more fully explained in the next sections.

### Datasets

The data employed in this paper is based on 12 sessions of 3 students studying engineering at a European University (average age 20 years old with 17 and 1 woman). Each student group used the system over 3 days completing 3 open-ended tasks. The students were introduced to the system and then their task was to prototype an interactive toy. No specific instructions about the timing of the sessions were given to students. Each session required between 60 and 80 minutes for the students to complete the task.

### Evaluation of CPS

The starting point of the automatic assessment is the expert coding of the students' sessions based on video captured with the LAS. The coding scheme makes use of three levels (0, 1 and 2) to represent passive, semi-active and active student states, one for each student of the team. The active code (2) is used whenever a student's hand is active on an object; the semi-active code (1) is used when a student is not physically active but his head is directed towards a peer, or teacher, whatever is active. Finally, the passive code (0) is used if student's hands are not physically active with any object and their head is directed away from the active position of the peers. The combination of the three codes of each student provides a synthetic representation of collaboration.

Students' behaviours have been coded in 30-second windows. To validate the coding actions, two researchers applied this coding scheme to all video data. This procedure has been used as a way of testing the validity of the coding system generated. Where there was disagreement, the researchers discussed the data and revised their coding accordingly.



**Figure 2.** Students from left to right have a coding of 122 meaning that leftmost student is looking, and the other two are working together.

The per time-window coding has been used to compute three aggregate scores for the whole session based on the semantics of the encoding. **Physical engagement** is measured by the percentage of code 2 over the total. **Synchronisation** is measured by the combined percentage of 222 and 111 codes. These are the two situations in which all students are active or all semi-active. The 111 code corresponds to the specific case in which a facilitator or teacher enters the scene and the students are looking (semi-active) at him/her. Finally, **individual accountability** has been measured by the total number of situations in which at least one student is looking at another student actively working: that is all the combinations of a 2 with two 1 code, and two 2 codes with one 1 code (e.g. 211, 221). These are the situations in which at least one student is paying attention to the physical actions of another student. Table 1 presents the summary of the CPS evaluation results based on this coding for the 12 sessions. Each score is presented as a percentage of the whole time.

**Table 1:** Researchers' Coding of the student sessions with the three CPS scores as percentages of the total session duration

Session No.	Physical Engagement PE Total	Synchronicity			Individual Accountability IA Total
		SYN 111 (teacher)	SYN 222	SYN Total	
1	46,4%	30,88%	0,00%	30,88%	66,18%
2	50,00%	0,00%	10,00%	10,00%	50,00%
3	53,7%	1,05%	16,84%	17,89%	46,32%
4	67,97%	7,1%	38,3%	45,45%	47,40%
5	66,0%	0,0%	25,9%	25,95%	55,06%
6	77,92%	0,6%	48,7%	49,35%	46,75%
7	51,9%	3,6%	12,3%	15,94%	64,49%
8	60,75%	0,0%	24,8%	24,80%	34,40%
9	62,7%	0,0%	30,1%	30,15%	34,56%
10	74,9%	3,0%	43,6%	46,53%	53,5%
11	46,4%	1,7%	11,7%	13,33%	44,17%
12	60,06%	0,0%	24,5%	24,53%	53,77%

**Acquired data**

For each of the sessions recorded, the LAS system collected data from the students comprising activity performed, user generated content (text and multimedia) and actions on the Arduino visual Integrated Development Environment (IDE). In particular, the following data has been acquired:

**Face Tracking** - Using a frontal camera and the Viola-Jones algorithm the face of students was tracked and thanks to camera calibration and assumptions about sizes it was possible to estimate the 3D position from the

camera. This means that the position of the face is computed in 3D coordinate. Two metrics have been identified: the first is the count of Faces Looking at the Screen (**FLS**), the second is the average between all the faces pairs providing an indicator called Distance between Learners (**DBL**). The measure **DBL** could be seen as a marker of collaboration obtained when **DBL** is a small value.

**Hand Tracking** - A top down camera, instead, monitored the motion of the hands of the students that were wearing fiducial markers that disambiguate each primary hand. Again, thanks to the calibration of the camera and the size of the markers the 3D pose of the hands was obtained. The resulting metrics were the Distance between Hands (DBH) and the Hand Motion Speed (HMS), respectively as the average distance between all the hands, and as the average motion speed.

**Other Data collected but not analysed in this paper** - The interface between the Arduino visual IDE and the data collection system provided information about the types of physical and software blocks used in the project and their connections. **Audio Level** - By means of the microphone included in one of the cameras and Fast Fourier Transformation (FFT) we compute the sound level during the session. The resulting feature is a value sampled at 2Hz called Audio Level (AUD).

## Methods

This section presents how we approached investigating which multimodal features are good predictors for CPS for the PELARS LAS. Data pre-processing and data analysis methods are explored based on the coding scheme results. The hypotheses rest on creating a set of regressors (independent variables) of the CPS scores and testing how the different observed features affect the quality of the regressors. This approach provided information on sensors and measures that can be as strong predictors for collaboration in the group.

## Preprocessing

Data has been collected at variable data rates (around 2 Hz) but without relative time offset of few seconds, and for this reason the pre-processing aggregates indicators from the different variables in large windows of same durations. The aggregation performed was based on counting for most of the variables, except for the distance/proximity functions for which we employed averaging, maximum and minimum. For taking into account the different durations of the sessions (average 63min, min 40, max 79min, std 15min). we employed zero padding for sessions that were too short. Additionally, the individual sessions were then broken down into three phases, planning, work, and reflection to help analyse the workflow of the student groups. Work phase lasted 42min in average, min 14, max 65, std 15. The overall recording time of 12.5 hours.

Additionally, these phases were coded by research observers during the sessions using the mobile component of the LAS (see Spikol et al., 2016). The data acquired by the PELARS LAS was exported from the server and then processed in Python using the sklearn toolkit<sup>2</sup> that provided state of the art machine learning techniques integrated with a common interface.

Fiducial markers were relatively reliable in positioning but they were subject to visual blockage. We considered the amount of time between marker presence greater than 2 seconds. This resulted visual occlusion in 65% of time in average during work phase (min 21%, max 94%, std 20%) in particular we are interested in the presence of all the three hands at once (every 20s in average, min 5s, max 33s, std 8s). Face tracking faces similar difficulties with an average of 55% of time in average during work phase (min 0%, max 97%, std 30%).

## Machine learning

This initial approach was based on regression task that used as inputs the features and as output the coding based scores (PE, SYN, IA) with the purpose of identifying which are the input features that can support the CPS framework. Among the different families of regressors we opted for Linear Regression (LR), Bayesian Ridge Regression (BRR) and Support Vector Machine Regression (SVR). A Ridge regression introduces parameters for keeping the size of the weights small, while BRR performs a regression based on the Bayesian framework so that it is possible to better handle ill posed problems. Then we statistically modelled the effect features over the outcomes using a General Linear Model provided by the Python state-model package: we use the GLM because we have a large number of variables with possible non-trivial interactions. Indeed, some of the employed regressors were not linear.

Data obtained from the sensors have been standardized to improve the learning rate. When considering a global window only 8 features were available, then 24 for 1800s, 32 for 1200s and 64 for 600s. We have manually explored the feature selection process by backward processing.

We used cross-validation ( $k=4$ ) for evaluation of the regression because due to the number of samples (12) the use of a leave-one-out scheme would bring to perfect regression. This means that 7 subjects were used for the training and 5 for testing. We compared the quality of the different regressors by using the  $R^2$  chi square metric, accepting regressions only in the range 0.1.

We also explored the effect of different parameters such as window size and the inclusion of different phases. Tested window sizes were 10,20 and 30 minutes (600,1200 and 1800 seconds), in addition to the case of one single window for the whole session. Such large window sizes allow to keep small the number of inputs to the regressor in comparison to the sample size. Fine-grained temporal analysis is discussed in the Next Steps section.

## Results

Phase information is useful for differentiating the different moments of the sessions, and as from the visual inspection, there are large differences in behaviour between different phases. In particular, the regressors scored badly when considering the Reflect phase, while the Planning phase can be aggregated with the Work phase without major disruptions. Relying less on phases was a good strategy because it avoided the need for an automatic segmentation tool. Table 2 shows the conditions for which there is a reliable regression for the given data ( $R^2 > 0.1$ ), this means that PE Total never provided reliable regression, and only the listed windows were successful (e.g. no regression when using the whole window).

Table 2: Results of the Regression Analysis – Scores of  $R^2$  for features Hand Distance, Speed and Face Count. Only the reliable regressions are reported.

	Window (s)	SYN222	IA211
Linear	1200	0.28	0.17
Bayes Ridge	1200	0.28	-
SVMR	1200	-	-
Linear	1800	0.48	-
Bayes Ridge	1800	-	-
SVMR	1800	-	-

The regression tests gave unsatisfactory results, so we proceeded with statistical modelling using Generalized Linear Models. And we obtained the following results for window of 1800 seconds:

- IA 222 has Hand Distance (Max and Min) as regressor (with significance  $p < 0.05$ )
- SYN Total has Hand Distance (Min) as regressor (with significance  $p < 0.05$ )
- SYN 111 has Face Count and Hand Distance (Min and Max) as regressor (with significance  $p < 0.05$ )

If instead we look at the overall window duration, we obtain:

- PE level can be regressed by Hand Max Distance (significance  $p < 0.005$ )

## Discussion

In this paper, we investigated what multimodal learning analytics features could be identified and used to support CSCL through the use of the PELARA LAS. Our primary aim was to identify the MMLA features that can be used to determine aspects of collaboration in project-based learning. The purpose of this work is to develop CSCL that can aid in the assessment of CPS and to determine how different MMLA interactions can be automated to support and understanding collaborative learning. As discussed by Blikstein and Worsley (2016) existing evidence about the effectiveness of constructivist learning activities including project-based learning is rare. These teaching and learning approaches are notoriously hard to be evaluated via standardised measurements due to open-ended and dynamic nature of their implementations. However, MMLA provides new methods and methodologies with relevant potentials to provide evidence about the impact of such teaching approaches. In this research study, we presented that where the students are looking, the distance between them, the motion of their hands our key features for a learning analytics system to be effectively used to identify collaboration in small groups of Engineering students. These results are significant for the CSCL community as a starting point to investigate further what features of MMLA can be used to support collaborative learning providing insights about the physical and embodied processes involved in hands-on learning and how.

Amount of occluded hand pose due to the orientation of the fiducial markers will require to approach the hand tracking in a different way with the aim of extracting hand pose directly from video without the use of



such fiducial markers. The distance in the video space will be sufficient. This operation can be applied on these same sessions for which we have collected the original video stream.

Starting from the results of this work we are moving toward the reduction of the window size down to the fine-grained coding of the sessions at 30-seconds intervals. This coding will allow to train a machine learning classifier to recognize, from the video recording and the other multimodal data, the 3-number coding. For this purpose, we employed a deep neural networks (DNN) that are composed of a long sequence of linear matrix multiplications followed by non-linear activation functions. This supervised learning approach will be important for automating the scoring of students' session and providing assessment in the context of the CPS framework.

## Conclusion

The PELARS project is limited to the context of the study engineering students performing an open-ended task about physical computing and the relatively small sample size currently make it hard to generalise. However, the finding show similar results to other findings in MMLA (Blikstein & Worsely 2016; Ochoa et al., 2013; Grover et al., 2016) that begin to show that physical aspect of collaborative is an important part of this type of learning and that learning analytics systems can identify features that are relevant for helping researchers, teachers, and learners unpack what is happening.

Several questions are raised about the PELARS LAS exploratory approach on the physicality of the learners, the log files of the hardware and software, and the user generated data without a deeper connection to the students' conversation or an inquiry-like system to support the learning. However, the project and this paper have focused on the investigating other processes around face-to-face collaboration, which have can now be collected as demonstrated. Additionally, in classroom and lab contexts collection audio from learners provide new challenges for accurate audio processing (ASR) and the act of making things needs to be balanced with more traditional textual input. The next steps for the project are to explore how to integrate more user generated content from learning activities like virtual internships (Arastoopour & Shaffer, 2015) and how directional audio and automatic speech recognition could be utilised.

## Endnotes

- (1) <http://pelars.eu/>
- (2) <http://scikit-learn.org/stable/>

## References

- Arastoopour, G., & Shaffer, D. W. (2015). Epistemography and Professional CSCL Environment Design. In *International Conference on Computer Supported Collaborative Learning*. Gothenberg, Sweden.
- Banks, F. and Barlex, D. (2014). *Teaching STEM in the secondary school: Helping teachers meet the challenge*. Routledge.
- Blikstein, P. and Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*. Available at: <http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/4383/5596>.
- Chen, L., Feng, G., Joe, J., Leong, C.W., Kitchen, C. and Lee, C.M. (2014). Towards Automated Assessment of Public Speaking Skills Using Multimodal Cues. In: *Proceedings of the 16th International Conference on Multimodal Interaction - ICMI '14*. New York, New York, USA: ACM Press, pp. 200–203.
- Cukurova, M., Avramides, K., Spikol, D., Luckin, R. and Mavrikis, M. (2016). An analysis framework for collaborative problem solving in practice-based learning activities: A mixed-method approach. In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*. New York, New York, USA: ACM Press, pp. 84–88.
- Damon, W. and Phelps, E. (1989). Critical distinctions among three approaches to peer education. *International Journal of Educational Research* 13(1), pp. 9–19.
- Dillenbourg, P. 1999. What do you mean by collaborative learning? In: Dillenbourg, P. ed. *Collaborative-learning: Cognitive and Computational Approaches*. Oxford: Elsevier, pp. 1–19.
- Dillenbourg, P., & Jermann, P. (2007). Designing integrative scripts. In *Scripting Computer-supported Collaborative Learning*. (pp. 275-301). Springer US.
- Grover, S., Bienkowski, M., Tamrakar, A., Siddiquie, B., Salter, D. and Divakaran, A. (2016). Multimodal analytics to study collaborative problem solving in pair programming. In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*. New York, New York, USA: ACM Press, pp. 516–517.

- Halverson, E.R. and Sheridan, K. 2014. The Maker Movement in Education. In *Harvard Educational Review* 84(4), pp. 495–504.
- Johnson, D. W., & Johnson, R. T. (2003). *Assessing students in groups: Promoting group responsibility and individual accountability*. Corwin Press.
- Kirschner, P.A., Sweller, J. and Clark, R.E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychology* 41(2), pp. 75–86.
- Klahr, D. and Nigam, M. 2004. The equivalence of learning paths in early science instruction: effect of direct instruction and discovery learning. *Psychological Science* 15(10), pp. 661–667.
- Kotsiantis, S., Kiriakos Patriarcheas, and M. Xenos. (2010). "A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education." *Knowledge-Based Systems* 23(6). pp. 529--535.
- Lakens, D., & Stel, M. (2011). If they move in sync, they must feel in sync: Movement synchrony leads to attributions of rapport and entitativity. *Social Cognition*, 29(1), 1-14.
- Luckin, R. (2010). *Re-designing Learning Contexts: Technology-Rich, Learner-Centred Ecologies*. London: Routledge.
- Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B. and Castells, J. (2013). Expertise estimation based on simple multimodal features. In: *Proceedings of the 15th ACM on International conference on multimodal interaction - ICMI '13*. New York, New York, USA: ACM Press, pp. 583–590.
- Xavier Ochoa, Marcelo Worsley, Katherine Chiluiza, and Saturnino Luz. (2014). MLA'14: Third Multimodal Learning Analytics Workshop and Grand Challenges. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*. ACM, New York, NY, USA, 531-532. DOI: <http://dx.doi.org/10.1145/2663204.2668318>
- OECD (2015). PISA Draft Collaborative Problem Solving Framework 2015. Available at: <http://f1000.com/work/item/2523209/resources/1902358/pdf>.
- Spikol, D., Avramides, K., & Cukurova, M. (2016). Exploring the interplay between human and machine annotated multimodal learning analytics in hands-on STEM activities. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16* (pp. 522–523). New York, New York, USA: ACM Press. <http://doi.org/10.1145/2883851.2883920>
- Raca, M., Tormey, R., & Dillenbourg, P. (2014). Sleepers' lag-study on motion and attention. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge LAK'14* (pp. 36-43). ACM.
- Ruffaldi, E., Dabisias, G., Landolfi, L. and Spikol, D. (2016) . Data collection and processing for a multimodal learning analytic system. In: *2016 SAI Computing Conference (SAI)*. IEEE, pp. 858–863.
- Schneider, B., & Blikstein, P. (2015). Unraveling students' interaction around a tangible interface using multimodal learning analytics. *Journal of Educational Data Mining*, 7(3). Pp. 89-116.
- Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, 8(4), 375-397.
- Sharma, K., Jermann, P., & Dillenbourg, P. (2014). "With-me-ness": A gaze-measure for students' attention in MOOCs. In *International conference of the learning sciences* (No. EPFL-CONF-201918).
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of educational research*, 69(1), 21-51.
- Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological science*, 20(1), 1-5.
- Worsley, M. and Blikstein, P. (2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics* 1(2), pp. 151–186.

## Acknowledgments

We would like to thank all the teachers, students and facilitators involved in the data collection of this research study. This work was funded by European Commission, FP7Technology Enhanced Learning (GA No. 619738).

# Student Re-Design of Deprived Neighbourhoods in *Minecraft*: Community-Driven Urban Development

Rikke Magnussen, ResearchLab: ICT and Design for Learning, Department of Communication, Aalborg University, Copenhagen, Denmark, rikkem@hum.aau.dk  
Anna Lindenhoff Elming, Social Services Department, Copenhagen City Council, Copenhagen, Denmark, BS8I@sof.kk.dk

**Abstract:** This paper presents results from the three-year research and development project *Cities at Play: Community Drive*, where students in five 7<sup>th</sup>–9<sup>th</sup> grade classes, in collaboration with the Copenhagen City Council urban planners, used *Minecraft* and other digital and physical tools to redesign their deprived neighbourhood to generate solutions to problems in their local area. The overall aim was to understand how to integrate community-driven science and urban development in school education, in addition to identifying the knowledge and practices that emerged from the process. Our study presents survey data from two 7<sup>th</sup> grade classes (50 students 13–14 years of age) that developed *Minecraft* and LEGO models that re-designed their neighbourhood to solve local problems. The results show that, over the course of the project, students became aware of the authentic knowledge they possess about their urban area and that they are able to connect it to structural changes.

## Introduction and background

For the past decade citizen science has been defined as involving laypeople in the production of science outside formal research institutions, with opportunities to do so growing due to web 2.0 technologies and other types of digital platforms (Delfanti, 2010). New digital platforms are changing the way laypeople are involved in technical and scientific processes and production that affect their lives. Games, for example, have played a central role in the study and development of platforms for authentic collaboration between laypeople and scientists (Good & Su, 2011).

Some major trends in the study and development of learning game formats in the past decade include exploration of how the game media can facilitate new approaches to authentic science and technical education. Examples of this are games in which players become urban planners, biologists or forensic experts drawing on the authentic tools, processes and values of specific professions (Shaffer, 2006; Squire & Klopfer, 2007; Magnussen, 2007). The motivation for developing this type of games stems from a critique of the teaching of standardised skills to children in today's school system (Gee, 2003). It is argued that few schools teach students how to create knowledge; instead, students are taught that knowledge is static and complete. This means that they become experts at consuming rather than producing new knowledge (Sawyer, 2006).

Digital platforms that invite students to participate in professional processes bring up the matter of whether students learn to work as a scientific expert or whether they learn how to be an expert. The answer may depend on various design elements of profession simulation games. First, the clients and experts students collaborate with in the games are fictional characters with fictional problems that need to be solved to play the game in school but that do not have relevance in the world outside school. Second, the fictional problems to be solved in these games often follow a linear path and have a clear starting and end point. This is clearly different from real-life professional problem solving, where the processes are more multidimensional. Finally, even though these types of games have been shown to support student creation of new process tools (Magnussen, 2007), the solutions are often pre-defined and already known by the teachers. This stands in contrast to the real-life open-ended tasks professionals face and that can be carried out in various ways, the chance of success or failure always an issue to be considered. Scientific discovery games address issues that exist outside a formal learning setting. The main goal of this type of game is to create a platform that motivates players to contribute to solving scientific problems. Yet the question remains open as to what students learn when they take part in authentic problem solving processes.

Where major trends have involved simulating authentic technical or scientific processes in game environments to encourage student learning about authentic science, scientific discovery games or citizen science games now focus on gamifying professional research or technical processes, allowing and motivating players to take part in generating results for authentic application in scientific research (Cooper, 2015; Dean et al., 2015).

The development of scientific discovery games within the past couple of years introduces new elements into game-based participation in a science classroom setting (Magnussen et al., 2014). The main goal

of this category of games is to create a platform that enables and motivates players to contribute to solving scientific problems (Cooper et al., 2010). The most well recognised example of this class of games is *Foldit*, which is an online puzzle game where players participate in folding amino acid chains to form new protein structures (Good & Su, 2011). Studies show, however, that integrating games in authentic knowledge creation in science education can be complex. Previous studies on these types of learning environments (Magnussen et al., 2014) clearly indicate that using game media as a tool for conducting real-life problem solving and collaboration with real-life experts is potentially highly motivating for science students but can also be perceived as not addressing learning, leaving students feeling confused about the meaningfulness of being involved in a scientific field for which they do not possess the same expert skills as their professional scientific collaborators. The approach has the potential to introduce real-life problem solving into schools but it must be kept in mind that the initial expertise of students within the area of knowledge development and how this knowledge develops are a key aspect of community-driven science learning in school.

The goal of the City at Play project, where students collaborated with architects on developing models in the game *Minecraft* for redesigning their local urban space, was to further expand the scientific discovery game concept in an educational context. During our study students gained first-hand experience with creating new technical knowledge within the framework of professional architects. This community-driven science learning environment was based on the students' local knowledge and expertise about their under-privileged neighbourhood, as well as the theory and methods of the participating urban planners. Our main research objective was also to understand how student learning practices and knowledge developed during the City at Play course.

## Methods in Cities at Play

The project Cities at Play described in this paper was developed in close collaboration with the Copenhagen City Council Social Services Department and ResearchLab: ICT and Design for Learning at Aalborg University in Copenhagen. The purpose of Cities at Play was to involve young people in deprived areas as experts in their own living environments and to educate them on the influence of structural factors on their welfare and well-being and in how to use game tools to apply their knowledge and ideas to recreate and strengthen their neighbourhoods. From the start the project was launched to define problems and introduce game-based methodological solutions for developing structural changes in neighbourhoods in deprived areas in Copenhagen, therefore including both social and educational objectives. The project aimed to provide authentic contributions to City Council urban development and planning as a starting point and, ultimately, in the realisation of some of the ideas contributed.

The methodology used in developing the components of Cities at Play followed a design-based research process involving various design cycles, interventions, analyses and redesign (Brown, 1992). Design-based research was applied as a methodological framework and various methods were employed in the development and study of the game-based community-driven urban planning environment. The project moved through two iterations in a design-based research process (Brown, 1992), which included involving an increasing number of school classes and departments in the Copenhagen City Council. The first iteration is described elsewhere (Magnussen & Elming, 2015), hence the focus of this paper is the second iteration, Cities at Play 2.

## Study design, methods and data analysis

Cities at Play 2 included four teachers, two 7<sup>th</sup> grade classes and two 9<sup>th</sup> grade classes, thus involving 90 students aged 13–15 from a school in a deprived area in south Copenhagen, which was chosen as an area due to the high rate of unemployment and non-existing or low level of education of residents. The school is located in an area with older council housing that suffers problems involving gangs and drugs. A library, kindergartens and a nursing home are in the vicinity of the school. The project was conducted in the local library over a three-week period. The 7<sup>th</sup> and 9<sup>th</sup> grade classes each worked separately for one week on their models and then worked in parallel during the third week to finish their models for presentation. A mixed method approach where conducted where video observations were used to document the three weeks of student design sessions and focused on documenting the students' dialogue in the design process in order to understand how the various models were developed based on types of local technical knowledge (Onwuegbuzie & Leech, 2006). Video observations specifically focused on documenting the work of the students who were socially weak in school but showed a high degree of motivation for participating in changing their local area.

Pre and post-surveys were conducted to register: 1) student motivation for participation in the project, 2) local knowledge about their area and urban planning, 3) whether the course supported the 21<sup>st</sup> learning skills Real-world Problem Solving and Collaboration compared to what students defined as "every day school", 4) what the students viewed as making the course different from "everyday school" and 5) student understanding

of their ability to structurally change their living conditions. The digital surveys contained opportunities to provide quantitative answers in order to create an overview of the students' knowledge and experiences, as well as the opportunity to supply qualitative answers in order to understand the background for the quantitative answers (ref survey method). The teachers gave their classes the surveys the day before the course started and on the day the course ended. We also conducted semi-structured qualitative interviews with teachers and students that focused on further understanding the possible outcomes and challenges of the project (Kvale, 1996). Qualitative data was analysed applying grounded theory as a data categorisation method, where themes were defined based on participant-defined concepts in perceived knowledge generation and learning practices (Strauss & Corbin, 1998).

This paper focuses on analysing the motivation and forms of knowledge in the two 7<sup>th</sup> grade classes, which worked on changing their neighbourhood, while the two 9<sup>th</sup> grade classes worked on changing a neighbourhood they were unfamiliar with. This paper specifically focuses on the 7<sup>th</sup> grade students' authentic local knowledge and the role it plays in the collaboration with professional urban planners.

## Findings

The overall goal was to understand how to integrate community-driven science in school education with a focus on urban development. The research of the current study was to understand what motivation, knowledge and language emerged from the encounter between the principles of professional city changers and students with local authentic knowledge about the deprived area they lived in.

The design of Cities at Play: Community Drive comprises five phases and is based on experiences and results from previous studies of game-based innovation education and community-driven science games (Magnussen, 2011; Magnussen et al., 2014). As described in Table 1 the participating students went through various phases that involved finding inspiration, defining the potential and problems in their local area, developing ideas, building models in the game *Minecraft* and with other materials and presenting with feedback from professional architects and urban planners in various Copenhagen City Council departments.

**Table 1: Phases in the students' development process in City at Play 2 (modification of IDEO, 2009)**

Phases	Activity	Process
Phase 1 (Week 1)	Inspiration	Field trips to newly developed areas in the city, introduction to core concepts in urban planning by architects and urban planners
Phase 2 (Week 1)	Defining the potential and problems	Definition of core strengths and challenges in their local area
Phase 3 (Week 1)	Developing ideas	Development of ideas for solving local problems and strengthening its potential
Phase 4 (Weeks 1 & 2)	Modelling	Building models in <i>Minecraft</i> , LEGO and other digital and physical tools
Phase 5 (Week 2)	Presentation	Presentation of models for the head of the Department of Transport, Technology and Environment and Copenhagen City Council urban planners

### Motivation: Identification of local problems and potentials

A primary research objective was to understand what motivated students in the course. In the initial part of phase 1 urban planners from Copenhagen city council presented students with the overall objective of their development work: to redesign their neighborhood to create more life and more connections to other parts of the city. It is well-known that fundamentally rethinking existing formats is challenging, even for experienced innovators and designers (IDEO, 2009). As a result this phase focused on bringing students out of their everyday setting to inspire them to think about new structural changes. In addition to being presented with professional urban architectural methods and theories for changing cities (Gehl, 2010), students participated in field trips to other neighborhoods that had undergone an urban transition. During phase 2, the problem definition phase of the four-day development process (see Table 1), students were asked to photo document their area, focusing on its strengths and weaknesses. This process, planned in collaboration with the urban architects, included using a photo mapping method of areas well-known in the field of professional urban planning.

The two 7<sup>th</sup> grade classes generally worked extremely intensively during the course of City at Play. When asked "What do you think about the City at Play course?" in the survey, the two 7<sup>th</sup> grade classes gave a

positive response overall, with a total of 90% answering either “Excellent” (16%), “Really good” (58%) or “Good” (16%). The reasons behind their motivation were identified in the initial phase of the development process, where students had to identify the problems and potentials of their local area. In response to the question “Would you like to be a part of determining what Folehaven should look like and be like?”, 90% answered “Yes, a little” (14%), “Yes” (31%) or “Yes, very much” (45%). The majority of students were thus motivated from the very beginning of the project to take part in the redesign of their area. In that regard, finding out what motivated the students to want to take part in changing their area was a central goal. Two additional exploratory questions were also asked about living in the neighbourhood Folehaven. In response to the first one, “How much do you agree with the statement: Folehaven is fine as it is and does not need to be changed?”, a total of 65% either slightly disagreed (42%), disagreed (21%) or strongly disagreed (2%) with the statement. The percentage increased to 84% in the post-survey, which presented students with the same statement.

When asked why Folehaven was good or bad the majority of the qualitative answers focused on the criminal activities or feeling unsafe, which is indicated by statements such as: “Too much violence and trouble making”, “There are many rumours about drugs and stabbings. It’s a little unsafe at the moment” and “There are lots of criminals”. There are, however, also replies that address the more positive social aspects of living in Folehaven, for example: “It depends on who what your personality is like or who you’re with” and “Because I have some lovely people around me”.

These results indicate that the students were highly motivated from the beginning of the project to change their area and that the overwhelming majority saw problems in their area that they were motivated to change.

### Creating models: Real-world problem solving and community-driven urban development

The focus of phase 3, the idea phase, and phase 4, the modelling phase (see Table 1) was for students to develop ideas and game-based models of structural solutions to the problems identified in phase 2. Students were asked to list and draw designs to strengthen identified potentials and solve the identified problems of Folehaven. The questions for this task were: “How can you change Folehaven so potentials are strengthened or problems are solved?” Additionally the development of ideas and later models also (as described above) focused on creating life and new connections in the students’ neighbourhood. In phase 4 students selected and developed physical LEGO models and digital game models. Overall these models were characterised either by an explicit focus on structurally solving specifically identified problem areas, such as rundown playgrounds, or by being more theoretical with conceptual plans and examples for changing the social life of or cohesion in Folehaven. One example of the latter type of model was mixed housing designed to lure socioeconomically advantaged families and singles or new types of experiences, such as a network of bike paths connecting areas with cafes and play and sports facilities, to create lively areas that would draw visitors from other parts of the city. The different models proved to have a variety of potential in the urban planning activities. The physical LEGO models were well-suited for creating an initial overview of the developed area. The game tools, on the other hand, allowed students to communicate first-hand experience about the area being developed (see Figure 1). The Danish Geodata Agency offers a complete open-source, geographical model of Denmark (<http://eng.gst.dk/maps-topography/denmark-in-minecraft/>), allowing students to download a 10x10 km area with all existing buildings and streets to redesign.



Figure 1. Students developing models in *Minecraft* and LEGO.

The research focus was to understand what new types of practices and knowledge developed in City at Play. As a result our study was designed to document whether the students experienced working with 21<sup>st</sup>

Century Learning Skills and a more open design intended to document types of knowledge or practices specific to City at Play and community-driven science. As a result the study design also centred on understanding if students felt they were working with the 21<sup>st</sup> Century Learning Skills Collaboration and Real World Problem Solving and whether we could detect a difference compared to what students defined as “everyday school”. The pre-survey contained questions on whether the students experienced collaborating and solving real-world problems on a variety of levels in everyday school. They were asked whether more than two people worked together when they collaborated, if they made important decisions collaboratively and if the result of group work reflects a collaborative process. For real-world problem solving questions were asked whether they felt they experienced working with real-world problem solving in everyday school or in City at Play. Hence the pre-survey asked if they worked with “solving problems for people in the real world outside school” in everyday school and on what level, in addition to whether this required developing their own ideas or following teacher-defined stepwise processes. The post-survey included the same questions on collaboration and real-world problem solving but covered their participation in City at Play instead.

Our findings show that no significant difference exists between group work and collaborative practices in everyday school work versus City at Play, which may be due to how highly integrated group work and collaboration are in most subjects in primary school in Denmark. Dealing with real-world problems in everyday school education, however, differed significantly compared to practices in City at Play. At total of 7.5% of the students thus either strongly agreed (2.5%) or agreed (5%) that everyday school work often involved “solving problems for people in the real world outside school” compared to 76% for City at Play, where they either strongly agreed (23%) or agreed (53%) with this. For City at Play, 78% of students that either strongly agreed (26%) or agreed (52%) that they “worked with developing new ideas that they presented to people outside school that needed them” while the percentage was 17.5% for everyday school education. These results indicate that a majority of students experienced working with real-world problem solving practices in City at Play compared to everyday school education. Student experience with City at Play practices were further examined based on the qualitative post-survey questions: “Were the problems you worked with in City at Play different from the problems you normally work with at school?” and “What was different in City at Play compared to everyday teaching?”. Our analysis of their responses grouped them into a variety of themes that are presented in Table 2.

**Table 2: Themes that emerged based on responses to the two post-survey questions on City at Play: “Were the problems you worked with in City at Play different from the problems you normally work with at school?” and “What was different in City at Play compared to everyday teaching?”**

<b>Themes</b>	<b>Examples of student responses</b>
Changing things	“Yes, because we normally don’t work with changing things”, “Yes because we were working with changing something in our city, which is something we don’t do in class”
Something in the real world	“It was something that could happen in the real world”, “Yes, a lot, because it concerns the real world and it involved problems we could solve for the entire Folehaven neighbourhood”, “Yes, because in school we do, for instance, grammar and math, while in City at Play we were supposed to help others make Folehaven a better place to be”, “Yes, because in a way it did not involve problems related to school subjects but something in the real world”
About helping people, not just working for your own benefit	“In school we work more for our own benefit. In City at Play we made something that everybody could benefit from”, “In school you need to improve your grades, here we needed to help other people ... #Thatwasnew”, “Yes, because we had to consider whether it would work because here it’s all about people”, “Yes, here you can do something for a group of people and not just do math”, “We helped other people and not just ourselves”, “We had to make something that would benefit other people”
Decided more	“We decided more”, “What we had to make was not predetermined”, “It’s kind of good because we had to decide on what we needed to build and so on. It’s not like that in daily teaching, where teachers have the right to decide”, “We were allowed to determine/decide most things”
Using one’s imagination and inventiveness	“We had to use our imaginations”, “We don’t usually talk to architects and invent things”
Being active	“We didn’t sit down all the time”, “You were free to choose what to do”, “We got to move around and independently decide things, “We were active in City at Play”
Other tools	“We used other tools”, “We had to play a game to do our assignment”, “We were building with LEGO blocks and made models with them”, “No books, a lot of collaboration”

During our analysis of their responses a variety of themes emerged concerning the project’s methodological and educational approaches, which will be discussed further in the discussion section. A central aspect of understanding what knowledge the students brought into the project involved gaining an understanding of how students viewed their own knowledge about their local area in comparison with the local knowledge that urban planners participating in City at Play had. The pre-survey asked “Do you possess knowledge about Folehaven that the architects redeveloping Folehaven do not have?” In their responses, 9% answered either “Yes, I know a lot that they don’t know” or “Yes, I know a bit more”. This percentage changed to 45% in the post-survey (see figure 2).

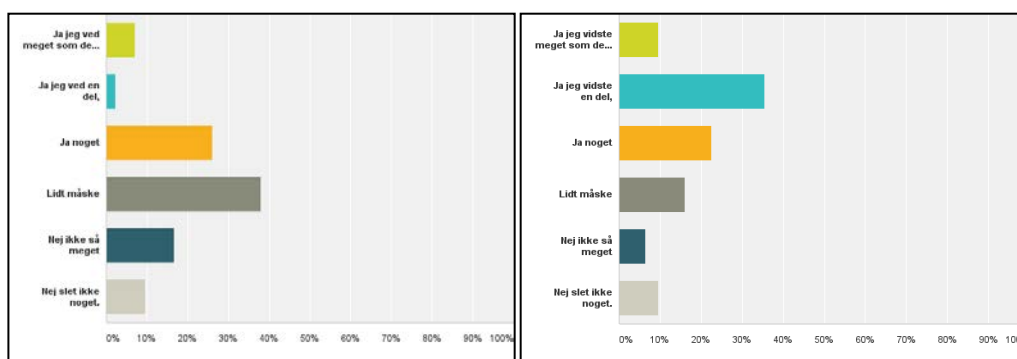


Figure 2. The bar chart on the left shows the pre-survey results and the bar chart of the right shows the post-survey results. The pre-survey question was: Do you have knowledge about Folehaven that the architects redeveloping Folehaven do not have? In the response, 7% (green) answered “Yes, I know a lot that they don’t know”, 2% (light blue) answered “Yes, I know a bit more”, 26% (orange) answered “Yes, some”, 38% (dark grey) answered “Maybe a little”, 17% (dark blue) answered “No, not very much” and 10% (light grey) answered “No, not at all”. Students answered a similar question in the post-survey: “Think about the City at Play course: Did you possess knowledge about Folehaven that the architects redeveloping Folehaven did not have?”. A total of 10% of the students answered “Yes, I knew a lot that they didn’t know”, 35% answered “Yes, I knew quite a bit more”, 23% answered “Yes, some”, 16% answered “Maybe a little”, 6% answered “No, not a lot” and 10% answered “No, none at all”.

These results indicate that the students’ perception of whether they have knowledge about their neighbourhood that the professional urban planners do not have changes due to their participation in City at Play. The study examined this change in their perception of their knowledge or local expertise more closely by asking students to qualitatively specify what specific knowledge they felt they had that the urban planners did not have. The qualitative part of the pre-survey conducted prior to participating in City at Play provided a picture of the specific local knowledge students believed they had. The majority of the answers to the question “What for instance do you know more about?” covered reasons why the students knew more than the urban planners, for example: “I know Folehaven very well because I grew up here, so I’m sure I can work out something with some of my friends or alone” and “I grew up in Folehaven and know almost everyone. I think my friends and I can find something good to build here in Folehaven”. The student responses to two post-survey qualitative questions: “Think about the City at Play course: Did you possess knowledge about Folehaven that the architects redeveloping Folehaven did not have?” and “What for instance did you know more about?” can be grouped into four types of knowledge:

- 1) Physical buildings or facilities in the area
  - E.g. “I know a little about Folehaven and the buildings”, “supermarkets and lighting”
- 2) Experiences or feelings
  - E.g. “What the atmosphere is like, what’s good and what’s bad, what it’s like in general to be here”, “I can find my way around Folehaven with my eyes closed, I’m part of it”.
- 3) Experiences or feelings concerning locations or facilities in the neighbourhood
  - E.g. “That it’s boring to be here/live here. They couldn’t know that there isn’t much light in the evening, which makes it scary.”, “Where it’s safe and unsafe”, “Safe and unsafe places. What needs to be changed.”



- 4) Social aspects of the community in the neighbourhood
  - E.g. “I know more about the things that some people need.”, “I know, for instance, what it’s like to live here and what most people want /don’t want”.

These results indicate that students became more aware of their knowledge about their neighbourhood (figure 3). The variety and diversity of the answers to the post-survey show that they also became aware of the different types of knowledge and its value in the development process. The answers concerning the first type of knowledge, “Physical buildings or facilities in the area”, suggest that students developed an understanding of how one’s experiences of an urban area are related to its structural elements and the urban development choice made in the area (Gehl, 2010). This understanding was further explored in the pre and post-survey, where two major issues in City at Play: 1) creating more life in Folehaven and 2) the creation of better connections between Folehaven and the rest of Copenhagen, were addressed with, for example, the questions: “Do you have any ideas about how more life can be created?” and “What ideas do you have about how better connections between Folehaven and the rest of the city can be created?” In the pre-survey 25% of the students’ qualitative answers did not contain any ideas or stated that they had not considered any. In the post-survey, however, only 3% said they did not have any ideas. Responses concerning the second big issue in the pre-survey addressed more general issues with overall suggestions, such as: “No crime”, “No crime, no drugs. You should be able to feel safe in Folehaven” and “Things that get people’s attention”. The concrete ideas that appeared in the post-survey responses indicated that students had worked intensively with ideas for creating connections and life: “Make a central hub or make streets, e.g. bike paths, that are a little more interesting so you feel like biking on them”, “Build new apartments so new people move to Folehaven” and “For example, make a blue bike path that would hopefully attract people or that would break down the wall around Folehaven”. In addition to suggesting specific ideas for redesigning the urban area, the students also integrated concepts introduced by the urban planners and architects, such as mixing the types of houses and apartments for new groups of people to move into, developing connections that would attract people to a central hub by building streets and paths that would give them a better experience. The discussion section will look at the knowledge types in further detail.

## Discussion

This paper presented results from the project City at Play: Community Drive, where young people in deprived areas built models in Minecraft with the aim of redesigning their neighbourhood based on what they define as the potential and problems in their local area. Game-based community-driven science is a growing field (Cooper et al., 2010; Good & Su, 2011) but more attention needs to be given to understanding what knowledge and learning processes authentic problem-solving in collaboration with professional partners can lead to in a school context. The overall research aim of this paper was to understand how to integrate community-driven science and urban development in school education, and to identify what knowledge and practices emerge from this integration. Data and findings from pre and post-surveys given to the two participating 7th grade classes showed that various aspects are involved in how students perceive how the approach used in City at Play differs from their everyday school practices. The findings on learning practices showed that students experienced working with real-world problem solving more than they did compared to everyday school. Their qualitative responses clearly demonstrated that problem solving in a context outside school was a central theme, as was focusing on the community by helping people, i.e. not just doing work for their own benefit, but also being in control of their own decision making and problem solving. During our analysis of the knowledge generation part of the above described practices, it became evident that the participating students’ perception of the knowledge they possessed changed significantly after the course compared to before. An assessment of the students’ qualitative answers regarding what local authentic knowledge they had that the professional urban planners did not have showed that it was quite specific and covered, for example, aspects such as buildings, facilities and environments. In addition the knowledge categories our analysis identified also showed that after the course the students had a better understanding of the physical and structural elements in their local urban space and of the effect of physical elements such as buildings and lighting. They had realised that all of these various aspects affected whether they felt safe or scared, in addition to influencing the broader community in Folehaven. We also examined their responses in the post-survey concerning ideas for creating a greater connection to the city and more life in their neighbourhood. After participating in City at Play the students’ ideas became more detailed with regard to the specific changes they wanted to implement in the area and incorporated technical terms related to the principles of professional urban planning. These results indicate that over the course of two weeks the students developed and changed from defining challenges in their neighbourhood in terms of larger concepts such as “crime” and “feeling unsafe” to concrete problems such as poor lightning, worn down housing, streets in need of repair and the scattered nature of community activities. These structural, well-defined problems created a framework for the student to develop redesigns of their neighbourhood, leading to

suggestions and plans aimed at structural changes to meet different needs and problems of the various types of residents. The authentic framework of the models presented to and commented on by real urban planners proved to be highly motivating for both resourceful and less-resourceful students.

Overall the above results indicate that over the two-week development process the students were able to expand their knowledge using their local expertise on their area by combining it and relating their authentic knowledge about the problems and potential of their area to its physical and structural features, but also to the professional principles of developing design solutions. Helping other people in the community, solving problems in the real world and being in control of their own process and decision making proved to be important practices they experienced during the City at Play course compared to everyday schooling.

The current study examined a brief two-week period and produced promising results. This paper presents the first results, but future studies should focus on generating a deeper understanding of the mentioned practices and types of knowledge. A future research aim is also to do a similar study over an extended period to determine if the results are similar or different and to look at the long-term effects of this type of community-driven development when integrated in schools over a longer period.

## References

- Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of Learning Sciences*, 2(2), 141-178.
- Dean, D., Guarino, S., Eusebi, L., Keplinger, A., Pavlik, T., Cammarata, A., Murrari, J., McLaughlin, K., Cheng, J. & Maddern, T., (2015). Lessons learned in game development for crowdsourced software formal verification. In: 2015 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE'15).
- Cooper, S. (2015). Massively Multiplayer Research: Gamifying and (Citizen) Science. In: *The Gameful World: Approaches, Issues, Applications*. Ed. Steffen P. Walz, Sebastian Deterding. pp. 487-500.
- Cooper, S., Adrien Treuille, Janos Barbero, Andrew Leaver-Fay, Kathleen Tuite, Firas Khatib, Alex Cho Snyder, Michael Beenen, David Salesin, David Baker, Zoran Popović & Foldit players (2010). The challenge of designing scientific discovery games. In: *Proceedings of Foundations of Digital Games, FDG 2010*, Monterey, CA, SA.
- Delfanti, A. (2010). Users and peers: From citizen science to P2P science. *JCOM: Journal of Science Communication*, 9(1), pp. 1-5.
- Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave/Macmillan.
- Gehl, J. (2010). *Cities for People*. Washington - Covelò – London: Island Press.
- Good, B. M. & Su, A. I. (2011). Games with a scientific purpose. *Genome Biol*, 12(135), pp. 1 - 3.
- IDEO. (2009). Human centered design toolkit (2nd ed.). Retrieved from <http://www.ideo.com/work/item/human-centered-design-toolkit/> 28 May 2015.
- Kvale, S. (1996). *Interviews—An introduction to qualitative research interviewing*, Thousand Oaks, CA: Sage.
- Magnussen, R. (2011). Game-like Technology Innovation Education. *International Journal of Virtual and Personal Learning Environments*, 2(2), pp.30-39.
- Magnussen, R. (2007). Games as a platform for situated science practice. In: de Castell, S., & Jenson, J. (Eds.), *Worlds in Play: International Perspectives on Digital Games Research* (301–311). NY: Peter Lang.
- Magnussen, R. & Elming, A. L. (2015). Cities at Play: Children's redesign of deprived neighborhoods in Minecraft. In: Robin Munkvold and Line Kolås (ed). *Proceedings of the 9th European Conference on Game-Based Learning*. Steinkjer, Norway. Pp. 331 – 337.
- Magnussen, R., Hansen, S. D., Planke, T. & Sherson J. F. (2014). Games as a Platform for Student Participation in Authentic Scientific Research. *The Electronic Journal of E-learning (EJEL)*, 12 (3), pp. 258 – 269.
- Onwuegbuzie, A. J., & Leech, N. L. (2006). Linking research questions to mixed methods data analysis procedures. *The Qualitative Report*, 11(3), 474–498.
- Sawyer, R.K. (2006). Educating for innovation. *Thinking Skills and Creativity*, 1(1), pp. 41-48.
- Shaffer, D. W., (2006) Epistemic frames for epistemic games. *Computers & Education*, 46:3, pp. 223-234.
- Squire, K., & Klopfer, E. (2007). Augmented reality simulations on handheld computers. *Journal of the Learning Sciences*, 16(3), pp. 371-413.
- Strauss A. & J. Corbin (1998) *Basics of Qualitative Research – Techniques and Procedures for Developing Grounded Theory*, second edition, London, Sage Publications.

## Acknowledgments

This work was supported by the Social Services Department at the Copenhagen City Council.

# How Technology and Collaboration Promote Formative Feedback: A Role for CSCL Research in Active Learning Interventions

Sally P. W. Wu, University of Wisconsin – Madison, [pwwu@wisc.edu](mailto:pwwu@wisc.edu)  
Martina A. Rau, University of Wisconsin – Madison, [marau@wisc.edu](mailto:marau@wisc.edu)

**Abstract:** Recent evidence for the effectiveness of active learning interventions has led educators to advocate for widespread adoption of active learning in undergraduate science, technology, engineering, and mathematics courses. Active learning interventions implement technology and collaboration to engage students actively with the content. Yet, it is unclear how these features contribute to their effectiveness. Research suggests that these features may enhance learning by providing formative feedback. To understand how technology and collaboration support learning by providing formative feedback, we conducted an observational study in a traditional and an active learning version of an undergraduate chemistry course. Results suggest that technology-provided feedback in the active learning intervention enhanced collaboration. We identify specific challenges and opportunities in technology design and active learning interventions for computer-supported collaborative learning research to address.

## Introduction

Recent research in undergraduate science, technology, engineering, and mathematics (STEM) education shows that *active learning* is more effective than traditional lectures (Freeman et al., 2014). This research describes active learning as a broad range of interventions in which students learn through activities and/or discussion in class, whereas in traditional instruction, students learn through passively listening to a lecture (Freeman et al., 2014). In accordance with the disciplinary-based STEM education research, recent interventions often involve the use of technology and collaboration to foster student learning from peers (Eddy, Converse, & Wenderoth, 2015; Lom, 2012). However, this active learning research has produced little theory about why and how technologies and collaboration can support student learning when compared to traditional undergraduate STEM courses.

One potential reason why active learning interventions are effective is that technologies and collaboration provide *formative feedback*. In traditional STEM courses, instructors typically provide summative correctness feedback based on the content, but they provide little formative feedback on how to learn the content. According to Sadler (1989), formative feedback helps students (1) understand what expert performance looks like and (2) assess their current performance, so that they can take measures to (3) bridge the gap between current and expert performance. Indeed, formative feedback has been shown to enhance student learning by providing correctness and corrective guidance about progress towards expert performance (Hattie & Timperley, 2007). Active learning interventions produce more formative feedback because of their use of technology and collaboration (Eddy et al., 2015), yet it is an open question whether students benefit from the additional feedback. Investigating how students use feedback provided in active learning interventions may provide insight into why active learning is effective and inform the design of technology-enhanced collaborative learning in such interventions.

Gaps in prior research gives rise to the question we examine in this paper: how does technology use and collaboration in active learning interventions support students' learning by providing formative feedback, as compared to common practice in traditional instruction? To address this question, we conducted an observational study of a traditional and an active learning version of discussion sections in an undergraduate chemistry course.

## Active learning in STEM instruction

Recent evidence for the effectiveness of active learning interventions has instigated widespread interest in implementing *active learning* in STEM courses, instead of passive lectures (Eddy et al., 2015). The editor of *Journal of Chemical Education* stated, "to put it bluntly, everyone should be taken off the control (i.e., traditional lecture) and switched to the treatment (i.e., carefully considered active learning methodologies)" (Pienta, 2015, p. 1). Indeed, chemistry education research has shown that active learning interventions lead to significantly higher learning outcomes than traditional lectures (Mahalingam, Schaefer, & Morlino, 2008; Paulson, 1999).

Active learning interventions emphasize learning by doing (Chi, 2009). Typical active learning interventions involve *technologies* such as clickers, online practice problems, or simulations, and *collaboration* such as discussions, problem-based workshops, and roundtables (Eddy et al., 2015). Prior research on active learning has focused on the effects of implementing these features in lectures held in auditorium-style classrooms (Lom, 2012).

To supplement lectures, STEM courses often include *discussion sections* (also known as recitations or review sessions). Discussion sections provide opportunity for an instructor to engage with a smaller group of

students and answer students' questions about the course content. Discussion sections are assumed to naturally foster discussion, but research shows that they often foster passive learning found in traditional lectures because students can copy solutions as the instructor solves them (Mahalingam et al., 2008; Paulson, 1999). Therefore, it is important that we understand how to design discussion sections so that they actively engage students with course content. Most research on active learning interventions has focused on lecture settings, not on discussion settings (Eddy et al., 2015; Lom, 2012). Therefore, we know little about how active learning interventions change learning processes in *discussion sections*. To address this gap, our study compares traditional and active learning interventions implemented in discussion sections, particularly on the formative feedback provided.

### Components of formative feedback

Formative feedback can guide student interactions by indicating *correctness* on students' performance and providing *corrective* redirection for students' progress. In general, "[f]eedback is commonly defined in terms of information given to the student about the quality of performance" (Sadler, 1989, p. 142). This definition is in line with other definitions of feedback that convey passing of information to the learner by instructors or instructional technologies (Shute, 2008; Van der Kleij, Feskens, & Eggen, 2015). Effective forms of formative feedback take into account students' needs for redirection (Hattie & Timperley, 2007). Sadler (1989) proposed that formative feedback to improve student performance should support students in: (1) understanding expert performance, (2) assessing their current performance, and (3) bridging the gap between current and expert performance.

Unless these three components are supported, students cannot act upon the feedback to improve learning. Specifically, without (1) an understanding of expert performance, students may proceed in unproductive directions. Without (2) an assessment of their current performance, students may not identify the gaps in their understanding. Without (3) the ability to bridge the gap, students cannot improve their current performance towards expert performance. The latter component may explain why "even when teachers provide students with valid and reliable judgments about the quality of their work, improvement does not necessarily follow" (Sadler, 1989, p. 119). Instructors must provide correctness and corrective feedback that help students improve their performance (Mahalingam et al., 2008; Michael, 2006), and students must act upon the feedback (Hattie & Timperley, 2007).

*Technologies* offer an effective way to provide formative feedback. For instance, technologies can provide practice problems with immediate feedback and detailed explanations that help students assess and bridge their performance in relation to expert performance. Immediate feedback from technologies has shown to be effective at enhancing student learning (Van der Kleij et al., 2015). Further, *collaboration* offers opportunities for formative feedback. When students work with peers, they have numerous opportunities to give and receive correctness and corrective feedback (Hattie & Timperley, 2007). Our present study seeks to understand the role of technologies and collaboration for providing formative feedback in traditional and active learning interventions.

### Active learning interventions and computer-supported collaborative learning

Research on active learning has been mostly situated in discipline-based education literature. These literatures focus on practice-oriented recommendations as to how best to implement active learning interventions. They recommend incorporating technology into STEM courses because it can increase accountability to specific tasks, reduce apprehension by anonymizing responses, and address multiple student responses all at once (Eddy et al., 2015; Lom, 2012). Further, they recommend incorporating collaboration into STEM courses because working in groups can foster deeper understanding of the content (Eddy et al., 2015; Mahalingam et al., 2008). However, discipline-based education research has failed to provide a theoretical explanation of why and how technology and collaboration interact in enhancing students' ability to actively engage with learning materials.

Therefore, our research examines how technologies and collaboration interact by providing formative feedback that can help students learn. This research is of relevance to the field of computer-supported collaborative learning (CSCL) for several reasons. First, our research seeks to uncover how the use of technologies in active learning interventions helps students collaborate. This may reveal insight into technology development for CSCL that meets the demands of active learning interventions. Second, our research seeks to uncover how collaboration helps students benefit from technology feedback. This may reveal additional opportunities for structuring collaborative activities in a way that helps students bridge the gap between their own and expert performance. To address these questions, we conducted an observational study to examine how technologies and collaboration support students' learning through formative feedback in traditional and active learning discussion sections.

### Method

We situated our observational study in an introductory accelerated chemistry course taught at a large Midwestern university in Fall 2015. The course involved three weekly lectures, a weekly laboratory, and a weekly discussion section. We observed two discussion sections taught by teaching assistants (TAs): a traditional discussion section

that involved problem-solving activities without technology support and an active learning discussion section that involved problem-solving activities and incorporated technology support and collaboration.

## Setting

*Traditional discussion sections* were held in classrooms in the Chemistry building with a table for the TA at the front, individual desks in rows oriented towards the front, and chalkboards and periodic tables on the walls. Twelve of the 16 discussion sections of the chemistry course were held in this traditional setting. The traditional sections emphasized problem solving through activities that asked students to engage with course content (e.g., worksheets). The TA circulated the room to monitor students' progress, provide feedback, and answer questions. Students could collaborate on the problems, but received no particular support for collaboration. At the end of the discussion section, the TA reviewed answers to the worksheets or emailed the answer key to the students.

The professor of the course designated four sections as *active learning discussion sections*. These sections were held in a nearby building that provided active learning spaces. These spaces provided large circular tables with outlets, rolling seats, and whiteboards to facilitate collaboration and use of technology (see Figure 1). The four active learning sections emphasized collaborative problem solving. In these sections, students completed paper worksheets that guided student activity. Worksheets included six to nine questions related to concepts discussed in a previous lecture. Each question directed students to complete a set of online problems followed by a worksheet question. Online problems provided correctness feedback and detailed explanations. The follow-up worksheet question asked about more complex concepts and encouraged students to collaboratively discuss concepts with their partners, table groups, or the TA.



Figure 1. Example classroom space designed for active learning. Image from <https://teachingacademy.wisc.edu/teach-in-sterling-clc-this-spring/>.

## Data collection and analysis

The first author observed a traditional section taught by Ted and an active learning section taught by Addie (all names are pseudonyms) from week two to week eight of the semester. In the traditional section, she sat at a desk in a back corner of the class and observed groups of students who worked together near her desk. In the active learning section, she sat at a table in the back corner of the classroom and observed a group of six students who sat at the table each week. While observing, she did not participate in student discussions but conducted fly-on-the-wall observations using a pen and notebook. She typed up observations following the discussion sections.

In week nine of the semester, the first author conducted interviews with the professor, the TAs (Addie and Ted), and five students (three from Ted's and two from Addie's section). Two student interviewees were members of observed groups, and the other three sat in other areas of the discussion sections that the first author did not observe, to provide a comparison of student experiences. Interviews were audio recorded and transcribed.

We conducted a bottom-up and top-down analysis of the field notes and interviews (Miles & Huberman, 1994). Our bottom-up analysis yielded 33 codes describing student interactions with technology and peers in three co-occurring categories: who was involved in the interactions (people), how they interacted (social interactions), and what they used (materials). Then, we applied the framework proposed by Sadler (1989) for formative feedback to co-occurring codes (e.g., TA + give-explanation + chalkboard/whiteboard).

## Results

We organize our results from interviews and observations in relation to components of formative feedback (Sadler, 1989): (1) understanding expert performance, (2) assessing current performance, and (3) bridging the gap.

## Feedback component 1: Understanding expert performance

According to Sadler (1989), students first need to understand expert performance. Our observations revealed that information about expert performance was often provided in the form of *correctness explanations* by the TA or by instructional materials (e.g., textbook, answer keys, and online problems). In the active learning section, we observed students reading explanations in online problems. In both the traditional and active learning sections, we observed TAs giving explanations to the whole class using the board or worksheets. For example, the TA of the traditional section, Ted, spent 15 minutes during one session at the chalkboard explaining crystalline solids, a topic that he said was not discussed in lecture. He said that “exposure will come from [the discussion section] and lab” and advised students to “read more about it.” In each discussion section, Ted provided answers and explanations to the whole class, individual students, and groups. Ted said he used his discussion section to “connect [worksheets] with what we [instructors] [are] trying to talk about—probably for the exam, for what [students] are going to use in the future.” He believed the goal of “the discussion session is more like trying to deliver what the class is talking about.” He wanted to provide feedback that helps students understand expert performance because students will be tested on it on exams and potentially in future courses.

Similarly, we observed Addie, the TA of the active learning discussion section, circulating the room to monitor collaboration, troubleshoot issues, check understanding of concepts, and explain concepts that students did not understand. Each time she gave feedback on students’ explanations, she also provided her own explanation. For example, in an interview, a student explained, “sometimes the worksheet does have you like tell [Addie] what your answer was and why, which is good because she is usually very critical about that and if you’re sort of vaguely explaining something, she wants you to do a better job and she’ll go through [the explanation] too.” Thus, both Ted and Addie viewed providing expert feedback as a key aspect of their role as TA.

However, Addie believed that her role differed slightly from TAs in traditional sections. As she explained in her interview: “The setup of the discussion section does impact my role somewhat, so I guess if it wasn’t structured in the way that it currently was, I would probably spend more time like answering questions [...] to see like [students] had any general questions on material in lecture, or on... like their pre-discussion worksheets and such.” In Addie’s section, feedback from online problems helped to answer student questions. The feedback provided expert performance that Addie otherwise would have provided if she was in a traditional section.

Students’ understanding of expert performance seemed to be a key aspect of the course. The professor stated in his interview that he and the other instructors try to “make [the course] better preparation for students who are going to engineering or bio-medical areas.” He then listed the concepts and skills important to the course. His focus on course content highlights how expert performance is emphasized in the design of the course.

Our results suggest that understanding expert performance is prevalent in course design and instructor interactions. The key difference between traditional and active learning sections was that online problems provided extra support for expert feedback by indicating correctness of responses. Thus, students in the active learning sections received more feedback for understanding expert performance than students in the traditional section.

## Feedback component 2: Assessing current performance

Next, students need to assess their current performance (Sadler, 1989). Our observations revealed that students and TAs assessed students’ performance by *checking answers for correctness* on worksheets or online problems.

Both TAs circulated the room to check students’ answers. As Addie explained in her interview: “If I notice that [...] they got the question wrong, then I might ask them if they understand what’s going on, and [...] if they don’t have any questions, then I would move onto the next [...] table to see if they have any questions.” Addie explained that she stopped to check students’ answers, but that she had to move on to other students to make sure she addresses all student questions. This suggests that she had limited time for each student and may therefore not be able to help all students assess their current performance.

Our observations suggested that peers can augment TA feedback by helping each other assess their current performance. For example, in Ted’s section, we observed a student, Tammy, helping another student, Tom:

Tammy: “What about phosphorus?” (*looking over at Tom’s worksheet*)

Tom: “It’s ...” (*trying to explain his answer*) “Oh no, no it’s not... ah phosphorus, why you do that to me?” (*he erases work on his paper*)

This excerpt illustrates that Tom did not realize his error until Tammy asked about his answer and he attempted to explain it. Our observations showed that feedback from peers often triggered students’ assessment of their current performance. This example also suggests that collaboration can help students self-assess their current performance through explaining. In an interview, a student in the active learning discussion section, Colleen,

stated that explaining is “one way [she] like[s] to self-assess, it’s being able to explain it.” The following example illustrates a common observation from the active learning discussion section where students explained concepts to each other and used feedback from online problems to assess their performance on a specific problem:

Carl: “2-2NO”

Colleen: “Wait”

Carl: “I don’t think we need this other box”

Colleen: “But we aren’t allowed to put intermediates in the rate law”

Carl: “No...”

*Colleen and Carl stare at the computer.*

Colleen: “I’ll try one...” (*clicks on the computer*)

Carl: “So then the reaction is dependent on that, right? I think...” (*starts to explain his reasoning while drawing on his paper, and then suggests an answer*)

Colleen: “Oh, I’ll try it” (*clicks on the computer*)

Colleen: “Yes” (*announcing that they were correct*)

In this excerpt, Carl and Colleen were unsure how to approach a problem. Carl found a solution but needed feedback from the online problem to assess whether his understanding and explanation was correct.

Further, correctness feedback from online problems can help students address gaps in their understanding. For example, a student explained that online problems helped her because “if you got [a problem] wrong, that’s where you learn a lot because I can go back and [...] if I just know the numerical answer, like if it’s 8.3, and my answer is way off, then I could just go back and just try different ways to do it and then the one way that works, then I’ll know, ok, this is the technique that I need to use to do this problem. Or I can see where I went wrong, I guess, and that really helps me ‘cause the next time I do it, I make sure I don’t make that same mistake.” This quote illustrates that correctness feedback from online problems may help students identify *what* they do not understand. Further, the feedback may also help students identify *how* to resolve such gaps.

In sum, our observations showed that feedback assessing students’ current performance can be provided by TAs, peers, and online problems. Because TAs must manage many students and can hence not always be readily available, students had to rely on additional feedback sources to assess their current performance. In the traditional discussion section, peers served as feedback sources when they checked each other’s answers for correctness. In the active learning discussion section, peers and online problems served as feedback sources. Particularly, online problems provided correctness feedback that augments students’ assessment of current performance.

### Feedback Component 3: Bridging the Gap Between Current and Expert Performance

Third, students need to bridge the gap between current and expert performance (Sadler, 1989). Our observations revealed that TAs and students bridged the gap by providing *corrective explanations*. For instance, students may explain concepts not covered in the course to support *what* students do not understand and *how* to address it:

Colleen: “Ok, I’m completely unfamiliar” (*referring to the problem regarding the heat formula and how to calculate it*) “It was not in the textbook or the lecture”

Carl: “You probably remember  $q = MC\Delta T$ ” (*the heat formula*)

Colleen: “I didn’t take AP Chem”

Carl: “I can explain it if you want”

Aaron: “Explain it please” (*looks over from his computer*)

Carl: “You too?” (*turns to Aaron*)

Aaron: “I like how he explains it” (*looks at both Colleen and Carl*)

*Carl explains the use of the heat formula by drawing a diagram on his paper.*

*Addie comes over and hears the end of the explanation. She suggests paying attention to the units because they are different (joules vs. kilojoules).*

*Addie then looks at the paper and realizes students do not understand the heat formula. She suggests that Aaron and Colleen do question 3 on the worksheet to get the background.*

In this example, Colleen and Aaron asked Carl to explain a concept that was not covered in existing instruction. Colleen mentioned in her interview that Carl often explained concepts as they worked together (see also the excerpt of Carl and Colleen in the above section). She finds that Carl is “really helpful. He took AP Chemistry so I think he just knows some of the basics better.” Hence, Aaron and Colleen seem to value his explanations. Further, Carl’s explanation helped the TA, Addie, realize what students did not understand and provide corrective feedback needed to bridge the gap. Hence, collaboration not only helped students, but also the TA, identify and bridge gaps.

This observation also suggests that TAs may not be able to resolve the gap easily because their thinking differs from students. Interviews provided further evidence for this observation. For instance, Ted said that he “can see the answers most of the time, but [students] don’t, so [...] if we [TAs] work it out too fast, [students] are [...] confused... I’m not sure if they do understand or don’t understand it.” A student in Addie’s active learning section also identified this difference: “there’s peers who [...] try to explain it in a way that you understand, rather than teachers explaining it in the way that they think people understand maybe, which is sometimes right.” She appreciated peer explanations because TA explanations were not always useful to her own understanding.

Addie’s student also added that, “it’s also helpful to listen to other people’s questions because a lot of times like you haven’t quite gotten to that yet, or like they have a different insight [...] and then um it’s also helpful like to try to explain to other people too, so like if you think you have a good aspect on that, it would help them out.” This student found collaboration useful because it allowed her to ask questions, explain to peers, and to listen to exchanges among other students.

Our observations suggest that collaboration and student explanations occurred less frequently in traditional sections. A student in Ted’s traditional section articulates this observation: “[Ted] says, you can work together but a lot of people don’t work together that much, and I’m probably one of them too, because I don’t know, I wish there was a different way to have us work together than just ok, here’s a worksheet [...] no one has seen the materials before, so then, I don’t know it’s hard to just be able to work together on it because you have to really understand it personally. A lot of times working together, it’s usually like ok, here’s the answer.” This quote illustrates that many students in the traditional section did not work together. Those who worked together checked answers for correctness to assess their current performance, but did not work with each other to bridge gaps.

The student further explained she was hesitant to collaborate because there is risk in working with peers: “I would rather get help from a student, but I probably go to the TA to get help because it’s more convenient [...] students could tell me the wrong way to do problems. You have to be careful of that, make sure it’s the right way and the right answer.” This student worried that corrective feedback from peer explanations included incorrect information. Hence, she relied on the TA for feedback. This suggests that, without *correctness* feedback from an expert on “the right way and the right answer,” students may be hesitant to collaborate and help each other bridge the gap between current and expert performance because they perceived peers’ corrective explanations as risky.

In sum, our findings suggest that feedback from TAs may not sufficiently bridge the gaps that students identify. Peer feedback seemed to be most effective in helping students bridge the gap between current and expert performance because students have a better understanding of their peer’s gaps. Such peer collaboration was prevalent in the active learning discussion section, but not in the traditional discussion section because students did not trust corrective feedback from peers without correctness feedback.

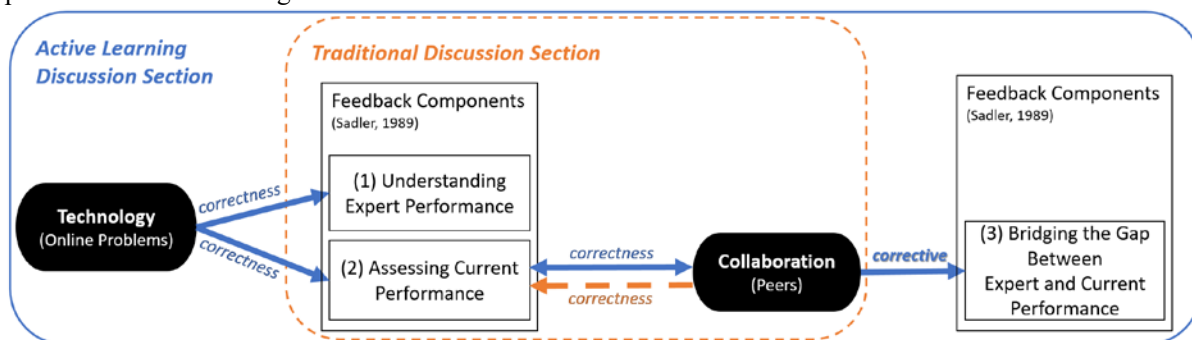
## Discussion

This paper presents an observational study of active learning and traditional discussion sections in an undergraduate chemistry course. Interview and observational data showed that students received formative feedback from: (1) TAs and online problems to help students understand expert performance, (2) TAs, online problems, and peers to help students assess their current performance, and (3) TAs and peers to bridge the gap between expert and current performance. In regards to bridging the gap, results suggest that students wanted collaboration and more feedback from peers, because peers “explain it in a way that [they] understand.” Such peer explanations were more prevalent in the active learning sections than in the traditional sections.

One key difference between the active learning and traditional discussion sections was the availability of online problems. In the active learning discussion section, students received correctness feedback from online problems. The correctness feedback may have helped students address confusions and identify gaps between their current and expert performance. Then, students could ask their TAs and peers to help them bridge specific gaps with *corrective* feedback. In the traditional discussion section, students were not provided feedback from online problems to assess whether their answers were correct. These students checked answers with peers to assess their current performance, but they did not trust the correctness feedback or corrective explanations from peers because they might provide incorrect information. Hence, they did not engage in peer explanations to help them bridge the gap. Therefore, one possible explanation of how technology supported collaboration in active learning interventions is that correctness feedback from the online problems helps students trust corrective feedback from peers,



particularly in bridging gaps between expert and current performance. Figure 2 shows a theoretical model of this process for active learning and traditional discussion sections.



**Figure 2.** Theoretical model of the interactions between technology, collaboration, and feedback in active learning discussion section (blue, solid lines) and traditional discussion section (orange, dashed lines).

From a CSCL perspective, it is striking that the online problems did not directly support collaboration. Rather, they *indirectly* supported collaborative learning by providing feedback that students elaborated on in peer collaboration. Specifically, the online problems in our study focused on content, not on collaboration. This finding extends CSCL research, which has typically investigated technologies designed to directly affect student collaboration and learning (Strijbos, Kirschner, & Martens, 2004). On the one hand, this indicates that CSCL research should investigate how content-focused technologies enhance collaboration without explicitly being intended to do so. Because active learning interventions describe a broad range of interventions that use various types of educational technologies and collaborative interventions, these interventions provide a rich context to investigate how technologies enhance learning through collaboration. In this context, CSCL provides a useful perspective that can help explain which mechanisms account for the effectiveness of active learning interventions (Strijbos et al., 2004). On the other hand, CSCL research may further improve active learning interventions. If indirect supports for collaboration are effective, can *direct* computer-based supports with elaborated feedback further enhance collaborative learning and yield even more effective active learning interventions?

The fact that content-focused correctness feedback supported collaboration by fostering trust in peers' corrective feedback highlights the need to consider *trust* when designing technology to support collaborative learning in STEM courses. While students preferred peer feedback, students did not trust it unless it was informed by correctness feedback. Students' apprehension may result from the emphasis of exams in STEM courses on *one* correct answer as an indicator of expert performance. Thus, students require content-focused correctness feedback that assure progress towards expert performance. This has important implications for technology design and implementation. For instance, technology could enhance collaboration by providing access to peer explanations, but students may not trust that feedback without confirmation from an expert that the explanations are correct. Once students received correctness feedback that they trust, they may engage with peer explanations that enable them to bridge the gap. Future investigations should confirm our findings that correctness feedback makes active learning interventions effective for collaboration in order to inform the implementation of active learning in STEM courses. For instance, a follow-up study could investigate whether providing online problems that only provide correctness feedback to students in the traditional discussion section can also promote collaboration.

Our study also raises questions regarding the role of instructors and peers with collaboration and technology. Feedback supports students in understanding expert performance, assessing current performance, and bridging the gap between them. Instructors typically provide all three types of feedback. However, our findings suggest that they may not be able to effectively do so for all the students. Hence, instructors and students may both benefit from technology support that adapts feedback to instructor and student needs. Investigating how instructors and students use technology feedback can provide insights into how educational technologies might provide more specific, tailored guidance that supports student learning (Van der Kleij et al., 2015).

## Limitations

Our study should be interpreted in light of the following limitations. In general, qualitative studies serve to reveal causal mechanisms in the specific study context. They do not attempt to prove generalizable causal relationships. As with all qualitative studies, our study provides an account of a specific sample, context, and setting. A variety of factors may contribute to the feedback and collaboration between peers and TAs, such as motivation and ability of students to provide quality feedback to each other (Nicol & Macfarlane-Dick, 2006). Future research will

investigate how these factors influenced interactions among students, instructors, and computer-supported instructional materials and whether they contribute to the effectiveness of active learning interventions. Further, this study focused on formative feedback. Many other aspects affect student learning such as the group dynamics in which interactions are situated. Future analyses of this data will investigate the sociocultural factors in the discussion sections that may explain the effectiveness of active learning interventions and provide further insight into the mechanisms underlying how technology and collaboration support student learning.

## Conclusion

In sum, an observational study showed an indirect role for technology on collaboration between students in an active learning discussion section. The technology provided formative feedback that helped students understand expert performance and assess their current performance. By supporting these two key components of feedback, the technology indirectly enhanced students' ability to collaboratively bridge gaps between current and expert performance. The technology also enhanced the instructors' ability to provide appropriate feedback. Our results provide a first attempt at building a theoretical model describing the mechanisms that account for the effectiveness of active learning interventions in STEM courses. Further, our findings provide directions for further CSCL research that should test whether enhanced technology-supported feedback or direct support for collaboration may further enhance the effectiveness of active learning interventions.

## References

- Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73–105. <http://doi.org/10.1111/j.1756-8765.2008.01005.x>
- Eddy, S. L., Converse, M., & Wenderoth, M. P. (2015). PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes. *CBE-Life Sciences Education*, 14(2), 1–16. <http://doi.org/10.1187/cbe-14-06-0095>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23), 8410–5.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Lom, B. (2012). Classroom activities: Simple strategies to incorporate student-centered activities within undergraduate science lectures. *Journal of Undergraduate Neuroscience Education*, 11(1), A64–A71.
- Mahalingam, M., Schaefer, F., & Morlino, E. (2008). Promoting student learning through group problem solving in general chemistry recitations. *Journal of Chemical Education*, 85(11), 1577.
- Michael, J. (2006). Where's the evidence that active learning works? *Advances in Physiology Education*, 30(4), 159–167. <http://doi.org/10.1152/advan.00053.2006>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. SAGE Publications.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <http://doi.org/10.1080/03075070600572090>
- Paulson, D. R. (1999). Active learning and cooperative learning in the organic chemistry lecture class. *Journal of Chemical Education*, 76(8), 1136. <http://doi.org/10.1021/ed076p1136>
- Pienta, N. J. (2015). Understanding our students in general chemistry. *Journal of Chemical Education*, 92(6), 963–964. <http://doi.org/10.1021/acs.jchemed.5b00330>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <http://doi.org/10.3102/0034654307313795>
- Strijbos, J.-W., Kirschner, P. A., & Martens, R. L. (2004). *What We Know About CSCL-And Implementing It In Higher Education. Computer-Supported Collaborative Learning*.
- Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, 85(4), 475–511. <http://doi.org/10.3102/0034654314564881>

## Acknowledgments

This research was funded by the Institute of Education Sciences, U.S. Department of Education, through Award #R305B150003 to the University of Wisconsin-Madison. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education. We thank John Moore, Rachel Bain, Hannah Bowman, Mark Bollom, Kristopher Kennedy, Luke Oxtoby, Abe Wu, and Katie Ziebarth for their help.

# On the Adoption of Social Network Analysis Methods in CSCL Research – A Network Analysis

Marielle Dado, Tobias Hecking, Daniel Bodemer, and H. Ulrich Hoppe  
marielle.dado@uni-due.de, hecking@collide.info, bodemer@uni-due.de, hoppe@collide.info  
University of Duisburg-Essen

**Abstract:** Originating from mathematical sociology, social network analysis (SNA) is a method for analyzing and representing relational structures in online communities. SNA applications in learning settings and CSCL scenarios are growing in popularity, which is in-line with new trends in learning analytics. For the CSCL research community, the adoption of SNA techniques as part of the methodological repertoire requires adequate understanding of the core concepts, their potential contributions, and limitations. We started from the hypotheses that (1) most applications of SNA in CSCL research make use of a small set of basic methods; and (2) the discourse related to SNA is partly inadequate or imprecise. To further analyze and corroborate these “issue hypotheses” we have used network analysis techniques in order to reveal relations between SNA measures and specific aspects of CSCL research (activities, contexts, research methods) based on a corpus of 90 published studies. Based on the results we pinpoint specific issues and outline new opportunities.

## Introduction

The methodological foundation of CSCL creates a genuine interest in computational methods to analyze and formally represent relevant characteristics of learning groups and communities based on large amounts of digital data (e.g., log files) collected in technology-enhanced learning settings (Jeong, Hmelo-Silver, & Yu, 2014). Rooted in sociological studies of communities, social network analysis (SNA) provides a well-defined and elaborate mathematical apparatus that resonates with theoretical models based on actor-actor and actor-artefact relations (Wasserman & Faust, 1994). The adoption of SNA in CSCL started more than 15 years ago (Nurmela, Lehtinen & Palonen, 1999; Reffay & Chanier, 2003). Originally, networks derived from email and discussion boards were the most prominent type studied, such as the study of cohesion in learning groups using a shared forum (Reffay & Chanier, 2003). Martínez et al. (2003) present an evaluation method that combines SNA with traditional sources of data and analyses in blended collaborative learning scenarios. More recently, the interest in network analysis techniques related to the study of learning with CSCL and online learning as particular cases has been strengthened by the emergence of learning analytics as a new research paradigm (Haya et al., 2015).

Given the existing usage of SNA in CSCL research, it is not surprising that such methods have been subject to meta-level analyses. One example is a co-citation network analysis of CSCL studies from 2006 to 2013 by Tang, Tsai, & Lin (2014) which found that the most widely cited works in the field pertain to issues regarding CSCL communication and interaction patterns. Social interaction has been referred to as the “key” to collaborative learning: “If there is collaboration then social interaction can be found in it... if there is no social interaction then there is also no real collaboration” (Kreijns, Kirschner, & Jochems, 2003, p. 338). Interactions characterized as “collaborative” do not merely refer to how frequently peers in a joint activity interact, but also to how influential these interactions are in their cognitive processes (Dillenbourg, 1999). Suthers and colleagues consider the fundamental basis of CSCL interactions as the relationship present when one actor’s learning activities builds upon that of another actor (Suthers et al., 2010).

Similarly, from the social network perspective, social relations are more important for understanding the behavior of groups and communities than individual attributes. In a group, actions and beliefs are strongly determined by social contexts and conditions (Wasserman & Faust, 1994). In this sense, the perspectives of CSCL and SNA coincide in focusing on social relations that go beyond individual characteristics. CSCL considers learning as a social endeavor, occurring as a result of relationships among learners and between learners and objects in the learning environment (Jones, 2015). By interacting, communicating, and sharing knowledge via computers, learners form “computer-supported social networks” from which learning emerges through constructive information exchange (Haythornthwaite, 1999). A social network approach to CSCL would thus help researchers and practitioners answer the questions: How is information distributed among learners? How much does the group share its information? What media supports this collaboration?

CSCL research has a vibrant, interdisciplinary research tradition that incorporate both qualitative and quantitative techniques (Jeong et al, 2014). SNA itself is also not exclusively one or the other, a trait that has been

an attractive incentive for educational researchers with regards to quantifying social phenomena (Carolan, 2014). Nevertheless, the adoption of SNA methods requires a certain degree of familiarity with mathematical concepts from graph theory and network algorithms, similar to foundations of quantitative-empirical research in statistics. Yet, even after overcoming this first barrier, the new concepts have to be appropriated and integrated with the “CSCL epistemology” so that “technical” SNA are smoothly inter-related with pedagogical interpretations. The bridges that we use in this context should be sound and precise, not just surface-level shortcuts. One example of a problematic bridge is the notion of “interaction pattern” that is frequently mentioned in conjunction with SNA studies. Also, without a deeper understanding of the formal-analytic background a large part of the potential of SNA may remain unexploited.

To empirically study and corroborate these “issue hypotheses”, we have applied network analysis techniques to a corpus of publications at the intersection of CSCL and SNA to detect and visualize relations between SNA measures and procedures and specific aspects of CSCL research. These networks encode concept-concept relations based on co-occurrence extracted from the abstracts and the results of publications which were previously subjected to a qualitative literature review. The concepts have been categorized, which allows for introducing a cross-category perspective in the form of multi-mode networks. The themes and trends that emerged from this analysis are then evaluated based on technical definitions of the SNA techniques and may serve as a springboard for researchers to (1) understand how CSCL research is currently viewed through the lens of SNA and (2) identify opportunities to apply underexplored SNA techniques to expand the current knowledge base in CSCL.

## Data sample and qualitative findings

The application of SNA in education research has grown in the last decade. However, relative to other research methods, the use of SNA in CSCL settings is not as well-established. A recent review of SNA in a related field, e-learning research (Cela, Sicilia, & Sanchez, 2015), revealed that SNA is mostly applied to study direct interactions between learners collaborating in online discussion forums based on communication patterns; these are usually measured using density and centrality indices. SNA is also often combined with qualitative content analysis to provide a deeper understanding of the nature of learner interaction within the network. The review however is preliminary and limited in that the general search term “e-learning” may have excluded relevant studies that use more specific terminology (e.g., CSCL).

In order to uncover trends in the application of SNA in CSCL research, a qualitative literature review was conducted. Ninety full-text studies published as peer-reviewed journal articles, book chapters, and conference papers were collected from October to November 2015 using the following keywords: social network analysis" AND "computer-supported collaborative learning" online OR computer OR collaborat\* (e.g., “collaborative” and “collaboration”) OR learning. To be included in the analysis, studies must: (1) use primary data; (2) be set in an instructional course/program up to the postgraduate (Masters) level; (3) use SNA techniques, explicitly mentioned in the Methods section; (4) report SNA findings in the Results section; and (5) analyze collaborative learning activities between learners using computers. Information on the general methodology (research design, learning setting, collaborative activity, non-SNA methods) and SNA features (actor type, relational tie, SNA measures and analysis on SNA data) were identified in each paper and quantified using content analysis.

Similar to the findings of Cela et al (2015) in the e-learning literature, between 50% to 70% of the analyzed studies measured interaction as direct communication between learners during project or task-based activities in blended learning settings, primarily using centrality and density indices. About a quarter of papers conducted content analysis to supplement SNA findings, although analyses of SNA data in most studies were limited to a descriptive report of the SNA indices. More sophisticated SNA procedures, such as identifying network positions and detecting cliques and subgroups, appeared in less than 20% of studies. A handful of studies conducted correlational analysis or inferential statistics on SNA and learner characteristics to enhance the implications of network data on learning.

These results suggest that applications of SNA in CSCL research are rather homogenous, dominated by the basic local and global measures of centrality and density as indicators of social interaction in CSCL environments. Although SNA is a promising method for analyzing collaborative learning in computer-mediated settings, the qualitative literature review lends support to the hypothesis that the analysis of CSCL interactions using SNA may not be adequate due to the limited range of applied SNA measures. In the present paper, we extend the qualitative results by exploring which CSCL activities, contexts, and research methods are associated with which SNA measures and procedures, using network analysis as a technique for meta-level literature analysis (cf. Tang et al, 2014).

## Methodology

In network text analysis (NTA) words or concepts are linked by relations based on proximity of occurrence in a text, manual coding, or grammar relations. By extracting terms in a set of texts and constructing a network based on how these terms relate to each other, NTA aims at preserving the conceptual structure of texts. More recently, NTA has also been applied to model and visualize the conceptual structure of learners within a knowledge domain based on questions and answers (Daems, Erkens, Malzahn, & Hoppe, 2014). When used in this manner, NTA could be considered automated technique for classical content analysis (Diesner & Carley, 2005). Our NTA approach is based on the network extraction pipeline used in state of the art NTA tools, such as Tools Automap (Diesner & Carley, 2005) and ConText (Diesner, 2014), and comprises three main steps: (1) concept identification, (2) concept normalization and classification, and (3) relation extraction.

First, the abstracts and the results section of the papers were prepared for analysis by removing non-relevant words (articles, auxiliary verbs, etc.) and stemming by removing suffixes. Then, parts-of-speech identification was done to identify nouns, adverbs and adjectives that represent concepts in the literature. The extraction process produced 3,057 unigrams (single nouns) and 38,934 bi-grams (two terms, combination of nouns, adjectives and adverbs), from which the top 150 unigrams and bigrams were included in the next step.

Table 1: Codebook example

Term	Concept	Categories
knowledge construction	KNOWLEDGE_CONSTRUCTION	CSCL_ACTIVITY
construction of knowledge	KNOWLEDGE_CONSTRUCTION	CSCL_ACTIVITY
online course	ONLINE_COURSE	CSCL_CONTEXT
content analysis	CONTENT_ANALYSIS	CSCL_METHOD
betweenness centrality	BETWEENNESS_CENTRALITY	SNA

An excerpt of the codebook used to identify and classify concepts in the pool of studies is displayed on Table 1. The first column contains concrete terms occurring in the texts. To account for different spellings and synonyms, the second column maps the specific terms in the first column to a general concept. The third column assigns each concept to one of the four categories: (1) “CSCL activity” for terms associated with aspects of a collaborative learning activity (e.g., “knowledge building”, “score”); (2) “CSCL context” for terms pertaining to physical/virtual settings or platforms where CSCL activities take place (e.g., “class”, “forum”); (3) “CSCL method” for terms pertaining to other analysis methods applied alongside SNA (e.g., “correlation”); (4) “SNA”, for terms related to SNA procedures and techniques (e.g., “centrality”). After combining synonyms and spelling variations, the final analysis included 101 concepts: 22 SNA, 45 CSCL activity, 23 CSCL context, 11 CSCL method. Figure 1 shows how the codebook is used to automatically identify and classify the concepts in text.

This paper took the social **KNOWLEDGE CONSTRUCTION** as the perspective and “Introduction to Educational Technology” **ONLINE COURSE** as an example to analyze **KNOWLEDGE CONSTRUCTION** level implied in those **POSTS** contributed by learners with the method of **CONTENT ANALYSIS**. Meanwhile, social network analysis (SNA) was adopted to explore the **DENSITY, CENTRALITY, COHESION** in this online **LEARNING COMMUNITY** and to discuss strategies for effective collaborative learning in virtual **LEARNING COMMUNITY**. Results indicated that the entire network **CENTRALIZATION** is comparatively low but still some points with higher **BETWEENNESS CENTRALITY** and some points functioned as bridges exist in our sample network.

Figure 1. Concept identification, normalization and classification.

(purple: CSCL\_ACTIVITY, blue: CSCL\_CONTEXT, orange: CSCL\_METHOD, green: SNA)

Once the CSCL and SNA concepts have been identified, the next step is to extract a concept network that reflects the associations made between the different concepts in the selected CSCL literature. First, for each publication the identified concepts from the codebook are interlinked to a fully connected network (or concept clique). Second, the concept cliques corresponding to particular publications are merged by overlaying these networks such that the result is a single network which contains all concepts and all links from the original networks. It is important to mention that in order to account for the relationships between SNA and CSCL a bipartite version of the network was used, which restricts the original one to having edges solely between SNA and CSCL concepts. The weight of each edge between two concepts corresponds to the number of concept cliques containing this edge, i.e., the number of papers in which the concepts co-occur.

The resulting edge-weighted and bipartite network together with the category attributes of nodes is the basis for our further analyses. In particular, we analyze the network in terms of distance between concepts and cohesive clusters, which will be described in more detail in the following section.

## Analysis and results

For a first overview we provide a frequency count of the number of publications mentioning the 22 SNA concepts. The result shown in Figure 2 supports the findings of the qualitative literature review that the usage of SNA in CSCL research is mostly restricted to centrality analysis. The most basic measure “degree centrality” appears in the abstracts and result sections of 70 of the 90 papers. In contrast, more advanced techniques such as modularity of sub-communities, positional analysis using blockmodels, or network simulation are rarely mentioned.

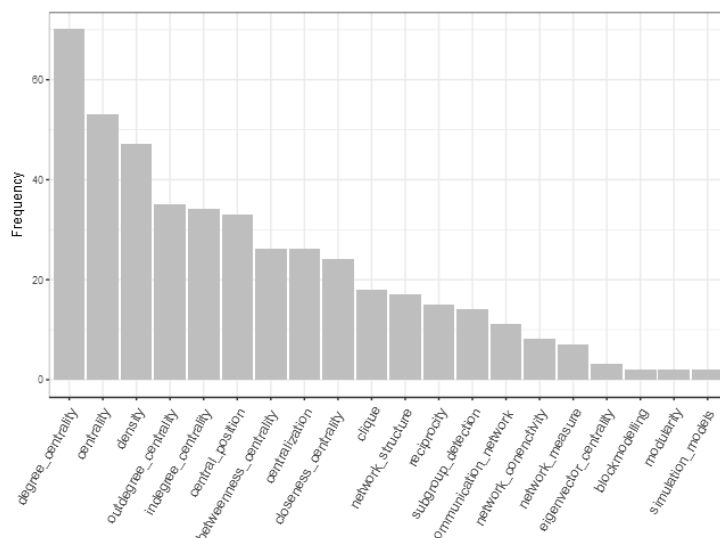


Figure 2. Number of occurrences of different SNA concepts in the 90 documents.

To further investigate the contexts in which particular SNA concepts are applied, the previous findings are extended by identifying for each SNA concept the closest CSCL concepts based on geodesic distance in the concept network. The edge weights (i.e., number of co-occurrences) of concepts were taken into account by setting the edge distance between connected concepts to the inverse of its weight.

Table 2: Profiles of SNA concepts based on proximities to different types of CSCL concepts

SNA	CSCL Activity	CSCL Context	CSCL Method
Degree Centrality	groupwork, interaction, help	course, message, class	interaction_pattern, correlation, questionnaire
Centrality			correlation, content_analysis, interaction_pattern
Density			interaction_pattern, correlation, questionnaire
Outdegree Centrality			content_analysis, correlation, interaction_pattern
Indegree Centrality			correlation, interaction_pattern, questionnaire
Central Position	groupwork, interaction, role	message, course, post	content_analysis, correlation, interaction_pattern
Betweenness Centrality			correlation, interaction_pattern, questionnaire
Centralization	groupwork, interaction, communication	course, message, class	content_analysis, interaction_pattern, correlation
Closeness Centrality			correlation, interaction_pattern, questionnaire
Clique			course, message, post
Network Structure			course, message, class
Reciprocity	groupwork, interaction, help		correlation, interaction_pattern, questionnaire
Subgroup Detection			
Communication Network	communication, groupwork, interaction		

Network Connectivity	groupwork, communication, help		
Network Measure	groupwork, discussion, interaction		
Eigenvector Centrality	groupwork, help, interaction		
Blockmodelling	communication, groupwork, discussion	course, discussion_forum, message	
Modularity		class, post, course	
Simulation Models	communication, groupwork, participation	course, message, class	time_period, interaction_pattern, correlation

Table 2 lists for all SNA concepts the three closest CSCL concepts of each category (Activity, Context, Method) yielding characteristic profiles of the SNA concepts in terms of their proximity to CSCL concepts. “Groupwork”, “interaction”, and “communication” as CSCL activities are the closest to most SNA terms, which suggests that interactions within CSCL group activities are often characterized by communication. The proximity of the concept “help” to the node- and group-centric SNA concepts (e.g., centrality, clique) suggests that help or support are mostly associated with positions of individuals in a social network. In the CSCL method category, “correlation analysis” and “questionnaire” alongside “interaction pattern” appeared in almost every profile. As was found in the qualitative literature review, this indicates that SNA is commonly used in combination with empirical data collected from questionnaires to relate structural network properties to quantitative measures of learning. Simulation models for social networks constitute the only SNA concept that is closely related to time (“time\_period” in the category CSCL Method).

Next, we have specifically analyzed the relation between SNA concepts and CSCL concepts in terms of cohesive bipartite network clusters to reveal the inherent organization of the complex network. Cohesive clusters in such a network stand for subgroups or clouds of terms that are more densely connected among each other than the average of the network. This cohesiveness can be characterized by the “modularity” measure (cf. Barber, 2007), here particularly using *bipartite modularity maximization* (Hecking, Steinert, Göhnert, & Hoppe, 2014). The clustered network representation shown in Figure 3 results from the following workflow: (1) filtering out of “weak” edges with weight below 8; (2) further reduction of the network to its 2-core to ensure that the remaining nodes are connected to at least 2 others (no singular or satellite nodes); (3) identification of mixed clusters of SNA and CSCL concepts based on bipartite modularity; (4) visualization, as presented in Figure 3. The color of the nodes indicates cluster association and node size is scaled according to its degree in the weighted network. The shapes of the nodes represent the different categories.

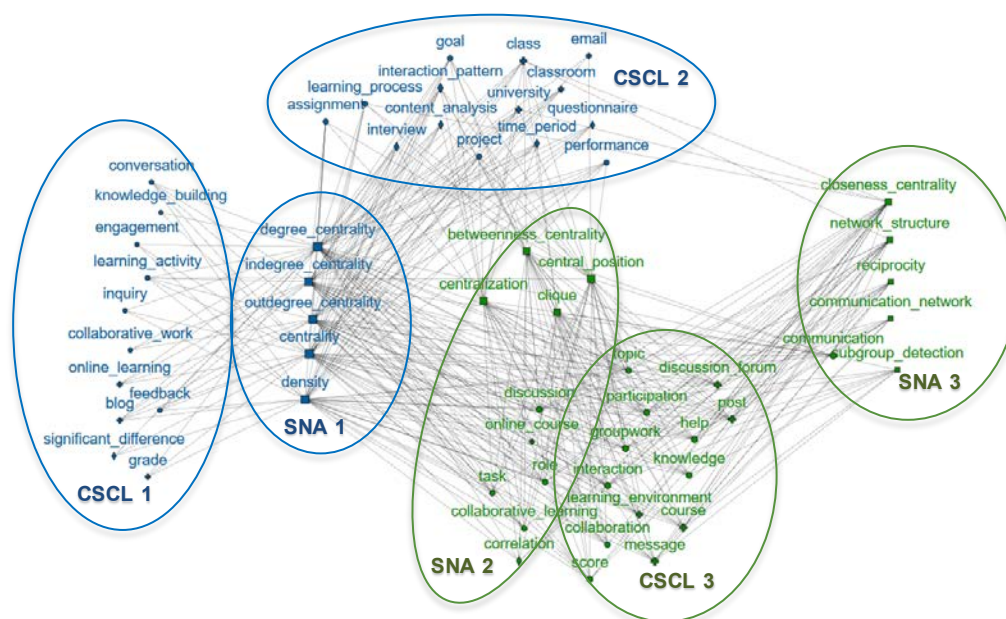


Figure 3. Result of the bipartite modularity optimization.

(• - CSCL activities, ♦ - CSCL research methods, † - CSCL context, ■ - SNA concepts)

The applied clustering algorithm identifies two clusters. The network is very densely connected, however interesting relational patterns between different parts of the network are salient. The blue cluster can be further divided into three groups of concepts indicated by the ellipses in Figure 3. The group “CSCL 1” contains general CSCL activities that are almost solely linked to the most common SNA concepts (SNA 1) and not to third blue group (CSCL 2), which in turn is strongly connected to the SNA concepts in its cluster (SNA 1) as well as in the other cluster (SNA 2). A similar pattern can be discovered for the green cluster. Here there is a group of SNA concepts (SNA 3) that is almost completely separated from the blue cluster. The concepts in this group can be considered as less common compared to the other SNA concepts in the network. The green cluster further contains SNA concepts (SNA 2) and CSCL concepts (CSCL 3) that act as bridges between the two large concept clusters. The distinction between groups of concepts based on “strong” connections to groups within the same cluster and “weak” connections to groups of the other cluster can be used to reveal the overall relational structure of SNA and CSCL concepts in the investigated literature. This macro-structure is shown in Table 3, which is also called an image matrix (Wasserman & Faust, 1994). Image matrices depict the presence or absence of pre-defined relations (here weak, strong, or absent) between different parts of a network and can be considered as a characteristic to help interpret the inherent organization of a given network

Table 3: Relationships between different parts of the concept network as image matrix

	CSCL 1 (blue cluster)	CSCL 2 (blue cluster)	CSCL 3 (green cluster)
SNA 1 (blue cluster)	strong	strong	weak
SNA 2 (green cluster)	absent	weak	strong
SNA 3 (green cluster)	absent	absent	strong

SNA concepts in SNA 1 (indegree/outdegree centrality, centrality, degree centrality) and CSCL 3 (e.g., course, post, collaborative\_learning, groupwork), are interlinked to most parts of the network, and can be considered as the conceptual core of the usage of SNA in CSCL research. Note that the terms in CSCL 3 are the same as the CSCL terms that appear in the proximity analysis in Table 2. The least connected CSCL group (CSCL 2) pertain to specific environments (e.g., blogs), pedagogical approaches (e.g., knowledge building), or processes (e.g., engagement); these are studied using most common SNA indices of centrality and density (SNA 1). Similarly, the least connected SNA group (SNA 3) contains techniques that are less frequently applied; this cluster is only associated with the most prominent CSCL concepts (CSCL 3).

## Discussion

Overall the analyses presented in this paper demonstrate a visible relational structure and conceptual core of SNA and CSCL activities, contexts, and methods that are present in the CSCL literature. The profiles generated using network analysis techniques corroborate the findings of the qualitative literature review, namely that SNA applications in CSCL research aim to understand CSCL “interaction patterns” based on communication. This was the case for all SNA indices included in the analysis and not only the most frequently used indices of centrality and density. The results also show that there have been efforts to associate these interaction patterns with learning-related variables using statistical methods (correlation), which could be seen as a way of bridging research perspectives between CSCL and education traditions (Carolan, 2014). In this section we critically discuss the core CSCL concept of “interaction patterns” in relation to the technical underpinnings of the core SNA indices.

Given the high connectivity of “interaction” and “communication” to the most prominent SNA concepts in our pool of studies, one might think that “interaction pattern” is a technical term derived from SNA. However, the basic SNA measures that were used frequently in our sample, especially centrality measures but also subgroup detection methods, only characterize relational attributes based on the connectivity of single nodes (actors), sub-networks, or the entire network. Centralization and reciprocity are global measures that indeed reveal certain general network characteristics and can characterize the topology of communication in networked learning environments. This can be conceived as a certain type of structural-relational pattern, but it is not about repeated concrete constellations, especially not in a temporal sense. The SNA technique of blockmodeling would be a means of detecting roles and role models based on consistent network structures and relational similarity. However, as our results show, it has been rarely used in CSCL studies.

Furthermore, single instances of social networks do not represent time-dependent relations, but rather capture and aggregate relations harvested during a given time window. Hoppe, Harrer, Göhnert, Hecking (2016) have made the point that the choice of different time windows as a step prior to the network generation can have systematic effects on the ensuing analysis results. While SNA research has developed several ways to handle dynamic networks that evolve over time (Aggarwal & Subbian, 2014), such techniques are not widely adopted in



CSCL: the only indication of considering time dependencies in combination with CSCL concepts was related to network simulation models, which is one of the least common SNA methods in our results. Since it has been argued that the explicit consideration of temporal processes is crucial to make sense of data produced in CSCL environments (Reimann, 2009), a future advancement of SNA in CSCL can be to consider dynamic network analysis methods based on time series of graphs.

All this indicates that the notion of “interaction patterns” as measured in SNA is indeed not very specific in a technical sense: what constitutes “interaction patterns” is not strictly operationalized in SNA. Thus, the actual technical definitions of SNA measures should be clarified when interpreting SNA findings. The SNA concept of “density”, the third frequent in our list of SNA concepts is another example for potential issues: There is a general caveat concerning the usage of density as a comparative measure applied to networks of different sizes (including growing networks). *Density* as a general measure for graphs is equal to the *average degree divided by the number of nodes* in the graph. On the other hand, in most naturally evolving networks the *average degree* will grow (if at all) at a much lower rate than the number of nodes. For scale-free networks (Barabasi & Bonabeau, 2003) the average degree is inherently constant. This means that the smallest networks will have the highest density. Hoppe, Engler & Weinbrenner (2012) have discussed this effect when studying student-generated concept maps. This makes it difficult to definitively identify an “optimal” density level of communication in CSCL research, where units of analysis tend to vary in size (Stahl, 2015). Density can be reasonably used as a comparative measure only with networks that have an identical number of nodes, otherwise the ratio of *number of edges per number of nodes* (which is proportional to the *average degree*) should be used.

In sum, the results indicate mismatches between the intended aims of applying SNA in CSCL (i.e., to investigate interaction patterns) and the actual technical definitions of SNA concepts even at the most basic level. SNA *describes* characteristics of network structures based on several common indices, and those descriptions do not necessarily capture consistent patterns of interaction that persist over time and onto other contexts. A number of sophisticated SNA techniques that are able to accomplish this, such as blockmodeling, measures on time series of networks, and network simulations, are largely underexplored.

## Conclusion

The adoption of SNA techniques in CSCL is very much focused on understanding and modeling “interaction patterns”. However, we also argue that CSCL “interaction patterns” do not necessarily become apparent from the basic and most commonly used SNA indices. However, despite SNA originating from a different analytical tradition, its use in combination with statistical analysis shows how SNA is able to cut through disciplinary boundaries. We hope that interested CSCL researchers will use our analysis as a basis for expanding the current knowledge body to include advanced SNA techniques for exploring other network dimensions such as time. This challenge would not only contribute to a more nuanced analysis on CSCL interactions, but it would also enrich the interdisciplinarity of research methods and the skill sets of CSCL researchers. As long as the CSCL community is interested in the study of social interactions, there is a place for SNA in this research field.

## References

- Aggarwal, C., & Subbian, K. (2014). Evolutionary network analysis: A survey. *ACM Computing Surveys (CSUR)*, 47(1), 10. doi:10.1145/2601412.
- Barabási, A. L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American*, 288(5), 50-59.
- Barber, M. J. (2007). Modularity and community detection in bipartite networks. *Physical Review E*, 76(6), 066102. <https://doi.org/10.1103/PhysRevE.76.066102>
- Carolan, B. V. (2014). *Social network analysis and education: Theory, methods & applications*. California: Sage Publications.
- Cela, K. L., Sicilia, M. A., & Sanchez, S. (2015). Social network analysis in e-Learning environments: A preliminary systematic review. *Educational Psychology Review*, 27(1), 219–246. doi:10.1007/s10648-014-9276-0.
- Dillenbourg, P. (1999). What do you mean by “collaborative learning”? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (pp. 1–19). Oxford, UK: Elsevier.
- Daems, O., Erkens, M., Malzahn, N., & Hoppe, H. U. (2014). Using content analysis and domain ontologies to check learners’ understanding of science concepts. *Journal of Computers in Education*, 1(2), 113-131. doi:10.1007/s40692-014-0013-y
- Diesner, J. (2014). *ConText: Software for the integrated analysis of text data and network data*. Paper presented at the Social and Semantic Networks in Communication Research. Preconference at Conference of International Communication Association (ICA), Seattle, WA.

- Diesner, J., & Carley, K. M. (2005). Revealing social structure from texts: Meta-matrix text analysis as a novel method for network text analysis. In V.K. Narayanan and D. Armstrong (eds.), *Causal mapping for information systems and technology research* (p. 81-108). Hershey, PA: IGI Global.
- Haythornthwaite, C. (1999). Collaborative work networks among distributed learners. In R. Sprague (ed.), *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences, 1999. HICSS-32, Hawaii, US* (pp. 1023-39). Virginia, US: IEEE Computer Society Press. doi:10.1109/HICSS.1999.772707
- Haya, P. A., Daems, O., Malzahn, N., Castellanos, J., & Hoppe, H. U. (2015). Analysing content and patterns of interaction for improving the learning design of networked learning environments. *British Journal of Educational Technology*, 46(2), 300-316. doi:10.1111/bjet.12264
- Hecking, T., Steinert, L., Göhnert, T., & Hoppe, H. U. (2014). Incremental clustering of dynamic bipartite networks. In *Proceedings of the 2nd European Network Intelligence Conference (ENIC 2014) Wroclaw, Poland* (pp. 9-16). Los Alamitos, CA : Conference Publishing Services. doi:10.1109/ENIC.2014.15
- Hoppe, H. U., Engler, J., & Weinbrenner, S. (2012). The impact of structural characteristics of concept maps on automatic quality measurement. In J. van Aalst, K. Thompson, M. Jacobson, and P. Reimann (eds.), *Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012), Sydney, Australia* (pp. 308-316). Sydney: International Society of the Learning Sciences. www.isls.org/icls2012
- Hoppe, H. U., Harrer, A., Göhnert, T., & Hecking, T. (2016). Applying network models and network analysis techniques to the study of online communities. In U. Cress, J. Moskaliuk, and H. Jeong (eds.), *Mass Collaboration and Education* (pp. 347-366). Switzerland: Springer International Publishing.
- Jeong, H., Hmelo-Silver, C. E., & Yu, Y. (2014). An examination of CSCL methodological practices and the influence of theoretical frameworks 2005–2009. *International Journal of Computer-Supported Collaborative Learning*, 9(3), 305–334. doi:10.1007/s11412-014-9198-3.
- Jones, C. (2015). *Networked learning: an educational paradigm for the age of digital networks*. Switzerland: Springer International Publishing.
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, 19(3), 335–353. doi:10.1016/S0747-5632(02)00057-2.
- Martínez, A., Dimitriadis, Y., Rubia, B., Gómez, E., & De La Fuente, P. (2003). Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education*, 41(4), 353-368. doi:10.1016/j.compedu.2003.06.001
- Nurmela, K., Lehtinen, E., & Palonen, T. (1999). Evaluating CSCL log files by social network analysis. In C. Hoadley and J. Roschelle (eds.), *Proceedings of the 1999 Conference on Computer Support for Collaborative Learning*, article no. 42. International Society of the Learning Sciences. <http://dl.acm.org/citation.cfm?id=1150294>.
- Suthers, D. D., Dwyer, N., Medina, R., & Vatrappu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, 5(1), 5–42. doi:10.1007/s11412-009-9081-9.
- Stahl, G. (2015). A decade of CSCL. *International Journal of Computer-Supported Collaborative Learning*, 10(4), 337–344. doi:10.1007/s11412-015-9222-2.
- Tang, K. Y., Tsai, C. C., & Lin, T. C. (2014). Contemporary intellectual structure of CSCL research (2006–2013): a co-citation network analysis with an education focus. *International Journal of Computer-Supported Collaborative Learning*, 9(3), 335–363. doi:10.1007/s11412-014-9196-5.
- Reffay, C., & Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. In B. Wasson, S. Ludvigsen, and H. U. Hoppe (eds.), *Designing for change in networked learning environments* (pp. 343-352). Springer Netherlands. doi:10.1007/978-94-017-0195-2
- Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, 4(3), 239-257. doi:10.1007/s11412-009-9070-z.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 1). Cambridge University Press.

## Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under Grant No. GRK 2167, Research Training Group “User-Centred Social Media”.

# Think First: Fostering Substantive Contributions in Collaborative Problem-Solving Dialogues

Mehmet Celepkolu, Joseph B. Wiggins, Kristy Elizabeth Boyer, and Kyla McMullen  
mckolu@ufl.edu, jbwiggi3@ufl.edu, keboyer@ufl.edu, drkyla@ufl.edu  
University of Florida

**Abstract:** Working collaboratively holds many benefits for learners. However, varying incoming knowledge and attitudes toward collaboration present challenges and can lead to frustration for students. An important open question is how to support effective collaboration and foster equity for students with different levels of incoming preparation. In this study, we compared two collaborative instructional approaches for computer science problem solving, in which students participated in one of two conditions: The *Baseline* condition featured collaborative problem solving in which students worked in dyads from the beginning of the collaboration; in the other condition, called *Think-First*, students first worked on the problem individually for a short time and then began collaborating to produce a common solution. The results from 190 students from an introductory programming class working in 95 pair-programming teams demonstrate that this simple modification to pair programming had a significant positive effect on test scores and on substantive contributions in collaborative dialogue.

## Introduction

The CSCL community has long worked to understand important facets of collaboration, such as the importance of sharing and constructing knowledge together (Stahl, 2002; Ludvigsen et al., 2011). A central idea that has emerged from this research is that explicit and/or implicit structure for learning activities can significantly improve the quality of collaboration (Beers et al., 2007; Kirschner et al., 2008). One example of a collaboration structure that has been actively used in the domain of computer science is *pair programming*, in which two learners work side by side at the same computer and collaborate on the same code. While one of the learners, the *driver*, is typing at the computer, the other partner, the *navigator*, assists in planning, strategizing, and pointing out errors. Through pair programming, students produce higher quality code, derive greater enjoyment from the programming experience, and learn more from each other compared to individual programming (Cockburn & Williams, 2000; Dybå et al., 2007; McDowell et al., 2002; Nagappan et al., 2003; Preston, 2006; Williams et al., 2000). Previous research has also provided insight into how students should be paired together, using approaches such as personality tests and previous test scores (Chao & Atli, 2006; Werner et al., 2004). However, there has been very little investigation of how to support collaboration after students are paired.

When individual learners within a dyad have substantially different levels of knowledge or motivation, support for collaboration is particularly important. For example, when students with different knowledge levels work on the same problem, information usually flows from the *high-performing* student to the *low-performing* student rather than through a balanced dialogue (McCarthy & McMahon, 1992). Similarly, some students, when paired with a less knowledgeable peer, can feel that collaboration is not worthwhile (Bevan, Werner & McDowell, 2002). These challenges highlight a pressing open research question: ***How can we effectively support collaborative problem-solving dialogue between two students with contrasting levels of incoming preparation?***

To investigate this research question, we conducted a study of collaborative learning for computer science, in which there were two conditions: one in which students collaborated in standard pair programming (the *Baseline* condition), and one in which students worked on the problem individually for fifteen minutes and then began collaborating with their partner to construct a common solution (the *Think-First* condition). We hypothesized that students in the *Think-First* condition would achieve higher learning outcomes than students in the *Baseline* condition while reporting equal or higher levels of enjoyment. The results support these hypotheses. Without the *Think-First* structure, stronger students often started out more prepared and took increasing control, while the less prepared students became increasingly marginalized as time went on. The results show that reserving some time for students to think individually before they began to collaborate improved learning outcomes without adding extra total time, and without negatively affecting their enjoyment of the collaborative process.

## Prior work

A widely used instructional design approach is for students to be given some initial time to engage with the learning activity individually and then be expected to work in a small-group or share their knowledge with the whole-class. This approach has been shown to increase engagement in computer science courses (Fitzgerald, 2013; Kothiyal et al., 2013) and improve student achievement (Kaddoura, 2013; Siburian, 2013; Sugiarto & Sumarsono, 2014; Azlina & Nik 2010). This instructional method can potentially help computer science courses as it gives students time, which allows them to judge what they know, prepares them for collaboration, and eventually improves their higher-order thinking skills (Yerigan, 2008). Examining the dialogues between dyads can shed light on the underlying phenomena that lead to success in collaborative learning, such as negotiation of meaning and construction of shared conceptions (Stahl, Koschmann & Suthers 2006). Stahl (2006) states "meaning is created across the utterances of different people." Studies of spoken or textual dialogues help us to understand the changing needs of individuals in dyads or groups (Davidsen & Ryberg, 2015).

Socio-Constructivist theory suggests that learning is a social, interactive and collaborative process through which students develop higher-order thinking skills, such as reasoning and problem solving (Kozulin et al., 2003). Vygotsky (1978) emphasizes the significance of pairs remaining within each other's "Zone of Proximal Development" (ZPD) because a large difference can create too much or too little challenge, leading to confusion or demotivation. Matching students with partners who operate within their own ZPD is very important; however, practical challenges in classrooms usually make it difficult to match students perfectly in this way. Structured support for collaborative problem solving can be helpful even in the presence of contrasting incoming knowledge levels. In this study, we examine the simple strategy of setting aside time for individual thinking before collaboration begins.

## Methods

We conducted a quasi-experimental study to investigate differences between the *Baseline* and *Think-First* conditions and examined case studies to make qualitative observations. Pre- and post-surveys and content knowledge posttests were analyzed using descriptive and inferential statistics. We also examined five Baseline and five Think-First dyads' collaborative learning activities during pair programming, which we video recorded and transcribed manually.

## Participants

The study was conducted with students who were actively enrolled in the first computer science class for computer science majors at a university in the southeastern United States. The class was taught in the Java programming language and had a total enrollment of 471. Class met three times per week in a large lecture hall. In addition to lectures, each student attended one two-hour lab per week in which they collaboratively solved lab exercises. Undergraduate and graduate teaching assistants led these labs. All students completed the same lab exercises (including students who were not part of the study). We only collected data on the 190 students who consented to participate in research, which was voluntary. Of the 190 students, there were 85 freshmen (45%), 56 sophomores (29%), 41 juniors (21%), 5 seniors (3%), and 3 with other designations (2%). 125 students (66%) reported having prior programming experience and 91 (48%) reported prior experience in the Java programming language. There were 46 (24%) women and 144 (76%) men. Of all participants, 54% identified as White/Caucasian, 17% as Hispanic/Latino, 14% as Asian/Pacific Islander, 4% as Black/African-American, and 11% as Other (including multiracial and some races not listed in the above categories).

One week prior to the study, students took a midterm exam as part of regular course activities. The average score on the exam was 75.6 (*median* = 77; *stdev* = 15.2). We grouped consenting students based on whether they were above the median (High) or below the median (Low). Students were assigned into "High-High", "High-Low" and "Low-Low" dyads randomly. We used the midterm scores as the pre-test to gauge levels of understanding before the lab to compare to the post-test after the lab. There are no significant pre-test differences between the *Baseline* and *Think-First* conditions (High performing students in High-High dyads  $p = 0.2370$ ; high performing students in High-Low dyads  $p = 0.1302$ , Low performing students in High-Low dyads  $p = 0.7068$  and Low performing students in Low-Low dyads  $p = 0.4331$ ). After pairing was complete, each lab section (consisting of between 5 to 10 dyads) was randomly assigned to either the *Baseline* or the *Think-First* condition.

Table 1: Number of dyads in each category considered in this study

Dyad Type	Baseline Condition	Think-First Condition	TOTAL
High - High	15	17	32 pairs
High - Low	20	11	31 pairs
Low - Low	17	15	32 pairs
TOTAL	52	43	95 pairs

## Procedure

Students participated in the study as part of their regular two-hour lab section. At the beginning of the lab, students were placed into their pre-arranged dyads. All dyads completed a learning task involving one-dimensional arrays. In the *Baseline* condition, students began pair programming immediately. In the *Think-First* condition, students worked individually for fifteen minutes and then began pair programming to construct a collaborative solution. After they completed the assignment, or when the allocated time was over (two lab hours), students individually completed a ten-question pair programming attitude survey. Then, they completed a posttest consisting of ten multiple-choice questions on one-dimensional arrays. We randomly selected five dyads from each condition to be video and audio-recorded throughout their collaboration.

## Results and discussion

We compared learning outcomes in the *Baseline* and *Think-First* conditions. The results show that students performed significantly better on the posttest in the *Think-First* than in the *Baseline* condition. This result held for both low-performing and high-performing students (as measured on the preceding midterm exam). As shown in Figure 2, high-performing students in the *Baseline* condition had a mean posttest score of 4.22 ( $n = 50$ ;  $stdev = 2.24$ ) and high-performing students in the *Think-First* condition had a mean posttest score of 5.20 ( $n = 45$ ;  $stdev = 1.94$ ). This difference is significant ( $p = 0.0257$ , ANOVA). For only the high-performing students, we examined the effect of the *Think-First* condition on different dyad types. The effect of *Think-First* was not significant for high-performing students paired with other high-performing students, but high-performing students paired with low-performing students achieved significantly higher posttest scores ( $p = 0.0429$ , Wilcoxon rank-sum test) in the *Think-First* condition compared to *Baseline*. Table 2 shows the outcomes for high-performing students.

Low-performing students also clearly benefitted from the *Think-First* condition. As shown in Figure 2, low-performing students in the *Baseline* condition had a mean posttest score of 3.74 ( $n = 54$ ;  $stdev = 2.25$ ) and those in the *Think-First* condition had a mean posttest score of 5.12 ( $n = 41$ ;  $stdev = 2.24$ ). This difference is also significant ( $p = 0.0038$ , ANOVA). For low-performing students paired with another low-performing student, the effect of *Think-First* was significantly positive ( $p = 0.012$  Wilcoxon rank-sum test). The benefit for low-performing students paired with high-performing students was not statistically significant ( $p = 0.16$ ). Table 3 shows the outcomes for low-performing students.

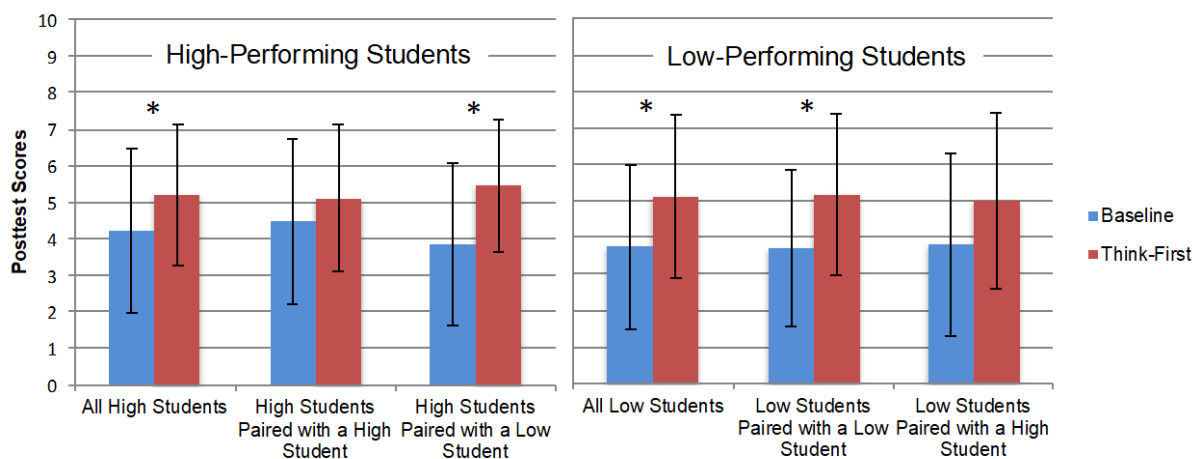


Figure 2. *Think-First* effect on students' individual posttest score by collaborative condition (\* indicates significant difference between *Baseline* and *Think-First* conditions).

Table 2: Summary of pre-test scores and outcomes for high-performing students

	<i>High-performing students paired with another High-performing student</i> mean (stdev)		<i>High-performing students paired with a Low-performing student</i> mean (stdev)	
	<i>Baseline</i> (n = 30)	<i>Think-First</i> (n = 34)	<i>Baseline</i> (n = 20)	<i>Think-First</i> (n = 11)
<b>Preceding Midterm (out of 100)</b>	88.7 (6.34)	86.97 (5.07)	86.7 (6.07)	90.45 (6.47)
<b>Post Test (out of 10)</b>	4.47 (2.27)	5.12 (2.00)	3.85 (2.21)	5.45 (1.81)
<b>Enjoyment (out of 50)</b>	37.20 (10.16)	35.06 (9.25)	37.50 (8.79)	31.72 (11.13)

Table 3: Summary of pre-test scores and outcomes for low-performing students

	<i>Low-performing students paired with another Low-performing student</i> mean (stdev)		<i>Low-performing students paired with a High-performing student</i> mean (stdev)	
	<i>Baseline</i> (n = 34)	<i>Think-First</i> (n = 30)	<i>Baseline</i> (n = 20)	<i>Think-First</i> (n = 11)
<b>Preceding Midterm (out of 100)</b>	60.64 (10.91)	62.93 (12.10)	67.05 (10.97)	65.18 (14.00)
<b>Post Test (out of 10)</b>	3.71 (2.13)	5.17 (2.21)	3.80 (2.50)	5.00 (2.41)
<b>Enjoyment (out of 50)</b>	38.06 (8.31)	39.83 (9.90)	35.90 (11.54)	39.55 (10.23)

Our hypothesis was that students in the *Think-First* condition would achieve higher learning outcomes compared to students in the *Baseline* condition. Indeed, this was the case. *Low-performing* students benefitted more overall: there was a medium effect size for the low-performing students (Cohen's  $d = 0.59$ ) and a small effect size for the high-performing students (Cohen's  $d = 0.46$ ).

To explore the ways in which the *Think-First* condition influenced collaborative problem-solving dialogue compared to the *Baseline* condition, we now consider several excerpts. Figure 3 left shows a conversation between a low-performing student (whom we call Lindsay) and a high-performing student (whom we call Hector). In this interaction, Lindsay took on the role of driver first, and she began writing code. As soon as Lindsay began coding, the pair asked questions of each other. But soon, Lindsay made mistakes and Hector corrected her several times in less than one minute. Within two minutes of working on the programming code, Lindsay asked Hector if he wanted to be the driver. Hector accepted the offer and took control. This example from the *Baseline* condition illustrates a case of a less-prepared student losing confidence and giving up control quickly. In typical programming, we see drivers and navigators switch roles every fifteen to thirty minutes, but in Figure 3 left, Lindsay gave up control after less than two minutes of active coding.

We contrast Hector and Lindsay's interaction with that of a *High-Low* dyad from the *Think-First* condition. Figure 3 right shows dialogue between Luke (a low-performing performing student) and Hilda (a high-performing student). Hilda began as the driver and Luke contributed to the process of code writing. As soon as the collaboration started, Luke made a suggestion about how to start solving the problem, and Hilda

agreed. At the 04:41 mark, Luke referred to a previous lab assignment, which he had looked up individually during the *Think-First* time. Although he was the low-performing student in the dyad, he was able to make this substantive contribution because he used the *Think-First* time to refresh his memory and prepare for collaboration. At the 05:13 mark, Hilda made a suggestion and Luke responded by accepting it. After this step, Hilda offered for Luke to do the next step and Luke agreed. In contrast to Lindsay in the *Baseline* condition, who was eager to give up control quickly, Luke accepted control willingly.

<i>Baseline session</i>	<i>Think-First session</i>
<b>Lindsay (06:10)</b> Ok, we could take an input first.	<b>Luke (04:33)</b> So I guess the first part is the random number.
<b>Hector (06:13)</b> No, I think you need a scanner, right?	<b>Hilda (04:40)</b> Yeah
<b>Lindsay (06:15)</b> Yes, new scanner?	<b>Luke (04:41)</b> we need to, although... So I looked up from our last lab, we could make random number by saying randNumber. They say they call it randnumber
<b>Hector (06:17)</b> Not scanner. I use Command key + B. It is like keyboard	<b>Hilda (04:51)</b> Yeah
<b>Lindsay (06:23)</b> Alright, let's call it console, do you care?	<b>Luke (04:52)</b> So if you do randnumber = math.random
<b>Hector (06:25)</b> No.	<b>Luke (05:00)</b> Capital A. You have to do Math. Capital M. Uh... Times one thousand
<b>Lindsay (06:25)</b> I am calling it console. Console equals to new? I am definitely doing it wrong.	<b>Hilda (05:08)</b> So, that is on the outside
<b>Hector (06:30)</b> No, it is new, it is true. New scanner. Create a new object so you are making a new scanner object and then like whatever parameter just system.in	<b>Luke (05:11)</b> and we have not imported it yet so it is gonna give an error.
...	<b>Hilda (05:13)</b> Alright. Let's just do that. However it is an Int, I guess
<b>Lindsay (07:54)</b> Do you wanna do this? (Offering Hector to become driver.)	<b>Luke (05:18)</b> Yeah it will be an int. We haven't imported it yet so we gotta go top to do import
	<b>Hilda (05:24)</b> Ok. Do you want to do this part?
	<b>Luke (05:26)</b> Yes, sure

**Figure 3.** Illustrative excerpts from *High-Low* dyad in a *Baseline* condition on the left and a *High-Low* dyad from the *Think-First* condition on the right.

The previous two excerpts illustrate collaboration between *High-Low* dyads. Next we examine dialogues between *Low-Low* dyads. Figure 4 left shows a conversation between Lloyd (*low-performing*) and Larry (*low-performing*) in the *Baseline* condition. Larry started as the driver and Lloyd first suggested creating a new Scanner (to receive user input) but Larry was unaware of how to do this, so he asked what they needed. Larry continued typing, but he was not sure how to import a scanner, so he asked Lloyd to clarify it. At 07:35, Larry said they needed to create a scanner and Lloyd told him that he usually names it "input". However, Larry struggled to create the scanner and Lloyd ultimately dictated the syntax to Larry. Lloyd was visibly frustrated and becoming increasingly disengaged.

We contrast Lloyd and Larry's experience with a *Low-Low* dyad from the *Think-First* condition. Figure 4 right shows a conversation between Laura (*low-performing*) and Lewis (*low-performing*) in the *Think-First* condition. At first, Laura and Lewis agreed on creating a Scanner, but, like Lloyd and Larry, they did not recall how to do it. However, they spent two minutes checking their own notes made during the *Think-First* time, and collaborated productively to complete the needed line of code. Neither student appeared to become frustrated during this time, and they both made substantive contributions as they worked through the challenges they faced.

<i>Baseline session</i>	<i>Think-First session</i>
<b>Lloyd (06:07)</b> We need to import scanner first.	<b>Laura (03:11)</b> Alright. First thing first.. Oh yeah, we need a scanner
<b>Larry (06:10)</b> We need what?	<b>Lewis (03:12)</b> Hmm hmm
<b>Lloyd (06:12)</b> Import scanner	<b>Laura (03:31)</b> Oh, ok. Hmmm. How does that one go there?
<b>Larry (06:13)</b> Oh, you are right. Java utils and... Is it space scanner or dot scanner?	<b>Lewis (03:32)</b> Scanner console equals System dot in?
<b>Lloyd (06:23)</b> Dot scanner	<b>Laura (03:43)</b> I think there was something else too
<b>Larry (06:24)</b> oh ok	<b>Lewis (03:45)</b> Yeah there is something before
<b>Lloyd (06:26)</b> Also, uppercase	<b>Laura (03:47)</b> Is it a "new" or something?
<b>Lloyd (07:05)</b> ... so the first one should probably import line as well.	<b>Lewis (03:51)</b> new scanner?
<b>Larry (07:10)</b> Oh, you are right. My bad.	<b>Laura (03:53)</b> yeah something like that
<b>Larry (07:35)</b> We should also create a scanner	<b>Lewis (03:55)</b> I always forget that
<b>Lloyd (07:36)</b> Usually, I call mine "input"	<b>Laura (03:56)</b> Yeah me too! Let's see.
<b>Larry (07:43)</b> What is to uhh...	<b>Lewis (05:46)</b> new Scanner and then parenthesis system dot in
<b>Lloyd (07:45)</b> Scanner space input	<b>Laura (05:52)</b> Oh! new Scanner...
	<b>Lewis (06:04)</b> Like this. That's on the right
	<b>Laura (06:07)</b> New scanner. Oh ok! Gotcha!

Figure 4. Illustrative excerpts from an interaction between a *Low-Low dyad* in a *Baseline* session on the left and *Think-First* session on the right.

To test the second part of our hypothesis, we compared the two conditions in terms of enjoyment of the programming activity. There were no significant differences in students' pair programming enjoyment ratings (Figure 5). In high-performing students paired with high-performing students, we see a decrease in enjoyment from *Baseline* ( $mean = 37.20; stdev = 10.16$ ) to the *Think-First* condition ( $mean = 35.06; stdev = 9.25$ ) with small effect size (Cohen's  $d = 0.2215$ ). In high-performing students paired with low-performing students, we see a decrease in enjoyment from *Baseline* ( $mean = 37.50; stdev = 8.79$ ) to the *Think-First* condition ( $mean = 31.72; stdev = 11.13$ ) with medium effect size (Cohen's  $d = 0.5828$ ). However, in low-performing students paired with low-performing students, we see an increase in enjoyment from *Baseline* ( $mean = 38.06; stdev = 8.31$ ) to the *Think-First* condition ( $mean = 39.83; stdev = 9.90$ ) with small effect size (Cohen's  $d = 0.1958$ ). In low-performing students paired with high-performing students, we see an increase in enjoyment from *Baseline* ( $mean = 35.90; stdev = 11.54$ ) to the *Think-First* condition ( $mean = 39.55; stdev = 10.23$ ) with small effect size (Cohen's  $d = 0.3296$ ).

Taken together, the quantitative findings and case studies indicate that the *Think-First* approach provided an opportunity for low-performing students to form goals and to reflect on their previous learning experiences before beginning collaboration. This opportunity improved their collaborative outcomes, with one benefit being that the *Think-First* time decreases how apparent the knowledge gap between students is to each teammate, fostering more equitable collaboration. At the same time, there was no significant decrease in students' overall enjoyment of pair programming (Figure 5).

**Limitations:** It is important to acknowledge that this study was not conducted in a controlled environment, but in an actual lab for a CS1 class. There were some students who did not fully complete the lab assignment within the allotted time (two lab hours). These students still completed the posttest and were included in the analysis presented here. Additionally, in this naturalistic classroom environment, some students interacted with classmates from other dyads, and also sought help from the teaching assistant. We did not intervene in these cases. Finally, the overall posttest scores were low (averaging 4.5 out of 10) in both the *Baseline* and *Think-First* conditions. This limitation is due primarily to the fact that students had been scheduled to cover the relevant new material on one-dimensional arrays in their lecture classes before attending this lab, but an unexpected severe weather incident cancelled classes. As a result, students attended their lab before seeing the relevant concepts presented in lecture. Teaching assistants conducted brief mini-lectures in labs to attempt to mitigate the impact, but students still received less exposure to the concepts prior to the lab than was originally intended.



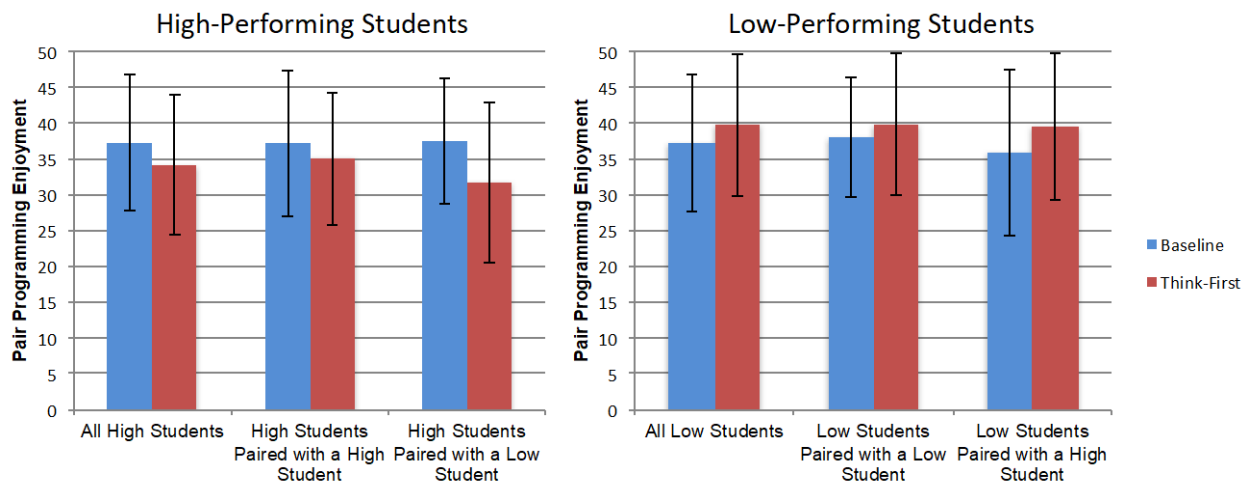


Figure 5. Enjoyment levels across conditions (no significant differences between conditions)

## Conclusions and future work

The CSCL community has long aimed at supporting equity in collaborative problem solving. In this study, we examined the effect of the *Think-First* approach on learning outcomes and collaborative dialogue. We found that it had a significant benefit for learning. Moreover, examination of the dialogues suggests that the individual thinking time enabled students, particularly low-performing students, to formulate their own ideas and build confidence before sharing their ideas with their partners. In the *Think-First* condition, substantive contributions from both partners characterized the dialogues, while in the *Baseline* condition, the low-performing student in a dyad would often become increasingly marginalized over time, turning over control and disengaging from the problem-solving process. The *Think-First* approach holds the benefit of decreasing how apparent the difference in student's previous knowledge is, and may help to foster engagement in both students early in the conversation. As depicted above, the *Think-First* condition seems to change the nature of the students' interactions. In the case of *High-Low* dyads, it seems that the low-performing student was less dependent on the high-performing student, making contributions to both the conversation and the problem-solving task, which promotes healthy collaboration despite one contributor having more incoming knowledge.

There are several important directions for future work. First, while this work has examined one simple type of collaboration structure, there are many types of potential support structures and scaffolds that are important to explore further, such as peer teaching, completing a peer's half-written program, code dividing, and error hunting. Additionally, it is important to more deeply investigate the affective and social components of effective collaboration. Finally, as we move toward supporting diverse learners in problem solving, developing adaptive techniques that consider a rich set of learner characteristics is a promising research direction for computer-supported collaborative learning technologies.

## References

- Azlina, N. N., & Nik, A. (2010). CETLs: Supporting collaborative activities among students and teachers through the use of Think-Pair-Share techniques. *International Journal of Computer Science Issues*, 7(5), 18-29.
- Beers, P., Boshuizen, H., Kirschner, P. A., & Gijssels, W. H. (2007). The analysis of negotiation of common ground in CSCL. *Learning and Instruction*, 17(4), 427-435.
- Bevan, J., Werner, L., & McDowell, C. (2002). Guidelines for the use of pair programming in a freshman programming class. In *Software Engineering Education and Training, 2002.(CSEE&T 2002)*. Proceedings. 15th Conference on (pp. 100-107). IEEE.
- Chao, J., & Atli, G. (2006, July). Critical personality traits in successful pair programming. In *AGILE 2006 (AGILE'06)* (pp. 5-pp). IEEE.
- Cockburn, A., & Williams, L. (2000). The costs and benefits of pair programming. *Extreme programming examined*, 223-247.
- Davidson, J., & Ryberg, T. (2015). "This Is the Size of One Meter": Children's Bodily-Material Collaboration and Understanding of Scale around Touchscreens. *Exploring the Material Conditions of Learning: Opportunities and Challenges for Cscl*.

- Dybå, T., Arisholm, E., Sjøberg, D. I., Hannay, J. E., & Shull, F. (2007). Are two heads better than one? On the effectiveness of pair programming. *IEEE software*, 24(6), 12-15.
- Fitzgerald, D. (2013). Employing think-pair-share in associate degree nursing curriculum. *Teaching and Learning in Nursing*, 8(3), 88-90.
- Kaddoura, M. (2013). Think pair share: A teaching learning strategy to enhance students' critical thinking. *Educational Research Quarterly*, 36(4), 3.
- Kirschner, P. A., Beers, P., Boshuizen, H., & Gijselaers, W. H. (2008). Coercing shared knowledge in collaborative learning environments. *Computers in Human Behavior*, 24(2), 403-420.
- Kothiyal, A., Majumdar, R., Murthy, S., & Iyer, S. (2013, August). Effect of think-pair-share in a large CS1 class: 83% sustained engagement. In *Proceedings of the ninth annual international ACM conference on International computing education research* (pp. 137-144). ACM.
- Kozulin, A., Gindis, B., Ageyev, Vladimir S., Miller, Suzanne M. (Ed.) 2003. *Vygotsky's Educational Theory In Cultural Context*: Cambridge.
- Ludvigsen, S. R., Lund, A., Rasmussen, I., & Säljö, R. (Eds.). (2010). *Learning across sites: New tools, infrastructures and practices*. Routledge.
- McCarthy, S. J., & McMahon, S. (1992). From convention to invention: Three approaches to peer interactions during writing. *Interaction in cooperative groups: The theoretical anatomy of group learning*, 17-35.
- McDowell, C., Werner, L., Bullock, H., & Fernald, J. (2002). The effects of pair-programming on performance in an introductory programming course. *ACM SIGCSE Bulletin*, 34(1), 38-42.
- Nagappan, N., Williams, L., Ferzli, M., Wiebe, E., Yang, K., Miller, C., & Balik, S. (2003). Improving the CS1 experience with pair programming. *ACM SIGCSE Bulletin*, 35(1), 359-362.
- Preston, D. (2006). Using collaborative learning research to enhance pair programming pedagogy. *ACM SIGITE Newsletter*, 3(1), 16-21.
- Sibirian, T. A. (2013). Improving students achievement on writing descriptive text through think pair share. *IJLLALW*, 3(03), 30-43.
- Stahl, G. (2002). *Computer Support for Collaborative Learning: Foundations for a Csel Community* (Csel 2002 Proceedings). Psychology Press.
- Stahl, G. (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge* (Acting with Technology).
- Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. *Cambridge Handbook of the Learning Sciences*, 2006.
- Sugiarto, D., & Sumarsono, P. (2014). The Implementation of Think-Pair-Share Model to Improve Students' Ability in Reading Narrative Texts. *International Journal of English and Education*, 3(3), 206-215.
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children*, 23(3), 34-41.
- Werner, L. L., Hanks, B., & McDowell, C. (2004). Pair-programming helps female computer science students. *Journal on Educational Resources in Computing (JERIC)*, 4(1), 4.
- Williams, L., Kessler, R. R., Cunningham, W., & Jeffries, R. (2000). Strengthening the case for pair programming. *IEEE software*, 17(4), 19.
- Yerigan, T. (2011). Getting active in the classroom. *Journal of College Teaching & Learning (TLC)*, 5(6).

## Acknowledgments

The authors wish to thank the members of the LearnDialogue group at the University of Florida for their helpful input. This work is supported in part by Google through a CS Capacity Research Award and by the National Science Foundation through grant CNS-1622438. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the authors, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

# Finding the Community in Online Education: It's in the Instructors' Eyes

Na Sun, Pennsylvania State University, nzs162@psu.edu  
Mary Beth Rosson, Pennsylvania State University, mrosson@ist.psu.edu

**Abstract:** An instructor's teaching presence affects students' feelings of social presence and connectivity in distance education, pointing to an opportunity for intentional orchestrations of class members' social experience. To explore such opportunities, we interviewed instructors about what and how they see community from their online teaching experiences. In this paper, we characterize the presence and nature of the community that instructors see in their distance education classes. We highlighted the strategies instructors used in an online environment that support both synchronous and asynchronous collaborative learning. Design implications for sociable learning environments are discussed.

## Introduction

With advances in technology and broadened access to educational resources, university degree programs provided via online delivery have drawn considerable attention and effort, and have been rewarded with regular increases in enrollment (Kena et al., 2015) and comparable achievements in terms of learning outcomes (Shachar & Neumann, 2003). However, distance education continues to suffer from the mere fact of *distance*: separation of instructors and students in terms of time and space leads to feelings of being socially removed from the situation (Guo, Tan, & Cheung, 2010). This social isolation is a major contributor to the retention problems common in online education (Carr, 2000; Tinto, 1975). Ashar and Skenes (1993) suggest that learning goals attract adults *to* an online program, but it is the presence of a social environment that makes them *persist*. Meanwhile, productive social interaction is also necessary for collaborative learning (Clegg et al., 2013), whereas finding communities leads to fuller engagement with the class and dialogue (Brown, 2001).

As a step in a larger project studying community-building in distance education, this study investigated whether and how distance education instructors see evidence of community amongst their students. We started with the experiences and views of teachers because they are central agents in orchestrating online collaborative learning experiences. In the following we discuss how instructors perceive student connections with respect to both their class settings and the host institution, and what this might mean for building community online.

## Sociality in online education

Social presence and social interaction are strong predictors of students' learning performance in CSCL (Xing, Kim, & Goggins, 2015). For example, online learners who report higher social belonging also report better learning outcomes (Kizilcec & Halawa, 2015). Furthermore, learners with higher social skills exhibit greater social interaction, which in turn mediates the impact of system functionality on learning (Xing et al., 2015).

Researchers investigating the nature and impacts of social behaviors in CSCL have studied students' perceptions (Shelley, 2008), communication artifacts (Oliver Ferschke, Iris Howley, Gaurav Tomar, Diyi Yang, 2015) and activity logs (Soller, Wiebe, & Lesgold, 2002). For example, discussion posts and chat interactions are often used to enact social behaviors, such as lurking (Mustafaraj & Bu, 2015) and peer support (Appiah-Kubi & Rowland, 2016). To some extent, however, studies of discourse capture only part of the collaboration in an online environment, because the analysis of utterances ignores the larger social context within which a specific interaction takes place. In addition, class-based discourse lasts for only one course, whereas computer-mediated interpersonal connections would typically emerge over longer periods of time (Joseph B Walther, 2002). We seek to add further insight to the question of social behavior in CSCL by first asking what online instructors are observing about student interconnections as well as how they might be promoting this.

As the central coordinators of class-wide activities, online course instructors have the opportunity to promote exchanges among class members, for instance through synchronous or asynchronous interaction during lectures or question and answer, over a considerable period of time, and even across multiple semesters. However, little attention has been paid to whether and how instructor-led activities might affect the sociality of distance education. One counterexample is the reflections of two online professors (McElrath & McDowell, 2008) who offered suggestions about how to create a supportive environment, including course chats, interactive introductions, and student stories as the examples to illustrate the core concepts. Their work was leveraging Brown's framework for community-building in distance education (Brown, 2001); this work suggests a process of making online acquaintances, gaining community acceptance and experiencing camaraderie. However, this

strand of work focused on text-based asynchronous learning in graduate teaching contexts that involved only experienced online instructors who attended to community building carefully. Along with the general expansion of online technology, we are working toward a more expansive view of how instructors with varying experiences and disciplinary background may be doing to enrich online learners' social experience.

### Instructors' beliefs and attitudes about social construction of online learning

Bandura (1993) describes college as a place where teachers' interpretations of students' needs and actions affect the social construction of what is happening in the learning community, including how students perceive their own performance, sense of belonging and expected behaviors. Schools are complex social environments where students share beliefs, fears, values and norms (Hofman, Adriaan Hofman, & Guldmond, 2001), but many of these experiences are guided by the instruction team. For instance, research on distance education has shown that instructors' attitudes towards technology, teaching styles, and technology control will affect what students report about learning outcomes (Sun, Tsai, Finger, Chen, & Yeh, 2008; Webster & Hackley, 1997).

Meanwhile, one pitfall of CSCL environments may be that students' social experiences are taken for granted (Kreijns, Kirschner, & Jochems, 2003). MOOC instructors report that they attend most to performance (Stephens-Martinez, Hearst, & Fox, 2014) and content interaction; social data are of little interest to them (Zinn & Scheuer, 2006). One study of novice online instructors suggests that they overlook the presence or level of community in their classes (Conrad, 2004). Underscoring the inclination to overlook the social psychological aspects of CSCL, Kreijns et al. (2003) argue that possible benefits of *social but non-task* interaction in online learning have been ignored. Thus we see a gap between the expected benefits of social interchange in online learning and the tendency for instructors to ignore such behaviors. This leads us to the following questions:

1. What sorts of social connections do instructors perceive among their distance learners?
2. What techniques do online instructors use to promote such connections among students?
3. What implications do such findings have for design – both of online pedagogy and technology support?

### Method

We conducted an exploratory interview study aimed at unearthing patterns for further research and discussion, not to test or generalize differences among courses or instructors. The semi-structured interviews were carried out at a large northeastern university (MyUni). We recruited the distance instructors through convenience sampling and personal connections within an iSchool at MyUni, while also seeking a relatively diverse sample with respect to online teaching experience, subject areas and teaching styles. We analyzed the interviews through inductive thematic analysis (Braun & Clarke, 2006; Strauss & Corbin, 1998). During the data collection, both authors took notes and discussed emerging themes after each interview. Once we did not discover new themes, we concluded that we had reached the point of theoretical saturation and stopped recruiting instructors (n=11).

### Participants

We interviewed 11 instructors in the spring of 2016; all teach both in residence and online at MyUni, with a range of prior teaching experience (Table 1). Eight are American; two are from outside the United States. Five are females. The courses taught online by these instructors mirror what they teach in residence; our organization within MyUni has an interdisciplinary education mission, so the courses intermix a wide range of disciplines related to the information sciences, including computer programming, application and database design, security policies and laws, and enterprise architecture. Most of the online courses were for undergraduates; a few instructors teach both graduate (e.g., professional masters) and undergraduate courses.

### Interview process

We began each interview with questions about prior teaching experience (e.g. courses, class size, how long online). We next asked for general reflections about connections among online students in three different social contexts: a) among students in the class; b) between the students and the instructor; and c) with MyUni. We focused primarily on online teaching, but at the end asked for a comparison of online and residential teaching. In the process of gathering data, instructors often interleaved comments about their own connections to students with those that the students have with each other, without explicitly placing themselves as "outsiders" to the student milieu; this is not surprising given that we were asking for personal impressions. Therefore, there is considerable overlap in the first two contexts with respect to the presence of communities. However, perceived connections at the level of MyUni or the overall distance education program were viewed as more distinct; these connections are inherently more abstract, diffuse and to a great extent invisible.

Table 1: Instructors' Teaching Profiles

Pseudonym	Teaching years (online/general)	Recent Online Class(es)	Class size
Lily	3/26	Undergraduate intermediate course in human-centered design	35~57
Kyle	3/24	Undergraduate intermediate course in programming	N/A
Matthew	3/22	Undergraduate introductory course & senior technical course in information security	50
Rebecca	2/20	Graduate project course in enterprise architecture	44
George	7/16	Undergraduate intermediate course in databases, project management, and human-centered design class	50
Eric	3/15	Undergrad introductory class in information security; graduate capstone class	50 undergrads, 20 graduates
Scott	1/15	Undergraduate introductory course on programming	35~55
Kristen	3/6	Intermediate undergraduate classes in statistical methods	50~60
Amy	1/5	Undergraduate intermediate course in information security	50
Max	3/4	Undergraduate introductory and senior-level courses in information security concepts and cyber law	50~60
Julie	1/1	Undergraduate introductory course in information security; graduate course in social network analysis	50 undergrads, 20 graduates

*Pseudonyms are chosen to signify gender; instructor profiles are ordered roughly by teaching experience.*

We recognize that the instructors' comments about student community serve only as indirect evidence, but no instructors found the questions odd and were quite able to answer, often in considerable detail. We also inquired about strategies and technology being used; student data they can access or would like to access; comparison between online and residential instruction; and visions or suggestions to improve sense of community. We interviewed 10 instructors in person; the other participated via a Google Hangout video call. We recorded 9.6 hours of audio data, with an average of 52 minutes per interview session.

## Data analysis

We first used open coding to obtain categories related to our research questions. The first author organized these codes into a table that consisted of code names, code memos defining the codes, and sample quotes. We asked a researcher outside the project to review this table using the method of constant comparison (Hallberg, 2006). Coding issues were resolved by discussion with the first author, with changes to categories as needed. The first and second authors then searched for semantic themes and examined similarities and differences, followed by a pruning of codes seen as irrelevant to the central issue of social connections. We organized the remaining high-level themes — indicators of perceived connections and techniques to foster connection — into a thematic map. Each theme was articulated into the subthemes that form our primary findings. This analytic approach is not intended to yield conclusive arguments, but to provide exploratory insights that can guide further in-depth investigations and inform design thinking.

## What connections do instructors see?

All instructors reported social connections among students, though some offered more examples than others. In this exploratory study, we cannot be certain whether this is due to personal characteristics of the instructor, course design or content, or other course-specific variables. Thus we report a mix of general patterns and specific illustrative examples, in an effort to shed light on whether and how instructors find the presence and degree of students' connections. Instructors shared a wide range of "evidence" when they discussed how their students felt connected to one another. In the discussion of their comments we focus primarily on positive examples, as these help to paint a broad picture of what instructors notice about how their students interact with each other.

One type of evidence cited regarding student community was students' *real-time engagement*. Instructors who reported strong connections among students tended to see their students as highly engaged, for example participating in class-wide online chats. Rebecca once attracted ~40 students to a single evening chat and estimates that a large percentage of her students regularly attend her live sessions (e.g. 27 out of 28). George reported a similar success: *"I invite them to come. Of course, I record everything so that they can watch it at a later time, but a significant number of them do come. That's part of building some of this community, too."*

Instructors also reported that students' real-time engagement with one another could enrich the live session because their shared discussion might uncover some hidden knowledge gap that instructors would not otherwise know to address. The result is a benefit for a broader audience that includes not only the witnesses, but also the students who may not have been there but viewed the recorded session later.

*Someone asked one question about the next assignment... It was an interesting question, and I explained it. The other student said, "thank you, Nelson, for asking this question. It was very good and helped me understand it." It is hard to know what students want to know, but the live session is an opportunity for real students asking real questions...It is not just my presentation; it has the student's question. I think it enriched the session a lot. (Kyle)*

Participating regularly in synchronous sessions gives students a chance to find out more about one another and feel more connection. Rebecca offers many such sessions in her online teaching, and in one class students displayed a strong attachment to the class at the end of the semester as evidenced by *reluctance to end their semester together*: "I have students that say: 'It's Tuesday night and there's no class, I really looked forward to class'. There was real camaraderie... I think they really bonded. I think there are some good connections that... some good networking connections that have been made." (Rebecca)

In contrast, Scott shared negative results for attempts to engage online students in real-time sessions. Reacting to low attendance rates (less than 5%) at "live with the professor" events, he questioned the need for fostering community, wondering whether people prefer to focus on connections in their real life settings:

*I provide some fairly rudimentary mechanisms to be more connected, you know discussion forums, online live sessions, and they, and both have been, um very sparsely attended. So, that's my first reaction is that maybe the whole nature of, of at least some online learning is, 'you know I have enough community, at work. I have enough community with my family. Um, I don't want any more community with my online courses' (Scott)*

Some instructors reported students' efforts in sustaining or strengthening interpersonal ties as evidence of student connectivity. Examples of such efforts for *continued collaboration* included requests to join a team with someone; calling one another by name in a post; or referencing previous shared experiences. They also observed that students inquired about upcoming classes and anticipated future interactions: "I've noticed that there can be almost cliques among students, where students that know each other very well and tend to take their classes in the same order, take them the same semester. They know each other already." (Lily)

Interwoven within the normal rhythm of classwork, instructors observed evidence of *mutual support groups*, where student ties include emotional connections. Sometimes these social ties are born of team-based collaborations, but they seem to go beyond assigned projects: team members cover for others who may be suffering from personal issues at the cost of doing more than their own share of the work. At a broader scope, online students root for one another through hardships from their encounters in an online class setting, sending encouragement that addresses worries expressed in someone else's message. For example, Julie shared this heartening support offered in response to a self-deprecating post: "A lot of people were rooting for him, 'I think you'll do good, I hope you'll get a lot out of this degree.'"

Beyond the context of a class, students also seem to bond around and benefit from ties to a host university (Zhang, Jiang, & Carroll, 2012). We asked instructors whether and how their students feel connected with MyUni. A majority (9 of 11) held a positive impression of students' feelings of connection to MyUni, although not all offered evidence. When it comes to institutional bonds, some instructors felt that online students may even feel more *institutional pride* than their residential counterparts. Eric said that his online students use Reddit (a popular forum) to publicize useful online offerings from MyUni; other students use the university mascot as a profile picture, conveying a tie with MyUni; still others drive over one thousand miles to visit the physical campus or attend commencement, suggesting that "they are very proud of" the institution (Eric).

Finally, George shared his own piece of evidence for connections among students at the university level: some of his online students have taken the initiative to create *student clubs*, reflecting an urge to experience social ties that extend beyond a class and are similar to those built by residential student populations: "I'm also involved with the MyUni Online Ed club. I'm just an adviser, so once again...I let them have their club. I'm just an adviser. They have a lot of pride in being MyUni students. Overall, I get a sense of that". (George)

## What techniques may help to build community?

After asking the instructors about their perceptions of how student connections are present in their online classes, we asked what they do that may contribute to a sense of community. Not surprisingly, *public introductions*

are a common technique for increasing the visibility of information about online peers (McElrath & McDowell, 2008); some of our participants appropriated this as a socializing activity to find compatible team members. When used in combination with self-organizing their teams, students are encouraged to purposefully read information about their peers to select ideal collaborators. The self-organizing orientation instigates a social process in which students glean knowledge about their peers, find commonality and other implicit ties, and may thus build feelings of trust and community more quickly: *“As part of the discussion that takes place in this class introduction, I do suggest that they talk about their expectations for teammates and their general idea of their availability to meet as a team, if they're available in the evenings or weekends or whatever. That's a very important part of any {deleted} class, is having a good team.”* (Lily)

Some instructors award points for replies to peers' introductions, but students also exhibit voluntary interest in engaging this way. As one instructor explained, social exchanges that are oriented towards knowing more about each other seem to evoke reciprocal sharing, perhaps making them feel good about themselves and others: *“I require them to comment on each other's posts, but you could just see that they enjoyed it, that they were curious about what others were doing ... When they see the benefits of getting comments, there's some generosity in the way that they are giving them out. It's this culture of giving and receiving....”* (Julie).

Another technique for raising students' visibility within the class relates to the work that they produce. For example, Rebecca **publicizes bits of student assignments** in the shared forum, inviting responses from the others. Later, she reviews the responses and generates themes as fuel for a live session, where a high percentage of students attend and discuss the topics. This reuse of content promotes attention to peer contributions; students regularly follow up on others' thoughts and experiences, often voicing gratitude that they are learning new solutions from peers who are working in different corporate settings. Similar strategies are employed by Kyle, who uses bits of student programs to demonstrate effective and ineffective coding techniques. Publicizing students' work products not only makes the online peers more socially visible (i.e. as seen in their work products) but also invites students to make comparisons to and reflect on their own work products.

Some instructors told us that they work to **maintain a teaching presence** to promote feelings of connection throughout their classes. For example, they hold one-to-many broadcast sessions to gather and attend to students' needs, or just offer general availability: *“I think the first thing to make sure that they understand that you're a real person ... I don't like doing pre-recording stuff. I like the live thing. I went to the live things more. I think that's the teacher connection, the teaching presence that helps students to relate to me.”* (George)

We found two related approaches for enhancing one's teaching presence. Lily wants to make students believe that she is **real**, so she deliberately communicates using a rich modality (e.g. video camera); Max offers to take phone calls at any time, implying a real-time availability. Another approach is for instructors to provide **personal** cues about themselves. For instance, Rebecca injects chit-chat about ongoing university activities or her own life at the start of her live sessions (e.g., while students are “arriving”). This sets up a sort of casual socializing and can evoke student social chit-chat in return: *“... Oh gosh, hold on. Or the cat. Well, they love this. Then they'll say, “say nighty nighty to kitty,” or “how is your dog”? “Is your son on the Xbox”?”* (Rebecca)

Another technique that may enhance an instructor's presence is the **personalization** of their communications to students. George regularly sends out custom messages to individual students: he keeps a student roster and mines information from ongoing class activities to write a personal email for five students he chooses at random each week. His goal is to nurture connection with them:

*What I try to do, and I don't do this all the time, it depends on my workload, my availability, I try to pick out about five students a week that I will send a personalized email by getting personal, but I know a little bit about them and I'll ask them, especially if they're working professionals, I try to connect with them. Actually, that's some of my success has been being able to do that, connecting with some of these adult learners.”* (George)

Personalized communication might be experienced as a social overture that makes students feel cared for and embraced. Regarding student-instructor interaction, one instructor with expertise in sociology shared her happy surprise about reactions to simply asking “How's everything else going” in a class-wide email. Most students dismissed it, but some saw it as an opportunity to “unload”; they appreciated this gesture so much that this instructor was subsequently rated as a student's best professor ever:

*... for some students apparently, a couple of small emails meant a lot. ... Some person reads it as being very casual, another person reads it as in I really care about their life and the way it was going. They write a long email and they start like, “Thank you so much for asking. I have this thing going on. I had another thing going on with this ...”* (Julie)

## Discussion

Our study of online instructors' perceptions on building and sustaining community fills a research gap between the expected benefits of enhancing learners' social experience (Kreijns et al., 2003), and the tendency for it to be neglected in distance education practice (Stephens-Martinez et al., 2014).

### Toward more sociable CSCL environments

Instructors demonstrated their capability to perceive and infer the presence of community based on their direct experiences and observation. In particular, for relationships between instructors and learners, they affirmed students' appreciation for interpersonal outreach though they also conceded that such outreach often comes with a substantial amount of extra time and coordinating efforts. For student interconnections, their one-to-one connections may grow during a class together, as evidenced by the emotional supports they offer or simply calling one another by name. However, even if strong ties are built there are not obvious channels for continuing these relationships, short of starting an extra curricular organization or trying to schedule courses together. At the institution level of community, ties are difficult to discern or celebrate due to the absence of institution-wide activities and interactions, which also partly accounts for why social connections are hard to maintain once a class session is over. Looking across these findings, it is important to support the desires for interpersonal communication initiated and managed by students on their own, instead of depending on what an instructor can organize or lead.

We suggest that more "sociable" CSCL environments for distance education might store a history of students' social interactions or online encounters (e.g. attending the same online activity in the same class), reinforcing students' inclination to strengthen and maintain well-developed interpersonal connections that grow out of their interactions as classmates (we recognize that such histories and resulting access would raise many issues of privacy, but the general concept is valuable nonetheless). More persistent social venues, such as cross-course or cross-program forums or chat rooms, could be used to reinforce and sustain the relationships developed through the camaraderie of shared coursework, which otherwise would dissipate at the end of a course. The continuation of relationships that originate from (virtual) classroom encounters has been proved to be possible over chat even in MOOC settings, in which alumni in programming classes linger as part of an extended learning community to both contribute and learn from others (Nelmarkka & Vihavainen, 2015).

### Helping instructors nurture student engagement

We found that instructors may be promoting feelings of community by publicizing bits of students' work, by injecting more of their "real" selves into their communications and by personalizing the communications they use. Class-level connections seemed to be most strongly created and reinforced through real-time activities that attracted good attendance and high engagement. In particular, instructors can enhance the visibility of implicit social content by publicizing cognitive and socio-emotional knowledge in a larger context. For example, organizing a discussion about someone's submitted project not only brings social attention to that student at that moment, but also invites peer-based evaluation, reflection, and learning. In this sense, instructors may shift interactional resources across different levels of connected learning group units, assembling isolated individual learning processes towards collective knowledge construction at a community level (Stahl & Öner, 2013). Whether it is students' knowledge artifacts (e.g. assignments) or personal background, instructors' public sharing action serves as a social overture that may provoke synthetic discussions or other collaborative learning opportunities.

With respect to CSCL design implications, our interviewees' practices suggest specific enhancements of technology support to attract students socially into the learning subject and with one another. For instance, a synchronous meeting room might include channels to convey contextual cues on what is happening on the other end to make it feel real, such as status updates on specific activities. Text-mining techniques can be applied to self-introductions or other coursework for instructors or students to customize communications with socio-emotional background and predictions. Another design option is to increase the visibility of individual contribution based on instructors' endorsement or teaching assistants' grading outcomes.

### Shooting in the dark?

Taking another perspective, we heard concerns from some instructors about students who are "silent" online or fail to even show up for online activities. As a result, these teachers questioned the benefits of organizing class-wide activities in the sense "if we built it, will they come?" Researchers know that heavy reliance on remote communication may encourage students to filter out the cues that indicate presence or attention (J B Walther & Parks, 2002). Just as there will always be students who "take a back seat" in residential courses, similar students will enroll in online courses. At the same time, computer-mediated communications often leave a historical record so that students who do not attend a live session can engage later on. Nonetheless, lurking behaviors in



online courses are even less visible than sitting in the back of a resident course; they require extra efforts by other students and instructors to assess and respond (Tynan, Ryan, & Lamont-Mills, 2015).

Interestingly, the *time* over which computer-mediated interactions take place, and the *visibility* of personal cues within those records may contribute to the growth of social schemas about others (Liebman & Gergle, 2016); these then can be used to draw inferences about affinity (Joseph B. Walther, 1993). Without records of students' time investment and social presence, instructors and peer learners are less able to respond to individual needs. An individual's silence (e.g. not attending a real-time session) may be due to time constraints or other reasons, but attendees at the session are left in the dark about these factors. We suggest that CSCL design enhancements are needed to convey online students' real-time constraints and subsequent activities (e.g., viewing or commenting a recorded session) in ways that make their presence visible and accountable (Erickson & Kellogg, 2000). As an example of a specific design direction, the sharing of explicit declarations about time availability may help both instructors and peers adopt a more nuanced set of participation expectations.

### Limitations and future work

Because we are part of an interdisciplinary college, our instructor-participants represent a range of course topics, teaching experiences and approaches; even so, this exploratory study was conducted within one college at MyUni, and we call for future research to generalize and extend our exploratory analysis. We also recognize that ours was a descriptive study that relied extensively on instructors' opinions and our own interpretation of the same; we cannot draw firm conclusions about causal factors or mechanisms in building community in online courses. We are now analyzing data from students' reflections about community to triangulate our findings, as well as iteratively designing online tools for evoking and enhancing student community.

### References

- Appiah-Kubi, K., & Rowland, D. (2016). PEER Support In MOOCs: The Role Of Social Presence. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale* (pp. 237–240). CONF, ACM. <http://doi.org/10.1145/2876034.2893423>
- Ashar, H., & Skenes, R. (1993). Can Tinto's student departure model be applied to nontraditional students? *Adult Education Quarterly*, 43(2), 90–100. JOUR.
- Bandura, A. (1993). Perceived Self-Efficacy in Cognitive Development and Functioning. *Educational Psychologist*. [http://doi.org/10.1207/s15326985ep2802\\_3](http://doi.org/10.1207/s15326985ep2802_3)
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. JOUR.
- Brown, R. E. (2001). The process of community-building in distance learning classes. *Journal of Asynchronous Learning Network*, 5(2), 18–35.
- Carr, S. (2000). As distance education comes of age, the challenge is keeping the students. *Chronicle of Higher Education*, A39–A41.
- Clegg, T., Yip, J. C., Ahn, J., Bonsignore, E., Gubbels, M., Lewittes, B., & Rhodes, E. (2013). When face-to-face fails: Opportunities for social media to foster collaborative learning. In *Tenth international conference on computer supported collaborative learning*. CONF, Citeseer.
- Conrad, D. (2004). University instructors' reflections on their first online teaching experiences. *Journal of Asynchronous Learning Network*, 8(2), 31–44.
- Erickson, T., & Kellogg, W. A. (2000). Social translucence: an approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(1), 59–83. JOUR.
- Guo, Z., Tan, F. B., & Cheung, K. (2010). Students' Uses and Gratifications for Using Computer-Mediated Communication Media in Learning Contexts. *Communications of the Association for Information Systems*, 27(20), 339–378.
- Hallberg, L. R. M. (2006). The "core category" of grounded theory: Making constant comparisons. *International Journal of Qualitative Studies on Health and Well-Being*, 1(3), 141–148. <http://doi.org/10.1080/17482620600858399>
- Hofman, R. H., Adriaan Hofman, W. H., & Guldmond, H. (2001). Social context effects on pupils' perception of school. *Learning and Instruction*, 11(3), 171–194. [http://doi.org/10.1016/S0959-4752\(00\)00016-5](http://doi.org/10.1016/S0959-4752(00)00016-5)
- Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., ... Velez, E. D. V. (2015). The Condition of Education 2015. NCES 2015-144. *National Center for Education Statistics*. JOUR.
- Kizilcec, R. F., & Halawa, S. (2015). Attrition and Achievement Gaps in Online Learning. *Proceedings of the Second (2015) ACM Conference on Learning@ Scale - L@S '15*, 57–66.
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-

- supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, 19(3), 335–353. [http://doi.org/10.1016/S0747-5632\(02\)00057-2](http://doi.org/10.1016/S0747-5632(02)00057-2)
- Liebman, N., & Gergle, D. (2016). It's (Not) Simply a Matter of Time: The Relationship Between CMC Cues and Interpersonal Affinity. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 570–581. <http://doi.org/10.1145/2818048.2819945>
- Mcelrath, E., & McDowell, K. (2008). Pedagogical Strategies for Building Community in Graduate Level Distance Education Courses. *Journal of Online Learning and Teaching*, 4(1), 117–127.
- Mustafaraj, E., & Bu, J. (2015). The Visible and Invisible in a MOOC Discussion Forum. *Proceedings of the Second (2015) ACM Conference on Learning @ Scale - L@S '15*, 351–354.
- Nelimarkka, M., & Vihavainen, A. (2015). Alumni & Tenured Participants in MOOCs. *Proceedings of the Second (2015) ACM Conference on Learning @ Scale - L@S '15*, 85–93. <http://doi.org/10.1145/2724660.2724671>
- Oliver Ferschke, Iris Howley, Gaurav Tomar, Diyi Yang, C. R. (2015). Fostering Discussion across Communication Media in Massive Open Online Courses. *Proceedings of the 11th International Conference on Computer Supported Collaborative Learning*, in review.
- Shachar, M., & Neumann, Y. (2003). Differences Between Traditional and Distance Education Academic Performances : A meta- analytic approach. *International Review of Research in Open and Distance Learning*, 4(2).
- Shelley, M. (2008). Considering Students ' Perceptions : The Distance Education Student Satisfaction Model Ismail Sahin. *Educational Technology & Society*, 11, 216–223.
- Soller, A., Wiebe, J., & Lesgold, A. (2002). A machine learning approach to assessing knowledge sharing during collaborative learning activities. *Proceedings of the Conference on Computer Support for Collaborative Learning Foundations for a CSCL Community - CSCL '02*, (1), 128. <http://doi.org/10.3115/1658616.1658635>
- Stephens-Martinez, K., Hearst, M. a., & Fox, A. (2014). Monitoring MOOCs. *Proceedings of the First ACM Conference on Learning @ Scale Conference - L@S '14*, 79–88.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. BOOK, Sage Publications, Inc.
- Sun, P. C., Tsai, R. J., Finger, G., Chen, Y. Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers and Education*, 50(4), 1183–1202. <http://doi.org/10.1016/j.compedu.2006.11.007>
- Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research Winter*, 45(1), 89–125.
- Tynan, B., Ryan, Y., & Lamont-Mills, A. (2015). Examining workload models in online and blended teaching. *British Journal of Educational Technology*, 46(1), 5–15.
- Walther, J. B. (1993). Impression development in computer-mediated interaction. *Western Journal of Communication*, 57(Fall), 381–398.
- Walther, J. B. (2002). Time effects in computer-mediated groups: Past, present, and future. *Distributed Work*, 235–257. JOUR.
- Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filter in: computer mediated communication and relationships. *The Handbook of Interpersonal Communication*, (January 2002), 529–563.
- Webster, J., & Hackley, P. (1997). Teaching effectiveness in technology-mediated distance learning. *Academy of Management Journal*, 40(6), 1282–1309. JOUR.
- Xing, W., Kim, S., & Goggins, S. (2015). Modeling performance in asynchronous CSCL: an exploration of social ability, collective efficacy and social interaction. *Exploring the Material Conditions of Learning: Proceedings of the Computer Supported Collaborative Learning (CSCL 2015), International Society of the Learning Sciences*, 276–283.
- Zhang, S., Jiang, H., & Carroll, J. M. (2012). Babel or Great Wall: Social Media Use Among Chinese Students in the United States. *Proceedings of the 30th ACM International Conference on Design of Communication - SIGDOC '12*, 37.
- Zinn, C., & Scheuer, O. (2006). Getting to know your student in distance learning contexts. *European Conference on Technology Enhanced Learning: Innovative Approaches for Learning and Knowledge Sharing*, 437–451.

# Revealing Interaction Patterns Among Youth in an Online Social Learning Network Using Markov Chain Principles

Sarah Bishara, Jennifer Baltes, Taha Hamid, Taihua Li, Denise C. Nacu, Caitlin K. Martin, Jonathan Gemmell, Chris MacArthur, Daniela Raicu, and Nichole Pinkard  
sbishara44@gmail.com, jenniferbaltes1@gmail.com, tahahamiid@gmail.com, taihua.ray.li@gmail.com, dnacu@depaul.edu, cmartin@digitalyouthnetwork.org, jgemmell@depaul.edu, chrismacarthur@mac.com, draicu@depaul.edu, npinkard@depaul.edu  
DePaul University

**Abstract:** The problem of the digital divide has shifted attention from access to inequities of participation and opportunities to develop 21st century skills in online learning platforms. In this paper, we explore Markov chain principles in a time-based probabilistic graphical approach to analyze a multi-year data set of log data generated by students from one urban middle school and coded using a framework aligned with 21<sup>st</sup> century learning activities. Results showed the efficacy of applying Markov chain principles in helping reveal similar and distinct usage patterns of the learners in this community across different time spans. This work has implications for the design and analysis of online learning platforms and for creating opportunities to help youth build 21<sup>st</sup> century skills using online learning platforms.

## Introduction

The notion of a “digital divide” has shifted from a focus on inequities of access to *equipment* toward one of inequities of access to *opportunities*. There is a recognized “opportunity gap” that characterizes differences in access to learning activities and networks, resulting in a “participation gap” between youth in underserved communities and their more affluent counterparts (Warschauer & Matuchniak, 2010; Watkins, 2011; Lenhart, 2015). Computational and technological learning experiences have been linked to the development of skills and dispositions that are viewed as critical for participation in the 21st century, such as communication, creative production, problem solving, and collaboration (Ito et al., 2009; Barron, Gomez, Pinkard & Martin, 2014). A concerning consequence of this existing gap, then, is that certain populations of youth have fewer opportunities to develop these skills necessary for productive life and citizenship in today's world (Levy & Murnane, 2012).

Online social learning networks have affordances for supporting 21st century learning in ways that can potentially bridge inequities by offering learning opportunities, resources, and social supports beyond the physical boundaries that demarcate underserved populations (AITF, 2014; Hamid, Waycott, Kurnia, & Chang, 2015; Jenkins, Purushotma, Weigel, Clinton, & Robinson, 2009). As teachers in formal and informal learning settings are increasingly willing to make use of such systems (MMS Education, 2012), and as schools and districts increasingly adopt and require the use of online platforms (Burch & Good, 2014), there is anticipation that the gap will be minimized. These systems generate massive use data, and there is excitement around the potential of harnessing this data to reveal insights about learning and to design interventions that can level the playing field. *Who is participating? What resources are being used? Who is supporting whom?* To make progress on understanding learning patterns in user logs generated from online social learning networks, collaborative work is necessary to bring together learning theory, deep contextual understanding, and mathematical algorithms (Bienkowski, Feng, & Means, 2012; Pea, 2013; Siemens, 2012).

This study joins learning sciences and data science methods to explore patterns in user trace log data from students in a public urban middle school using an online social learning network over multiple years. We used a sequential pattern data mining technique, Markov chains, to explore student actions as opportunities for building 21st century skills. We asked, *how can Markov chain principles be used to reveal patterns of online activity over time in the areas of creative production, social learning, and self-directed learning?* In addressing this question, we aim to inform research methodologies used to study online learning platforms and to identify use patterns that can be later explored for their potential to provide evidence of 21st century learning. While attention to 21st century skills is increasing, we don't yet have practical ways to measure them and to understand how experiences using online social learning networks may contribute to building those skills. To address the “participation gap,” developing ways to reveal how youth may be using these systems differently is needed.

Markov chains work on the assumption that the probability of the next action taken is exclusively dependent on the current action. For this study, a Markov chain model represents each action as a node in a network graph and any existent relationship between two actions as an edge connecting the nodes corresponding

to those actions. Probabilities are associated with each edge to encode the strength of the relationships based on the temporal log data over a period of 28 months.

## Related work

We contextualize this work within ongoing efforts to design and study socio-technical systems that help youth develop skills and competencies by interacting with peers and adult educators around the creation, sharing, and communication of digital artifacts. We define these systems as online social learning networks, web-based environments that use features of social network sites and learning management systems to support and develop an online learning community and the individual participants within it (Martin, Nacu & Pinkard, 2016). Youth participation in online networks has been linked to fostering 21st century skills such as self-directed learning, creativity, and communication (Hamid, et al., 2015; Ito et al. 2009; Jenkins et al. 2009). A focus on these types of skills is becoming increasingly emphasized in K-12 standards striving to prepare youth for future workforce needs (Common Core State Standards Initiative, 2010; NGSS, 2013; Pellegrino & Hilton, 2013).

Inequities exist in terms of who has the opportunities to participate in programs and activities that can build technological competencies related to 21st century learning. While youth are increasingly participating in social network sites (Blair, Millard, & Woolard, 2015; Lenhart, 2015; Watkins, 2011), studies have revealed that contributors of online content, in general, are a small subset of the population using technical systems (Rideout, 2015), and that this subset is not representative of the larger population (Glott, Schmidt, & Ghosh, 2010). Youth from areas with fewer socioeconomic resources are especially underserved, demonstrating inequities in more sophisticated forms of participation such as interest-driven practices involving creating, sharing, communicating, and critiquing (Margolis, Estrella, Goode, Holme, & Nao, 2010; Warschauer & Matuchniak, 2010).

In the last five years, a growing body of research studies have applied data mining techniques to examine log data generated by online learning platforms. These data have been analyzed in the context of user-to-user social interaction (Cela, Sicilia, & Sánchez, 2015; Xu, 2011), user-to-platform interaction (Kardan, Roll, & Conati, 2014), and user-to-content learning behaviors (Jeong et al., 2008). Some of these studies focus their analysis on Markov model chains (Faucon, Kidzinski, Dillenbourg, 2016; Marques & Bello, 2011). Rodríguez and Boyer (2015) analyzed probabilistic chains to compare problem-solving approaches by individual versus collaborative users. Within the educational domain, some prior studies have applied Markov chains to improve intelligent tutoring systems while others have used students' answers to multiplication problems to generate Markov chains that informed a question recommendation system (Taragari, Saranti, Ebner, & Schön, 2014).

## Research context

This work is part of a multi-year study examining interactions among youth and adults in online social learning networks. We used data from an urban middle school ELA (English Language Arts) teacher and his 54 students. The teacher remained their ELA teacher for three academic years. Students in this study were in 6th grade in the first year, and, by the third year, were 8th graders. The K-8 school draws the majority of students from a predominantly Latino community: 91% of students are Latino, 8% black, and 1% white. In the school, almost half of students (43.3%) are classified as having limited English (English Language Learners) and 95% are classified as coming from low income households. The student sample in this work reflected the larger school demographics, with 47.2% girls, 52.8% boys, 89% self-reporting as Latino, 6% black, 2% white, and 2% Chinese. Over three quarters of students in our sample (83%) reported being part of Spanish-speaking households.

The study involved iRemix, which has an interface and functionality similar to popular online social networks (Barron et al., 2014; Martin, Nacu & Pinkard, 2016; Zywicki, Richards & Gomez, 2011). Students and teachers can share digital artifacts such as blog posts, photos, and videos. They can respond to posts using comments and reactions, browse activity through a feed, edit profile pages, and link to peers. iRemix is intended to support the development of 21st century skills through production, reflection, critique, and revision. While the system is meant to be youth-driven and reflects youth interests, teachers can post challenges to prompt activities.

## Method

### Log data

Logs of student actions were pulled for the period from 1/1/2014 through 4/15/2016, spanning three academic years. Each observation in the data generates a row with the student's user ID, the associated action code, the timestamp, and other information not covered in this paper (Nacu, Martin, Schutzenhofer & Pinkard, 2016).

The framework for coding student actions builds on prior work which conceptualizes the activities the system was designed to support as opportunities for 21st century online learning. Specifically, the learning

opportunities on iRemix were categorized into three themes that reflect the platform’s learning goals: *creative production*, *self-directed learning*, and *social learning* (Martin, Nacu & Pinkard, 2016). *Creative production* involves developing identity as a creator, creating media, and revising work. *Self-directed learning* involves using online resources, monitoring one’s progress, and seeking support and learning opportunities. *Social learning* involves communicating with others and observing the work of peers. For example, sending a message to another user, commenting, and posting a reaction on another user’s work are considered *communicate* actions, and relate to the theme of *social learning*. Table 1 shows the action codes present in the dataset (with abbreviations), mapped onto the three focal themes of 21st century learning opportunities. After extensive data cleaning and preparation, there were 32,895 rows of coded data covering 10,193 sessions.

Table 1: Student action codes logged from user trace activities on iRemix

Abbreviation	Action Code	Description
<b>Creative Production</b>		
<i>per</i>	personalize	Change user profile information or avatar picture
<i>cre</i>	create	Submit original work (media artifact)
<i>sha</i>	sharing	Share a media artifact
<i>bw</i>	begin work	Start a set of scaffolded learning activities
<i>eow</i>	edit own work	Edit a submitted artifact created by self
<i>gb</i>	get badge	Complete set of scaffolded learning activities
<i>row</i>	review own work	View artifact created by self
<i>cw</i>	complete work	Submit a final draft of project
<b>Self-Directed Learning</b>		
<i>vr</i>	view resource	View a resource associated with a learning activity
<i>vpa</i>	view potential activity	View a description of a potential learning activity or set of activities
<b>Social Learning</b>		
<i>com</i>	communicate	Send a message, leave a comment or reaction on work, or reply to forum
<i>cri</i>	critique	Provide star-ratings for submitted work
<i>vpo</i>	view profile of others	View a user profile of another user
<i>vwo</i>	view work of others	View work submitted by another user
<i>jc</i>	join community	Join an interest group
<i>vc</i>	view community	View content of interest group
<i>inv</i>	invite	Invite other users to an interest group
<i>qc</i>	quit community	Leave an interest group
<b>Other</b>		
<i>lgi</i>	login	Login to system
<i>lgo</i>	logout	Artificial action indicating end of session

## Constructing Markov models

With sequential coded data, a transition matrix can be computed. Each entry,  $T(i,j)$ , in the transition matrix represents the probability of the action  $A_i$  to the action  $A_j$ . For example,  $T(row|eow) = 0.36$  shows that the percentage of times a student who has just completed a “review own work” (*row*) action will complete an edit own work (*eow*) action next is 0.36. Likewise,  $T(per|per) = 0.13$  shows that a student who has just completed a *personalize* (*per*) action is likely to complete another *personalize* action 13 percent of the time. Four transition matrices were created for this study: one for the full dataset spanning all three years, and one for each of the three academic years. The size of each matrix was (20x20), representing the twenty action codes.

Building on transition matrices, we constructed Markov chain models. A Markov chain is a stochastic process where a future state prediction is dependent only on the present state and does not take into account any information about the previous states. Conceptually, the Markov chain can be perceived as users moving from one state to another, and it can be visualized by network graphs. Due to the abundance of action-to-action activity that occurred in the logs, we only include edges representing a probability of 1% or more to increase readability of the network graphs. The thickness of the node’s outline corresponds to the frequency of the action, highlighting the more frequent items. The thickness of the edges correlates to the probability of the action-to-action movement.

## Results

Here, we first discuss how the Markov models can be interpreted and present insights from a network graph aggregating students' online activities over the full 28 months. Then, a year-by-year analysis follows.

### Multi-year Markov model

The nodes of this Markov model represent the action types (see Table 1). The connecting edges between the nodes indicate both the direction and magnitude of the relationship between actions. For example, at the bottom right side of Figure 1, an edge exists from *eow* to *row* located near the left side of the graph, representing the 99% probability of students moving from editing their own work to reviewing their own work. In this case, the strong correlation is due to the internal workflow of iRemix, which automatically shows a user the work they have just edited. In comparison, *row* connects to *eow* with a thinner line, indicating a lower probability (36%) of a user returning to edit their work after reviewing it. Though the model shows any probability greater than 1%, actions occurring with a probability 10% or greater are significant given the number of possible transitions.

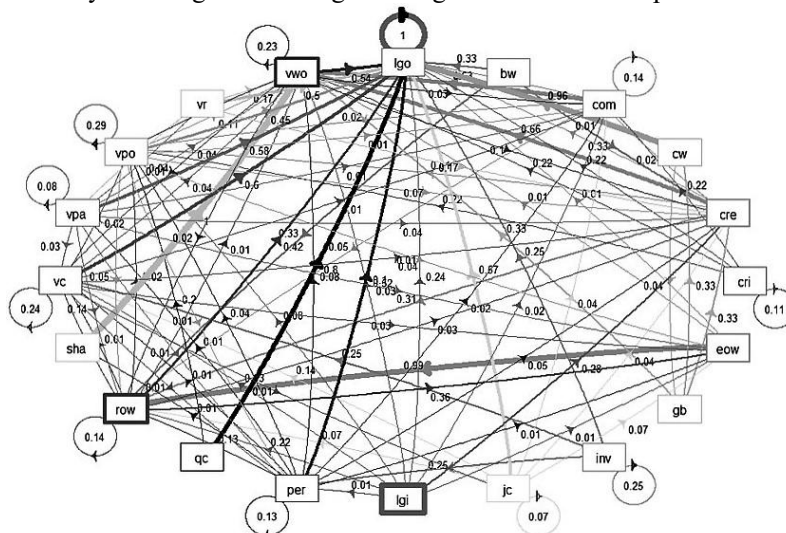


Figure 1: Multi-Year Network Graph.

Nodes can connect to themselves. For example, *row* transitions to itself (shown in the circle overlapping with the node) 14% of the time, indicating the likelihood of students reviewing their own work twice in a row. There are two unique states on the network graph: logging in (*lgi*) and logging out (*lgo*). *lgi* is the beginning state of every user sequence. Therefore, no other nodes point toward it. *lgo* is the absorbing state for the model; ultimately all users will end by logging out, thus no edges originate from it.

A key advantage of the Markov model is its capacity to describe the sum of actions taken over the entire program. Although there is an enormous amount of information captured by the graph, next we will focus on several of the larger patterns as they relate to our research questions.

### Opportunities for creative production

One of the most notable transitions between nodes occurs between *row* and *eow*. In iRemix, editing actions include both altering the content of a previously submitted post as well as deleting the post. The *row* actions represent a student looking over their own post or looking at the assignment description. The system automatically directs students who have submitted work to review it, thus creating an extremely strong edge, 99% from *eow* to *row*. The opposite pathway is not automatic and occurs 36% of the time, suggesting students may be editing their work, viewing it, and then returning to the editing options. While this cyclical pattern of editing and reviewing would not be surprising to educators, this method helps to reveal that students are using the system in a way that is related to creative production of work, and reflects a type of desired activity. While not shown in the graph, this type of analysis can lead to further examination about which students are exhibiting this pattern and which are not. From an equity perspective, it is important to discern if and how students are participating online as creative producers.

*Cre* also frequently leads to *vwo* with a probability of 22%. This indicates a pattern of students posting their work then comparing their own to those posted by their peers. Qualitative studies have shown that looking at the work of others is often associated with generating new ideas and pushing new possibilities for the quality of submitted content (Ito et al., 2009; Barron et al., 2014).

### Opportunities for social learning

One finding related to *social learning* is the connection between communicating and exploring the work of others in the community. Figure 2 shows that, across the 28 months, 62% of *com* actions end with a student transitioning to *vwo*. In iRemix, viewing others' work includes looking at another user's portfolio of work or individual posted artifact. Communicating can take many forms, including sending a message, responding to a user's work with a reaction, or contributing to a forum. This suggests that after a communication-related move, a student is very likely to keep exploring work submitted by others. This idea of an exploratory activity pattern is supported by the finding that 23% of *vwo* actions result in another *vwo* action. This pattern is encouraging as it reveals a social learning pattern in which communication actions lead to potential for learning by viewing work submitted by others. *Vwo* is also highly correlated to *gb* (33% probability) indicating that, after finishing a challenge, students are either looking for comparison pieces on the current challenge, or they are seeking inspiration for a new task.

In contrast, *vwo* to *com* does not have as strong of a connection in our data, as this chain of actions is only present for 3% of the pathways stemming from *vwo*. As a way to understand the potential for social learning observed in the data, we may interpret this finding to suggest that even when the work of others is viewed, users are not likely to provide feedback through comments or reactions, thereby missing learning opportunities for both the viewer and creator of posted work. While communication happened during the exploratory activity patterns, as evidenced above, communication actions, in general, were much less frequent in this dataset than other, more passive activity on the system. The Markov model easily revealed this disconnect in a way that has encouraged program educators and designers to think about ways to support youth to engage more in a communicative interaction around the work after viewing it.

One of the most telling sequences is between actions and themselves. These are referred to as "sticky states," actions which frequently lead to another instance of the same action. *Com*, *cri*, *inv*, *per*, *vc*, *vpo*, and *vwo* are all social actions. These are also states which have a self-referential probability of greater than 10%. This suggests that once engaged in social activity, users are likely to continue their social action. Although the connections between social actions are not substantial individually, edges originating from social actions are numerous. For example, *vpo* has a greater than 1% probability of leading to one of five social actions (including itself). These chains indicate that once social action has been initiated, it is likely to continue to another form.

### Opportunities for self-directed learning

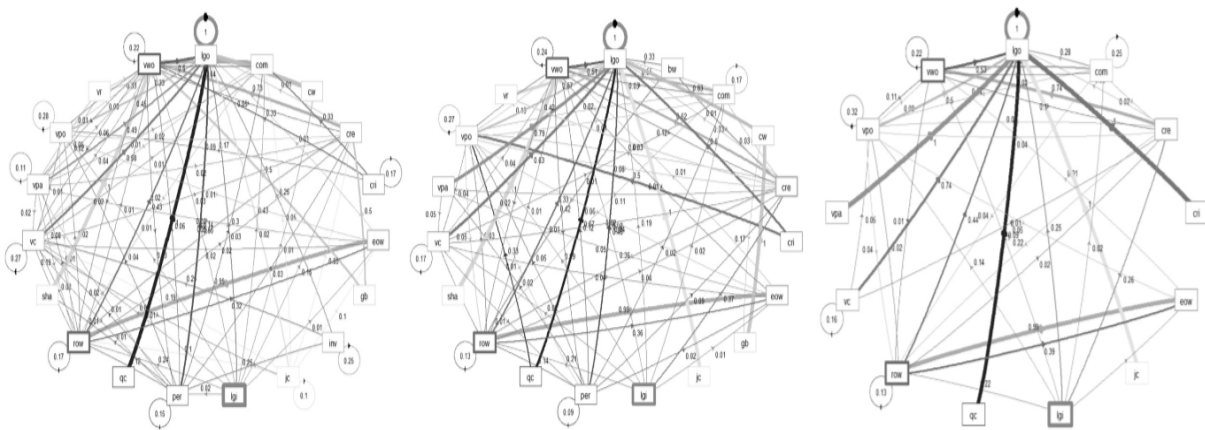
One goal for iRemix is to foster a space in which students have the social support and access to resources that allow students to take initiative in pursuing interests. In this dataset, we observed two actions aligned with this 21st century learning theme of *self-directed learning*: *vr* and *vpa*. We found that *vr* led to *cre*, *vpa* or *vwo* 16.7% of the time for these actions, suggesting that after viewing a resource associated with a learning activity, students submitted a work artifact, viewed a description of a potential learning activity or set of activities, or looked at others' work. Encouragingly, this finding highlights connections among actions cutting across the three 21st century learning themes. These observed connections among using resources, creating artifacts, viewing the work of peers, and exploring what challenges to take on next would indicate a sophisticated form of participation in the online system that is desired over other, more passive forms.

## Year-by-year Markov models

The Markov models in Figure 2 each represent a single year of sequential data and provide the opportunity to evaluate how activity on iRemix changed over time. For Year 1, the system was adopted in the winter of the first year and thus accounts for approximately 6 months of the school year. For Year 3, data was pulled at the end of the winter; thus, the third year has approximately 4 months' worth of activity. However, as Markov models deal with transition probabilities and not counts, the models can still be compared with one another.

Many of the same observations from the multi-year Markov model (see Figure 1) are also evident in the yearly models. A strong connection is still observed between *eow* and *row*. Likewise, *cre* actions often transition to *vpo* across all three years, as they do in the multi-year model.

One apparent difference in use patterns of students over time is that by year three there were fewer actions present in the model. For example, *per* does not occur at all in the third year. An action is coded as *personalize* when a user updates information in his/her profile or posts a status to be shared with the community. This pattern can be explained by the notion that users would be inclined to spend more time personalizing their profile as a community is first developing. By the eighth grade in Year 3, it is likely that personas are more solidly established through face-to-face and online interactions, reducing the need for updating online profiles. From an educator's point of view, being able to discern changes in personalization may trigger intervention in which students may be encouraged to update their profiles to reflect their changing interests and identities within the learning community.



**Figure 2.** Network graphs by academic year: Year 1, Year 2, Year 3 (left to right).

Activities reflective of *social learning* such as *vwo*, *vpo*, and *vc* played out similarly, whereas *inv*, an action logged when a user invited another user to a community, appears in the Year 1 graph but not in the subsequent year graphs. This may indicate the developmental change of the community; creating and participating in interest groups may have had novelty and momentum in Year 1, but later dropped off as the community evolved. Again, this suggests an opportunity for the teacher to encourage students to engage in interest-based activities.

The social trends visible in the network graph evolved over the course of the program. In Year 1, 97% of *com* actions led to *vwo*. In Year 2, students performed a wider variety of actions, including *cre*, *row*, *vpo*, and *lgo*. *Vwo* remained the most likely transition at 54%, while another *com* action occurred 17% of the time. By Year 3, *vwo* was likely in 37.5% of transactions, while the self-referential probability increased to 25%. *Cre*, *row*, *vpo*, and *lgo* were also visible, with *lgo* having a probability of 27%. These findings add another dimension to the patterns of social learning visible in the multi-year Markov model; although social actions frequently follow one another, the learning community used communicate actions differently across the three years.

## Discussion

The purpose of this study was to explore how Markov chain principles could be used to reveal patterns of students' online activity over time in the areas of *creative production*, *social learning*, and *self-directed learning*. Using a coding framework and Markov chain principles to discern frequent sequences of actions, this approach was useful for visualizing students' use of platform features with respect to goals to foster 21<sup>st</sup> century learning activities. We explored this approach with the notion that being able to understand how youth are making use of online social learning platforms is needed in order to attend to equity in terms of the quality of participation. Indeed, by analyzing at the level of a user session (rather than by single action, or by week, for example), the network graphs efficiently captured episodes of student online interaction in terms of types of desired activities.

In the three-year Markov model, we observed cyclical movement between editing and viewing own work, indicating the theme of *creative production*. We also saw a connection between *creative production* and *social learning* through the relationship between creation actions and viewing work own work and the work of others. *Social learning* occurred frequently with *creative production*, suggesting an underlying feedback loop between these two types of activities. Taken as a whole, the findings revealed by the Markov graphs show that students using iRemix were engaged in activity that was desired by the teacher: creative production characterized by iteration and supported by social interactions among the community. While the results do not present definitive evidence that learning actually occurred, they do demonstrate that the online learning platform provided opportunities to build 21<sup>st</sup> century learning skills for students (Reich, Murnane & Willett, 2012). The methodology presented here suggests one way to reveal those opportunities.

The network graphs produced for each year provided a concise way to examine how this particular learning community changed over the three year period. In particular, the graphs helped to show the continuity of core activities that were prompted and encouraged by the teacher including the posting of work, revising, and communicating, as well as types of actions that tended to drop off over time. For designers, this type of analysis can help generate insights that can inform how desired learning cycles can be strengthened by making the connections between these activities more obvious to students. The probabilities generated by these models have revealed existing patterns present in student work cycles, but providing personalized prompts to users could increase the likelihood of target behavior. Seeing how features are used or not over time may also suggest the need for teacher prompting to use desired features or may point to potential usability or logging issues that need



to be further investigated. In terms of implications for developing theory, examining differences across time can give researchers a developmental view of a changing learning community, and the evolution of how the teacher used iRemix as part of his class over the three academic years. More research analyzing longitudinal log data using this type of approach is needed to compare high-level patterns which may provide a view on how a community of learners evolves and matures. While not addressed in this paper, our qualitative data (e.g., field observations, interviews, and surveys) can aid interpretation of these long-term trends.

The method explored here resulted in graphs that could be used by researchers to study patterns of use; however, such visualizations are likely not usable for educators. Additional research and design with educators is needed to better communicate insights in a way that integrates with practice. Furthermore, although the network graph is useful for determining global, longitudinal patterns, it does not provide information about individual students' use in a way that is helpful to the classroom teacher. The current visualizations, then, represent an initial step to reveal patterns which can establish a baseline. Future applications will allow identification of groups that are distinguished by their use patterns in order to provide the teacher with recommendations for action.

One of the main limitations of log data analysis is accurately interpreting user intent. Although some of the interpretations may be accurate for a group of students, further investigation is required to more deeply understand why students are exhibiting certain sequences of actions in the online platform. A key implication of this work is the need for diverse research collaborations to connect an understanding of intent with analytic methods used. Educators, learning scientists, data analysts, and software designers are only a subset of the types of researchers whose insights will provide a well-rounded interpretation of these types of data.

## Conclusions

In this paper, we examined how Markov model principles could be applied to student action log data from the iRemix platform to visualize and better understand online behavior patterns with respect to 21st century learning opportunities. Network graphs generated based on the proposed Markov chain approach quantified the strengths of any existent pair of actions over different time spans, revealing distinctions and similarities in online behavior data. This approach is encouraging, as it shows potential for generating insights that can be useful for educators, designers, and researchers. Particularly for efforts aimed at addressing the "participation gap" characterized by inequities in terms of the development of 21<sup>st</sup> century skills among youth, this method suggests an approach that can reveal the nature of online interactions in a holistic and concise manner. Future work will explore new visualization techniques for predictive models of students' online interactions using Markov principles.

## References

- AITF on Learning and the Internet. (2014). *Learner at the center of a networked world*. Washington, DC: The Aspen Institute.
- Barron, B., Gomez, K., Pinkard, N., & Martin, C. K. (2014). *Digital Youth Network: Cultivating digital media citizenship in urban communities*. Cambridge, MA: MIT Press.
- Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *US Dept. of Education, Office of Educational Technology*, 1-57.
- Blair, R., Millard, D., & Woollard, J. (2015). Perceptions of school children of using social media for learning. *International Journal on E-Learning*.
- Burch, P., & Good, A. G. (2014). *Equal scrutiny: Privatization and accountability in digital education*. Harvard Education Press.
- Cela, K. L., Sicilia, M. Á., & Sánchez, S. (2015). Social network analysis in e-learning environments: A preliminary systematic review. *Educational Psychology Review*, 27(1), 219-246.
- Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics*. Wash., DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.
- Faucou, L., Kidzinski, L., & Dillenbourg, P. (2016). Semi-Markov model for simulating MOOC students. In Barnes, T., Chi, M., & Feng, M. (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining*. Paper presented at the 9<sup>th</sup> International Conference on Educational Data Mining.
- Glott, R., Schmidt, P., & Ghosh, R. (2010). *Wikipedia survey—overview of results*. United Nations University: Collaborative Creativity Group.
- Hamid, S., Waycott, J., Kurnia, S., & Chang, S. (2015). Understanding students' perceptions of the benefits of online social networking use for teaching and learning. *The Internet and Higher Education*, 26, 1-9.
- Ito, M., Baumer, S., Bittanti, M., Cody, R., Stephenson, B. H., Horst, H. A., Perkel, D., et al. (2009). *Hanging out, messing around, and geeking out: Kids living and learning with new media*. MIT press.
- Jenkins, H., Purushotma, R., Weigel, M., Clinton, K., & Robison, A. J. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century*. MIT Press.

- Jeong, H., Gupta, A., Roscoe, R., Wagster, J., Biswas, G., & Schwartz, D. (2008, June). Using hidden Markov models to characterize student behaviors in learning-by-teaching environments. In *International Conference on Intelligent Tutoring Systems* (pp. 614-625). Springer Berlin Heidelberg.
- Kardan, S., Roll, I., & Conati, C. (2014, June). The usefulness of log based clustering in a complex simulation environment. In Trausan-Matu, S., Boyer, K., Crosby, M., Panourgia, K. (Eds.), *Intelligent Tutoring Systems*. Paper presented at the 12<sup>th</sup> International Conference on Intelligent Tutoring Systems (pp. 168-177). Springer International Publishing.
- Lenhart, A. (2015). *Teens, social media & technology overview 2015*. Washington, DC.
- Levy, F., & Murnane, R. J. (2012). *The new division of labor: How computers are creating the next job market*. Princeton University Press.
- Margolis, J., Estrella, R., Goode, J., Holme, J. J., & Nao, K. (2010). *Stuck in the shallow end: Education, race, and computing*. Cambridge, MA: MIT Press.
- Marques, A., & Belo, O. (2011). *Discovering student web usage profiles using Markov chains*. *Electronic Journal of e-Learning*, 9(1), 63-74.
- Martin, C.K., Nacu, D., & Pinkard, N. (2016). Revealing Opportunities for 21st Century Learning: An Approach to Interpreting User Trace Log Data. *Journal of Learning Analytics*, 3(2): 37-87.
- Mitchell, C. M., Boyer, K. E., & Lester, J. C. (2013, July). A Markov decision process model of tutorial intervention in task-oriented dialogue. In *International Conference on Artificial Intelligence in Education* (pp. 828-831). Springer Berlin Heidelberg.
- MMS Education. (2012). Survey of K-12 educators on social networking, online communities, and web 2.0 tools. Retrieved from <http://www.edweb.net/fimages/op/reports/>
- Nacu, D., Martin, C. K., Schutzenhofer, M., & Pinkard, N. (2016). Beyond traditional metrics: Using automated log coding to understand 21st century learning online. *Proceedings of the 3rd ACM Conference on Learning@Scale*, 197-200.
- NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.
- Pea, R. (2013, February). Learning sciences and learning analytics: Time for a marriage. *EDUCAUSE Learning Initiative*. Denver, CO.
- Pellegrino, J. W., & Hilton, M. L. (Eds.). (2013). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. National Academies Press.
- Reich, J., Murnane, R., & Willett, J. (2012). The state of wiki usage in US K-12 schools leveraging web 2.0 data warehouses to assess quality and equity in online learning environments. *Ed. Researcher*, 41(1), 7-15.
- Rideout, V. (2015). *The common sense census: Media use by tweens and teens*. Common Sense Media. Retrieved from [https://www. commonsense media.org/](https://www.common sense media.org/)
- Rodríguez, F. J., & Boyer, K. E. (2015, June). Discovering individual and collaborative problem-solving modes with hidden Markov models. In Conati, C., et al. (Eds.), *Artificial Intelligence in Education*. Paper presented at the 17<sup>th</sup> International Conference on Artificial Intelligence in Education (pp. 408-418). Springer International Publishing.
- Siemens, G. (2012, April). Learning analytics: envisioning a research discipline and a domain of practice. In Shum, S., Gasevic, D., & Ferguson, R. (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. Paper presented at the 2<sup>nd</sup> International Conference on Learning Analytics and Knowledge (pp. 4-8). ACM.
- Taraghi, B., Saranti, A., Ebner, M., & Schön, M. (2014, June). Markov chain and classification of difficulty levels enhances the learning path in one-digit multiplication. In Zapharis, P., & Ioannou (Eds.), *Learning and Collaboration Technologies: Designing and Developing Novel Learning Experiences*. Paper presented at 1<sup>st</sup> Intl. Conference on Learning and Collaboration Technologies (pp. 322-333).
- Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education*, 34(1), 179-225.
- Watkins, C.S. (2011). *Digital divide: Navigating the digital edge*.
- Xu, B. (2011). *Clustering educational digital library usage data: Comparisons of latent class analysis and K-Means algorithms*.
- Zywica, J., Richards, K. A., & Gomez, K. (2011). Affordances of a scaffolded-social learning network. *On the Horizon*, 19(1), 33-42.

## Acknowledgments

This material is based on work supported by the National Science Foundation Cyberlearning and Future of Learning Technologies grant (DIP #1325004).

# Analyzing Students' Collaborative Regulation Behaviors in a Classroom-Integrated Open Ended Learning Environment

Mona Emara, Michael Tscholl, Yi Dong, and Gautam Biswas  
memara@isis.vanderbilt.edu, mtscholl@cs.ucl.ac.uk, yi.dong@vanderbilt.edu, Gautam.Biswas@vanderbilt.edu  
Institute for Software Integrated Systems, Vanderbilt University

**Abstract:** Identifying the effects of students' collaborative regulation behavior when working on a task is an important step towards a better understanding of how collaboration supports learning. We discuss a study where we combined analysis of students' dialogues with an automated analysis of their action patterns as they constructed science models in an open-ended learning environment. Our results show that students use different types of collaborative regulation behaviors, and that these behaviors affect their performance on the system as well as their pre-post learning gains. We also showed that groups, which adopt more shared regulation used different learning strategies than groups that did not.

## Introduction

Research on collaborative learning has established that metacognitive processes are critical for group performance and individual learning (Volet, Vauras & Salonen, 2009; Saab, 2012). Computer-supported collaborative learning systems derived from open-ended learning environments (OELEs) (Land, 2000; Segedy, Kinnebrew, & Biswas, 2015) provide effective pedagogical contexts to support these processes, which, in turn allows researchers to examine and analyze these processes in greater detail. By providing students with tools and other resources to construct science models, combined with data logging capabilities we can identify behavioral patterns and make inferences on learners' cognitive and metacognitive processes.

The research presented in this paper assumes learners' action patterns (i.e. frequent action sub sequences) interpreted in the context of the learning tasks facilitated by the OELE, are indicative of their learning behaviors and strategies. Thus, a contextual analysis of these action patterns may allow us making inferences on students' metacognitive processes (Kinnebrew, Segedy, & Biswas, 2014). There is growing consensus that metacognitive processes evolve as a series of events that can be derived from students' learning behaviors (Winne, 2010; Hadwin, et al. 2007).

In the study reported in this paper, 6<sup>th</sup> grade students worked in pairs using an OELE called Betty's Brain (Leelawong & Biswas, 2008) to construct a causal map of a complex science topic (climate change). The pair worked on a laptop, with shared input devices (mouse and keyboard) to build their models. The classroom teacher asked students to work together to build their models, and take turns in controlling the input devices. Betty's Brain adopts an open-ended approach to learning by modeling, and this prompts students to develop learning and problem solving strategies, along with metacognitive processes that help them monitor their learning and model building tasks. The system collects rich log data of students' action sequences and their performance on the system, thus allowing more detailed analyses of their behavior and performance.

Also, we leverage analytics and data mining techniques (Kinnebrew, Loretz & Biswas, 2013) to detect behavioral patterns, and combine them with an analysis of the dialogue between the students to study three forms of regulation: (1) *self*-, (2) *shared*-, and (3) *other-regulation*. We hypothesize that groups that exhibit more shared regulation show greater awareness and better understanding of the task. As a result, they learn more and perform better. We perform a case study to determine if the data supports our claims.

## Background

### Metacognition and collaborative learning

In collaborative situations, one's metacognitive awareness and understanding may be extrapolated and projected for the active regulation of shared problem solving processes, such as contributing to goal setting, planning, progress monitoring, and reflection (Brown & Palinscar, 1989; Kinnebrew, Segedy, & Biswas, 2014). Collaborative work brings out the reciprocal nature of metacognitive activity and processes that is governed by dialogue structures that may include self-disclosure, feedback requests, and other-monitoring. A few studies have examined socially shared regulation or how a social setting affects individual metacognition. Goos et al. (2002), for example, argue that the social setting mediates metacognition by prompting an individual's metacognitive awareness, a process where transactive reasoning (i.e., reasoning on the reasoning of others) is critical. Other authors (e.g., Volet, Summar & Thurman, 2009; Iiskala, Vauras, Lehtinen & Salonen, 2011)

define socially shared metacognition as inter-individual metacognition. Similar to an individual controlling her cognitive activity, socially shared metacognition “is consensual monitoring and regulation of joint cognitive processes” (Iiskala et al., 2011, pg. 379). Primers for metacognitive processes becoming shared are verbalizations of metacognitive experiences that prompt other group members to contribute to the regulation of the task. That individuals often profit from external regulation of their problem solving task is well established (e.g., Azevedo, Cromley & Seibert, 2004), and the assumption is that in collaboration, such regulation is partially accomplished by the collaborators. In consideration of the additional tasks related to coordinating collaborative activity, the question is how socially shared regulation affects group performance and individual learning. Like the study presented in this paper, recent work used mixed methods, combining the analysis of verbal interaction with trace methodologies to measure collaborative metacognitive regulation. Winters & Alexander (2011) found a significant positive relationship between students’ regulatory behaviors and their learning gains. Schoor & Bannert (2012), in contrast, found no difference in the frequency of regulatory activities between high and low performing dyads. Järvelä, Malmberg & Koivuniemi, (2016) differentiated between self- and shared-regulatory activities related to task understanding, planning, strategy use, and motivation (cf. Winne and Hadwin, 1998), finding positive relations among socially shared regulation processes (planning and motivation) and performance. Importantly, Järvelä et al (2016) applied a temporal process analysis of student actions, and identified correlations between process patterns and high-performance groups: socially shared regulation of tasks is a critical factor for performance, but it needs to be supported by prior practice of individual regulation. The study also showed that students’ sharing regulation of learning accrued more learning gains, especially if regulation was shared during the solution construction phase.

## Analyzing students’ actions

As learners’ cognitive and metacognitive processes are not directly visible, a common approach is to infer these processes from patterns of learners’ actions and the context in which they invoke these actions (Winne, 2010, Kinnebrew, Segedy, & Biswas, 2016). In OELEs, such as Betty’s Brain (see Figure 1), actions occur on explicitly designed interface features that enable the analysis of their activities in more specific task-related contexts. For example, students may access a *Science Book* by clicking on a tab of the main interface, an action that is an indicator of acquiring information to learn about the science topic; accessing the *Quiz* tab is linked to the more general task of solution assessment and monitoring; by working in the *Causal Map* page, students construct their solution (Leelawong & Biswas, 2008; Segedy, Kinnebrew, & Biswas, 2015).

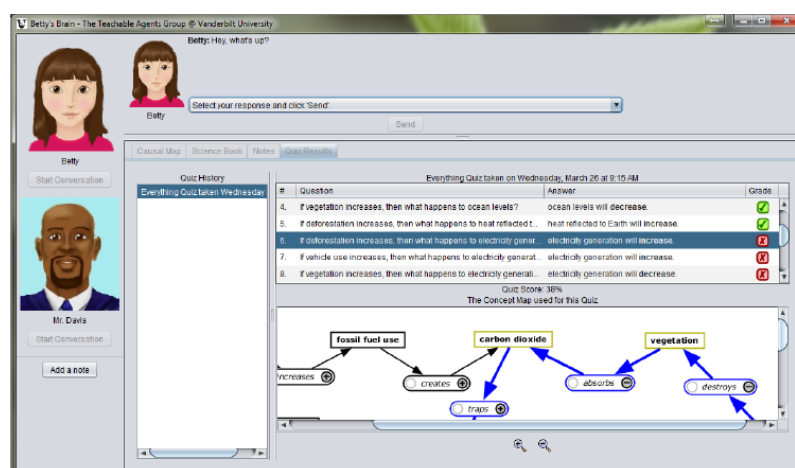


Figure 1. The Betty’s Brain system. The page shown in the figure is the Quiz page.

## Analyzing students’ action sequences in learning research

Increasingly, researchers are utilizing trace methods to analyze learners’ actions in an effort to make inferences on their metacognitive processes. The underlying rationale is that metacognitive processes are best studied by examining the effects of metacognitive control. Other methods, such as questionnaires, typically provide post-hoc accounts of metacognitive experiences and processes, and are less suited to capture the dynamic nature of regulation processes.

A new development in this respect is using sequence mining techniques to detect patterns of recurring actions. These techniques are especially applicable to examine differential use of actions and action patterns in

varying groupings of learners. For example, high-performing students are much more likely to read information and use that information to make progress in a problem solving task than are their lower-performing peers who may find it difficult to connect their reading to their problem solving task requirements (Kinnebrew, Segedy, & Biswas, 2014).

Sequential characteristics describe the learning activities a student typically uses and they inform the actual process of self-regulated learning (Johnson, Azevedo, & D'Mello, 2011). To examine the links between actions produced by shared or individual regulation, action sequences have to be examined in context and their relation to each other. Weak or no relations between actions that make up a pattern suggests that learners' cognition is broadly un-regulated. This may imply the learners are applying a trial and error approach without a specific focus or direction. In contrast, regularity and coherence relations among actions and frequent use of action patterns suggests that the learner(s) are exerting control, derived from strategies (Kinnebrew, Segedy, & Biswas, 2016). Thus, a first important analysis concerns whether regularly occurring action sequences can be detected; subsequently we examine whether action patterns resulting from shared, other-, or self-regulation differ.

### Actions in Betty's Brain

In Betty's Brain, students' information acquisition, map building, and map checking actions are recorded in log files. In addition, students can also add and view notes (NOTESVIEW; a 'note' is a text box in which to collect or summarize information from the science book deemed to be relevant); and ask Betty to explain her answers to quiz questions (EXPL). The most common actions and their labels are listed in Table 1.

Table 1: Actions in Betty's Brain and their log labels

Action description	Action log label
Reading the resources to learn about the domain of study	<i>READ</i>
Adding or Viewing a Note	<i>NOTESVIEW</i>
Adding a concept	<i>CONCADD</i>
Adding a causal link	<i>LINKADD</i>
Removing a concept	<i>CONCREM</i>
Removing a link	<i>LINKREM</i>
Indicating that a link is believed to be 'correct';	<i>CLMARKCORR</i>
Have Betty take a quiz about the map	<i>QUIZTAKEN</i>
Highlight the part of the causal map that Betty used to answer a quiz question	<i>QUIZVIEW</i>
Ask Betty to explain her answer	<i>EXPL</i>

Some logged actions are further characterized by attributes. Relevant for this study are 2 attributes. If the period between a READ action and an immediately following action is below 3 seconds, the READ action is logged as READ-SHRT (for 'short reading'). Short reads are interpreted as shallow skimming over the learning material. Actions on links (adding or removing) are associated with the change in map score that these actions entail. A link action increasing the map score is logged with the suffix -EFF (for 'effective'), an action decreasing it with the suffix -INEFF (for 'ineffective').

All actions are classified in higher-level categories corresponding to cognitive process that, when combined, constitute metacognitive strategies. Building on prior work (Kinnebrew, Segedy, & Biswas, 2016), actions are classified as the cognitive processes of Information Acquisition (READ, NOTESVIEW), Solution Construction (all actions on the causal map, e.g. LINKADD) and Solution Assessment (QUIZ actions). The actions listed in table 1 constitute the elements of action sequences; in the Results section we will present the action sequences in the terms they are recorded by the system.

### The study

The focus of the study conducted in this paper revolves around two primary research questions:

*RQ1: Do the extent of groups' self-, other and shared regulatory activities correlate with their*

performance? (RQ1.1); and do students' self-, other and shared regulatory activities correlate with their pre-posttest learning gains? (RQ1.2)

**RQ2:** Do students' action patterns differ when they adopt socially shared regulation strategies as opposed to working more on their own?

We discuss our methods to answer these questions below.

## Method

### Participants

14 middle school students (6<sup>th</sup> grade) from an academic magnet school in Southern USA participated in the study. To enroll in this school, students need an average grade of B+ during the previous academic year. All participating students were in a self-contained class taught by the same teacher. Overall, the study lasted 7 days. On day 1, students were introduced to the science topic (climate change), provided an overview of the idea of concept maps, and given hands-on training with the system. On day 2, students were tested on their knowledge of climate change, of reading comprehension and introduced to causal maps, and reasoning with causal maps. These tests constituted the pre-test. Then students of equal academic capability who were deemed compatible following the teacher's suggestions were paired to work together. The grouping yielded 3 single-sex dyads (all male-male dyads), and 4 mixed-sex dyads. On days 3 to 6, the dyads then worked on the system constructing the causal map. We call these days *study days*. They lasted on average 50 minutes per day. Students were instructed to collaborate, i.e. to discuss and take decisions jointly. These instructions were occasionally repeated during the study sessions. On day 7, students took a post-test consisting of the same items as the pre-test.

### Data collection

Groups worked at a single computer with one student controlling the mouse. The teacher explained that one student in the group should have control of the mouse, the primary input device, for half a study day, and then let the second student take over control of the mouse. Talk and behavior (e.g., attention to the other student) was recorded using Camtasia as audio-video data from web-cams synced to a screen capture video. Student-system interactions were automatically logged by the system. The log files recorded the action (e.g. adding a causal link, accessing a specific page in the Science Book) with the group ID with corresponding time stamps on a central server.

### Transcript analysis

Students' talk was transcribed with the turn-of-talk as the smallest unit of analysis. The time stamp of each turn was recorded manually to allow synchronization with action logs. When a student-system interaction was carried out immediately after or during talk, talk and action were transcribed as a single turn-of-talk. Transcription yielded a total of 5774 turns-of-talk (mean: 825, *SD*: 157).

A small set of transcribed talk was first coded by two authors of this article. Verbal data were the primary source for coding, although visual gaze, nodding or other non-verbal behavior was occasionally also examined for coding. Differences in coding were discussed until a consensus on the definition of the coding categories was reached. Then, one author coded all remaining transcripts, while the other coded 20% of transcripts selected from different groups and study days. Inter-rater agreement (Cohen's kappa) was .87. By most accounts (e.g., Bakeman, Quera, McArthur, & Robinson, 1997), a kappa value above .81 is considered "excellent agreement."

Coding proceeded by first identifying initiative-response relationships between turns of talk (Kneser, Pilkington & Treasure-Jones, 2001). We argue that this lens is helpful to capture whether a discussion is reciprocal (i.e. both students initiate ideas and contribute constructively to the discussion) or not (when one student is dominant, the other is silent or may carrying out the instruction).

We relied to prior analysis schemes as a guide to develop our coding categories (Volet, Summer, & Thurman, 2009; Hadwin et al. 2011):

**Socially Shared Regulation** (SSR, figure 3): turns-of-talk were coded as SSR when the students question or elaborate on their partner's initiative or with an alternative. This is the case, for example, when proposals are not passively accepted, but negotiated and may lead to more protracted discussions. Discussions include various topics, such as what to do next, how a proposed action fits into an overall plan, how to proceed after an assessment, and so on. In these discussion, students are likely to verbalize their strategies (e.g. "how do we link incorrect quiz answers to incorrect links in our map?") and demonstrate metacognitive awareness and processes

(“Have we read about sea level rise?”, “Should we check our map now?”). Goals, plans, and strategies are co-constructed, and regulation is distributed and shared with multiple ideas being weighed and negotiated (Miller and Hadwin, 2015).

**Other-Regulation (OR, figure 2):** in this form of regulation, a student is temporarily dominant and instructs the other. This form of regulation occurs frequently when the less dominant student is in control of the mouse. When this student just performs suggested actions without discussion, we code it as OR. Importantly, instructions are almost never accompanied by verbalizations of rationales, plans or strategies, or metacognitive experiences.

**Self-Regulation (SR):** Self-Regulation occurs when a student is temporarily in full control of problem solving, with no contribution from the other student. The Self-Regulation code can only be applied to a student in control of the mouse. The partner may be absent, may not be contributing or interested in contributing, or the student controlling the mouse may disregard the partner’s attempted contributions and suggestions.

S Talk and Actions	Code	S Talk and Actions	Code
Adam: "We should add a link"	OR	Hana: "We got 4 right and 2 wrong" (students are viewing quiz results)	SSR
John: [Activates link button]	OR	Sam: "Let's edit it" [marks the causal link between 'ice and ocean level' as 'maybe wrong']	SSR
Adam: "No, go back to the science book"	OR	Hana: "So, we should fix all of these links?" (points to the links)	SSR
John: [Opens science book]"What should we read?"	OR	Sam: "Yes, we need to do more work" "Let's go to the science book"	SSR
Adam: [Reads from the page and tells John]	OR	Hana: [Opens science book to search for 'ocean levels']	SSR

Figures 2 and 3. Transcript extracts illustrating other-regulation (figure 2, left figure) and socially shared regulation (figure 3, right figure). Interface actions were noted in the transcript in square brackets (e.g. [Opens Science Book]). These actions are automatically recorded in the log files as well and are the data on which the sequence mining algorithm is applied.

## Results

We report briefly the quantitative profiles of all groups. In two groups, students jointly carried out the task – groups 1 and 7; mean percentages: 68% SSR ( $SD = 4.2\%$ ), 23% SR ( $SD = 2.1\%$ ), 9% OR ( $SD = 3.5\%$ ); in two others it was often only one student – groups 5 and 6; mean percentages: 21% SSR ( $SD = 2.8\%$ ), 73% SR ( $SD = 9.1\%$ ), 6% OR ( $SD = 6.3\%$ ). In the three remaining groups, joint and individual regulation is balanced – groups 2, 3 and 4; mean percentages: 40% SSR ( $SD = 9\%$ ), 48% SR ( $SD = 6.8\%$ ), 12% OR ( $SD = 4.5\%$ ). OR occurs infrequently in all groups.

### *RQ1.1: Do groups’ self-, other and shared regulatory activities correlate with their performance?*

To address this question, the Pearson correlation coefficient between self-, other- and socially shared regulation of learning per group (sum of all days) was calculated. A significant positive relationship between the frequency of SSR and the map scores ( $r = 0.75, p < .05, SE = 0.31$ ) was found.

### *RQ1.2: Do students’ self-, other and shared regulatory activities correlate with their learning gains?*

To answer this question, the Pearson correlation coefficient was calculated between the frequency of SSR and SR of a group, and the students’ learning gains. There was a significant positive relationship between individual learning gains and frequency of socially shared regulation ( $r = 0.59, p < .05, SE = 0.32$ ), and a significant negative relationship between individual learning gains and the frequency of self-regulated actions ( $r = - 0.54, p < .05, SE = 0.31$ ). These results lend some support to the claim that individual students learn more when they jointly regulate the task.

### *RQ2: Do action patterns differ when a group jointly regulates the task compared to when in a group individual students often work on their own?*

To answer this question, we employed a differential sequence mining technique (Kinnebrew, Loretz & Biswas, 2013) that detects recurrent action patterns in a set of action sequences. We selected two groups (1, 7) which consistently demonstrated high frequencies of shared regulation (the *bothStudents* profile) and two groups (5, 6)

where one student worked often by themselves (self-regulation), or instructed the other (other-regulation). This set is called the *oneStudent* profile. We justify aggregating instances of Self- and Other-Regulation by pointing out that in Other-Regulation students rarely verbalize their metacognitive experiences or provide rationales for their instruction.

The sequences of logged actions of all groupings (oneStudent and bothStudents) were analyzed with a gap size constraint of 1, yielding a very large set of action patterns (~ 60,000). A gap size constraint of 1 means that between each consecutive pair of actions in a frequent pattern, the mining algorithm allows up to 1 of additional actions. A sequential pattern with such an optional (small) gap (e.g., the pattern  $A \rightarrow B \rightarrow C$ ) means that each subsequent action in the pattern is performed shortly after the previous action even if there were intervening actions that are not part of the pattern. For example, a student performs A, followed by B, then D, then C; i.e. the actions A, B and C are followed in close, but not necessarily direct, succession, and the single action D is treated to be irrelevant.

Associated with each pattern are the following values: for each pattern, a value representing in how many groups was the pattern detected regardless of frequency (S-frequency), and a value of how frequently a pattern occurred by an 'average' group (I-frequency). The 'average' group is calculated using the actual data used in this study. Patterns are ordered by I-Frequency of the oneStudent grouping and assigned an index starting from 1 and increasing as that I-frequency decreases. Patterns with low indices occur very frequently in the oneStudent condition; patterns with a very high index (40k-60k) occur with high frequency in the bothStudent condition. We computed the difference in I-frequencies and their ratio to identify 1) frequently occurring action patterns and 2) action patterns occurring with high I-frequency differences or ratio differences relative to the two groupings.

We first report the most and least frequent action patterns in both groupings. The most frequent patterns for the oneStudent grouping are short or longer map edits (e.g. ADDCONC; ADDCONC; ADDLINK; ADDCONC) (1). and explanations (e.g. pattern 96, table 2). Sequences of only explanation actions and only map edits (up to 7 consecutive actions) occur at much higher frequencies in the oneStudent grouping than in the bothStudents grouping. The I-frequency ratios (oneStudent:bothStudents) averaged over all action patterns with only map edit or only explanation actions, regardless of length, are 5.1:1 and 5:1, respectively. The most frequent action patterns in the bothStudent grouping involve READ actions. Action patterns with only READ actions (up to 7 consecutive actions) occur on average 2.6 times more often in the bothStudent grouping. For this grouping we also identify distinct action patterns that involve taking and viewing notes (VIEWNOTES), and marking a causal link as 'correct' or 'incorrect' on the causal map (CLMARKCORR). This latter action is an expression of the belief that a link is correct (or not), and thus can be interpreted as a progress marker (if marked 'correct') or as an organizer of future activity (if marked 'incorrect'). Action patterns involving the VIEWNOTES action occur 5 times more often in the bothStudents groups; those involving marking actions occur almost never in the oneStudent grouping.

We then examined the sequence mining results to identify action patterns that span distinct high-level process elements, such as patterns that involve reading *after* taking a quiz, or taking a quiz *after* adding elements on the map. With regard to strategies involving information acquisition and subsequent actions, we found that students' in the bothStudents grouping identify important information and then work only briefly on the map more often those in the oneStudent grouping. More in general, in the bothStudents grouping, the students read the Science Book or their notes, and then add one or very few concepts or links (e.g. pattern 51919). Students in the oneStudent grouping, in contrast, read in the Science Book, then work on the map adding several concepts or links without returning to the Science Book (pattern 635). This finding generalizes to several action patterns involving READ followed by MAPEDIT actions.

Of particular importance are strategies carried out after a progress assessment. For strategies involving the QUIZ as the first action we find that in the bothStudents grouping, the students return to the Science Book after a QUIZ much more frequently than students in the oneStudent grouping (e.g. pattern 51581). This strategy indicates that the students in the bothStudent grouping were better able to interpret the results of a quiz in terms of missing understanding, which they seek to address by reading. Students in the oneStudent grouping, in contrast, either immediately work on the map (including by removing concept, see pattern 76) or by asking Betty for an explanation (pattern 71). Working on the map after a quiz results is very likely indicative of a trial-and-error strategy.

Students' application of information and actions after assessment generalizes to strategies involving all multiple different high-level process elements. Students in the bothStudents grouping are more likely to carry out an IA, SC, SA strategy (e.g. pattern 52028 and 49062); students in the oneStudent grouping demonstrate a strategy that involves multiple quiz views and map actions (pattern 1376), a validation of our earlier suggestion that these students are more likely to solve the task through trial-and-error.



Table 2: Action patterns with high differences in I-frequency (reported as ratio) relative to the two groupings. Patterns are listed by ascending index

Index	Pattern	1S:2S ratio	Index	Pattern	1S:2S ratio
49062	LINKADD-INEFF; READ; QUIZTAKEN; READ;	1:1.9	71	QUIZTAKEN; QUIZVIEW; EXPL; EXPL;	5:1
51581	QUIZTAKEN; QUIZVIEW; REA; READ; READ; LINKADD;	1:2.4	76	QUIZTAKEN; QUIZVIEW; CONCREM; CONCREM;	7.6:1
51897	LINKADD-EFF; QUIZTAKEN; QUIZVIEW; READ;	1:2.1	96	EXPL; EXPL; EXPL	4.3:1
51919	READ-SHRT; READ; LINKADD;	1:1.9	635	READ; CONCADD; CONCADD; LINKADD; LINKADD; LINKADD;	5.3:1
52028	READ; LINKADD; QUIZTAKEN; QUIZVIEW;	1:2.6	1376	QUIZVIEW; CONCADD; LINKADD; QUIZTAKEN; QUIZVIEW;	2.2:1

## Discussion

This study set out with the aim of analyzing students' collaborative regulation behaviors in an OELE and linking the behaviors to students' metacognitive strategies. Our quantitative analyses show that students who adopted more shared-regulation strategies demonstrated better task performance and learning gains than those that do not. We thus confirm other's findings on the importance of collaborative metacognitive regulation (Winters and Alexander, 2011) and socially shared regulation (Järvelä et al., 2016).

By detecting frequent action patterns and how they occur with differential frequency in the two groupings, our study contributes to the question of *how* socially shared regulation affects performance and learning gains. Our results indicate that sharing regulation affects students' strategies with important implications on when and how they acquire knowledge, apply acquired knowledge, assess progress and react to that assessment. Evaluating differential frequencies of strategies based on Winne & Hadwin's (1998) model, we find that students sharing regulation adopt a) improved information acquisition and application strategies (reading followed by working on a single or only a few element of the map, then returning to the Science Book) b) better monitoring strategies (when to check the correctness of the current causal map), and c) improved strategies on how to react to assessments (reading followed by working on the map, instead of immediately altering the map).

We also identify action patterns that are distinct for each grouping, which further the question on the construct of socially shared regulation. As note taking and marking are much more frequent in the both Student groups, these actions can be interpreted as organizing collaborative activity. These groups seek external markers to summarize what they have read together, to mark what they have accomplished and to organize future joint activity. With regards to the identification of potential benefits of sharing metacognitive processes, the effect of such markers will need to be taken into account. Benefits may result from activities intended to organize collaboration – rather than strictly organizing problem solving. Iiskala, Vauras & Lehtinen (2004) argue that socially shared metacognition is a distinct process that cannot be reduced to the contribution of individuals to joint activity. Our study allows to qualify this assumption. Activities aimed at regulating and coordinating group activity prompt sharing of metacognitive awareness. Whether sharing metacognitive processes emerges from such coordination activities, from transactive processes (Goos et al., 2002) or from the projection of individual metacognitive awareness onto the group activity (Iiskala et al., 2011) is a question that warrants further examination.

## Endnotes

(1) ADDCONC: added a concept to the causal map; ADDLINK: added a link to the causal map

## References

- Azevedo, R., Cromley, J.G. & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, 29, 344-370
- Bakeman, R., Quera, V., McArthur, D., & Robinson, B. F. (1997). Detecting sequential patterns and determining their reliability with fallible observers. *Psychological Methods*, 2 (4), 357-370.
- Brown, A., & Palinscar, A. (1989). Guided, cooperative learning and individualized knowledge acquisition. In Resnick, L. (Ed.), *Knowing, learning, an instruction*, Hillsdale, NJ: LEA.

- Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Ed. studies in Mathematics*, 49(2).
- Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 65-84). New York: Routledge.
- Hadwin, A., Nesbit, J., Jamieson-Noel, D., Code, J., & Winne, P. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, 2(2), 107–124.
- Iiskala, T., Vauras, M., & Lehtinen, E. (2004). Socially-shared metacognition in peer learning? *Hellenic Journal of Psychology*, 1, 147–178.
- Iiskala, T., Vauras, M., Lehtinen, E., & Salonen, P. (2011). Socially Shared metacognition within primary school pupil dyads' collaborative processes. *Learning and Instruction*, 21, 379–393.
- Järvelä, S., Malmberg, J., & Koivuniemi, M. (2016). Recognizing socially shared regulation by using the temporal sequences of online chat and logs in CSCL. *Learning and Instruction*, 1-11
- Johnson, A. M., Azevedo, R., & D'Mello, S. K. (2011). The temporal and dynamic nature of self-regulatory processes during independent and externally assisted hypermedia learning. *Cognition and Instruction*, 29(4), 471-504.
- Kinnebrew, J., Loretz, K., & Biswas, G. (2013). A contextualized, differential sequence mining method to derive students' learning behavior patterns. *Journal of Educational Data Mining*, 4 (1), 190-219.
- Kinnebrew, J., Segedy, J.R., & Biswas, G. (2016). Integrating Model-Driven and Data-Driven Techniques for Analyzing Learning Behaviors in Open-Ended Learning Environments. *IEEE Trans. LT*.
- Kinnebrew, J.S., Segedy, J.R., & Biswas, G. (2014). Analyzing the Temporal Evolution of Students' Behaviors in Open-Ended Learning Environments. *Metacognition and Learning*, 9(2), 187-215.
- Kneser, C., Pilkington, R., & Treasure-Jones, T. (2001). The tutor's role: an investigation of the power of exchange structure analysis to identify different roles in CMC seminars. *International Journal of Artificial Intelligence in Education*, 12, 63-84.
- Land, S. (2000). Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research and Development*, 48(3), 61–78.
- Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The betty's brain system. *International Journal of Artificial Intelligence in Education*, 18(3), 181–208.
- Miller, M., & Hadwin, A. (2015). Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL. *Computers in Human Behavior*, 52, 573-588
- Saab, N. (2012). Team regulation, regulation of social activities or co-regulation: different labels for effective regulation of learning in CSCL. *Metacognition and Learning*, 7(1), 1-6.
- Schoor, C., & Bannert, M. (2012). Exploring regulatory processes during a computer supported collaborative learning task using process mining. *Computers in Human Behavior*, 28(4).
- Segedy, J.R., Kinnebrew, J.S., & Biswas, G. (2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, 2(1), 13-48.
- Volet, S. E., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? *Learning and Instruction*, 19(2), 128-143.
- Volet, S. E., Vauras, M., & Salonen, P. (2009). Self- and social regulation in learning contexts: an integrative perspective. *Educational Psychologist*, 44(4), 215-226.
- Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist*, 45, 267–276.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Hillsdale, NJ: Lawrence Erlbaum.
- Winters, F. I., & Alexander, P. (2011). Peer collaboration: the relation of regulatory behaviors to learning with hypermedia. *Instructional Science*, 39, 407-427.

# Integrating Physical and Virtual Models in Biology: A Study of Students' Reasoning While Solving a Design Challenge

Nicole D. Martin, Dana Gnesdilow, and Sadhana Puntambekar  
ndmartin@wisc.edu, gnesdilow@wisc.edu, puntambekar@education.wisc.edu  
University of Wisconsin – Madison

**Abstract:** Using models to explain phenomena is important in science. Virtual and physical models have different affordances that can be integrated to foster students' learning. Integrating evidence from multiple models to justify explanations is challenging, and we know little about how students coordinate such information, especially in biology. This study investigated how students' integrated information from virtual and physical models in a design-based, biology curriculum. Some students used information from virtual simulations in written explanations of changes they would make to their physical models. However, one-third of students did not use the virtual model to justify their revisions, despite prompts from instructional materials, the teacher, and other group members. Even some students who integrated these different models did not initially do this without support from external prompts. This study provided deeper understanding of how students integrated physical and virtual models, which can help identify the kinds of support students may need.

## Introduction

Constructing and using models is an important practice in science (Lehrer & Schauble, 2006; National Research Council, 2012; Windschitl, Thompson, & Braaten, 2008). Scientific models are representations of natural phenomena that are “testable, revisable, explanatory, conjectural, and generative” (Windschitl et al., 2008, p. 3). They can be generated by students (e.g., drawings, physical representations) or can be generated by others and manipulated or utilized by students (e.g., simulations). Inquiry centered around models strives to develop and defend explanations, which importantly includes using evidence to justify and inform these models (Lehrer & Schauble, 2006; Windschitl et al., 2008). Inquiry with virtual models such as simulations offers unique opportunities for students' learning, allowing students to investigate phenomena that might not otherwise be possible in a classroom. Research has investigated the affordances and learning benefits of virtual versus physical models and experiments, but this has predominately been done in physics and engineering contexts (e.g., de Jong, Linn, Zacharia, 2013; Zacharia, de Jong, 2014). Findings from research in these contexts suggest that virtual models can allow students to conduct experiments about unobservable phenomena and alter the time scale of experiments that would take a long time in the real world, whereas physical models can directly expose students to real phenomena, authentic problems, and measurement error they would not face in a virtual experiment (de Jong et al., 2013; Olympiou & Zacharia, 2012). Further, studies suggest that investigations with virtual models may better support conceptual understanding than with physical models, because investigations with virtual models allow students to more easily isolate variables, produce cleaner data to analyze, and offer more time for students to engage in experimentation (de Jong et al., 2013).

Previous research in this area has primarily explored comparing students' learning on content-based tests from participating in experiments using either physical *or* virtual models or participating in different sequences of physical and virtual experiments (e.g., Zacharia & Olympiou, 2011; Zacharia, Olympiou, & Papaevripidou, 2008). This research has produced mixed findings on the benefits of one modality over the other and ideal sequencing. Given the unique affordances of physical and virtual models, integrating or blending the usage of these models based on the particular learning objectives of an experiment has the potential to optimize these affordances for student learning (Olympiou & Zacharia, 2012). While some researchers have begun exploring such integration, there is still little research examining how students integrate and reason about information from one model to inform their experimentation with another. Further, little research has focused on how students collaborate while engaging in such tasks.

Additionally, while these findings from physics and engineering contexts are informative, it is unclear how they extend to biology, as less research has been done on physical and virtual models in biology. The complexity of biological processes offers an important context to explore the integration of these models. Many processes in biology are microscopic and take time to observe (e.g., decomposition), so virtual simulations can be valuable for quickly conducting experiments of such processes in the context of a classroom. The affordances of both physical and virtual models offer different learning opportunities that could be utilized to deepen students' understanding of biological processes. Similar to physics and engineering, virtual models in biology can let students conduct experiments that would take a long time in the real world, and physical models can let

students experience real phenomena and encounter problems they would not face in a virtual experiment. Based on these affordances, both types of models can be integrated to help foster students' learning. For example, students could use simulated experiments to gain understanding about how to improve a physical model, or the conditions of a physical model could influence how they use a virtual simulation. To successfully do this, students would need to integrate and use evidence from multiple models to appropriately justify an explanation. Previous research has shown that learning from multiple representations (e.g., Ainsworth, 2006) and using evidence to support scientific explanations (e.g., Sandoval & Millwood, 2005) are both challenging for students. Thus, utilizing information from virtual and physical models to solve a design challenge is a complex task for which students likely need support. Recent research in the context of engineering design has shown that using virtual models to plan future design decisions and to reflect on previous decisions can both help students integrate information for their designs and understand science concepts (McBride, Vitale, Applebaum, & Linn, 2016). However, we still do not know much about how students might integrate information from virtual and physical models in this way and, thus, know little about how to best support students to do this, especially in the field of biology. Our research begins to explore i. whether students integrated information from virtual and physical models; and ii. how students collaborated to try to successfully integrate this information to work towards solving a design challenge in the context of a design-based, biology curriculum.

## Methods

This was an exploratory study investigating students' ability to integrate information from virtual and physical models in the context of a design-based, biology curriculum. The curriculum challenged students to work in teams to design a compost that would break down quickly and contain a lot of nutrients to reduce the amount of waste going into landfills. Students learned key concepts related to energy transformation and matter cycling in ecosystems to solve their challenge. To study decomposition and collect data to justify for their designs, students built, monitored, and refined a physical bio-reactor throughout the unit. Since it takes several weeks for compost to break down, students also used virtual compost simulations to better understand how abiotic factors influence decomposers' ability to break down matter. They conducted four virtual compost experiments related to the carbon to nitrogen ratio of materials, the amount of moisture, the size of the particles, and the combination of all these factors. Based on what they learned in the virtual experiments, students were required to revisit their physical bio-reactors and use the science ideas, data, and evidence from the virtual experiments to decide on and justify one change that would increase decomposition in their physical bio-reactors. The data for this study focused on students' written explanations in their journals and audio-recorded conversations from three groups that occurred during this activity, further described below.

## Participants

The participants were twenty-six students (N=26), working in seven small groups, from an 8<sup>th</sup> grade science class at an urban middle school in a Midwestern city in the United States. This was a "talented and gifted" class taught by an experienced teacher. The composition of the small groups was determined by the teacher. We chose this class as a potential exemplar of how high-achieving students might integrate models. We further chose this class to see how an experienced teacher might facilitate this process in a student-centered way that could provide information about useful teacher supports that helped students integrate different models in a design-based curriculum.

## Data sources

We identified a *Change Page* in the students' journals where they needed to provide written responses about what to change in their physical bio-reactors after conducting four virtual experiments using a compost simulation. Students were asked to work with their group to provide evidence for a current problem in their physical bio-reactors, write a potential change they could make, and then supply evidence for why the proposed change would cause improved decomposition. To do this, students needed to observe their bio-reactors to identify problems and use evidence from their virtual experiments to *explain* the change they wanted to make. The only data students had to provide evidence was from these virtual experiments, and thus they should have provided data from the simulation about the ideal conditions for decomposition in compost to support their changes. We inductively developed a coding rubric to explore how students provided evidence for the physical change from their virtual experiments (see Table 1). Twenty percent of the students' written explanations on their *Change Pages* were coded by the first and second author separately and a 90% agreement was achieved. All disagreements were resolved with discussion. The remaining written responses were coded by the second author.

To begin understanding *how* students worked together to construct the explanations recorded on their *Change Pages* and to identify if and how students talked about data from their simulations in making their decisions, we qualitatively analyzed conversations from the three groups (A, B, C) that we had audio-recorded during the unit. These groups were chosen for recording throughout the unit by the teacher as being representative groups of academically average-performing students within the class.

**Table 1: Coding scheme to evaluate quality of students' written explanations to support their proposed change**

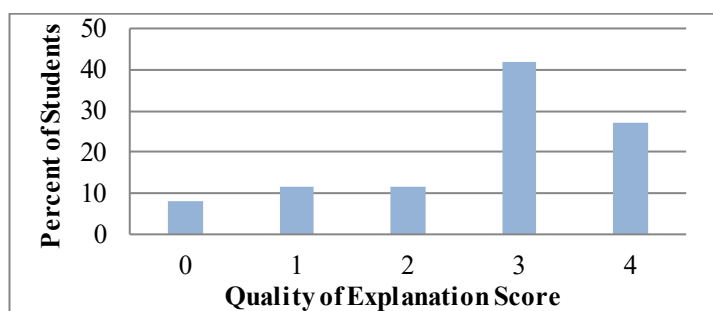
Level	Example	Point Value
None / Incoherent	Blank, no explanation	0
Partial explanation without evidence from the simulation	We are adding browns like oats, leaves, newspaper, and sawdust. This will make decomposition fast.	1
General explanation without evidence from the simulation	We believe that mixing up our compost will distribute the moisture evenly. Currently the moisture is only in the middle of the compost. If we mix it up then the moisture should be distributed evenly.	2
Explanation refers to simulation	I think the C:N ratio is the most important change. We must increase nitrogen for large ratio and more moisture. If we don't have enough browns nothing will be able to break down. Our greens are living so nothing will decompose. We decided to add greens to give it more moisture.	3
Explanation refers to and uses evidence from the simulation	Our carbon to nitrogen ratio is too small. [We could] add browns to increase decomposition speed. The closer the C:N ratio is to 30:1 the better. Adding more browns 1 & 2 (63.7 grams) or more carbon. It will improve our carbon to nitrogen ratio. In turn, this will improve our smell and moisture level.	4

## Results

### Students' written explanations

We found that some students successfully integrated information from the virtual simulation to inform or justify the changes to their physical bio-reactors. However, the quality of students' written explanations for the proposed changes to their physical bio-reactors varied. We found that eight percent of students did not include a coherent explanation at all, while 23 percent of students gave a partial (score = 1, 11.5%) or general (score = 2, 11.5%) explanation without providing evidence (see Figure 1 below). Forty-two percent of the students referred to the simulation in their explanation, but did not include specific evidence (score = 3), and 27 percent of students offered an explanation that referred to the simulation and included evidence (score = 4). This means that 31 percent of students did not refer to the simulation in their explanation despite scaffolding in the students' journal, teacher prompting, and working with their group for support.

Some groups were more successful than others at collaborating to integrate information from the virtual simulation to provide evidence for making a change in their physical bio-reactor. We calculated an average quality of explanation score for each group and found that Groups B and C had the lowest average score ( $M = 2$ ) and Groups D, E, and F had the highest average scores ( $M = 3.3, 3.2, 3$  respectively). Groups A ( $M = 2.75$ ) and G ( $M = 2.5$ ) had scores in the middle. See Figure 2.



**Figure 1.** Percent of individual students' (N=26) responses within each quality of explanation score.

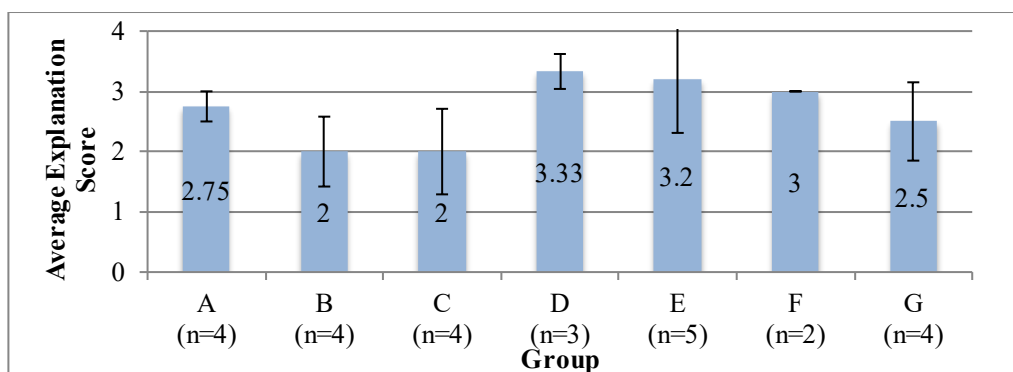


Figure 2. Groups' average quality of explanation scores, number of students per group shown below each group.

### Groups' conversations

We examined groups' conversations to better understand how students collaborated to integrate information from their virtual model to make a change to their physical model. These three groups (A, B, C) were chosen by the teacher at the beginning of the unit as representative groups of academically average-performing students to be audio-recorded throughout. The analysis of Group A, B, and C's conversations revealed several differences in how these groups integrated information from their physical and virtual compost models. Group A made more use of prompts within the instructional materials and from the teacher to have richer conversations about using data from the simulation to inform the changes to their bio-reactor than Groups B and C. Groups B and C ignored, missed, or misinterpreted several prompts that Group A utilized. Vignettes of each group's interactions are presented below to illustrate interactions that were more or less successful.

#### Group A

Of the three audio-recorded groups, Group A's students made more references to the simulation in their explanations of changes for their bio-reactor than Groups B and C, with a group average score of 2.75. In their negotiation of changes to make in their bio-reactor, they first identified that the compost in their physical bio-reactor was too wet. Initially, they did not use any evidence from their simulation experiments to justify their decisions. However, when the group was explicitly prompted in the instructional materials to give "evidence for why this change will cause improved decomposition" on their *Change Page*, they utilized data they gathered from the virtual simulation to justify their change to their physical bio-reactor:

Student 1: Ok, so what's the evidence?

Student 2: We don't really have evidence.

Student 1: How do we know it's too moist?

Student 3: Because.

Student 1: Maybe that's the right range, how do you know that's the right range?

Student 3: Because that's not the right range.

Student 4: Because we did the simulation and the right range was 40-60% moisture.

Here, students 1 and 2 were unsure about how to justify their proposed change, or even why they thought their compost was too moist. Student 3 alluded to their moisture level not being in the "right range," but student 4 took this a step further and reminded the group of the results from their virtual experiment investigating the ideal moisture content of compost. At this point, the group decided to add oats to their bio-reactor because it was too wet, and they knew that the ideal amount of moisture should be 40-60% (by weight) based on data from the virtual simulation. The teacher then suggested to the whole class that they might want to use the virtual simulation to try their proposed change before actually making a change to their bio-reactor.

Teacher: Make sure you think out the changes that you are going to make. If that requires running through one of the simulators again, do it. Maybe run yours through the simulator.

It was not until this direction from the teacher to utilize the virtual model before making physical changes that this group began to meaningfully integrate information from both models. As a result of this prompt, Group A decided to input the conditions of their own physical bio-reactor into the virtual compost simulation and experimented with adding different amounts of oats to achieve their desired moisture content:

Student 4: What was the C to N ratio [in our bio-reactor]?

Student 3: 16.9 to 1.

...

Student 4: So try some browns [in the simulation].

...

Student 4: If it doesn't work, then we shouldn't add anything.

...

Student 4: Ok, now try it.

Student 3: 20 to 1 ratio, with 50% moisture. So add 10 grams of oats.

Student 4: Yeah just a little bit. Just to balance out the moisture.

...

Student 3: So the more, like 10 is about the same, but since ours is moister, I think we need to add a little bit more, like 15 to dry some of it up.

Student 3: So we need like 15 grams of oats or something.

Student 4: Yeah that's it.

This excerpt showed how students 3 and 4 integrated information about the conditions in their physical bio-reactors with data from the virtual simulation to refine their proposed design change. Instead of just simply "adding oats," they determined that they needed to specifically add 15 grams of oats to solve their problem. Further, student 4 insisted that they must base their final decision for the physical model on the evidence from the virtual model when he said, "If it doesn't work, then we shouldn't add anything." Through this discussion, the group was eventually able to use observations of their bio-reactor and data from both the virtual simulation to inform and justify their design change.

#### Group B

Group B was less successful than Group A in making reference to the simulation in both their written explanations in their journals (average group score = 2 for their quality of explanations) and during their conversations about making changes to their bio-reactor. Overall, this group of students got caught up in the "doing" of making the change, rather than thoughtfully planning and justifying their change based on data from the simulation. Unlike Group A, they did not discuss the instructional prompt in their journals that asked them to provide evidence for their change. They quickly decided on the change for their bio-reactor and left the classroom to make it outside. When they were outside, they missed the teacher's suggestion (written above) that students could use the simulation to try out their change, which prompted Group A to have a deeper discussion about how data from the simulation could inform revising their bio-reactor. Instead, Group B solely focused on discussing the state of their physical model and how the materials inside it needed to be mixed:

Student 2: Just mix it up and shake it.

Student 3: It's not evenly distributed (inaudible) and it's moist.

Student 1: Yeah but that could also cause problems if we don't do it correctly.

Student 3: Yeah cus we don't want to have the stuff growing on it.

Student 1: Because if, let's say we don't distribute everything evenly, it could cause problems. How are we gonna mix it without taking it out?

Student 2: I think if we shake it, I shook it a little bit this morning and everything moved so I think we'll be able to-

Student 1: Yeah we can only move it up and down.

Student 2: No it was-

Student 3: I think it's a little compact right here, like we pressed it down firmly, and now that area is opened, so that area is decomposed and now that area is opened, so I think it's decomposed enough so like a little bit more space. If we need to grab something to mix it with like a pencil that we're not gonna use again or something.

From this excerpt we can see that the students only focused on their physical bio-reactor when deciding to make their change. This focus on the physical model continued. They then took measurements of the temperature, pH, and moisture and made other qualitative observations of their bio-reactor. But they never connected these measurements of the current state of their bio-reactor to information they learned in the simulation to provide evidence for why their change would improve decomposition. They simply went outside and worked on mixing the materials up in their bio-reactor, and their conversation focused around how to accomplish this. This lack of

using evidence from the simulation may not have been entirely the students' fault though. The *Change Page* in the students' journal had been intentionally designed to help the teacher monitor students' explanations. Since the teacher needed to approve students' proposed changes, the teacher had the opportunity to check students' supporting evidence prior to making their change. In this instance, the teacher allowed Group B to proceed with their change without providing evidence from the virtual simulation, perhaps due to the many groups needing his attention at that time.

### Group C

Like the students in Group B, the students in Group C's average written explanation score was a 2, slightly lower than Group A's average score. We found that, unlike Group B, the students in Group C did discuss the prompt in the students' journals that asked them to provide evidence for why their change would cause increased decomposition. However, their discussion around this prompt was less focused on using data from the simulation to provide evidence than Group A's discussion, described above. Group C discussed two different potential changes: reducing the moisture in the bio-reactor because it was too wet and adding more carbon rich materials because there were not enough. Group C appeared to misinterpret the instructional prompt in their journal that asked them to provide evidence to support their change: they gave a prediction for how their change would help their bio-reactor, rather than providing evidence from their simulation experiments to *explain why* they should make the change. For example, they could have discussed what they learned about the ideal C:N ratio range to promote decomposition from doing their virtual experiments and how the ratio in their bio-reactor was not in this ideal range.

Student 1: Evidence for why this change will cause improved decomposition.

Student 2: A better smell. A more normal odor, and faster compost.

Student 3: Well, no I know that, but why would this change be better? It'll create a better smell...

Student 2: Faster, faster decomposition, it just makes it slow and foul odor. Um, what else?

Student 3: I think that's it, right? Because we don't really need to say anything else right?

Student 2: That's good.

After this exchange, Group C decided to add more carbon to their bio-reactor. When the teacher mentioned to the entire class that students could use the virtual simulation to test their proposed change before making it, the students in Group C then more specifically discussed what materials they could add to increase the carbon to their bio-reactor. However, they seemed to ignore the teacher's suggestion and never tested their ideas in the simulation or mentioned data from the previous simulation experiments they ran about the ideal C:N ratio range, like Group A did with the ideal moisture range. It was not until the teacher visited with Group C individually and prompted them to further explain what they thought would happen by adding carbon rich materials that the group specifically discussed improving the C:N ratio of their bio-reactor:

Teacher: ...it says explain why, explain specifically what you expect to happen... Don't just say something like 'we think it'll help', or 'the process will work more efficiently', explain specifically what you expect to happen.

Student 2: If we say evening out the carbon to nitrogen ratio, would that be good? Ok.

...

Student 1: I bet, I bet if we add carbon we will be adding more materials and carbon to even out the nitrogen to carbon ratio.

Despite receiving multiple instructional prompts from the journal and the teacher, Group C never discussed data from their simulation to provide evidence for their change. These students had previously experimented with the virtual simulation to learn about the ideal C:N ratio range for decomposition, and they had previously calculated the C:N ratio in their own physical bio-reactors; however, they did not make connections between these two models to propose and support their revision to their physical model.

## **Conclusions and implications**

We were interested in exploring whether and how students integrated information from virtual and physical models to work towards solving a design challenge in the context of a design-based, biology curriculum. This study suggested that some students in a talented and gifted class were able to integrate information from these different models to work towards solving a challenge when provided with several supports, such as prompts from instructional materials, the teacher, and other students in their group. However, even with these supports, about one third of the students' written explanations and two of the three groups' conversations showed a lack of integrating information from the virtual and physical models. The findings from our qualitative analysis of



three groups additionally suggested that even the one group of students who provided information about the simulation in their explanations did not initially integrate this information from different models without the support from external prompts. These findings importantly contribute to our understanding of students' ability to reason about information from different models to inform and justify decisions. Previous research has theorized about integrating virtual and physical models and has shown enhanced conceptual learning from experimentation that blends the affordances of these models over using virtual *or* physical models (Olympiou & Zacharia, 2012). However, little is known about students' ability to reason about and coordinate these different models and the information they provide. Our findings shed light on the challenging nature of this task and suggest the need for additional support. Our findings also align with previous research that identified that students struggle to learn from multiple representations (e.g., Ainsworth, 2006) and justify explanations with evidence (e.g., Sandoval & Millwood, 2005).

Our work further extends the findings of prior research by exploring how students used evidence from different models to write explanations in biology. Students' use of multiple models in biology may be especially difficult, because the time scale between virtual and physical models in biology may add another layer of complexity. For example, many biological processes (such as decomposition) take weeks to observe whereas simulations can be run in seconds. To complicate things more, data collection in biology can be less straightforward than from a virtual simulation because many biological processes are complex systems that are hard to accurately measure. For example, the contents of a bio-reactor are varied, and temperature and moisture levels may be different depending on where students take their measurements. Therefore, the complexity of biological systems may present extra challenges for students to navigate between such diverse models (Hoskinson, Caballero, & Knight, 2013).

Additionally, we found that students seemed to pay more attention to their physical models in their written explanations and conversations about their potential revisions. We conjecture that students may focus more on physical models because they are more concrete and more familiar objects in a science classroom. Perhaps this is also due to the idea that more concrete representations have been shown to better facilitate students' understanding of scientific principles than abstract representations (Goldstone & Son, 2005). This may be because concrete representations offer information that is more salient and connected to the real world, thus better supporting students' reasoning (Goldstone & Son, 2005). Given that the students in our study were asked to make revisions to their physical bio-reactors, students may have seen these concrete physical models as providing more relevant and salient information for making their decisions. This may mean that teachers and instructional materials need to more explicitly discuss the relationships between virtual and physical models in biology to help students make useful connections between the information represented in both models.

The prompts within instructional materials and from the teacher appeared to be crucial in encouraging students to integrate data from virtual simulations to make a decision about revising their physical model. However, students did not utilize these prompts in the same way. First, explicitly pressing students for evidence in the student journal resulted in one group making a connection to the virtual simulation and utilizing that data to justify their change. This same prompt was not sufficient for supporting students in other groups to integrate information from their virtual and physical models, as either they ignored or misinterpreted the prompt. Second, the teacher's suggestion that students could use the simulation to try their proposed changes in the virtual model was instrumental for one group but ignored by another. For the group of students who decided to follow the teacher's suggestion, this facilitation was a key turning point to integrate information from their physical bio-reactors and the virtual simulation; this helped them refine their solution by utilizing the virtual model to test how their change would theoretically affect their compost. Even though this prompt was useful for one group, many students continued to focus only on their physical bio-reactors. The idea that various scaffolds are important for students when engaging in complex tasks is well known (e.g., Puntambekar & Kolodner, 2005); however, this study deepens our understanding of how particular prompts influenced students' reasoning and where students may need additional or different forms of support to successfully engage in this complex task.

While prompts from the teacher appeared to be influential, it seems that it may be difficult for teachers to support students in integrating information from multiple models. Even in a talented and gifted class, it appeared that more numerous and more explicit prompts were needed to support students to improve on such a challenging practice, given that many students did not use information from the virtual simulation in their explanations. Since this is such a difficult task for students and facilitating multiple groups of students is complex for the teacher, designing instructional materials to help students integrate virtual and physical models needs to be more intentional. Specifically, perhaps one way to better support students in making connections between different models would be to design activities that explicitly asked students to use information gathered from a virtual model to inform or revise a physical model, and vice versa. For example, students could be required to use a virtual model to test how changes might affect a physical model. Additionally, teachers likely

need more professional development to better support students' work in this area (Gilbert, 2004; Gnesdilow, Smith, & Puntambekar, 2010), as even an experienced teacher who understood the nature of models could have supported students more successfully.

This study offers an important contribution because it provided a deeper look at *how* groups of students reasoned and negotiated to make connections between physical and virtual models. This information can lead to a more thorough understanding of how many students struggle to accomplish this challenging integration of information. More work needs to be done to understand the kinds of support that students may need to be successful. Given that this study took place in a class of identified talented and gifted students with an experienced science teacher, we wonder how students in a more typical context would perform on the same task. More research is necessary to understand what types of prompts and scaffolds are necessary to support students' collaboration to integrate virtual and physical models. Our future research will examine how more teachers facilitate students' coordination of information from multiple models to solve a design challenge and how different facilitation strategies may impact students' learning.

## References

- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183-198.
- de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science, 340*(6130), 305-8.
- Gilbert, J. K. (2004). Models and modelling: Routes to more authentic science education. *International Journal of Science and Mathematics Education, 2*(2), 115-130.
- Gnesdilow, D., Smith, G.W., & Puntambekar, S., (2010). An analysis of science teachers' classroom discourse relating to the use of models and simulations in physics. In Zacharia, Z. C., Constantinou, M. P., Papaevripidou, M. (Eds.) *Application of New Technologies in Science Education: Proceedings of the International Conference of Computer Based Learning in Science, 141-152*. OELiZK, Warsaw, Poland.
- Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences, 14*(1), 69-110.
- Hoskinson, A. M., Caballero, M. D., & Knight, J. K. (2013). How can we improve problem solving in undergraduate biology? Applying lessons from 30 years of physics education research. *CBE-Life Sciences Education, 12*(2), 153-161.
- Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 371-388). New York: Cambridge University Press.
- McBride, E., Vitale, J., Applebaum, L., & Linn, M. C. (2016). Use of interactive computer models to promote integration of science concepts through the engineering design process. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016* (pp. 799-802). Singapore: ISLS.
- National Research Council. (2012). *A Framework for K-12 science education: practices, crosscutting concepts, and core ideas*. Washington, D.C.: National Academies Press.
- Puntambekar, S., & Kolodner, J. (2005). Toward implementing distributed scaffolding: Helping students learn science from design. *Journal of Research in Science Teaching, 42*(2), 185-217.
- Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23-55.
- Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education, 92*(5), 941-967.
- Zacharia, Z. C., & de Jong, T. (2014). The effects on students' conceptual understanding of electric circuits of introducing virtual manipulatives within a physical manipulatives-oriented curriculum. *Cognition and Instruction, 32*(2), 101-158.
- Zacharia, Z. C., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction, 21*(3), 317-331.
- Zacharia, Z. C., Olympiou, G., & Papaevripidou, M. (2008). Effects of experimenting with physical and virtual manipulatives on students' conceptual understanding in heat and temperature. *Journal of Research in Science Teaching, 45*(9), 1021-1035.

## Acknowledgments

We thank the students and the teacher who participated in this study. The research reported in this study is supported by a grant from the National Science Foundation to the last author.

# High Accuracy Detection of Collaboration From Log Data and Superficial Speech Features

Sree Aurovindh Viswanathan, Arizona State University, sviswa10@asu.edu  
Kurt VanLehn, Arizona State University, kurt.vanlehn@asu.edu

**Abstract:** Effective collaborative behavior between students is neither spontaneous nor continuous. A system that can measure collaboration in real-time may be useful. For instance, it could alert an instructor that a group needs attention. We tested whether superficial measures of speech and user interactions of students would suffice for measuring collaboration. As pairs of students solved complex math problems on tablets, their speech and tablet gestures were recorded. These data and multi-camera videos were used by humans to code episodes as collaborative vs. various kinds of non-collaboration. Using just the speech and tablet log data, several detectors were machine learned. The best had an overall accuracy of 96% ( $Kappa=0.92$ ), which is higher than earlier attempts to use speech and log data for detecting collaboration. The improved accuracy appears to be due both to our analytic methods and to the particular mathematical task, which involves moving objects.

**Keywords:** collaboration detection, cooperation detection, learning analytics

## Introduction

Many projects have worked on the challenge of automating the analysis of interaction among group members. These antecedents will be briefly reviewed by defining two dimensions, *purpose* and *input*, then describing the few systems whose position along these two dimension match the position of the project reported here. The two dimensions are excerpted from several similar multi-dimensional reviews (Diziol & Rummel, 2010; Magnisalis, Demetriadis, & Karakostas, 2011; Rummel, Walker, & Alevan, 2016; Soller, Martinez, Jermann, & Muehlenbrock, 2005; VanLehn, 2016). When a large number of projects could be cited as illustrations of a dimension, only those published most recently will be cited.

The first dimension concerns the purpose or function of the collaboration measure. That is, what does the system do with the output of the collaboration detector? This dimension has the following categories: Clustering, Classification, Mirroring, Meta-cognitive, Guiding, Orchestration and Restructuring. Our project fits into the *Classification* category. Projects in this category (e.g., Anaya & Boticario, 2011; Chounta & Avouris, 2012, 2014; Dragon, Floryan, Woolf, & Murray, 2010; Gweon, Agarawal, Raj, & Rose, 2011; Gweon, Jain, McDonough, Raj, & Rose, 2013; Martinez-Maldonado, Dimitriadis, Martinez-Mones, Kay, & Yacef, 2013; Martinez-Maldonado, Kay, & Yacef, 2013a; Martinez-Maldonado, Wallace, Kay, & Yacef, 2011; Martinez-Maldonado, Yacef, & Kay, 2013) used human judges to code group interactions into a variety of collaboration categories, then used supervised machine learning methods to induce classifiers (also called detectors) whose accuracy was measured and reported to researchers.

The second dimension classifies prior work by input to the detector. All projects had students working on a shared workspace, so they included the users' interactions (log data) as one input. Most projects also analyzed some form of communication among group members. The communication input can be classified as:

- Group members communicated in a formal language (Tedesco, 2003).
- Group members used a small set of buttons to express agreement/disagreement (de los Angeles Constantino-Gonzalez, Suthers, & de los Santos, 2003).
- Group members communicated by typing natural language and classifying their contribution using a menu of sentence openers or speech acts. Some systems ignored the text and used *only* the students' classifications of their text (e.g., Bravo, Redondo, Verdejo, & Ortega, 2008; Soller, Wiebe, & Lesgold, 2002)
- Group members communicated via typing (chat), with or without sentence openers. The text was analyzed by human "wizards" (Tsovaltzi et al., 2010), keywords (e.g., Adamson, Dyke, Jang, & Rose, 2014; Dragon et al., 2010; Martinez-Maldonado, Yacef, & Kay, 2015) or machine-learned text classifiers (e.g., McLaren, Scheuer, & Miksatko, 2010; Walker, Rummel, & Koedinger, 2014).
- Group members conversed in unconstrained speech, recorded by individual microphones (Bachour, Kaplan, & Dillenbourg, 2010; Gweon et al., 2011; Gweon et al., 2013; Martinez-Maldonado,

Dimitriadis, et al., 2013; Martinez-Maldonado, Kay, et al., 2013a; Martinez-Maldonado et al., 2011; Martinez-Maldonado et al., 2015; Roman, Mastrogiacomo, Mlotkowski, Kaplan, & Dillenbourg, 2012).

Our project falls into the last category, where group members' speech was analyzed. Two other projects also developed classifiers with speech input, so they will be described in more detail.

Although our work used machine-learned classifiers based on low level features of speech similar to Gweon et al. (2011; 2013), it differs in several ways. First, the Gweon et al classifiers used only the students' speech while ours used their actions as well. This allowed us to compare the accuracy attained from actions alone, speech alone and both actions and speech together. Second, whereas collaboration was the focal code in both projects, the two project chose different non-collaboration codes. This choice may impact accuracy, so we measured the accuracy of classifiers trained with different combinations of non-collaboration codes.

Our project is similar to one done by Martinez-Maldonado et al. (2015) in using both speech and actions. Their analysis of the participants' speech used a silence detector to convert the speech into a binary feature (present vs. absent). Machine-learned detectors induced from log data and the silence/talk data stream were only moderately accurate (Martinez-Maldonado, Dimitriadis, et al., 2013; Martinez-Maldonado et al., 2011). Increasing the length of the segments to 60 seconds or 90 seconds did not have much impact on accuracy (Martinez-Maldonado et al., 2011). The group also used differential sequence mining to find sequences of speech, silence and action that would reliably split groups into high and low collaboration (Martinez-Maldonado, Dimitriadis, et al., 2013; Martinez-Maldonado, Kay, et al., 2013a; Martinez-Maldonado, Yacef, et al., 2013). However, they did not convert their findings into a collaboration detector and measure its accuracy.

Generalizing across these two projects, it appears that a simple binary analysis of the speech signal into "talk" vs. "silence" may suffice for achieving moderate accuracy in collaboration detection. This was a Martinez-Maldonado et al. finding in several studies, and it is consistent with the strength of such features in the Gweon et al. detectors. However, we do not know how accuracy would change with the addition of other acoustic or prosodic features. Moreover, when speech is analyzed with this enriched feature set, we do not know how accuracy of a single modality, either speech or actions, compares with accuracy when using both modalities. Our project has addressed these questions, as well as studying classification accuracy in a different task domain from those studied earlier.

## Methodology

### System setup

This section describes the hardware and software setup that was used. Although students worked in pairs and sat beside each other at a table, they each had their own tablet. The tablet had active digitizer technology and a stylus that allowed students to write legibly on the 10-inch touchscreen. Students wore headset microphones.

The software used by participants is called FACT, an acronym for the Formative Assessment using Computational Technology (Cheema, VanLehn, Burkhart, Pead, & Schoenfeld, 2016). The FACT user interface mimics the original materials, which were a large poster (about 24" by 36") and paper cards (about 1" by 2") that are moved and eventually pasted onto the poster. Each group of students had one electronic poster. Group members could scroll and zoom independently, and they could edit different parts of the poster at the same time.

### Participants and procedure

The study enrolled 28 paid participants. They were a mix of graduate and undergraduate students from Arizona State University. They were run one pair at a time in a lab. After the experimenter briefly described the problem, the students mostly worked head-down on their own tablets, but occasionally would huddle over one of them. They generally spoke without looking at each other. They worked until the problem was solved, which took 30 to 40 minutes. The same set-up and procedure have been used in approximately 15 classroom studies, but without the microphones.

### The problem to be solved

Students were given a table with 3 columns and 9 rows, and a set of 27 cards to put into the 27 cells of the table. There were three types of cards: Graph cards, table cards and story cards. The graph cards were already positioned in the left column of the table. Students positioned the story cards in the middle column and the table cards in the right column. All cards in a row should describe the same process, which involves Tom making a short journey. (See Fig 1 for one such row).

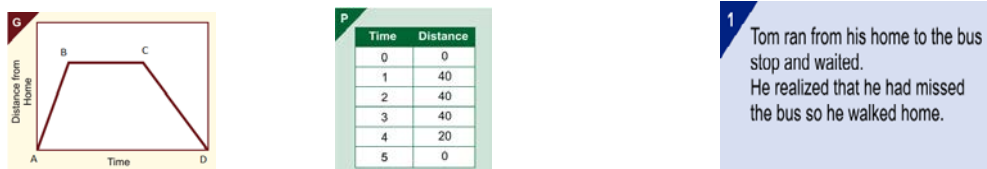


Figure 1. A row in the card matching task describing a story.

In building the row of Figure 1, students should make the following inferences about each points: a) *Point A on the graph card*: Tom is at his home, because the point is at zero on the vertical axis b) *Line segment A-B*: Tom moved away from home rapidly, because the segment has a steep slope c) *Line segment B-C*: Tom waited at the bus stop, because the segment is flat. d) *Line Segment C- D*: Tom returned home, because the segment has a negative slope. e) *Point D*: Tom reached home, because the point is on the x-axis.

The cards are designed around commonly observed misconceptions. For instance, students often view the graphs as a cross-section, so they often match the card in Figure 1 to the following story “*Opposite to Tom’s home is a hill. Tom climbed slowly up the hill, walked across the top and ran quickly towards the down the other side.*” Our subjects found this task rather difficult, but all were able to solve it in less than 90 minutes.

## Raw data collection

The recording setup combined several different input streams.

- Unidirectional headset microphones were used to capture each user’s speech.
- The tablet screen content was streamed to a desktop computer using an HDMI cable.
- Log data were collected at the FACT server.
- Web cameras, one per student, recorded the student’s head and shoulders. The video data were streamed to desktop computer.

The desktop computer showed all four videos on its screen: two tablet screen videos and two head and shoulder videos. It also received the two audio streams. Figure 2 shows a snapshot of the desktop computer’s screen. In order to synch all 6 streams, the desktop screen was saved as a single video. Thus, all the data sources except the log data were synched as they were recorded.

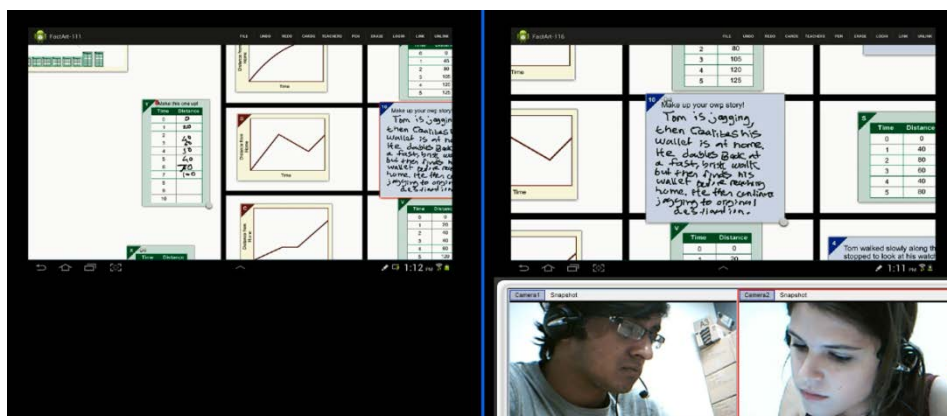


Figure 2. Snapshot of desktop recording student’s action along with gestures.

## Coding categories

When a collaboration monitoring system is used the classroom, it should probably help a teacher make a binary decision—whether to visit a group or not. For this, only two categories are essential: successful collaboration or not. Although Martinez-Maldonado and other prior researchers used a coding scheme developed by Meier, Spada and Rummel (2007), it requires thresholding in order to convert its scores into the binary collaboration/non-collaboration classification. We prefer to leave that decision to the coder. Nonetheless, we are interested in deeper level of granularity, so we defined four codes instead of just two:

- *Collaboration*. The interaction between the pair was considered collaboration when they both worked on placing the same card (i.e., they had the same immediate goal) and they often built on and extended

each other's reasoning. In a few situations, they disagreed with their partner's reasoning instead of extending it, and the pair engaged in argumentative co-construction until they reached agreement. This definition of collaboration includes attributes of joint problem solving such as common ground, knowledge convergence, co-construction, transactivity, scaffolding contributions and have a shared conception of the problem. The following is an example

Student A: It is S [a table card] because all other tables are ending at zero

Student B: No. This cannot be right. The distance can never be decreasing. In that [card] the distance is decreasing with time. In S

Student A: 40...80...60...40...80... [Reading the table card] He is never going back.

Student B: This one is for [the story card where] he has forgotten his watch

Student A: Oh! Okay hmm...

- *Asymmetric Contribution*: The interaction between a pair was considered asymmetric contribution when they were working on placing the same card but one student did most of the work. That is, one person led the conversation and the other person added at most a few reasoning statements. We define two different levels of asymmetric contribution:

- *Asymmetric Contribution (Low)*: The interaction between a dyad was characterized as "low" when no reasoning statements or exchanges occurred, but the human coder could tell from the videos that both students were attending to the same card. The following is an example:

Student A: For this card...

Student B: Yes Tom is... [B moves the card to the solution grid] Yeah done.

Student A : [Head Nod and both moves to the next problem]

- *Asymmetric Contribution (High)*: The interaction between a dyad was characterized as "high" when the active person expressed reasoning and the passive person accepted the reasoning without contributing additional reasoning. The following is an example:

Student A: Probably, the first one will be 20 40 40 [Reading a table card] and it goes to zero. So that table... Because the slope while going up is little longer than the slope while coming down. So T [a table card] goes with E [a graph card]

Student B: Yeah.. Yeah...

This definition of asymmetric contribution shares a few characteristics of joint problem solving sessions such as common ground and establishing a shared conception of a problem. However, it lacks other properties such as transactivity, scaffolding contributions or argumentative co-construction.

- *Cooperation*. The interaction between a dyad was considered cooperation when subjects have different immediate goals, that is, they were working on placing different cards. Although there was usually little or no conversation between the pair, sometimes one student idly chattered about the problem and the other student did not respond back. Since students worked on different immediate goals, cooperation episodes do not have any attributes of joint problem solving.

This coding scheme is similar to ones used by other projects. In addition to a Collaboration code, almost all have noted that students sometimes work independently (our Cooperation code) and that sometimes one student is passive while the other does most of the work (our Asymmetric Contribution code). The distinction between High and Low Asymmetric Contribution is included because episodes where one person is explaining their actions while the other appears to listen (coded as High Asymmetric Contribution) may be consider a form of collaboration. In several studies where asymmetric verbal collaboration was pointed out to participants, the low-verbal participants rejected it as a meaningful measure of their participation (DiMicco, Pandolfo, & Bender, 2004; Roman et al., 2012), which suggests that they considered listening intently to be a form of collaboration.

## Analysis methods

### Data cleaning and segmentation

This section briefly describes our data cleaning, synchronization and segmentation techniques. The unidimensional mics meant that the audio streams obtained from tablets corresponded to single speakers' voice. The background noise was removed by using Audacity. Log data were recorded separately from the audio and video, so all data streams were synched post-hoc manually at the millisecond level.

Segmentation refers to dividing chronological data into segments or episodes (Chi, 1997). It is necessary whenever assigning a code (e.g., “Collaboration” vs. “Cooperation”) to a whole session would be unreliable and perhaps even nonsensical. Although some projects have used overlapping segments (e.g., Gweon, Jain, McDonogh, Raj, & Rose, 2012; Rosé et al., 2008), most projects define segments to partition the whole session. Some projects use constant duration segments, such as 30 seconds (e.g., Martinez-Maldonado, Kay, & Yacef, 2013b; Martinez, Kay, Wallace, & Yacef, 2011). This can harm inter-coder reliability, so aligning segments with naturally occurring discontinuities is often preferred when possible. Our project’s goal was to divide the solution of time-distance problem into separate sub-problems. A sub-problem is considered “done”, and a segment boundary is placed, whenever a story or table card was placed in a table cell and the participant(s) moved on to another card. If the participants came back later and moved that card to a different cell, the new placement was considered a new segment.

### Human coding

Once the segmentation was performed, human annotators classified each segment as either cooperative (P), Low asymmetric contribution (L-A), High asymmetric contribution (H-A) or collaboration (C). In addition to the audio recordings and log data from each participant, which were used as inputs to the machine learner and detector, the human coders viewed a four-video stream (see Figure 2). Thus, the human coders had much richer data than the machine detector. Two human coders tagged a sample of 35% of the segments. Inter-rater agreement was considered acceptable with Cohen’s  $\kappa = 0.78$ . For consistency across the whole dataset, the classifications of one annotator (the first author) were used in subsequent analyses.

### Feature extraction and selection

The goal of the feature extraction process was to obtain superficial features from the students’ work that could potentially differentiate between collaboration, cooperation and asymmetric contribution. Features were extracted computationally from audio files and log files; video data (Fig. 2) were ignored.

The log data were sequences of timestamped events that included card movements, scrolling and zooming. Feature definitions were specific to the user interface and the task. For instance, a sequence of scrolls with limited time between them were categorized as “search” scrolls whereas sequences with larger inter-event times were seen as “reading” scrolls. In addition, some features referenced the past behavior of the students up to this segment. Examples include a card moved twice by the same person to two different cells, a card moved twice by two different students to two different cells, and the total number of card placements so far by the student.

Audio features were extracted using both Audacity (silence and sound features) as well as the OpenSMILE audio feature toolkit, which represents “the state of the art for affect and paralinguistic recognition” (Eyben et al., 2010). We extracted all the features using the Emobase feature set which consisted of 1582 features and has often been used for non-semantic analyses of speech.

Feature selection was performed because the number of features was greater than the number of observations. Pairwise correlations were performed on features likely to be redundant. Sets of highly correlated features (coefficient  $> 0.9$ ) were reduced to a single feature chosen arbitrarily from the set. Next, we applied resampling of the attributes in order to have uniform distribution across class labels. Finally, we applied an attribute selection algorithm using best first search in Weka in order to reduce the feature set further. This reduced set of features was used for inducing the collaboration detectors.

## Findings

The overall goal of the study is to induce a classifier that can distinguish collaboration from several types of non-collaboration. However, there is some ambiguity about how to treat the asymmetric contribution category, so three levels of granularity were defined and classifiers were induced for each.

- *Quaternary*: These classifiers were trained to distinguish all four categories coded by the human annotator. That is, their output was drawn from the set: Collaboration, Asymmetric contribution high, Asymmetric contribution low, and Cooperation.
- *Ternary*: These classifiers lumped together the two Asymmetric contribution categories, so their output was drawn from the set: Collaboration, Asymmetric contribution and Cooperation
- *Binary*: These classifiers lumped together Collaboration with Asymmetric contribution, so their output was drawn from the set: non-Cooperation (a broad definition of collaboration) vs. Cooperation.

### Results from binary classifier

This section reports on the binary classifier, which was trained to discriminate only two categories: Cooperation versus Non-cooperation. We built classifiers for both the audio data alone, the log data alone and both sources of data combined. Random forest yielded the best result for all three data sets. The models were validated using the tenfold cross validation. Contingency tables, accuracy, Cohen's Kappa and F are shown in Table 1.

Table 1: Results from Binary Classifier (NP: non-Cooperation; P: Cooperation)

		Predicted Class (Audio) 93% ( $\kappa= 0.85$ ; F=0.95)		Predicted Class (Log) 92% ( $\kappa= 0.83$ ; F=0.94)		Predicted Class (Combined) 96% ( $\kappa= 0.92$ ; F=0.97)	
		NP	P	NP	P	NP	P
True Class	NP	203	6	199	10	207	2
	P	16	100	15	101	10	106

### Results from ternary classifier

The ternary classifier distinguished between Collaboration, Cooperation and Asymmetric contribution, where low and high asymmetric contribution was lumped together into one category. Additive logistic regression performed the best for the audio and combined feature sets, while random forests yielded the best result for features extracted from logs. The models were validated using the tenfold cross validation. Table 2 shows the results.

Table 2: Ternary Classifier (C: Collaboration; A: Asymmetric contribution; P: Cooperation)

		Predicted Class (Audio) 88% ( $\kappa= 0.82$ )			Predicted Class (Log) 85% ( $\kappa= 0.78$ )			Predicted Class (Combined) 86% ( $\kappa= 0.79$ )		
		C	A	P	C	A	P	C	A	P
True Class	C	88	10	1	80	11	8	82	12	5
	A	12	87	8	7	87	13	12	89	6
	P	1	8	110	4	5	110	1	9	109

### Results from quaternary classifier

This classifier used the same codes as the human annotators. Random forests performed best for audio and combined feature sets while additive logistic regression performed best for log based feature sets. See Table 3.

Table 3: Quaternary classifier (C: Collaboration; H-A & L-A: High & Low Asymmetric; P: Cooperation)

		Predicted Class (Audio) 85% ( $\kappa= 0.79$ )				Predicted Class (Log) 77% ( $\kappa= 0.66$ )				Pred. Class (Combined) 87% ( $\kappa= 0.81$ )			
		Cc	H-A	L-A	P	C	H-A	L-A	P	C	H-A	L-A	P
True Class	C	64	10	5	3	58	1	3	20	74	7	1	0
	H-A	11	27	4	2	3	19	1	21	9	29	3	3
	L-A	5	2	57	3	2	0	45	20	5	1	55	6
	P	0	2	2	128	1	1	1	129	0	3	4	125

With one exception, accuracies were high for all classifiers. The exception is the quaternary classifier that used log data only, whose accuracy was 77%. This makes sense, because the difference between high and low asymmetric collaboration depends on the amount of conversation by the participants, and hence it would have been difficult for the log based features to detect this. However, this explanation predicts confusion between the H-A and L-A categories, and this did not occur with the Log-only detector. Instead, the confusion was spread evenly about the categories.

### Discussion and conclusion

When this project began, we did not think the induced detectors would be accurate because they used only a low-level audio analysis that could not understand what the participants are saying, and low-level log features that could not understand the participants' plans and goals. Against these low expectations, the results were surprisingly good, with accuracies between 85% and 96% (except for the quaternary log data detector, discussed



above). For comparison with prior work, Gweon et al. (2011) found  $F=0.35$  for detecting transactivity, and Martinez-Maldonado et al. (2011) found  $F=0.68$  for detecting collaboration, whereas all our binary detectors had  $F \geq 0.94$ . However, there are some caveats and limitations that should be mentioned.

First, the task used here involved moving objects. Much like the classic collaboration examples of moving furniture or assembling a jigsaw puzzle, when our participants were collaborating, they were moving the same physical object, or at least, one person was moving it and the other was watching and offering comments. On the other hand, when participants were cooperating, they were simultaneously moving different objects. We are not sure how well our method would work when the task does not align sub-problems with object movements. Thus, our next study (which is in progress) uses a task with no moving objects: Two people are sharing responsibility for providing written answers to questions.

A second limitation is that the use of an object-moving task allowed automation of segmentation. Subtask segmentation (i.e., not using segments of constant duration) is usually to be done by a human annotator who can understand the participants' speech, plans and goals (Chi, 1997). We are not sure if subtask-based segmentation can be automated with problems where the subtask boundaries are less salient.

The third limitation was that the audio collection and cleaning used here would not be robust enough for use in classrooms. We are currently working on better methods. Synchronization also needs to be improved, as it currently requires too much human attention.

A fourth limitation is use of log data features that probably do not generalize to other tasks. However, the audio features are task-general, and the audio-only detectors were nearly as accurate as the multi-modal detectors. This suggests that lower-level, task-independent log data features might work nearly as well as the task-specific features.

## References

- Adamson, D., Dyke, G., Jang, H., & Rose, C. P. (2014). Toward an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence and Education*, 24, 92-124.
- Anaya, A. R., & Boticario, J. g. (2011). Application of machine learning techniques to analyze student interactions and improve the collaboration process. *Expert Systems with Applications*, 38, 1171-1181.
- Bachour, K., Kaplan, F., & Dillenbourg, P. (2010). An interactive table for supporting participation balance in face-to-face collaborative learning. *IEEE Transactions on Learning Technologies*, 3, 203-213.
- Bravo, C., Redondo, M. A., Verdejo, M. F., & Ortega, M. (2008). A framework for process-solution analysis in collaborative learning environments. *International Journal of Human-Computer Studies*, 66, 812-832.
- Cheema, S., VanLehn, K., Burkhart, H., Pead, D., & Schoenfeld, A. H. (2016). *Electronic posters to support formative assessment*. Paper presented at the CHI'16: Extended Abstracts.
- Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The journal of the learning sciences*, 6(3), 271-315.
- Chounta, I.-A., & Avouris, N. (2012). Time series analysis of collaborative activities. In V. Herskovic, U. Hoppe, M. Jansen, & J. Ziegler (Eds.), *Collaboration and Technology*, vol. 7493 (pp. 145-152). Berlin: Springer.
- Chounta, I.-A., & Avouris, N. (2014). It's all about time: Towards the real-time evaluation of collaborative activities *IEEE 14th International Conference on Advanced Learning Technologies* (pp. 383-285): IEEE.
- de los Angeles Constantino-Gonzalez, M., Suthers, D. D., & de los Santos, J. G. E. (2003). Coaching web-based collaborative learning based on problem solution differences and participation. *International Journal of Artificial Intelligence in Education*, 13(2), 263-299.
- DiMicco, J. M., Pandolfo, A., & Bender, W. (2004). *Influencing group participation with a shared display*. Paper presented at the Computer Supported Collaborative Work, Chicago, IL.
- Diziol, D., & Rummel, N. (2010). How to design support for collaborative e-learning: A framework of relevant dimensions. In B. Ertl (Ed.), *E-collaborative knowledge construction: Learning from computer-supported and virtual environments* (pp. 162-179). Hershey, PA: IGI Global.
- Dragon, T., Floryan, M., Woolf, B. P., & Murray, T. (2010). *Recognizing dialogue content in student collaborative conversation*. Paper presented at the Intelligent Tutoring Systems, Berlin.
- Eyben, F., W, M., #246, Ilmer, B, #246, & Schuller, r. (2010). *Opensmile: the munich versatile and fast open-source audio feature extractor*. Paper presented at the Proceedings of the international conference on Multimedia, Firenze, Italy.
- Gweon, G., Agarawal, P., Raj, B., & Rose, C. P. (2011). *The automatic assessment of knowledge integration processes in project teams*. Paper presented at the Computer Supported Collaborative Learning.

- Gweon, G., Jain, M., McDonough, J., Raj, B., & Rose, C. P. (2012). Predicting idea co-construction in speech data using insights from sociolinguistics *Proceedings of the International Conference of the Learning Sciences. International Society of the Learning Sciences: Sydney, Australia.*
- Gweon, G., Jain, M., McDonough, J., Raj, B., & Rose, C. P. (2013). Measuring prevalence of other-oriented transactive contributions using an automated measure of speech style accommodation. *International Journal of Computer-Supported Collaborative Learning, 8*(2), 245-265.
- Magnisalis, I., Demetriadis, S., & Karakostas, A. (2011). Adaptive and intelligent systems for collaborative and learning support: A review of the field. *IEEE Transactions on Learning Technologies, 4*(1), 5-20.
- Martinez-Maldonado, R., Dimitriadis, Y., Martinez-Mones, A., Kay, J., & Yacef, K. (2013). Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop. *Computer-Supported Collaborative Learning, 8*, 455-485.
- Martinez-Maldonado, R., Kay, J., & Yacef, K. (2013a). An automatic approach for mining patterns of collaboration around an interactive tabletop. In K. Yacef (Ed.), *Artificial Intelligence in Education, AIED 2013* (pp. 101-110). Berlin: Springer-Verlag.
- Martinez-Maldonado, R., Kay, J., & Yacef, K. (2013b). An automatic approach for mining patterns of collaboration around an interactive tabletop *Artificial Intelligence in Education* (pp. 101-110).
- Martinez-Maldonado, R., Wallace, J., Kay, J., & Yacef, K. (2011). Modelling and identifying collaborative situations in a collocated multi-display groupware setting *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 196-204). Berlin: Springer.
- Martinez-Maldonado, R., Yacef, K., & Kay, J. (2013). *Data mining in the classroom: Discovering groups' strategies at a multi-tabletop environment*. Paper presented at the Educational Data Mining.
- Martinez-Maldonado, R., Yacef, K., & Kay, J. (2015). TSCL: A conceptual model to inform understanding of collaborative learning processes at interactive tabletops. *International Journal of Human-Computer Studies, 83*, 62-82.
- Martinez, R., Kay, J., Wallace, J., & Yacef, K. (2011). Modelling Symmetry of Activity as an Indicator of Collocated Group Collaboration. In J. Konstan, R. Conejo, J. Marzo, & N. Oliver (Eds.), *User Modeling, Adaption and Personalization* (Vol. 6787, pp. 207-218): Springer Berlin Heidelberg.
- McLaren, B., Scheuer, O., & Miksatko, J. (2010). Supporting collaborative learning and e-discussions using Artificial Intelligence techniques. *International Journal of Artificial Intelligence and Education, 20*, 1-46.
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *Computer-Supported Collaborative Learning, 2*, 63-86.
- Roman, F., Mastrogiacomo, S., Mlotkowski, D., Kaplan, F., & Dillenbourg, P. (2012). *Can a table regulate participation in top level managers' meetings*. Paper presented at the Proceedings of the 17th ACM International Conference on Supporting Group Work
- Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 237. doi:10.1007/s11412-007-9034-0
- Rummel, N., Walker, E., & Alevin, V. (2016). Different futures of adaptive collaborative learning support. *International Journal of Artificial Intelligence and Education, 26*, 784-795.
- Soller, A., Martinez, A., Jermann, P., & Muehlenbrock, M. (2005). From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence and Education, 15*, 261-290.
- Soller, A., Wiebe, J., & Lesgold, A. (2002). A machine learning approach of assessing knowledge sharing during collaborative learning activities. In G. Stahl (Ed.), *Proceedings of Computer Supported Collaborative Learning 2002* (pp. 128-137). Hillsdale, NJ: Erlbaum.
- Tedesco, P. A. (2003). MArCo: Building an artificial conflict mediator to support group planning interactions. *International Journal of Artificial Intelligence in Education, 13*(1), 117-155.
- Tsovaltzi, D., Rummel, N., McLaren, B., Pinkwart, N., Scheuer, O., Harrer, A., & Braun, I. (2010). Extending a virtual chemistry laboratory with a collaboration script to promote conceptual learning. *International Journal of Technology Enhanced Learning, 2*(1/2), 91-110.
- VanLehn, K. (2016). Regulative loops, step loops and task loops. *International Journal of Artificial Intelligence in Education, 26*(1), 107-112. doi:10.1007/s40593-015-0056-x
- Walker, E., Rummel, N., & Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence and Education, 24*(1), 33-61.

# Scripting and Orchestrating Learning Communities: A Role for Learning Analytics

James D. Slotta, Boston College, [slotta@bc.edu](mailto:slotta@bc.edu)  
Alisa Acosta, University of Toronto, [alisa.acosta@utoronto.ca](mailto:alisa.acosta@utoronto.ca)

**Abstract:** This paper describes our efforts to add structure and formalism to the design of a CSCL curriculum for high school science—integrating individual, collaborative and whole-class inquiry activities into a coherent “learning community.” A pedagogical model called Knowledge Community and Inquiry (KCI) guided our design of a curricular sequence in which one activity feeds into the next, responding differentially to students, and scaffolding new forms of interaction. We include real-time analysis of student interaction data as a source of input into the orchestration of complex scripts, which can influence the assignment of students to groups, the distribution of materials or sequencing of activities. It can also be used to determine which groups may need help, to provide groups with formative feedback, and to provide the instructor with information concerning student groups. The primary outcome of this paper is the design itself, which is evaluated in terms of its theoretical coherence.

## Introduction

By now, most educators have heard about the need to foster “21st century knowledge skills,” such as critical thinking, collaborative problem solving, and evidence-based reasoning (Hargreaves, 2003; Pellegrino, Hilton, & others, 2013). The world of science, in particular, has become infused with new technologies and information practices, data-intensive methods and large, multidisciplinary collaborations distributed across space and time (e.g., the Human Genome Project, astronomical mapping, climate tracking). In response, many have argued that traditional modes of science instruction are inconsistent with the task demands of science and the wider STEM workplace (Collins & Halverson, 2010; diSessa, 2001). Thus, science education should help students develop relevant skills and literacies, in addition to basic skills and factual knowledge (NGSS, 2013).

Science educators have responded to this challenge, exploring new modes of learning and instruction. “Active Learning” (Bishop & Verleger, 2013; Charles et al., 2011; DeLozier & Rhodes, 2016), is one such approach that has now engaged many STEM educators, resulting in professional societies (e.g., SALTISE.ca) and university-based centers to support the design of active learning courses. Ruiz-Primo et al. (2011) summarize active learning as comprising four dimensions: (1) conceptually oriented tasks, (2) collaboration, (3) technology, and (4) inquiry based projects. Several studies have now measured the benefits of active learning (Code, Piccolo, Kohler, & MacLean, 2014; Dori & Belcher, 2005; Linton, Pangle, Wyatt, Powell, & Sherwood, 2014). Freeman et al. (2014) performed a meta-analysis of active learning in STEM, finding that exams scores improved by 6% and students were 1.5 times less likely to fail compared with traditional lecture approaches.

Despite this evidence of success, however, active learning remains largely ill-specified and difficult to study with any control (Brownell, Kloser, Fukami, & Shavelson, 2013; Ruiz-Primo et al., 2011). For example, while specific group strategies are often invoked (e.g., cooperative learning, groups, gallery walks, collaborative projects, problem solving or case based learning) very little definition is provided about the learning processes, materials or assessments, nor about the instructor’s role during the activities (Henderson & Dancy, 2007). Simply naming those collaborative approaches fails to provide sufficient detail about the content, structure or sequencing of activities. What makes a hands-on lab activity effective? When should it be used within the active learning sequence? How will students collaborate, and to what end? How should design projects be structured, and how should they be assessed? Practitioners and researchers require more detail about the curricular designs, in order to develop a deeper understanding of active learning.

This paper describes our recent efforts to add structure and formalism to the study of active learning, as we co-designed (i.e., with the teacher) a new high school biology curriculum that integrated individual, collaborative and whole-class inquiry activities into a coherent “learning community” design. While our curricular design is currently in the process of being enacted and studied as part of a broader research program, this paper focuses on the role of a pedagogical model in guiding a CSCL design. In that regard, the paper has a theoretical focus, although the specific design (i.e., of our active learning biology curriculum) can be considered an empirical outcome. We focus on the important questions of *how* a curriculum design can be constructed in a principled way that weaves together the different forms of activities into a coherent sequence, in which one activity feeds into the next, responding differentially to students, and scaffolding new forms of interaction for instructors.

Thus, the primary outcome of this paper is the design itself, which can be evaluated in terms of its theoretical coherence.

We begin with a discussion of the theoretical perspective of learning communities, including our own theoretical framework, called Knowledge Community and Inquiry (KCI). We then introduce important notions of scripting and orchestration, and the role of a formal model in guiding the design of CSCL curricular scripts. We include an emphasis on the real-time analysis or processing of student interaction data, as a source of input into the orchestration of complex scripts (e.g., where students' contributions on one activity may determine the condition or materials they are assigned to in a subsequent activity). We also focus on the important notion of *group process analytics*, for both scripting and orchestration processes. By introducing such real-time analytics of group process (e.g., whether the group is progressing according to the designed activities, whether all members are contributing, etc) we can add an important theoretical capacity to our scripting and orchestration of CSCL curriculum. This can influence the assignment of students to groups, the distribution of materials or sequencing of activities. It can also be used to evaluate, in real-time, the process of a scripted group interaction, in order to determine which groups may be on task, which may need help, to provide groups with formative inputs or feedback, and to provide the instructor with important information concerning the state of student groups. Our design-based research addresses the following questions:

1. What sequences of small and large group activities, including social media and technology-mediated learning, support a community of learners in our courses?
2. How can a learning community approach reinforce the lectures and other course activities, adding structure, coherence, and connections across topics?
3. What is the role of the instructor within these designs? Beyond simply acting as "guide on the side", what forms of classroom discourse must the instructor emphasize? What conditions or markers of progress should be monitored to determine needed discussions or activity transitions?

## Theoretical background

Active learning has become a movement amongst secondary and post-secondary educators (Freeman et al., 2014), founded on constructivist and social constructivist learning principles (Bransford, Brown, & Cocking, 1999), and informed by deep pedagogical expertise within the specific disciplines. In the life sciences, a surge of interest has driven a growing community of scholars, as evidenced by online communities like LifeSciEd.org and the Society for Advancement of Biology Education Research (SABER). While many undergraduate biology educators have advocated flipped classrooms (e.g., Gross, Pietri, Anderson, Moyano-Camihort, & Graham, 2015; van Vliet, Winnips, & Brouwer, 2015), others have cautioned that flipping alone will not improve student outcomes, unless accompanied by effective learning designs in the classroom (Jensen, Kummer, & Godoy, 2015).

One prominent form of active learning in biology education is concerned with the enhancement of whole-class discussion and lectures. The most effective lectures engage students in responding to questions, where the instructor "re-voices" their ideas, blending multiple responses, and bridging to new topics. The nature of instructor-led discourse, sometimes referred to as "accountable talk" (Michaels, O'Connor, & Resnick, 2008), has been a topic of growing interest for educational researchers. In biology as in other disciplines, the use of audience response systems ("clickers") has greatly increased the opportunities for instructor-led discussions that connect to student ideas (Smith et al., 2009). Following the wealth of work from the physics education community (i.e., the use of clickers and peer instruction methods), biology educators are also studying these methods. Giuliodori et al. (2006) incorporated peer instruction discussions four times during each 90-minute physiology lecture, resulting in statistically significant positive gains on qualitative questions. In a paper titled "*Teaching more by lecturing less*," Knight & Wood (2005) reported improved student outcomes in an upper level developmental biology course from the integration of collaborative problem solving and whole-class discussions. Similarly, Gardner & Belland (2012) observed that these various techniques work best when applied synergistically to create an active learning environment for students.

## Theoretical framework: Knowledge Community and Inquiry

Another promising approach to the design of active learning is to consider the entire classroom as a learning community, in which students draw upon their diverse interests and expertise with a common goal. They share the understanding that their learning activities will align to advance the community's cause, while at the same time helping individuals learn, and allowing everyone to benefit from the community's resources (Bielaczyc & Collins, 2005). In a review of learning community models, Slotta & Najafi (2013) articulated three common characteristics: (1) An epistemic commitment to collective advancement, (2) a shared community knowledge base, and (3) common modes of discourse. Several scholars have observed that it is challenging for teachers or researchers to coordinate a learning community approach (Slotta & Najafi, 2013; van Aalst & Chan, 2007). As

observed by Kling and Courtright (2003, p. 221) “developing a group into a community is a major accomplishment that requires special processes and practices, and the experience is often both frustrating and satisfying for the participants.” The limited success or uptake of this approach has been due to the pragmatic and epistemic challenges of shifting from a didactic mode of “knowledge transmission” into one of collective inquiry. But it is also due to the lack of explicit models to guide the design of curriculum where students are interconnected in a progression of individual, small group and whole class activities, creating and consuming materials from a community knowledge base (Slotta & Peters, 2008).

The Knowledge Community and Inquiry (KCI) model was developed to guide the design of such curricula, in which the whole class (or even multiple class sections) work together, with all students held accountable for content learning gains (Slotta, 2014; Slotta & Najafi, 2013; Slotta & Peters, 2008). The model includes principled requirements for (1) a knowledge base that is indexed to the targeted science domain (2) collective, collaborative and individual inquiry activities in which students co-construct the knowledge base and then use it as a resource for further inquiry, and (3) assessable learning outcomes that allow teachers to evaluate student progress. KCI curricula typically span multiple weeks or months, and are developed through a sustained process of co-design (Roschelle, Penuel, & Shechtman, 2006) amongst researchers, teachers and designers. Within KCI curriculum, inquiry activities are designed to engage students individually and in small groups where they make use of their community knowledge base as a resource. The designed curriculum constitutes a “script” that includes student-contributed content, social media, and small-group activities such as design, debate, critique, argumentation and reflection. The script is “orchestrated” by the instructor, who is enabled, in turn, by features within the *physical environment* (e.g., large screen projections of students’ pooled votes, resources or other products) as well as the *technology environment*, which can help track student progress, distribute instructions and prompts, pause students for planned or spontaneous discussions, etc. The orchestration of the script often depends upon in-the-moment decisions by the instructor, whose role is one of collaborator and mentor, responding to student ideas as they emerge, and orchestrating the flow of activities. Teachers are not just a “guide on the side” but rather have an explicitly scripted role at all times, as well as responsibility for overall coordination.

Prior KCI studies have investigated various forms of learning content, activities and environments, including mobile technology applications for student-contributed observations (e.g., forms, photographs, notes, votes, tags), large, projected “emergent representations” of the collective knowledge, and various forms of classroom instrumentation (Cober, McCann, Moher, & Slotta, 2013; Fong et al., 2013; Moher et al., 2015). Students are typically engaged in computer-supported inquiry activities, including note taking, observations, brainstorming, problem solving, modeling and simulation, design and argumentation (Slotta, Tissenbaum, & Lui, 2013). Large projected displays help teachers identify pedagogically meaningful signals from amidst the noise of student contributions, and track the community’s learning progress.

KCI research has produced a technology environment called *Common Knowledge (CK)* that includes server software that captures student contributions (i.e., the knowledge base), and a wide range of Web applications for students and teachers that support the collection, distribution, curation and application of that content. CK is a “bespoke technology,” meaning that it was developed in close alignment with the epistemological commitments of the model, for purposes of the research, and so provides a good fit for the complex activity sequences and dependencies on student interactions that are required by KCI designs. In recent versions of CK, the technology architecture has been improved to allow interoperability with many other platforms, including shared authentication (i.e., using the LTI standard). This allows our designs to include a variety of tools or other platforms, as well as the existing functionality offered by CK or new features that can be readily developed. CK provides a flexible foundation for technology-mediated collective inquiry, which has been extended and applied in the current work, supporting a rich array of biology learning materials, activities and interactions.

## Scripting and orchestration

One area of research from the learning sciences that is central to our designs are the concepts of scripting and orchestration (Dillenbourg & Jermann, 2007; Kollar, Fischer, & Slotta, 2007). Similar to a theatrical script, which specifies all aspects of a play (i.e., stage, props, lines, actions, and attitudes), a *pedagogical script* explicates a learning design in terms of the participants, roles, goals, groups, activities, materials, and logical conditions or determinants of activity boundaries (Fischer et al., 2013). Like its theatrical counterpart, a pedagogical script is only an abstract or idealized description...until it is performed. *Orchestration* refers to the enactment or coordination of the script, binding it to the local context of learners, classrooms, curriculum and instructor, and giving it concrete form in terms of materials, activities and interactions amongst participants (Tchounikine, 2013). Pedagogical scripts are orchestrated in the classroom, online or across contexts (i.e., home, school, or mobile), with the “orchestrational load” shared or distributed across several agents: (1) *the instructor*, who can tell students what to do, pause activities to hold short discussions, or advance the lesson from one point in the script to another;

(2) *the materials*, including text or other media, instructions, or interactive Web sites; (3) *the technology environment*, including online portals, discussion forums, note sharing or wiki environments, Google Docs, etc.; and (4) *the physical learning environment* (i.e., classroom configuration, furniture, walls, lighting). The notions of scripting and orchestration can inform our design of active learning, encouraging specificity about the materials, activities and sequencing, as well as deep understanding about the role of the instructor, and any scaffolding environments (Slotta, 2014).

## Learning analytics

While technology is often invoked as an important ingredient of active learning, the specific role of technology is seldom explicated (i.e., in terms of how it scaffolds individual or group work, the role of technology-enhanced media in student learning, or the best practices for instructors in working with any given technology). Indeed, many current models of Active Learning de-emphasize commitments to specific technologies, focusing on flexible classroom configuration, table-group collaborations with whiteboard surfaces and paper-based problem solving. Students may have their own laptops, and engage with any number of tools and materials, but the technology itself is not intrinsic to the design. While these approaches may be practicable and engaging, they fail to capitalize on promising new media like social networks, learning analytics, user-contributed content, tangible and embodied interactions, and “gameful design” (Fishman & Deterding, 2013). In such approaches, technology can play a central mediating role, supporting functions or features that would not otherwise be possible, connecting students and enabling real-time processing of student interactions (e.g. to inform new groupings or distribution of materials). In the past few years, the field of learning analytics has grown as a specialized discipline focused on “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (Ferguson, 2012, p. 305). One application of learning analytics relevant to scripting and orchestration is in the support of adaptive learning designs (i.e., scripts) in which students’ interactions with technology environments (i.e., click logs, Web form data, uploaded content, tags, votes, etc) are processed in real time to inform their assignment of materials, groups or activities (Lockyer, Heathcote, & Dawson, 2013).

## Methodology

KCI informs the design of inquiry curriculum that engages a community of learners at three levels of granularity: (1) the individual level, (2) the small group level, and (3) the whole class (i.e. knowledge community). As described above, materials and activities at each level are carefully designed to promote the development and re-use of a community knowledge base. Individual activities may engage students in adding content to the knowledge base. Small group activities may divide students according to the levels of an important organizational variable, and ask them to sort and tag the elements in a knowledge base, or to apply the contents in some design project (e.g., designing a solution to some environmental problem). Whole class activities could entail brainstorming, sorting and tagging, or whole class discussions. The activities are all indexed to a common domain model that also provides the structure or indexing of the knowledge base itself. In this way, all activities and assessable outcomes are assured of promoting progress on the targeted science learning goals (Slotta et al., 2013).

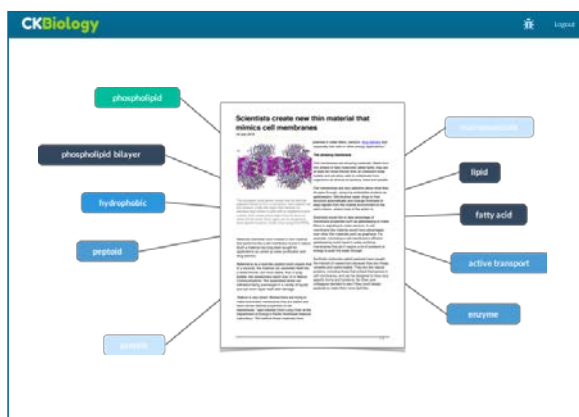
We employed a design-based research methodology (Brown, 1992; Collins, 1992; Edelson, 2002), wherein we worked closely with a high school biology teacher and team of technology developers to co-design an innovative active learning curriculum and corresponding technology environment called *CKBiology*. The structure of our designs was guided by KCI, and is comprised of three distinct elements: (1) *the content model* – specific forms of user-contributed content, Web form elements, votes and tags, photos or other media, emergent learning objects, and connections to course elements like lectures, homework, quizzes or exams; (2) *a process model* – how groups will be formed, roles for students and instructor, content logic, feedback and materials, generation of emergent learning objects, and specific bindings to the content model; (3) *a discourse narrative* – a detailed description of the expected forms of interaction between students, peers and instructors, relating to any materials or activities (i.e., expected discourse patterns and amongst students, peers and instructor, and orchestrational roles for instructor and technology environment).

The articulation of the content model began by defining and parametrizing the content domain of the course, including pertinent aspects of scientific inquiry (e.g., for molecular genetics, identifying the impacts of a mutation), as well as inquiry skills like collaboration and problem solving. Next, we defined a knowledge base, indexed to the domain parameters, to ensure that all student contributions are directly connected to targeted content areas. Finally, we designed the inquiry script, including materials, activities and tools that linked explicitly to the knowledge base, and a community of learners, including students, groups, roles, and any relevant metadata.

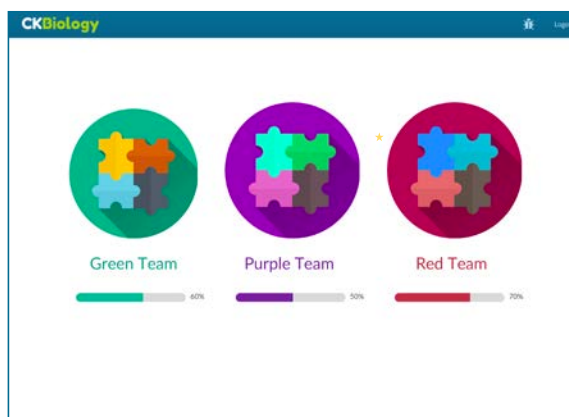
A substantial head start on the technology environment was gained from the existing CK technology, including the capacity for collecting, aggregating and re-distributing any form data (i.e., text entry fields, image



each review challenge team is represented by a group-level progress bar, which is also available on the teacher's dashboard (see Figure 4). Five distinct activity sequences were designed, each indexing to the core concepts, and engaging students in small group applications of that knowledge base. For each, we developed a group process model that tracked groups in terms of their overall process (completeness), successful coordination of the task (fidelity) and equity of participation.



**Figure 3.** Working in small groups, students negotiate how concepts from the knowledge base apply to their chosen article. Shades of blue represent levels of agreement among group members, with dark blue representing strongest agreement. Green tags reflect the end-product of the group's negotiation efforts.



**Figure 4.** Teacher dashboard showing group-level progress bars for the in-class review activity. Clicking on an individual team's icon will display the work that team has contributed so far, giving the teacher a deeper understanding of each group's progress and when/where to intervene.

### Discourse model

The teacher portal to supports at-a-glance information about the state of all groups in the various activities. We delineated specific determinants of teacher-led discourse, such as when no group had made any contributions for a specified amount of time, or when there were a given number of contested relationships in the knowledge base. We also expected extemporaneous discourse, as the teacher noticed opportunistic moments for intervention, based on information she received during small group visits, or by examining the teacher dashboard or student knowledge base. At present, our model for classroom discourse includes three primary dimensions: Small group discussion (students coordinating, with occasional teacher visits), whole class discussions (teacher initiated), and targeted mini-lectures, which emerge in response to revealed student misconceptions or lack of understanding.

### Implications and next steps

The formal specification of learning designs has been elusive. Yet we know from nearly all other science disciplines that formalisms lead to greater progress in research, allowing reliable communication and opening the doors to a wide range of applications. For learning scientists, formal descriptions could allow for comparison of learning designs, or they could inform the creation of taxonomies of pedagogical structure. Without them, we are reduced to deciphering the descriptions offered by course designers (which vary in detail and granularity), to infer the structure of the underlying script (including material design, activity sequencing, dependencies or conditions, etc.). For science educators, it is important that our designs are appropriated widely by colleagues, in part to ensure fidelity of adoption, but also to encourage experimentation and adaptation. This is how innovations spread and evolve. This project hopes to make a contribution to the growing community of biology educators, offering one complete course design that is equipped with an underlying formal structure, adheres to a central pedagogical perspective (learning communities and inquiry), and advances particular forms of collective and small group engagement. Ultimately, the goal would be to support the exchange, uptake, adaptation and critical evaluation of such design, nurturing a learning community of biology educators who build their own knowledge base of innovative designs, validated assessments and shared understandings about learning and instruction.

### References

Bielaczyc, K., & Collins, A. (2005). Fostering knowledge-creating communities. In A. M. O'Donnell, C. E. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology* (pp. 37–60). New York: Routledge.



- Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. In *Proceedings of the American Society for Engineering Education National Conference*. Atlanta, GA.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. National Academy Press.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2(2), 141–178.
- Brownell, S. E., Kloser, M. J., Fukami, T., & Shavelson, R. J. (2013). Context matters: volunteer bias, small sample size, and the value of comparison groups in the assessment of research-based undergraduate introductory biology lab courses. *Journal of Microbiology & Biology Education*, 14(2), 176–182.
- Charles, E., Tissenbaum, M., Whittaker, C., Lui, M., Dugdale, M., & Slotta, J. D. (2011). Co-design of Collaborative Collective Knowledge Environment. In *Proceedings of the 9th International Conference on Computer-Supported Collaborative Learning* (Vol. 1, pp. 641–645). Hong Kong: International Society of the Learning Sciences Inc.
- Cober, R., McCann, C., Moher, T., & Slotta, J. D. (2013). Aggregating students' observations in support of community knowledge and discourse. In *Proceedings of the 10th international conference on Computer-supported collaborative learning (CSCL)* (Vol. 1, pp. 121–128). Madison, WI: ISLS.
- Code, W., Piccolo, C., Kohler, D., & MacLean, M. (2014). Teaching methods comparison in a large calculus class. *ZDM*, 46(4), 589–601.
- Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). New York: Springer-Verlag.
- Collins, A., & Halverson, R. (2010). The second educational revolution: Rethinking education in the age of technology. *Journal of Computer Assisted Learning*, 26(1), 18–27.
- DeLozier, S. J., & Rhodes, M. G. (2016). Flipped Classrooms: a Review of Key Ideas and Recommendations for Practice. *Educational Psychology Review*, 1–11.
- Dillenbourg, P., & Jermann, P. (2007). Designing Integrative Scripts. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning* (pp. 275–301). Springer US.
- diSessa, A. A. (2001). *Changing minds: Computers, learning, and literacy*. MIT Press.
- Dori, Y. J., & Belcher, J. (2005). How does technology-enabled active learning affect undergraduate students' understanding of electromagnetism concepts? *The Journal of the Learning Sciences*, 14(2), 243–279.
- Edelson, D. C. (2002). Design Research: What We Learn when We Engage in Design. *Journal of the Learning Sciences*, 11(1), 105–121. [https://doi.org/10.1207/S15327809JLS1101\\_4](https://doi.org/10.1207/S15327809JLS1101_4)
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304–317.
- Fischer, F., Slotta, J. D., Dillenbourg, P., Tchounikine, P., Kollar, I., & Wecker, C. (2013). Scripting and orchestration: recent theoretical advances. In N. Rummel, M. Kapur, M. Nathan, & S. Puntambekar (Eds.), *Proceedings of the Tenth International Conference of Computer-Supported Collaborative Learning* (Vol. 1, pp. 564–571). Madison, WI: International Society of the Learning Sciences Inc.
- Fishman, B. J., & Deterding, S. (2013). Beyond badges & points: Gameful assessment systems for engagement in formal education. In *Proceedings of the Gameful Learning Symposium*. Madison, WI.
- Fong, C., Cober, R. M., Madeira, C. A., Messina, R., Murray, J., Peebles, B., & Slotta, J. D. (2013). Common Knowledge: Orchestrating Synchronously Blended F2F Discourse in the Elementary Classroom. In *Proceedings of the Tenth International Conference on Computer-Supported Collaborative Learning* (Vol. 2, pp. 26–29). Madison, WI.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415.
- Gardner, J., & Belland, B. R. (2012). A conceptual framework for organizing active learning experiences in biology instruction. *Journal of Science Education and Technology*, 21(4), 465–475.
- Ghadiri, K., Oayoumi, M. H., Junn, E., Hsu, P., & Sujitparapitaya, S. (2013). The transformative potential of blended learning using MIT edX's 6.002 x online MOOC content combined with student team-based learning in class. *JUCE Journal*, 8, 14.
- Giuliodori, M. J., Lujan, H. L., & DiCarlo, S. E. (2006). Peer instruction enhanced student performance on qualitative problem-solving questions. *Advances in Physiology Education*, 30(4), 168–173.
- Gross, D., Pietri, E. S., Anderson, G., Moyano-Camihort, K., & Graham, M. J. (2015). Increased preclass preparation underlies student outcome improvement in the flipped classroom. *CBE-Life Sciences Education*, 14(4), ar36.

- Hargreaves, A. (2003). *Teaching in the knowledge society: Education in the age of insecurity*. Teachers College Press.
- Henderson, C., & Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics-Physics Education Research*, 3(2), 20102.
- Jensen, J. L., Kummer, T. A., & Godoy, P. D. d M. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE-Life Sciences Education*, 14(1), ar5.
- Kling, R., & Courtright, C. (2003). Group behavior and learning in electronic forums: A sociotechnical approach. *The Information Society*, 19(3), 221–235.
- Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, 4(4), 298–310.
- Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction*, 17(6), 708–721.
- Linton, D. L., Pangle, W. M., Wyatt, K. H., Powell, K. N., & Sherwood, R. E. (2014). Identifying key features of effective active learning: The effects of writing and peer discussion. *CBE-Life Sci Ed*, 13(3), 469–477.
- Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, 2764213479367.
- Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education*, 27(4), 283–297.
- Moher, T., Slotta, J. D., Acosta, A., Cober, R., Dasgupta, C., Fong, C., ... Peppler, K. (2015). Knowledge construction in the instrumented classroom: Supporting student investigations of their physical learning environment. In *Knowledge construction in the instrumented classroom: Supporting student investigations of their physical learning environment* (Vol. 2, pp. 631–638). Gothenburg, Sweden: International Society of the Learning Sciences Inc.
- NGSS. (2013). *Next generation science standards: For states, by states*. National Academies Press.
- Pellegrino, J. W., Hilton, M. L., & others. (2013). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. National Academies Press.
- Roschelle, J., Penuel, W. R., & Shechtman, N. (2006). Co-design of Innovations with Teachers: Definition and Dynamics. In *Proceedings of the 7th International Conference on Learning Sciences* (pp. 606–612). Bloomington, Indiana: International Society of the Learning Sciences.
- Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impact of undergraduate science course innovations on learning. *Science*, 331(6022), 1269–1270.
- Slotta, J. D. (2014). *Knowledge Community and Inquiry*. Paper presented at the Network of Associated Programs in the Learning Sciences (NAPLES).
- Slotta, J. D., & Najafi, H. (2013). Supporting collaborative knowledge construction with web 2.0 technologies. In N. Lavigne (Ed.), *Emerging Technologies for the Classroom: A Learning Sciences Perspective* (pp. 93–112). New York: Springer.
- Slotta, J. D., & Peters, V. (2008). A Blended Model for Knowledge Communities: Embedding Scaffolded Inquiry. In *Proceedings of the 8th International Conference on International Conference for the Learning Sciences - Volume 2* (pp. 343–350). Utrecht, The Netherlands: International Society of the Learning Sciences. Retrieved from <http://dl.acm.org/citation.cfm?id=1599871.1599914>
- Slotta, J. D., Tissenbaum, M., & Lui, M. (2013). Orchestrating of Complex Inquiry: Three Roles for Learning Analytics in a Smart Classroom Infrastructure. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 270–274). New York, NY, USA: ACM.
- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, 323(5910), 122–124.
- Tchounikine, P. (2013). Clarifying design for orchestration: orchestration and orchestrable technology, scripting and conducting. *Computers & Education*, 69, 500–503.
- van Aalst, J., & Chan, C. K. (2007). Student-directed assessment of knowledge building using electronic portfolios. *Journal of the Learning Sciences*, 16(2), 175–220.
- van Vliet, E. A., Winnips, J. C., & Brouwer, N. (2015). Flipped-class pedagogy enhances student metacognition and collaborative-learning strategies in higher education but effect does not persist. *CBE-Life Sciences Education*, 14(3), ar26.

## Acknowledgments

This work was supported by the Social Sciences and Humanities Research Council of Canada.

# Individual Versus Shared Design Goals in a Graph Construction Activity

Jonathan Vitale, Lauren Applebaum, and Marcia Linn  
jonvitale@berkeley.edu, lauren.applebaum@berkeley.edu, mclinn@berkeley.edu  
University of California, Berkeley

**Abstract:** Technologies can help foster diverse ideas in collaborative learning activities by taking advantage of group members' unique ideas and perspectives. Assigning individual group members to specific tasks may promote this diversity. In this paper, we introduce a graphing challenge, in which student pairs construct graphs to represent the motion of an amusement park ride. We assigned pairs to experimental conditions with either *individual* design goals or *shared* design goals. Analysis revealed that students with *individual* design goals demonstrated deeper engagement with one of the design tasks (i.e., to create a "safe" ride), while this goal was relatively neglected when goals were *shared*. No impact of condition was found on posttest learning; however, students demonstrated overall gains.

## Introduction

In collaborative learning groups, technologies can elicit a diverse range of ideas by facilitating expression of individuals' unique perspectives. Supporting individual voices within groups enables equity, particularly for students from non-dominant cultural backgrounds who may otherwise be reluctant to engage (Rosebery, Ogonowski, Dischino, & Warren, 2010). Moreover, the expression of diverse ideas provides an opportunity for learning when students reconcile conflicting perspectives (Gijlers & de Jong, 2009). In complex, inquiry-based learning activities the process of distinguishing and reconciling conflicting ideas promotes coherent understanding (Linn & Eylon, 2011). Therefore, collaborative learning activities can take advantage of collaboration by eliciting conflicting beliefs, and then managing their resolution. Yet, if group members begin an activity with similar ideas or if a single member adopts a dominant role, these opportunities are limited. For example, Clark, D'Angelo, and Menekse (2009) found that groups manufactured to ensure conflicting ideas between students demonstrated greater learning than groups constructed at random. Alternatively, in cases of intact groups, technology can support more dynamic collaboration by assigning (personalizing) responsibilities for specific students (Kollar, Fischer, Hesse, & Media, 2006). By supporting unique engagement patterns for each group member, these activities provide opportunities for productive discussion of conflicting ideas. In this paper, we present a comparison between individualized and shared goals in a collaborative design activity in which students construct graphs to represent the motion of an amusement park ride.

## Goals in design activities

According to the knowledge integration framework (Linn & Eylon, 2011), activities support learning by guiding students to elicit ideas, discover new ideas, distinguish between these ideas, and reflect on these ideas to develop a coherent understanding. Collaborative activities fit well within this framework because they are able to expose a broad range of student ideas and provide opportunities for distinguishing between potential conflicts. For this reason, curricula developed within the knowledge integration framework are typically intended for small groups of students (Linn & Eylon, 2011). Because learning environments such as the Web-based Inquiry Science Environment (WISE) can assign diverse roles, by name, during knowledge integration activities, they represent an opportunity to investigate how collaborative patterns impact learning.

In particular, in design projects, where students are expected to generate artifacts creatively, patterns of collaboration are likely to have a substantial impact on how artifacts are generated. When materials are limited (e.g., a single computer keyboard), tasks must inevitably be divided among group members. Without external structuring, group members negotiate duties according to personal (i.e., "internal") collaboration scripts about how tasks should be divided (Kollar, Fischer, & Slotta, 2007). In some cases, this leads to equitable participation, but in other cases participation may be unbalanced. To promote equitable engagement the teacher or software may assign specific responsibilities to each group member (Kollar et al., 2006). A common approach is to assign roles, which divide the overall activity into distinct tasks. For example, in an online discussion activity (Cesareni, Cacciamani, & Fujita, 2016) assigned students to roles of "social tutor", "synthesizer", "concept mapper" and "skeptic". While this approach may ensure accountability for each of the group members, it may not take advantage of students' diverse ideas about shared content.

Alternatively, assigning unique priorities or goals within shared tasks is a common practice in business and engineering design activities to encourage diverse perspectives (Détienne, 2006). In cases where inherent tradeoffs exist in the design, assigning individuals to focus on alternative features may make these tradeoffs more salient. For example, in this study we assign group members to focus on conflicting “safety” and “thrill” concerns for an amusement park ride that they are designing. By helping students evaluate and resolve conflicts centering around important structural features of their design, we can direct attention to central issues. In contrast, students may pay attention to superficial characteristics in unguided projects (Hmelo-Silver, Duncan, & Chinn, 2007).

Yet, there is risk in “over-scripting” collaboration (Dillenbourg, 2002). If expectations for a task are prescribed in a step-by-step manner, students may miss the opportunity to recognize and evaluate competing ideas. Furthermore, if assigned responsibilities are not aligned to students’ preferences or internal collaboration scripts, they may resist external scaffolds (Kollar et al., 2007). As a result, creating productive roles for group members requires attention to students’ actual processes when given specific assignments.

## Study goals and significance

We investigate the personalization of design goals within an online inquiry unit conducted in a classroom setting. The online environment provides us with the opportunity to directly present a design goal to an individual student, by name. We compare groups who are randomly assigned to either *individual* or *shared* design goals. Our aim is to investigate whether assignment alters the artifacts that students build and the concepts they learn. In particular, we study the following two research questions:

1. How do *individual* and *shared* design goals impact the artifacts of design?
2. How do *individual* and *shared* design goals impact learning of underlying concepts?

By addressing these questions we seek to contribute a new approach to the collaboration scripts literature (Dillenbourg, 2002) in the context of student design projects.

## Methods

### Participants and procedures

We performed this study with five participating teachers from two schools. All but one of these teachers had prior experience running similar online inquiry projects as part of a research study. In spite of their similar prior experience with collaborative inquiry projects, the teachers differed substantially (by school) on their approach to and acceptance of collaborative activities.

In school A (38% White, 31% Asian, 17% Hispanic, 4% Black, 22% Reduced-price lunch, 12% ELL), three teachers participated in this study: A1 (male, 10+ years of experience, 103 students), A2 (female, 2<sup>nd</sup> year, 125 students), and A3 (female, 1<sup>st</sup> year, 26 students). In each of these teachers’ classrooms, students sat at four-person lab tables and were expected to engage in collaborative inquiry projects throughout the school year. Students in these classrooms performed all learning activities in dyads.

In school B (51% White, 9% Asian, 28% Hispanic, 3% Black, 32% Reduced-price lunch, 7% ELL), two teachers participated in this study: B1 (female, 10+ years of experience, 126 students), and B2 (male, 10+ years of experience, 103 students). In both of these teachers’ classrooms students sat in rows of desks facing the front of the classroom. These teachers both expressed skepticism about collaborative activities (e.g., assuming they promote off-task activity) and preferred to assign tasks individually. At the teachers’ request, students in these classrooms performed the first part of the learning activity (“Graphing Stories”) individually, and then were grouped in dyads to perform the collaborative challenge activity.

In all classes students completed a pretest and posttest individually.

### Materials

All materials were presented in the Web-based Inquiry Science Environment (WISE).

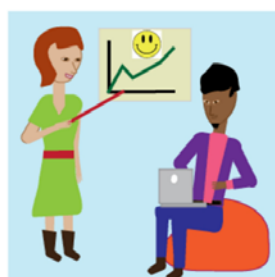
#### Graphing stories

In this set of activities students construct position vs. time graphs to correspond to simple stories of motion (e.g., a hike in the woods). See previous work for more details (Vitale, Lai, & Linn, 2015).

#### Amusement park challenge

In this challenge activity dyads of students are randomly assigned to either *shared* or *individual* conditions to complete the following (<http://wise.berkeley.edu/previewproject.html?projectId=18233>):

1. *Join the team.* In this challenge students design an amusement park ride by constructing graphs of position vs. time. In the *individual* condition, students are uniquely assigned to a single goal, embodied by either the *safety* or *thrill* team (Figure 1). In the *shared* condition, students read instruction that they will be working together for both teams. Following this introduction, students are asked either individually or as a group to describe how a thrilling and a safe ride would differ from each other.



**Thrill Team: Student Name(s)**

On the Thrill Team we try to make the fastest, most exciting rides! We want to get thrill seekers from miles around to visit our park. A rider might get sick once and a while, but that's the cost of making such a thrilling ride!

**Safety Team: Student Name(s)**

On the Safety Team we care about fun too, but we want to make sure everybody is safe. If riders get sick or hurt they won't want to visit our park. Also, we want to make sure our rides are accessible to all kinds of people.



**Figure 1.** Introduction to design goals in *Amusement Park Challenge*. In *individual* condition, each student in the workgroup pair is assigned to either the “thrill team” or the “safety team” by name. In the *shared* condition, both students are referred to, by name, next to each of the teams.

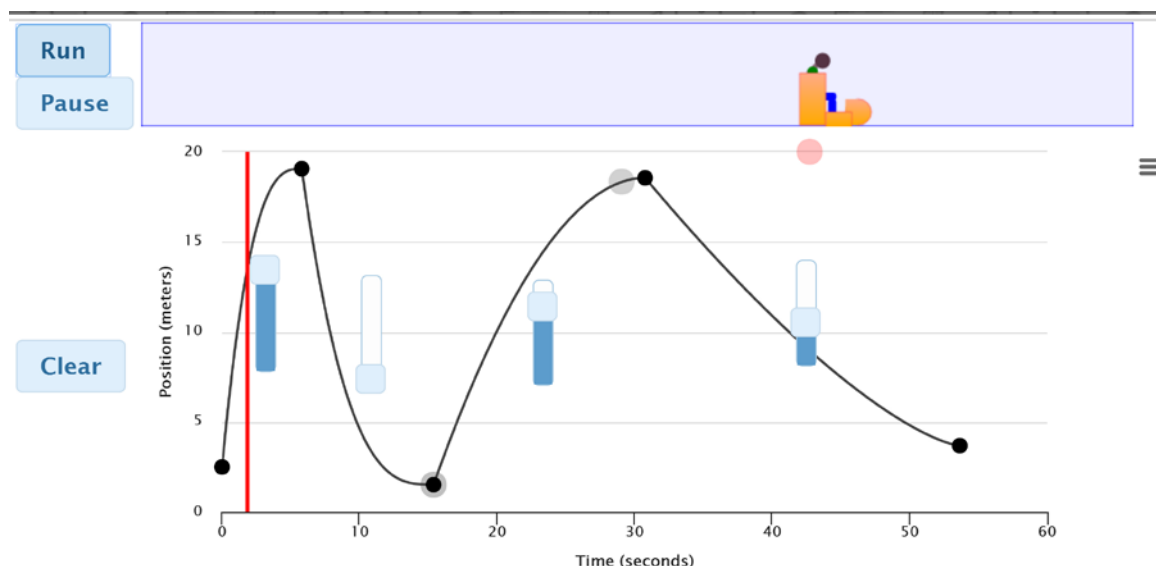
2. *Graphing curves.* Like the *Graphing Stories* curriculum, the primary activity in the *Challenge* is to construct a graph and observe the corresponding motion on a linked simulation. However, building on *Graphing Stories*, students are afforded the additional ability to modify the curvature of segments, thereby impacting acceleration. To introduce this new feature, students are asked to construct a single line segment, modify the curvature in both directions, and observe the impact on the animated ride. Additionally, to emphasize the significance of acceleration, movement of the head of the rider is accentuated forwards or backwards, based upon the acceleration. For example, in Figure 2, the rider is experiencing negative acceleration (slowing down), and is therefore leaning forwards. Individualized roles are not utilized for this step.

3. *Design each ride.* In this step groups design two rides: one that is “thrilling” and one that is “safe”. The students are not given precise criteria, but are expected to follow their own definition of each. Prior to constructing a graph, students are prompted with a question, “What are you trying to design?”, to which they could respond by selecting either “a thrill ride” or “a safe ride”. Following selection, students construct a graph with up to five segments, and then observe the corresponding ride (Figure 2). Following completion of a ride (once the animation was observed), students can press the “New” button to clear the graph. They would again be prompted to indicate the type of ride. In the personalized condition students are asked to construct each of these ride designs individually, although the partner was still available for assistance. In the *shared* condition students are expected to work on all tasks together.

The separate “thrill” and “safe” ride designs are intended to highlight critical relationships in the graphs, including the link between speed and slope, acceleration and curvature. By manipulating graphs for each of these design goals students produce contrasting cases, which illustrate these relationships (Schwartz, Chase, Oppezzo, & Chin, 2011).

When students feel satisfied with each design they are instructed how to download an image of the corresponding graphs. On two later pages, one for each design, they upload these graphs for public display

within a chat forum. Students are prompted to describe uploaded images (and their corresponding rides), and then comment on another groups' ride.



**Figure 2.** Amusement park challenge. Students plot points and adjust sliders to modify curvature. Upon pressing “Run” a simulation of the ride begins. A corresponding red vertical line displays the current time on the graph.

4. *Design your best ride.* Following their experience exploring designs for “safe” and “thrilling” rides, workgroups are prompted to construct a “best” ride that is both fun and safe. In this case, students could construct up to 10 segments. Students are, once again, prompted to save their favorite ride and upload it on a discussion board. There is no reference to individual design goals.

5. *Make a final report.* After completing all graphing steps students are prompted to make a recommendation about how to design rides, with graphs, that are both fun and safe.

#### Pretest-posttest graphing item: “Playing Pool”

Although the pretest and posttest consisted of multiple items, we focus on a single item, *playing pool* (Figure 3), which best aligns with conceptual themes of the *Challenge*. In this item (Figure 3), students are asked to select the graph that best represents a simple story about a pool (billiards) shot. Multiple choice items distinguish students’ understanding of the relationship between slope of segments and speed, and between curvature and acceleration. Distractor items featuring “graph-as-picture” representations are also included.

#### Posttest collaboration survey

To assess students’ understanding of their collaborative strategies we prompted students with the following open ended questions at the end of the posttest:

1. Describe how you and your partner worked together on the Amusement Park Challenge. Did you each take on different roles and responsibilities? If so, describe your role in the group.
2. Describe one example of a time during the Amusement Park Challenge where you and your partner had to make a decision, but each of you had different ideas. How did you make this decision?
3. Describe one example of something you would have done differently during the Amusement Park Challenge if you had been working on your own.

Mel placed a ball about one and a half feet from the edge and hit the ball across the table. The ball went all the way to the end, bounced all the way back, and fell into the hole in the corner pocket. The shot took about 3 seconds.

Which of the following Position vs. Time graphs best represents Mel's shot?

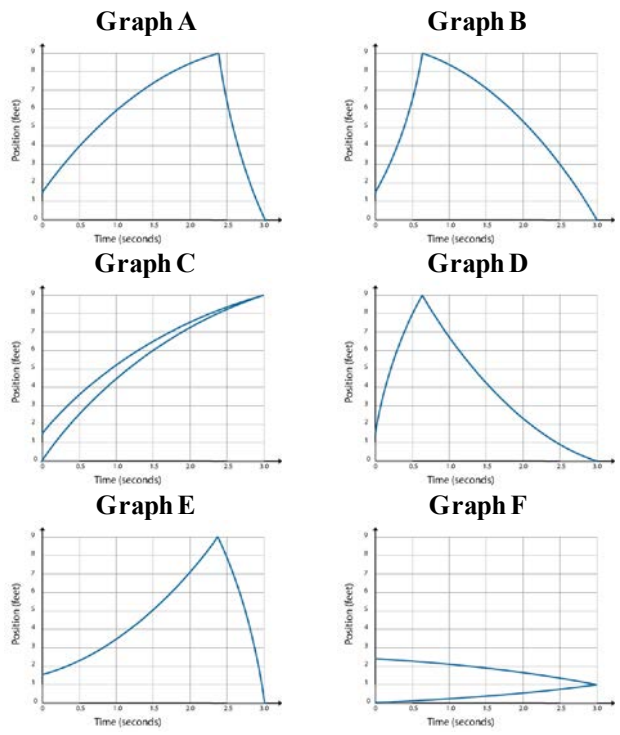
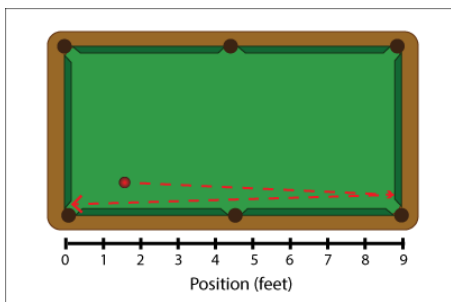


Figure 3. “Playing Pool” pretest and posttest item. The correct response (Graph D) illustrates a faster speed (steeper slope) on the path to the wall than back to the hole and negative acceleration (decreasing speed).

### Analysis method

The pretest-posttest item was scored according to a knowledge integration rubric (scores 1 – 6). This approach has been used in previous graphing applications (Vitale et al., 2015) to emphasize links between narrative elements of the item (e.g., the “speed”) and spatial elements (i.e., the slope).

Graph artifacts were stored digitally by tracking the position of each vertex and the curvature of the segment, as given by the value of the corresponding slider. A graph was marked as complete if students ran the corresponding animation. Using this information, we analyze several features of each graph, including the average angle of segments and the number of segments. Additionally, for *design each ride* data logs indicate whether a graph was intended to be a “thrill” or “safe” ride (i.e., “type”). We analyze the impact of ride type and condition on graph features (e.g., average segment angle). To reflect likely covariation of features for “thrill” and “safe” rides made by a single workgroup we use linear mixed effects models with a random intercept for workgroup, for both continuous and ordinal outcomes. We report on both the statistical significance of predictors as well as the standardized regression coefficients or odds-ratio (for ordinal variables).

### Findings and analysis

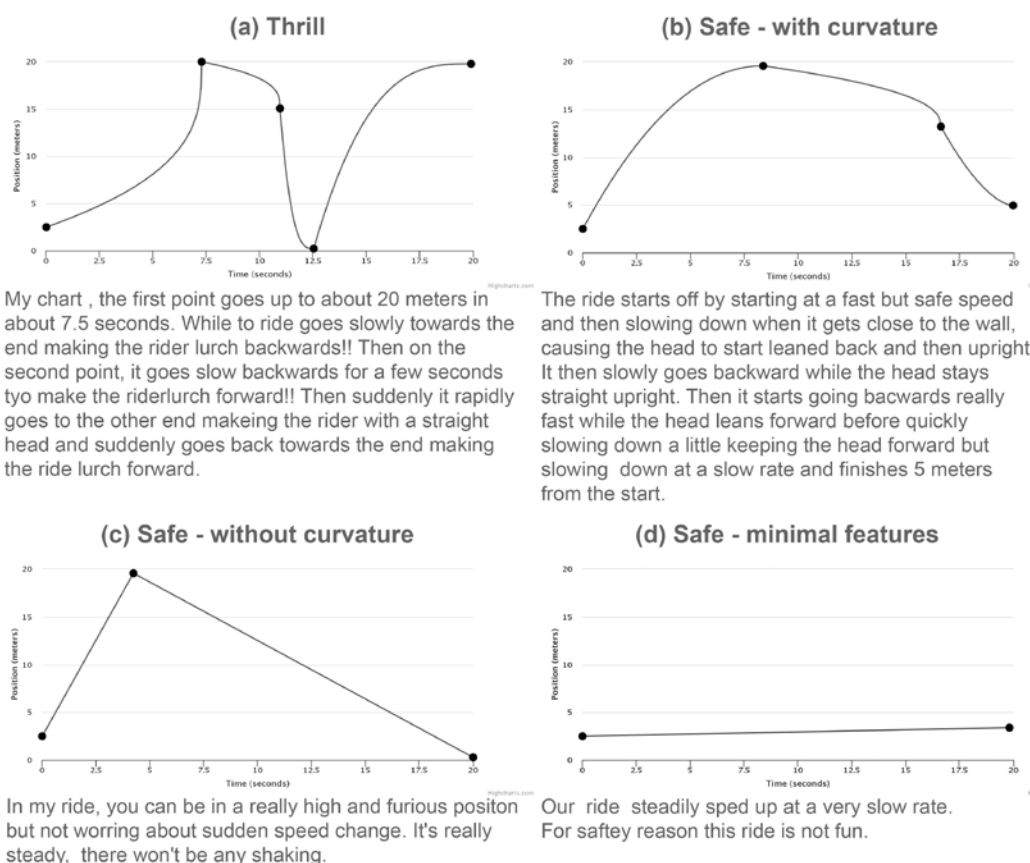
#### Artifact design

To get an initial sense of how students perceived the difference between “safe” and “thrill” rides we analyzed student descriptions of each type of ride in *Join the Team*. After processing the 275 responses to remove common words, we found that for “thrill” rides meaningful, frequent terms included, “fast” (185), “drop(s)” (136), “loop(s)” (124), “upside” (52), “turn(s)” (63), “speed(s)” (64), and “steep” (28). In this case students were often considering features of roller coasters that were not relevant to the ride they were designing (e.g. loops); however, words such as “fast” and “speed” were relevant. Very few terms that clearly relate to acceleration emerged, but perhaps include “sharp” (14) and “sudden” (11). Likewise, for safe rides, frequent terms included, “(seat)belt(s)” (151), “bar(s)” (63), “harness(es)” (24), “speed” (15), and “slow” (13). Clearly, students interpreted “safety” in terms of protecting riders from impact, although some did make reference to speed.

To compare emphasis on each design goal, we computed a count of the number of complete graphs made for each type, for each workgroup, and compared these by condition. In the *individual* condition group

members averaged 3.1 (SD = 2.9) “thrill” graphs and 2.5 (SD = 2.2) “safe” graphs. In the *shared* condition group members averaged 3.6 (SD = 2.8) “thrill” graphs and 1.9 (SD = 2.0) “safe” graphs. A mixed effects model of graph count, using 222 workgroups, reveals a significant effect of graph type (is safe) [ $\beta = -0.31$ ,  $t = -5.9$ ,  $p < .001$ ], no main effect of condition (is *individual*) [ $\beta = -0.08$ ,  $t = -1.2$ ,  $p = .2$ ], and a significant interaction of condition and graph type (is safe and *individual*) [ $\beta = 0.16$ ,  $t = 2.4$ ,  $p = .02$ ]. The interaction indicates that the proportion of safe rides was higher (but still less than ½ of all ride designs) when design goals were *individual* than when they were *shared*.

A higher proportion of safe rides in the *individual* condition suggests more engagement with the safe task. Another proxy measure of engagement is the complexity of the graphs. To analyze this, we restricted analysis to final graphs and categorized complexity by number of segments (0: none, 1: small, 2+: large). An ordinal mixed effects model of complexity, using 209 workgroups with completed graphs, reveals a main effect of graph type (is safe) [odds-ratio = 0.38,  $z = -2.6$ ,  $p = .009$ ], a trend towards a main effect of condition (is *individual*) [odds-ratio = 0.51,  $z = -1.7$ ,  $p = .09$ ], and a significant interaction of condition and graph type [ $\beta = 3.5$ ,  $z = 2.3$ ,  $p = .02$ ]. This indicates while safe rides were likely to be less complex than thrill rides overall, the odds of a more complex safe ride were higher in the *individual* condition than the *shared* condition.



**Figure 4.** Examples of final thrill (a) and safe rides (b, c, d), with descriptions. Rides a, b, and c were designed by students in the *individual* condition, while ride d was designed by a group in the *shared* condition.

Considering students' expressed emphasis on speed as a distinguishing feature of thrill and safe rides, we evaluated the mean (absolute) angle of segments in students' final graphs. A mixed effects model of mean angle, using 209 workgroups with completed graphs, reveals a significant effect of graph type (is safe) [ $\beta = -0.55$ ,  $t = -9.2$ ,  $p < .001$ ], no main effect of condition (is *individual*) [ $\beta = -0.01$ ,  $t = -0.1$ ,  $p > .2$ ], and a significant interaction of condition and graph type (is safe and *individual*) [ $\beta = 0.14$ ,  $t = 2.0$ ,  $p = .05$ ]. This indicates that while, overall, “safe” rides had less steep slopes, those in the *individual* condition made steeper segments for the *thrill* ride than those in the *shared* condition.

To make sense of these findings we selected examples of final designs from students taught by a single teacher. We chose teacher A2 because her students demonstrated a high level of enthusiasm (they frequently encouraged other students, outside their group, to view their designs), and her large number of students (125)



allowed us to explore a diverse range of artifacts. From these students, we chose four representative examples. The distinguishing features of thrill and safe rides in Figure 4 (a) and (b) are the slope and curvature of segments, not the number of segments. Students who produced these graphs were clearly engaged with the task. Figure 4 (c) also demonstrates a valid safe ride, although this workgroup did not take advantage of curvature. We do not know if it reflected less engagement than Figure 4 (b); however, the student authors note that the ride is “really steady”, perhaps referring to the lack of acceleration during most of the ride. On the other hand, Figure 4 (d) displays a graph that likely indicates superficial engagement with the task. In contrast to instructions, the ride did not progress back and forth at least once. Moreover, the author states that the ride “steadily sped up”, although the actual ride would move at a very slow, constant speed.

It may be the case that students in the *individual* condition produced fewer superficial designs because they took more personal ownership of the task, whereas those in the *shared* condition were more likely to spend their combined efforts on the more appealing task of designing a thrilling ride. Another possibility is that students who were assigned to design the safe ride abandoned their role and produced thrilling rides instead.

## Posttest performance

Overall students’ scores on *playing pool* rose from pretest ( $M = 2.7$ ,  $SD = 1.0$ ) to posttest ( $M = 3.3$ ,  $SD = 1.2$ ), significantly [ $t(431) = 11.0$ ,  $p < .001$ ]. To investigate the impact of condition on learning, we performed an ANCOVA on *playing pool* posttest scores, with Condition as an independent variable and pretest score as a covariate. This analysis shows a significant impact of pretest score [ $F(1, 429) = 167.5$ ,  $p < .001$ ], but no effect of condition [ $F(1, 429) = 0.1$ ,  $p > .2$ ], indicating that both collaborative conditions were equally effective.

To explore whether engagement in the challenge was related to learning we performed an ANCOVA on *playing pool* posttest scores with number of “safe” graphs produced by the students’ workgroup, controlling for pretest score. This analysis shows a significant effect of number of safe graphs [ $F(1, 429) = 26$ ,  $p < .001$ ]. This suggests that students who learned more were more likely to engage in the designing activity. Conversely, since the challenge activity is the only exploration of graph curvature in the instruction, it may be that deeper involvement in the design activity promoted better understanding of non-linear graphs. For example, this student illustrates how her experience during instruction informed her response to the *playing pool* posttest item:

I chose d because when we were doing the amusement park ride problem, the graphs looked the same... The only difference between the two [b and d] was the placement of their curves, which brought me back to the amusement park ride. I recalled that the cart went the fastest at the most inverted part of the curves, or the opposite, depending on whether they curved in or out. I assumed that the two fastest points should be when the ball is originally hit, and when it bounces off the wall. Graph d showed that the ball would start off, and slow down as it reached the top, from there, it would bounce off the wall and speed up for a bit, before slowing down again as it reached 0,0.

A lack of difference between conditions may be due to a number of factors. First, although *individual* goals produced a better balance of thrill and safe rides, it may be that designing either type of ride supported learning. In the case where students designing safe rides chose not to manipulate curvature (like Figure 4, c and d) then learning was more likely for thrilling rides so an imbalance in trials could result in more experience with curvature. Furthermore, while *individual* goals helped to structure collaboration, in many of the groups, collaborative roles emerged spontaneously. In the *shared* condition 19% of groups indicated that they divided up “safe” and “thrill” responsibilities. Additionally, 35% of participants indicated that they alternated turns or constructed alternative roles (e.g. “typer”, “grapher”). Only 6% of students indicated that one group member (themselves or partner), took a dominant role. Thus, spontaneous collaborative strategies may have mimicked the advantages of the collaborative strategies implemented in the treatment.

## Implications

This investigation suggests that personalized design goals can help direct engagement to specific instructional activities – including those that may be valuable, but less appealing to students. Directing students in the *individual* condition to focus on separate goals increased attention to building safe rides. Although students in the *individual* condition were more likely to engage in building safe rides, we did not find that additional focus on this task improved performance on the outcome measure. Future work is needed to determine whether personalizing priorities can foster learning by boosting engagement with tasks that are clearly aligned with learning goals. Future work can also investigate whether an even distribution of task attention could help both partners learn or whether gains are more likely for the assigned student.

More generally, the individualized design goals approach represents a tool by which teachers and designers can establish equitable student partnerships during collaborative activities. This stands in contrast to a division-of-labor strategy in which one student may select a less demanding or gender-stereotyped role. For example, in some studies boys take the role of primary computer user, particularly in game-like settings, to the disadvantage of others (Volman & van Eck, 2001). Rather, by personalizing goals, students are expected to perform tasks that engage in similar conceptual processes. By prompting them to then coordinate between two sets of goals, the students are required to take each other's contributions seriously. As complex projects become more integral in STEM classrooms (NGSS Lead States, 2013), helping students develop both individual responsibility for a project and sensitivity to their partners' ideas is essential to ensuring successful experiences.

## References

- Cesareni, D., Cacciamani, S., & Fujita, N. (2016). Role taking and knowledge building in a blended university course. *International Journal of Computer-Supported Collaborative Learning, 11*(1), 9–39.
- Clark, D., D'Angelo, C., & Menekse, M. (2009). Initial structuring of online discussions to improve learning and argumentation: Incorporating students' own explanations as seed comments versus an augmented-preset approach to seeding discussions. *Journal of Science Education and Technology, 18*(4), 321–333.
- Détienne, F. (2006). Collaborative design: managing task interdependencies and multiple perspectives. *Interacting with Computers, 18*(1), 1–20.
- Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 61–91). Heerlen: Open University of the Netherlands.
- Gijlers, H., & de Jong, T. (2009). *Sharing and confronting propositions in collaborative inquiry learning. Cognition and Instruction* (Vol. 27).
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist.*
- Kollar, I., Fischer, F., Hesse, F. W., & Media, K. (2006). Collaboration scripts — a conceptual analysis. *Educational Psychology Review, 18*(2), 159–185.
- Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708–721.
- Linn, M. C., & Eylon, B.-S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration*. New York: Routledge.
- NGSS Lead States (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academy Press.
- Rosebery, A. S., Ogonowski, M., Dischino, M., & Warren, B. (2010). “The coat traps all your body heat”: Heterogeneity as fundamental to learning. *Journal of the Learning Sciences, 19*(3), 322–357.
- Schwartz, D. L., Chase, C. C., Oppezzo, M. a., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*(4), 759–775.
- Vitale, J. M., Lai, K., & Linn, M. C. (2015). Taking advantage of automated assessment of student-constructed graphs in science. *Journal of Research in Science Teaching, 52*(10), 1426–1450.
- Volman, M., & van Eck, E. (2001). Gender equity and information technology in education: The second decade. *Review of Educational Research, 71*(4), 613–634.

# Who Signs Up and Who Stays? Attraction and Retention in an After-School Computer-Supported Program

Maggie Renken, Jonathan Cohen, Tugba Ayer, Brendan Calandra, and Aeslya Fuqua  
mrenken@gsu.edu, jcohen@gsu.edu, tayer1@student.gsu.edu, bcalandra@gsu.edu, afuqua2@student.gsu.edu  
Georgia State University

**Abstract:** We report findings from a study assessing computer-supported curriculum designed to engage low SES, underrepresented minority middle school students enrolled in an after-school program with collaborative tasks that build 21<sup>st</sup> century skills, particularly related to digital literacy. Early in the program, we collected survey data from participants and from a sample of after-school attendees who decided not to enroll in our program concerning their goals, feelings toward STEM, and experiences with and access to technology. Over the first 7 weeks of programming, we also have collected attendance records. We report findings relating students' individual factors at program onset to their attraction to and retention in our program. Our findings shed light on important issues relevant to the CSCL community and the conference theme, including identifying potential for attrition among students and engaging a diverse pool of students in computer-supported collaborative learning.

## Background

Computer-supported collaboration has the potential to impact student learning in a personalized and engaging way (Jeong & Hmelo-Silver, 2016). Such programs may be particularly suited to increase interest and broaden participation to include underrepresented groups in STEM fields (Margolis, Ryoo, Sandoval, Lee, Goode, & Chapman, 2012; Peterson & Britsch, 2013). Recent initiatives and reports from national funding agencies place emphasis on developing and evaluating the impact of such programs on student outcomes (e.g., *2014 Science and Engineering Indicators*, National Science Board, 2014; *Innovative Technology Experiences for Students and Teachers (ITEST)*, National Science Foundation, 2016). Before we can begin to consider the effect of computer-supported educational programs on students' interest and learning in STEM subjects, we must understand the complex issues associated with attracting and retaining students in these programs.

Attraction and retention are particularly challenging when educational programs are housed in informal settings (e.g., outside of the classroom), in which participation is not compulsory, and when working with adolescents from populations that are typically underrepresented or even marginalized in the targeted STEM domains (Bell, Lewenstein, Shouse, & Feder, 2009; Hernandez et al., 2013). Weisman and Gottfredson (2001) assessed 8 Maryland-based after school programs for youth in grades 4-8 from 1998 to 1999. 80% of their sample self-reported race as Black, or non-White. Although the focus of their work was on relations between at-risk behavior and retention in after school programming, they also found a third of program dropouts reported being bored. The implication is that to recruit and maintain enrollment in such programs, activities must hold participants' interest.

The research reported here starts at a crucial point. First, we analyze patterns in student attendance and determine factors associated with student retention in a computer-based after-school program. Second, we examine differences in factors across a subsample of students who chose to participate in our program and those who did not. Specifically, we consider students' gender, goals, prior experiences, and access to technology as factors that may influence students' decision to participate in our program and to continue attending over time. Prior work supports relations between gender and interest in STEM (e.g., Peterson & Britsch, 2013); between goal orientation and persistence in STEM programs (e.g., Hernandez et al., 2013); and between experiences with and access to technology and STEM achievement (e.g., Judge, 2005). We extend this work to consider these relations in an informal computer-supported program for Black or African American middle school students in an urban setting. We expect the findings we present here and any resulting discourse among researchers with similar aims to advance efforts in line with those of the CSCL 2017 Conference Theme, prioritizing equity and access in CSCL.

## Method

### Study context

Following two beta tests of an online learning environment (OLE), we are currently conducting a pilot test of a semester long curriculum that embeds the OLE. The pilot test described here is the first part of a design-based

study to take place over 3 years to develop and assess informal after-school educational programming that fosters students' 21<sup>st</sup> century skills through their participation in mock technology start-ups which develop products (e.g., mobile applications) addressing some culturally relevant problem space. Our programming is housed within a pre-existing, well-established after-school program at a single school site. The existing after-school program is structured so that students may select activities to participate in from a menu of activities. Our program was listed among others on a flyer describing program offerings and distributed to students. Students were free to self-select their activities for the semester. At the writing of this paper, the semester-long pilot program is still underway. Our data are derived from the first 7 weeks of programming.

## Participants

Twenty-seven students have attended at least one session of programming over the first 7 weeks of programming, and fifteen of these attended more than one session. Eighteen of the participating students completed an online survey and served as a treatment group for the purposes of the current study. Of those, 16 participants reported their race/ethnicity. Of these 16, 100% reported Black or African American as their race/ethnicity. Participants were allowed to select multiple race/ethnicity identifications, and one participant selected Hispanic or Latino in addition to Black or African American. 31% of participants with survey data were male and 69% were female. Of all 27 participating students, 9 (33%) were male and 18 (67%) were female. All students were in middle school grades. Participants' ages ranged from 11-14, which is typical for American middle school grades.

In addition to collecting data with participants enrolled in our program, we also collected data with 21 students in the broader afterschool program who were not enrolled in our program. These participants served as a comparison group. Comparison group participants were those who did not elect to be in our program, and selected another offering instead. 62% of these participants were male and 38% were female ranging in age from 11-14. Of the 18 students who reported their race/ethnicity, 16 self-reported as Black or African American; two students selected Native American in addition to Black or African American.

## Design and procedure

Over the course of 4 sessions for the treatment group and 3 sessions for the comparison group, a team of 1-6 researchers visited the school site to collect data. Students who provided assent to participate in research completed an online survey housed on Qualtrics. Items for each of the survey instruments described below were presented in blocks. Blocks of items were randomized across students. While students worked on the survey at individual computers in a group computer lab setting, students were pulled aside to work individually on a tablet-based Scratch Jr. task with a researcher observer. This task took 10 minutes to complete, and when students finished, they returned to where they left off in the online survey. This procedure was the same for treatment and comparison groups.

## Survey instruments

### Student goals for and interest in the program

Students answered a multiple-choice, multiple-select question that included the following goals for their participation in the program: "I want to have fun," "I want to understand how to do stuff," "I want to be better than my AMAYS groupmates," and "I do not want to fail." A second forced-choice question measured student interest in the program, with choices such as, "I can't wait to get started!" and "I'm not very interested, and I don't want to do it."

### STEM Semantic Survey

The STEM Semantic Survey includes 5 items concerning one's feelings about STEM domains (science, technology, engineering, math, and STEM careers) (see Christensen, Knezek, & Tyler-Wood, 2014). Item responses are presented as dichotomous word-pairs (e.g., "interesting/boring" and "exciting/unexciting"), and students were asked to make selections on a 7-point scale, in which 7 indicated the highest affinity toward the domain.

### Prior experience with technology

Students answered a series of questions about the extent of their prior experience with technology. These items measured the students' technological education and prior use of both software, such as app building, and hardware, ranging from scanners to tablets. Questions about experience (e.g., "Have you ever worked with computer design tools?") were presented as multiple choice questions with "yes," "no," or "I'm not sure," as answer choices. If the student selected "yes," the student then answered a clarifying question about where that experience occurred by

selecting one or more responses, including “at home for fun,” “at home for a project,” “at school for fun,” “at school for a project,” “I’m still not sure,” and/or “some other place.” Students also had the option to elaborate further in a text box. These questions were derived from Barron, Walter, Martin, and Schatz (2010).

### Access to technology

We also asked students where they had access to technology and to what degree. Items regarding electronic access (e.g., “Which tools and electronics do you have at home?”) were presented as multiple choice with the option to select more than one answer. Students were asked to answer questions measuring frequency (e.g. “How often do you use a computer in classes at school?”) by making selections on a 5-point scale in which 1 indicated “never” and 5 indicated “almost every day.” These questions also were derived from Barron, Walter, Martin, and Schatz (2010).

## Results

### Attraction to program: Individual factors in treatment vs. comparison groups

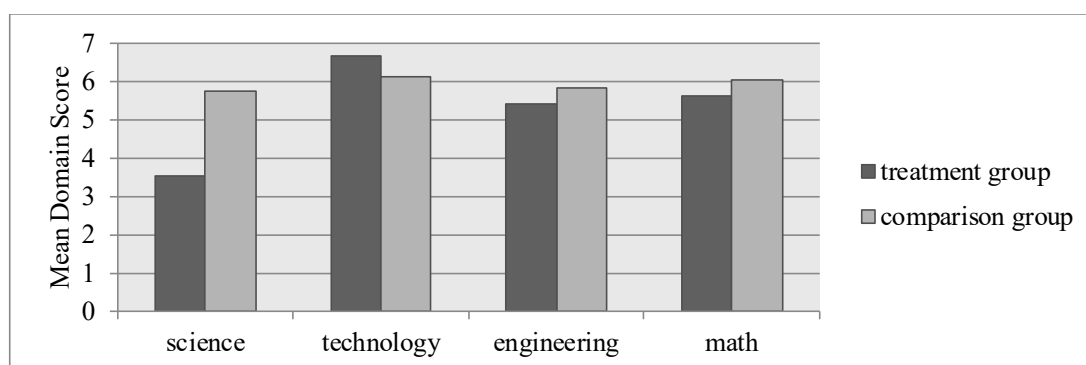
#### Gender

Because students could self-select into our program (i.e., treatment group) or some other afterschool activity (i.e., comparison group), we consider differences in individual factors across the treatment and comparison groups to examine factors related to attraction to our program. With regard to gender, among students who completed the online survey measures, 33% of the treatment group participants was male and 67% was female, while 55% of the comparison group was male and 45% was female. A  $\chi^2$  test of a 2 (condition) x 2 (gender) contingency table indicated that gender composition did not differ across the treatment and comparison groups ( $\chi^2 = 2.03, df = 1, p = \text{not significant}$ ).

#### Students’ feelings toward STEM and attraction to the program

Next, we considered students’ feelings toward STEM domains, according to STEM Semantic Survey, across condition (treatment vs. comparison). We assessed reliability of the domain scales within conditions. The items converged in every domain except the STEM careers domain. We excluded STEM careers from analysis. Cronbach’s alpha ranged from .56 to .94 for the 4 remaining domain subscales (science, technology, engineering, and math).

The mean technology domain score was 6.35, with a standard deviation of 1.54. One participant in the treatment group responded to all of the items in the technology domain with a 1. With this student removed as an outlier, the mean score on the technology scale rose to 6.68 ( $SD = .71$ ), with a minimum score of 4.6 and maximum of 7. Subsequent analysis was run with the outlier excluded. We ran independent samples *t*-tests with condition as the independent variable and mean domain score as the dependent variable for each of the four domains. Science was the only domain for which students’ feelings differed significantly as a function of condition (Figure 1). Participants in the treatment group had more negative feelings toward science than did those in the comparison group ( $t = -4.73, df = 34, p < .001$ ). Although not significantly different, we also point to the trend concerning the technology domain. This is the only domain for which the treatment group expressed more positive feelings than did the comparison group. Within the treatment group only, a *t*-test of domain score as a function of domain (science vs. technology) revealed that mean student feelings toward technology were significantly greater than student feelings toward science ( $t = 5.50, df = 33, p < .001$ ).



**Figure 1.** Students’ mean domain scores for both the treatment and comparison group in all 4 STEM domains.

### Prior experience and attraction to the program

To consider the relation between students' prior experiences with technology and their attraction to the program, we conducted chi-square tests of 2, 2 x 2 contingency tables: have you programmed before (yes vs. no) x condition (treatment vs. comparison) and have you tried to build an app before (yes vs. no) x condition (treatment vs. comparison). There was no difference in the treatment and comparison groups concerning prior experience programming. There was, however, a significant difference in prior experience trying to build apps across condition. The majority of participants in the treatment group (67%) had never tried to build an app before, while the majority of participants in the comparison group (61%) had ( $\chi^2 = 2.79, df = 1, p = .09$ ).

### Access to technology and attraction to the program

Participants in the treatment group reported the most often computer use occurred at their own homes (mean use, treatment group = 4.28,  $SD = 1.18$ ). The average number of computers treatment group participants reported having at home was 2.5 ( $SD = 1.04$ ). A *t*-test of independent samples revealed that neither of these means reported by the treatment group differed significantly from the self-reports of comparison group participants (mean computer use at home, comparison group = 4.06,  $SD = 1.09$ ; mean number of computers at home, comparison group = 2.88,  $SD = .928$ ).

### Patterns in attendance among treatment group

Beyond who decides to sign up for our program, we are interested in the profiles of students who continue to attend once they have signed up. At the writing of this paper, attendance data for the treatment group have been collected for seven weeks of programming. During these 7 weeks, the program staff and students have met 2 days per week for an hour and half per session. Programming has been interrupted with one holiday and one internal school-based conflict, for a total of 12 days, or sessions, of programming. Average attendance for those 12 sessions was 7 students (max 10, min 2, median 8).

Twenty-seven students have attended at least one session in the first 7 weeks of programming. Individual students' attendance ranged from 1 to 7 sessions of programming. In other words, the fewest number of sessions any student attended was 1, while the greatest number of sessions any student attended was 7. On average, students attended 3 days of programming ( $SD = 2.12$ ). Absenteeism between sessions was common for returning students. In other words, days in attendance did not always occur back-to-back. For the 15 students who attended more than 1 session, the average time spent in the program was between 5 and 6 sessions from start to finish ( $SD = 3$ ). Eighty-three percent of the students with online survey data attended more than one session. For students who completed the online survey, the mean total days of programming attended was 3.56 (standard deviation = 2.25, min 1, max 7). For the students who attended more than 1 session and completed the online survey, the spacing of sessions from first attended to last attended ranged from 2 to 12, with mean = 5.69 and standard deviation = 3.03. Overall, patterns in attendance for students who completed the online survey do not differ markedly from the entire sample of students.

To better understand patterns of attendance and how they related to individual factors, we considered who stayed in the program and who did not. This framing results in 3 categories of students: those who continue to attend the program, those who do not continue to attend the program, and those who started attending the program late. In what follows, we refer to these categories of students as *stayers*, *leavers*, and *late starters*, respectively. To quantify who stayed, left, and started late, we set a session midpoint between sessions 6 and 7. We computed the number of sessions students attended prior to the midpoint (pre-midpoint) and following the midpoint (post-midpoint). Stayers ( $n = 10$ ) were students who attended sessions pre- and post-midpoint. Late starters ( $n = 9$ ) were students with attendance only post-midpoint. Leavers ( $n = 8$ ) were students who attended sessions prior to the midpoint only (Figure 2). This distinction allows us to consider the relation between individual factors and attendance, in subsequent analysis, by not only considering total sessions attended, but by also considering differences in students who are retained in the program and those who are not.

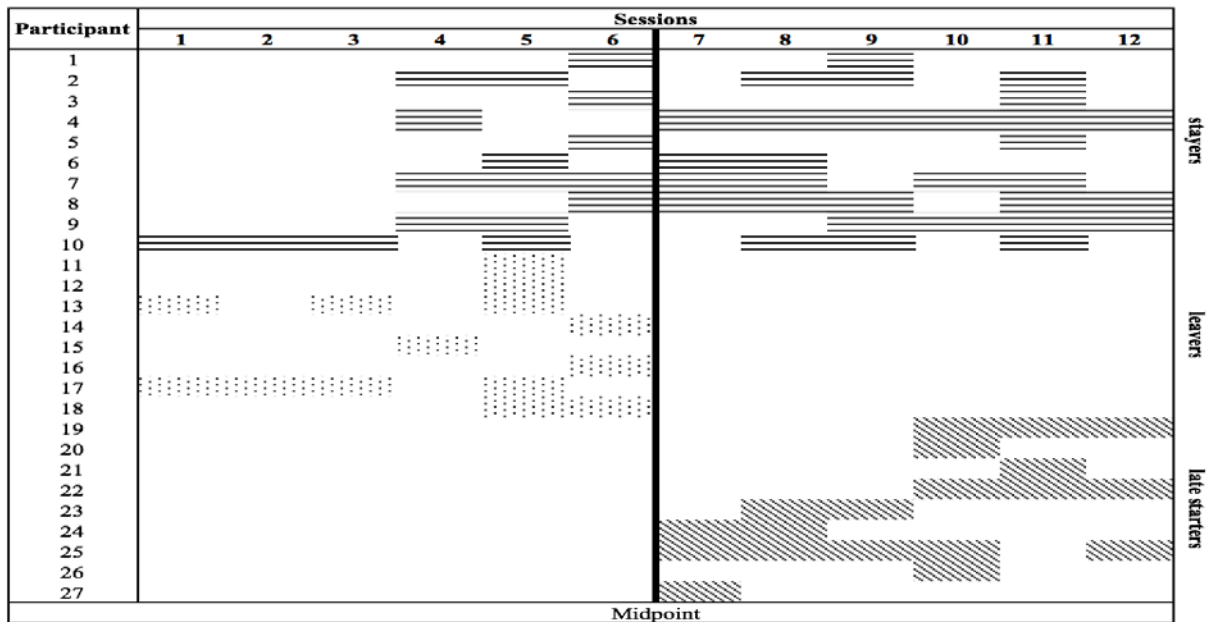


Figure 2. Participant attendance over a 12-session period.

### Retention in program: Individual factors and attendance among the treatment group

#### Gender

Including all 27 participants in the program, even those who did not complete the survey, we considered the relation between gender and retention (Figure 3). More females (18) participated in programming than males (9) overall. A  $\chi^2$  test of a 2 x 2 contingency table considering differences in leavers and stayers across gender revealed there was no significant difference in the proportion of females who stayed in the program and males who stayed in the program ( $\chi^2 = .06, df = 1, \text{Fisher's Exact test } p = 1.00, \text{not significant}$ ).

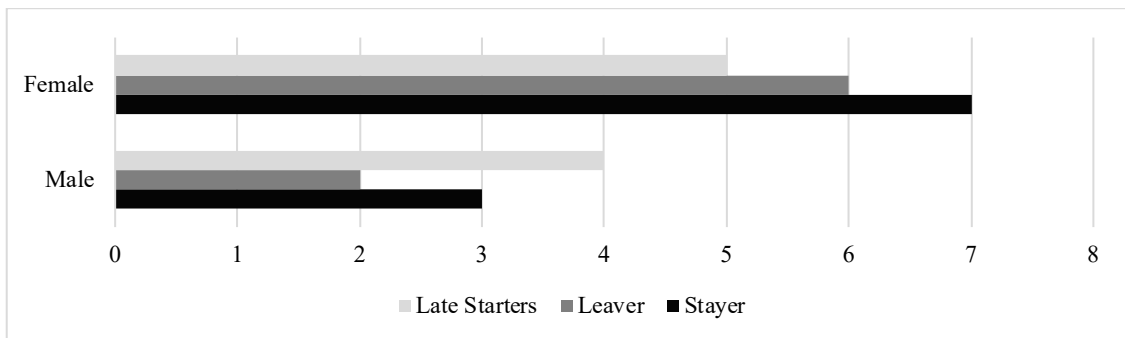


Figure 3. The number of females and males who stayed in the program, who left the program, and who started the program post midpoint.

#### Student goals for participation in program

72.2% of students reported more than one goal for their participation in the program. The 4 goals were not correlated. As demonstrated in Figure 4, the most highly endorsed goals were “I want to understand how to do stuff” (83% of sample endorsed) and “I want to have fun” (72% of sample endorsed), while “I want to be better at this than everyone else in the group” was least often endorsed (by only 33% of sample).

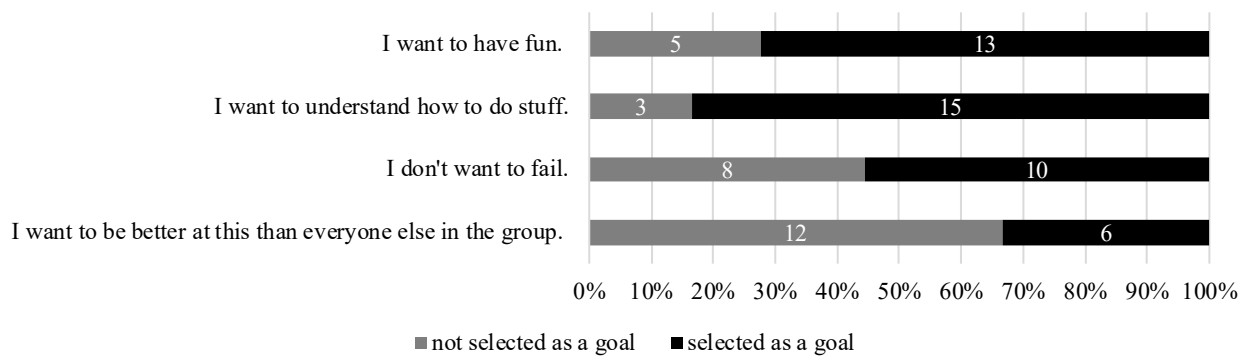


Figure 4. Students' goals for participation in the program.

To determine whether students' goals for participation were related to their retention in the program, we ran four independent *t*-tests to compare mean attendance across students' goals. None of the *t*-tests indicated statistically significant differences in mean attendance. We note that this may be an issue of power due to our small sample size. We also point out a practically important difference (despite no statistically significant difference) in mean attendance for students who indicate they want to understand how to do stuff versus those who do not. Students who want to understand attend almost 5 days on average, while those who do not want to understand only attend an average of 3 days (Figure 5). Hedges' *g* (an effect size measure designed to account for different sample sizes) confirms this is a moderate effect and is .71.

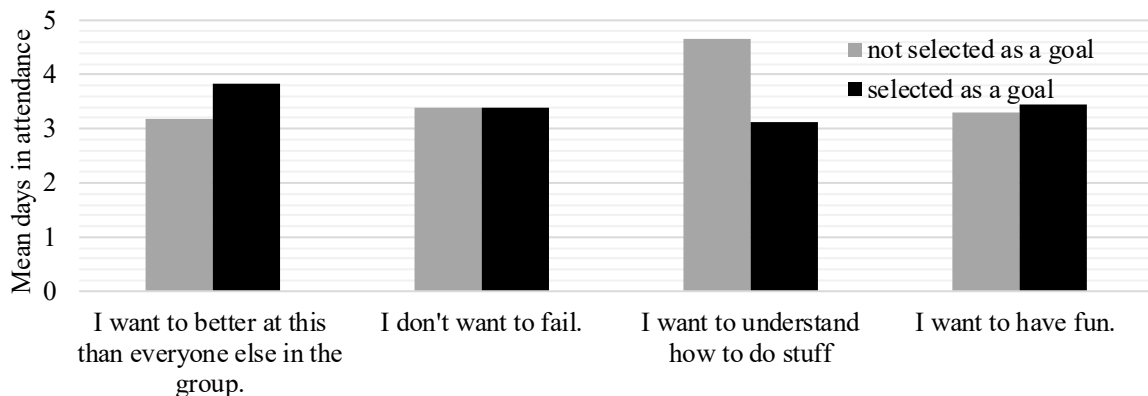


Figure 5. Mean attendance as a function of students' self-reported goals for participation in the program.

#### Students' feelings about the activities and participation

Students were asked to report how they felt about the activities in which they signed up to participate (Table 1). Only 1 of the 3 students who did not report being excited to get started was a leaver, one was a stayer, and one was a late starter. The leaver and stayer answered, "...a little interested, but okay doing something else."

Table 1. Students' feelings about the activities in which they signed up to participate.

Item Response	Frequency	% Selecting
Not very interested, and I don't want to do it.	0	0%
I'm not sure because I don't know enough about what I'll be doing.	1	6%
I'm a little interested, but I'm ok with doing something else.	2	12%
I can't wait to get started!	14	82%

#### Students' feelings toward STEM and participation



We ran one-way analysis of variance (ANOVA) for the 4 STEM domains assessed with the STEM Semantics Survey (science, technology, engineering, and math), with attendance category (leaver, stayer, late starter) as the factor and domain score as the dependent variable. Attendance category was significantly related to domain score for the technology and math domains (technology:  $F = 3.05, df = 2, p = .08$ ; math:  $F = 3.58, df = 2, p = .05$ ). According to a Tukey's posthoc test, late starters ( $Mean = 5.08, SD = 2.49$ ) had significantly less positive feelings toward technology than leavers ( $Mean = 7.00, SD = 0$ ) and stayers ( $Mean = 6.84, SD = .47$ ). Leavers ( $Mean = 3.60, SD = 3.08$ ) had significantly less positive feelings toward math than stayers ( $Mean = 6.38, SD = .76$ ) and late starters ( $Mean = 5.53, SD = 1.61$ ).

#### Prior experience with technology and attendance

We asked participants if they had ever tried to build an app before. 67% (12) said they had not, and 33% (6) said they had. We also asked them if they had programmed or coded before. 56% (10) said they had not, and 44% (8) said they had. We ran an independent samples *t*-test with attendance as the dependent variable and whether or not they had built an app before as the independent variable. Students who had never built an app before attended significantly fewer sessions (mean attendance = 1.67,  $SD = .52$ ) than those who had built an app before (mean attendance = 4.25,  $SD = 2.18$ ) ( $t = 2.82, df = 16, p = .01$ ). There was no difference in attendance as a function of prior experience programming (mean attendance, no programming = 3.50,  $SD = 2.46$ ; mean attendance, programming = 3.25,  $SD = 1.91$ ).

#### Access to technology and attendance

We asked students how often they used computers at home, at school, at a friend's house, or somewhere in the community on a scale of 1-5 (never to almost daily). We computed the mean across these items. The amount of computer use at home and the number of computers owned at home was not significantly correlated with the number of days students attended programming.

### **Conclusions and implications**

Students' feelings about STEM, their goals, and experiences are relevant to their attraction and retention in informal computer-based education programs. Participants in our program reported liking technology more than science. Participants wanted to understand how to "do stuff" and have fun. But they attended less often if their goal was to understand and if they had never built an app before. They were less likely to have built an app before than their peers who did not enroll in the program. However, participants were just as likely to have access to technology as were non-participants, and access was not related to their attendance in the program. We attracted more females than males, but participants were equally likely to leave, stay, or start the program late regardless of their gender. Overall, participants were excited to participate in the program, and those who were attracted to the program from the beginning had the most positive feelings about technology.

We often design programs with strong theoretical underpinnings without thinking carefully about our students as consumers. Such programs are designed to broaden participation but may fall short by not broadening *attraction*. Unfortunately, this means, especially in informal settings, we may be missing the very students such programs are in place to reach and impact. As our research continues, the question will be whether or not their interests deepen or expand over time. Given their positive attitudes coming into a self-selected program, we will need to be especially thoughtful about how to best measure changes in the quality or nature of their interests over time. Tracking emotional and cognitive interests coupled with explicit task meaningfulness descriptions could help preserve program interest and reduce attrition (Hidi and Renninger, 2006). Our findings also seem to suggest curriculum that is *novel* in addition to being interesting may be most attractive and engaging for these students who already have access to rich extra-curricular afterschool programming.

Although retention among females did not differ among males, we attracted twice as many females as males into the program. We expect this may have something to do with the collaborative nature of the program, which may have appealed to participants who have an affinity for communal goals (Diekmann & Steinberg, 2013). Further, students were charged with working in tech start-ups to address a socially relevant problem. As previous research has shown that females often have an affinity for social causes (Paulin, Ferguson, Schattke, & Jost, 2014), this may have been an additional factor in attracting female participants to the program.

Perhaps our most surprising finding was that students who want to understand "how to do stuff" are less likely to be retained. This may have to do with the rigor of the program. However, work avoidance stemming from unexpected increased rigor could negatively impact student engagement in students who want to have fun at the same time (Dowson & McInerney, 2001). Future research should address our loss of students with a more mastery orientation, for instance, by interviewing students who do not stay in the program.

Our study is not without limitations. We do not know about the activities that students who do not select our program enroll in. For instance, competing programs may be attracting males. We expect, given our knowledge of the setting and programming, that males were more likely involved in sports during the fall semester, which conflict with our program, but we do not have data to support this possibility. Programming conflicts are also likely responsible for the negative feelings toward math among leavers. We are aware of a math tutoring session that conflicts with our program, and students are encouraged to attend if they are doing poorly in math at school.

These findings contribute to our understanding of the complex issues associated with attracting and retaining students in an after-school, STEM-focused computer-supported program for urban, low-SES, Black, or African American, students. The national focus on these programs as on-ramps to the STEM pipeline is unlikely to diminish any time soon, and significant attention is being paid, quite correctly, to the design of these programs. However, the ultimate effectiveness of these programs is dependent on getting students in the door and keeping them there, and these findings can help designers to attend to issues of attraction and retention in their curricula.

## References

- Barron, B., Walter, S. E., Martin, C. K., & Schatz, C. (2010). Predictors of creative computing participation and profiles of experience in two Silicon Valley middle schools. *Computers & Education, 54*(1), 178–189. <https://doi.org/10.1016/j.compedu.2009.07.017>
- Bell, P., Lewenstein, B., Shouse, A.W., & Feder, M. (2009). *Learning Science in Informal Environments: People, Places, and Pursuits*. Washington, DC: The National Academies Press.
- Christensen, R., Knezek, G., & Tyler-Wood, T. (2014). Student perceptions of Science, Technology, Engineering and Mathematics (STEM) content and careers. *Computers in Human Behavior, 34*, 173–186. <https://doi.org/10.1016/j.chb.2014.01.046>
- Diekman, A.B., & Steinberg, M. (2013). Navigating social roles in pursuit of important goals: A communal goal congruity account of stem pursuits. *Social and Personality Psychology Compass, 7*(7), 487–501.
- Dowson, M., & McInerney, D. (2001). Psychological parameters of students' social and work avoidance goals: A qualitative investigation. *Journal of Educational Psychology, 93*(1), 35–42. doi: 10.1037//0022-0663.93.1.35
- Hernandez, P. R., Schultz, P. W., Estrada, M., Woodcock, A., & Chance, R. C. (2013). Sustaining optimal motivation: A longitudinal analysis of interventions to broaden participation of underrepresented students in STEM. *Journal of Educational Psychology, 105*(1), 89–107. <https://doi.org/10.1037/a0029691>
- Hidi, S., & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist, 41*(2), 111–127. [https://doi.org/10.1207/s15326985ep4102\\_4](https://doi.org/10.1207/s15326985ep4102_4)
- Jeong, H., & Hmelo-Silver, C. E. (2016). Seven Affordances of Computer-Supported Collaborative Learning: How to Support Collaborative Learning? How Can Technologies Help? *Educational Psychologist, 51*(2), 247–265. <https://doi.org/10.1080/00461520.2016.1158654>
- Judge, S. (2005). The Impact of Computer Technology on Academic Achievement of Young African American Children. *Journal of Research in Childhood Education, 20*(2), 91–101.
- Margolis, J., Ryoo, J. J., Sandoval, C. D. M., Lee, C., Goode, J., & Chapman, G. (2012). Beyond access: broadening participation in high school computer science. *ACM Inroads, 3*(4), 72. <https://doi.org/10.1145/2381083.2381102>
- National Science Foundation (2016). ITEST Program Solicitation. Retrieved from [https://www.nsf.gov/funding/pgm\\_summ.jsp?pims\\_id=5467](https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5467)
- National Science Board (2014). *Science and Engineering Indicators 2014*. Arlington VA: National Science Foundation.
- Paulin, M., Ferguson, R.J., Schattke, K., & Jost, N. (2014). Millennials, social media, prosocial emotions, and charitable causes: The paradox of gender differences. *Journal of Nonprofit & Public Sector Marketing, 26*. 335–353. doi: 10.1080/10495142.2014.965069
- Weisman, S. A., & Gottfredson, D. C. (2001). Attrition from after school programs: characteristics of students who drop out. *Prevention Science: The Official Journal of the Society for Prevention Research, 2*(3), 201–205.

## Acknowledgments

This research is based upon work supported by the National Science Foundation under Grant Number (NSF Grant Number 1433280).

# Creating Parentopia: Design-Based Research to Develop an Interface for Parent Learning Communities and Networks

Susan K Walker, University of Minnesota, [skwalker@umn.edu](mailto:skwalker@umn.edu)

**Abstract:** Participatory design research was employed to create a virtual support and learning interface for parents and staff in an educational program for families and young children. Social constructivist perspectives on learning specific to the parenting role provided theoretical foundation for development. Platform design research validated user interest a complementary space for learning within class and wider program communities, and that integrated involvement of parents' personal social networks. An iterative design process identified tools and activities for initial prototype development, implementation and redesign. The time to identify platform and implementation features necessary to ensure further research proved a benefit. The natural evolution of system roles and supports, comfort with an innovation by parents and staff, and communication about the innovation over time has resulted in a shift in expectations about how learning and support takes place in the program.

## Introduction

Information and communication technologies (ICT) offer parents new ways and a virtual environment for gathering information, problem-solving, forging alliances for support and connecting with experts (Wartella, Rideout, Lauricella, & Connell, 2013). Research suggests that personal and professional connections facilitated by technology benefit parents' emotional well-being, problem-resolution and skill acquisition (e.g., Plantin & Danebeck, 2009). Although short term online educational programs and specific applications (e.g., texts) are effective at strengthening parenting behaviors and knowledge (Hughes, Bowers, Mitchell, Curtiss, & Ebata, 2012; Nieuwboer, Fukkink & Hermanns, 2013), unexplored is the strategic use of technology as an interface for parents' learning communities and that integrates their natural networks that support the childrearing experience. This paper describes the design research behind Parentopia—an online complement to a community-based program for collaborative learning by parents and children in early childhood.

## Design considerations

### Theoretical foundations

Social constructivist, ecological orientations (Bronfenbrenner, 1995; Rogoff, 2003; Vygotsky, 1978) when applied to parent learning and adult development in the parenting role demonstrate the importance of meaningful relationships through which parents can reflect and explore cognitions about childrearing (Azar, 2004; Marineau & Segal, 2006). Learning to parent and about children's development is largely informal and experiential, fueled by cooperative interactions with trusted, familiar and proximal others who share childrearing roles, interests and identities. For some learning is aided by participation in informal parent support communities or more structured education opportunities. In the hands of skilled facilitators, guided interactions for collaborative learning with peers offer the parent critical reflection, validation of the parenting experience, and a deepened identity in the parenting role (Azar, 2004; Belsky, 1984), and identification of diverse perspectives and new strategies for parenting (Campbell & Palm, 2004; Thomas, 1996).

Ecological orientations also promote parent learning as occurring across multiple communities through networked associations (Barron, 2006). Within one's learning ecology the individual self-initiates activities that facilitate learning from one community to another. Repeated exposure to content in multi-community interactions deepens knowledge and improves confidence. And, although parents may rely primarily on proximal, shared identity communities for learning and support, peripheral, weaker tie connections provide novel perspectives and specialized information, and may broker connections to wider networks. Expanded network size, and diversified membership enhances parents' access to a breadth of information and access to social capital that promotes competent parenting (Cochran & Walker, 2005).

These socio-ecological foundations of parent learning offer new directions for technology design. Parents employ a range of technologies to maintain close ties, and to access a range of supports (e.g., Rudi, Dworkin, Walker & Doty, 2014; Stern & Messer, 2009). Connectivity functions of new media can also expose the parent's proximal communities to wider networks. Online communities for parents show promise for learning when it is facilitated in ways that promote safe spaces, and encourage deeper understanding through shared perspectives (Farmer & Reupert, 2013; Gray, 2004; Guldberg & Pilkington, 2006). For parents,

technology can play a critical role as an interface and mediator of multiple community and network member influences.

Hypothetically then, parent learning can be enhanced when social technology:

1. Provides continuity and enhances parents' experience with a community-based, program built on collaborative learning principles through online affordances,
2. Encourages parents' access to their personal social networks to share program content and engage participation in online activities, and
3. Diversifies, increases and strengthens the parents' wider program community for collaborative learning by expanding parents' program connections and affiliations.

### ECFE as a parent learning community for technology design research

Most group parent education delivery models are structured toward improving the parents' knowledge or behavior through short-term experiences (Campbell & Palm, 2004). Some however, are built on relationship-based principles that foster a culture of sharing and sustained support for parents (i.e., knowledge networks) as they evaluate their beliefs and practices (Thomas, 1996). Ties developed through community interactions and responsibilities to the group foster parent self-efficacy and high community collective efficacy. As a result, social bonds are stronger and individuals gain empowerment. Ties developed through learning community interactions also foster collaborative problem solving and a shift in perspective from deeper understanding of diverse viewpoints and experiences.

A program that exemplifies this type of community-orientation to parent learning and support is Early Childhood Family Education (ECFE) in Minnesota ([ecfe.info](http://ecfe.info)). Since 1984, ECFE has operated through local school districts, primarily through no/low cost adult learning and enrichment classes known as Community Education. ECFE is open to all parents of young children from birth through age five. Participation is encouraged throughout the children's first five years, providing families with continuity in learning, peer relationships for support and community resource connections. Weekly two-hour classes are held during the school year. Unlike other group parent education, ECFE does not use an established curriculum; learning goals are tailored to individual classes. Communication fostered through comfortable and trusting relationships moves the group toward synergy characteristic of a learning community (Wenger, 1998). Collaborative learning for parents in ECFE means transformation in perspective from reflective group dialogue. This technique brings the group to shared understanding of the problem, collective determination of alternatives, and individual application to fit unique parenting, child and family needs (Campbell & Palm, 2004; Thomas, 1996).

With multiple classes per site, parents who attend at one class time have the potential to expand their neighborhood/parenting communities through ties with those who attend other classes. ECFE's multi-faceted community-orientation and open systems framework to parent learning, and the program's own lack of presence on the Internet (other than for administrative or marketing purposes) offered a unique opportunity to create a platform that explores technology's role in extending community learning processes among parents, and capitalizes on networking affordances to build larger connections.

### Design tensions

To design an online space for ECFE that would offer, as Preece observes "the virtual as a continuity of the real" (2000, p. 249), means to address the tensions, or endemic dualities and challenges for parents and staff with using online tools in concert with program participation. Introducing an innovation means identifying factors and the process for successful adoption (Rogers, 2003). A significant tension rests with the culture of learning and participation in a 40 year old program that has only operated face to face and challenges to program delivery expectations by parents, staff and administration. A related tension rests with the trust and security felt in a face to face program that protects the sensitivity of parenting discussions and parent fear of judgment.

Another relevant tension is between individual participation and group cohesion important to community learning and engagement. For parents of young children especially, whose attention is divided with life demands, identification with a learning community alone, and one that continues in an online space, is highly variable. Participation in organized instruction by adults is voluntary, and information and socialization needs may be satisfied by other sources. Parent participation with an online platform will vary by perceptions of usefulness to meet needs, application experience (e.g., social media as a successful resource for learning about parenting), and sophistication of technology use (e.g., Rothbaum, Martland & Janssen, 2011). And while parents may be attracted to continue communication with their class communities, engaging with those in the program who are less familiar may take particular encouragement.

Therefore, given the hypothetical value of social technology to parent community-based collaborative learning, and the challenges posed by introducing innovation to an existing learning culture, the research questions driving this project were as follows:

1. Would a virtual platform for the engagement of communities and personal networks be viewed as complementary to existing ways that parents learn through a face to face program?
2. What are the design features that would make a platform useful and used by parents and staff?
3. What is necessary for the implementation and adoption of innovation?

## Design research

The innovative nature of this project required a design assessment method to identify platform components that would convey ECFE principles, reflect parents and staff as users, and accomplish theoretical aims. Design-based research for technology enhanced learning environments (e.g., Wang & Hannafin, 2005) was selected. This method is grounded in relevant research, theory and practice, works with participants interactively and through iterative cycles of analysis and redesign, integrates mixed methods for data collection and analysis and connects results to the specific context.

The neighborhood site for platform development is in an urban Minnesota city, selected due to its 'readiness' for virtual adaptation and adoption by participants. The selected site is an established, familiar presence in the neighborhood, and offers seven classes each week (two are in Spanish). Data collected for this project in year 1 revealed that the majority of those attending this ECFE program are women (26% are fathers), ranging in age 26-58 ( $M=34.2$ ,  $SD=5.5$ ), with slightly more than half (55%) possessing a college degree and most reporting family incomes between \$40K and \$80K. Families have between one and four children; half report two children. The site is in a racially and ethnically mixed neighborhood; 21% of parents report non-white racial identity and although most parents report speaking English (85%), this amount includes parents who are bilingual. Participation in ECFE ranges, with a fairly equal distribution of parents attending their first, second, third, or fourth year. Staff include two licensed parenting educators (each with an average of 20 years of experience), one early childhood educator (with 10 years experience), and three classroom assistants.

## Method

To answer the first research question and identify features for platform design, focus groups and a survey with parents, and staff interviews were conducted. These gathered in depth information on the ways that ECFE helps parents learn as a community, the intersections of parents' personal and ECFE worlds that provide learning and support, and ways that technology might extend ECFE's benefits. Focus group sessions were recorded then transcribed to text for coding, and transcription accuracy validated. Three coders independently coded transcript samples then cross-validated a coding scheme to be used for thematic analysis (Silverman, 1993). A primarily forced-choice survey elicited data on parenting supports, perceptions of value for ECFE, and on technology use. Quantitative analysis included descriptive reporting and internal comparisons.

Within the total sample 52 responded to the survey and 55 participated in the focus groups, representing 83% of the parents registered for classes at the site. Four instructional staff were interviewed. Results of all analyses were shared back for validation and interpretation with staff and parents.

Following build of the prototype, usability testing with 8 parents (5 English speaking and 3 Spanish speaking) provided data for the second research question. Further data on was secured through a survey given to parents in all classes at the end of a 9 month school year (year 3) on site use, and site analytics were tracked to examine pages visited. Forty eight (of 56, or 85.7%) parents completed surveys representing those in English speaking and Spanish speaking classes. Staff interviews provided information for the third research question. All three teachers and three program assistants participated in the interviews. Data from the interviews was transcribed and coded for thematic analysis.

## Results

### Validating assumptions about community learning and network involvement

Analysis revealed that parents do view ECFE as a community for learning and use their personal networks as proximal supports for parenting. Word frequency in focus group descriptions of ECFE indicate the program viewed for 'community' 'learning' and 'support.' Further analysis identify the setting as being 'comfortable,' 'trusted,' 'nonjudgmental' and 'safe,' and even for some, like their 'church.' More than half (55.6%) identified ECFE as a source of emotional support; nearly all (94%) identified it as a valuable source of parenting information. Survey data revealed that discussion with other parents (4.56 out of 5) and expressing views on

parenting (4.33) as most highly rated learning methods. Also valued was getting information on child development (4.15), information on community resources (4.10), and learning activities to do with the child (3.94). The parenting educator was notable for facilitating collaborative discussion and access to expert information. Half (29, 56%) reported knowing no other parents at the site, not in their class well enough to ask for information or advice. Others reported knowing between one and four other parents. Staff interviews indicated that a variety of program-wide activities (e.g., clothing donations for families), events (e.g., spring festival, fund raisers) documents (e.g., program handbook, area guide) and physical spaces (e.g., site building, parent and child classrooms, and the neighborhood) represent an culture of program community engagement.

ECFE parents' personal social network connections represent proximal, strong ties (e.g., family) and distant, weak ties (e.g., professionals). Tie strength was indicated by the frequency of network member mentions during the focus groups, and by survey responses on the range of types of supports offered. Family, the co-parent and friends offer emotional, practical and informational support. Parents verbally shared information from ECFE with others; 81% share with their partner, 57% with a friend and 43% with their own mother/father. Discussion with the co-parent and other parents (in real life and online) helped parents validate beliefs, resolve conflict, boost confidence and get new ideas. Experts (like the pediatrician, and books) were popular information sources for specific topics, though sometimes information conflicted with what was gathered from more familiar sources.

Parents indicated that these network sources provide information that intersect with what is learned in ECFE:

*[parent 1] "...and the same with the people that you come in contact outside of ECFE, they're bringing the knowledge that they have from other sources and experiences that they have..."*

*[parent 2]: And it's definitely expanding your knowledge because we're not talking here, we're not talking about right and wrong, we're talking about..."*

*[parent 1] Ways to deal with situations."*

*[parent 2]: "ways to deal, could be a different idea sometimes and we exchange that and we might go and we go enriched out of the class because what we do, we live relationships wherever we go, you know, we bring that in, in and out. Dynamic."*

These results reinforce the value of collaborative learning through membership in ECFE, and that parents' personal networks and ECFE operate as parallel and intersecting ecologies for their learning. They also validate parents' interest and their role as a conduit for information sharing across communities. And they suggest that the culture of the program offers a foundation for further program-community-family engagement.

## Identifying online engagement activities

Analysis of survey items regarding technology use revealed that nearly all parents and staff were comfortable with and had access to computers, the Internet and cell phones (over 90% reported each). Just under three quarters (74.5%) owned smart phones; few (23%) had tablets. Texting and email were daily activities, and about half (48%) reported videoconferencing weekly (especially with extended family). Social networking (specifically Facebook and Pinterest), sending pictures and using mapping tools daily was reported by just over half. Participants were positive about technology being easy to use (4.25 out of 5), the value of technology to connections with family and friends (4.12) and the general usefulness of technology to their lives as parents (3.96). Analysis of parent and staff technology use and comfort supported use of applications that would be accessible and easy to use, and with English/Spanish translation options. A high need for privacy features (e.g., account approval, restricted access to classes) safety, and convenience (integration with existing social media accounts, mobile access) were recurring themes from focus group analysis of items on technology preferences.

When asked how technology could enhance ECFE's benefit, parents resoundingly asked for ways to maintain peer connections between weekly classes, and to access content about parenting and their child's learning shared by teachers. Activity interests for extending group class-based learning online included open discussion and document sharing, and member lists to connect with other parents. Parents voiced that virtual attendance and participation by the non-ECFE parent is particularly important to boost familiarity with the program and experiences that benefit the child. Parents also wanted easy ways to share the documents and content with others close to the childrearing experience (e.g., child care providers, extended family members). And they wanted easy ways to share information *from* others *to* the ECFE community (such as a website advised by a pediatrician).

Staff admitted little experience or familiarity with using online platforms for parent or child engagement, though they expressed open attitudes to learning and trying new approaches. They were particularly interested in a platform that encouraged wider program engagement and that would include efficiencies to their practice (e.g., staff to parent communication). They reinforced the need for technology that would secure confidentiality and the privacy of conversations, and underscored the value for technology that was easy to use and language flexible. Adherence to school district policies (e.g., use of release forms) was also mentioned.

The findings on technology use and preferences validated the social mechanisms of learning by parents in ECFE, and the value of a virtual platform that would enhance social connectivity to the class learning community, and integrate involvement by personal social network members, offering a clear yes to the first research question.

### Feature usability and usefulness

Based on these findings, platform design considerations leaned on community orientations (Wenger, White and Smith, 2009) of *individual participation, relationship building, access to expertise, content, open-ended conversations, cultivating community and serving the context* to inform tool selection (Table 1).

Table 1: Design aims, considerations and components of Parentopia 1.0 and 2.0

Platform aims	Design Considerations	Community Orientations <sup>1</sup>	Platform components <sup>2</sup>
Facilitate class community learning and support	Maintain preferred activities Convey felt context of learning and community Provide easy connections to class members Privacy and confidentiality	Relationship building Open-ended conversation Content Access to expertise Individual participation	<ul style="list-style-type: none"> <li>Multi-topic discussion forum (R)</li> <li>Chat tool (text, video)</li> <li>Advice wiki: "Suggestion Circle"</li> <li>Member directory (R)</li> <li>Class identity (logo, pictures) (R)</li> <li>Instructor moderation and presence (R)</li> <li>Controlled access to class pages (R)</li> </ul>
Involve the personal social network	Facilitate participation by non-attending parent and close others Easy content sharing out and in	Relationship building Access to expertise Individual participation	<ul style="list-style-type: none"> <li>Policies on platform roles and access (R)</li> <li>Co-parent full access (R)</li> <li>Account Moderation (R)</li> <li>Visitor participation on program features</li> <li>Content sharing through email, social media interface</li> </ul>
Build the wider community	Create a public face for the community Replicate the look and feel of program participation Collaborative activities that appeal to range of personal interests Program information and organization Security	Relationship building Open-ended conversation Content Access to expertise Serving the context Individual participation	<ul style="list-style-type: none"> <li>Multi-topic discussion forum (all program) (R)</li> <li>Resource area (document library, wiki, marketplace/exchange)</li> <li>Filtered search tool (R)</li> <li>Shared (Google) calendar w/ mapping, personalization (R)</li> <li>Family photo sharing (R)</li> <li>Staff and program information (R)</li> </ul>

<sup>1</sup>. Wenger, White and Smith (2009).

<sup>2</sup>. R = Features included in Parentopia 2.0

Initial system design occurred through an iterative process with parents and staff over a nine month school year. Design features for the overall platform layout, individual components, visual appeal, and usability were created and refined with participant feedback and testing. The domain parentopia.org was selected for the platform to identify its unique role for the program, and as a potential domain home for future adaptations by ECFE sites (there are over 330 school districts with ECFE programs in the state). The designed prototype was built on Wordpress CMS with a customized social networking package to provide enhanced social interaction

capabilities (e.g., forums, chats, wikis, Fig 1a). A public landing page provided basic program information, account creation and log in; account moderation approves and directs further access to the site as visitor, parent, staff and administrator. All accounts access a program home page that presents program-wide images, information and distributed communication activities (e.g., calendar, access to class pages, announcements). Individual class pages include areas for open-ended discussion, member directory, and content contributions. Class page access requires a secondary approval to ensure privacy and safety. The site's static content can be automatically translated into English or Spanish. Parents and staff established policies on inviting non-ECFE members and partners to create accounts at visitor or parent levels thereby gaining read-only or interactive access to program-wide content.

The prototype was employed for an academic year to identify implementation issues and feature use and usability. Staff introduced the platform at the beginning of year and encouraged parent account creation and class membership. The platform was introduced to parents in classes, highlighting features of the full platform, individual class pages and user accounts. Staff were encouraged to begin using discussion forums to continue conversations about class topics, and parents were encouraged to use the class and full program discussion forums to share parenting information.



Figure 1a. Parentopia 1.0 main page site design.

Assessment of the prototype revealed that, as expected, the discussion forums were the component used most often; most other components (e.g., wikis, libraries) and individual content pages, were rarely visited. The majority reported using Parentopia a few times a month or a few times a year. Parents reported using specific features, like the calendar or viewing pictures of their children. Barriers to use included its lack of flexibility on mobile platforms, the distraction of many features, and users forgetting logins. Despite limited use, most however (62%) commented positively about the ability to connect outside of class. They expressed excitement over a mechanism that enabled their connections with other ECFE parents and staff and that could involve the co-parent and other family members.

Staff reported that parents varied greatly in their comfort with account creation and use, which prevented some from logging in. They felt that certain features hampered engagement (e.g., click through email notification of discussion posts). Because parents needed assistance and forgot logins, staff felt they needed to take time away from teaching to assist. Additional personnel or volunteers were unavailable to assist. Most staff also felt time-challenged to lead parent use of discussion forums and other features in their teaching, and on site management to maintain content. Despite these drawbacks, staff were very positive about having access to a complementary platform for engagement with parents.

Implementation and usability testing of the prototype informed a redesign of a streamlined, social networking prominent version (Table 1, Figure 1b), designed collaboratively with site staff and parents. Unique programming using JavaScript and web sockets mimics a Facebook like interface for familiarity, centralizing the main page as a news feed to feature posts, with side bar area for announcements, the shared calendar, photo album and filtered search tool. The user account retains avatar personalization, contacts, notifications, and private messaging. Access to class-only posts, announcements and images remains. While the 2.0 version heavily emphasizes the social media interface and mobile device integration, it retains the essential elements of the original prototype: integrity to user privacy and confidentiality, layout that facilitates ECFE program operations and the relationship-based, supportive nature of ECFE (e.g., affirmations to parents by staff).



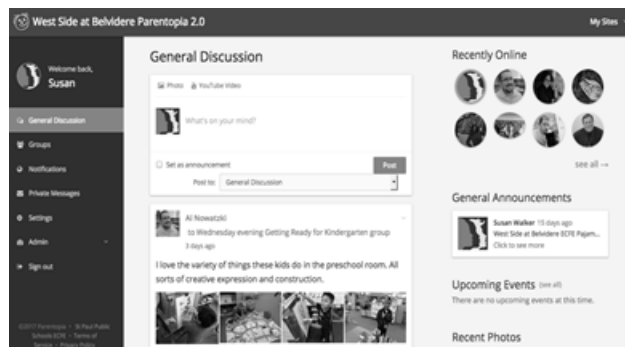


Figure 1b. Revised Parentopia (2.0) site design.

## Implementation insights

At the end of the school year of prototype use, teachers reported using a variety of techniques to encourage use by parents. The most frequently identified effort was posting in the class discussion forum following a class with a brief summary, question or sharing photos from the children's classroom. Instructors met together as a group through a Professional Learning Community monthly, and individually with assistants weekly and identified topics for posting. They also made an intentional effort to respond to posts made by parents to reinforce activity. By building in platform use as part of their regular instruction, over time the teachers felt that their interest and comfort in adopting the innovation increased. They observed that increasingly parents would note that an idea raised in class would be a good one to discuss through Parentopia. Or, parents would tell others to find a program event on the calendar. These indications of interest by parents were motivating to staff to continue, increase use and find additional ways to integrate into teaching.

## Conclusion

This design research to develop Parentopia as a platform to complement social, collaborative learning in ECFE is both promising and realistic. Exploration of learning and social connectivity by parents and staff in ECFE validated social constructivist theories of parent learning as members of ECFE, and as ECFE can integrate with members of the parents' personal social network. And it validated parent and staff interest in an online mechanism to maintain learning and build wider, stronger connections for learning and family life. The design process revealed essential elements to execute the theoretical aims and essential context features in the program for implementation: technology that places social interaction as the centerpiece of the application, is mobile friendly, and offers privacy. Although it is not a unique application in its construction – the revised application looks and feels a bit like Facebook – introducing the innovation of a virtual space for parent engagement to a program that reaches busy and distracted parent users requires that it start with something simple and familiar (Milheim, 2007). A user-friendly friendly interface is also an essential ingredient for a publicly funded nonformal education program with limited staff time, money and resources for technical support.

The participatory, iterative process of design and testing offered another critical dimension to adopting innovation in this traditional program: time (Rogers, 2003). In short, the design research over four years stimulated the cultural shift and expectations of how ECFE helps parents learn while at the program and away through virtual social interaction, and how ECFE needs to support use of a virtual platform. Parents' idea of a virtual space, interaction with it, and feeling its benefits, and staff comfort with how to integrate it into instruction on balance with costs and challenges developed over time. The resulting acceptance and desire for a complementary, virtual space for continued engagement as a class and as a community holds great promise for creating a true collaborative learning ecology to strengthen and support parenting. And with this capital, further research on the actual benefits to parent social learning through their communities and networks has a firmer foundation to occur.

## References

- Azar, S. T. (2003). Adult development and parenthood: A social-cognitive perspective. In J. Demick & C. Andreoletti (Eds.), *Handbook of adult development* (pp. 391-415). New York: Springer.
- Barron, B. (2006). Interest and self-sustained learning as catalysts of development: A Learning Ecology Perspective. *Human Development*, 49, 193-224
- Belsky, J. (1984.) The determinants of parenting. *Child Development*, 55 (1), 83-96.

- Bronfenbrenner, U. (1995). Developmental ecology through space and time: A future perspective. In P. Moen, G. Elder Jr. and K. Luscher (Eds), *Examining lives in context: Perspectives on the ecology of human development*. Washington, DC: American Psychological Association, pp. 619-647.
- Campbell, D & Palm, G. (2004) *Group parent education: Promoting parent learning and support*. NY: Sage.
- Cochran, M. & Walker, S. (2005). Parenting and Personal Social Networks. In T. Luster and L. Ogakaki, Eds. *Parenting: An Ecological Approach*. Mahwah, NJ: Earlbaum.
- Cui, G., Lockee, B. and Meng, C. (2013) Building modern online social presence: A review of social presence theory and its instructional design implications for future trends. *Education and information technologies 18 (4)*. 661-685.
- Farmer, J., & Reupert, A. (2013). Understanding autism and understanding my child with autism: An evaluation of a group parent education program in rural Australia. *Australian Journal of Rural Health, 21*, 20-27.
- Gray, B. (2004). Informal learning in an online community of practice. *Journal of Distance Education, 19*, No 1, 20-35.
- Guldberg, K. & Pilkington, R. (2006). A community of practice approach to the development of non-traditional learners through networked learning *Journal of Computer Assisted Learning 22*, 159-171.
- Hughes, R., Bowers, J., Mitchell, E., Curtiss, S. and Ebata, A. (2012). Developing online family life prevention and education programs. *Family Relations, 61*, 711 – 727.
- Marienau, C. & Segal, J. (2006). Parents as developing adult learners. *Child welfare, 85(5)*, 768-784.
- Milheim, K. 2007. Influence of technology on informal learning. *Adult Basic Education and Literacy Journal, 1 (1)*, 21-16.
- Nieuwboer, C. C., Fukkink, R. G., & Hermanns, J. M. A. (2013). Online programs as tools to improve parenting: A meta-analytic review. *Children and Youth Services Review, 35(11)*, 1823-1829.
- Plantin, L. & Daneback, K. (2009). Parenthood, information and support on the Internet. A literature review of parents and professionals online. *BMC Family Practice, 10*
- Preece, J. (2000). *Online communities: Designing usability, supporting sociability*. Wiley: NY.
- Rogers, E. (2003). *Diffusion of Innovations, 5th Edition*. New York, NY: Simon and Schuster.
- Rogoff, B. (2003). *The cultural nature of human development*. New York, NY: Oxford University Press.
- Rothbaum, F., Martland, N, & Janssen, J. (2008). "Parents' reliance on the Web to find information about children and families: Socio-economic differences in use, skills and satisfaction." *Journal of Applied Developmental Psychology 29*, 118-128
- Rudi, J., Dworkin, J., Walker, S., & Doty, J. (2014). Parents' use of information and communications technologies for family communication: Differences by age of children. *Information, Communication & Society, 18*, 78-93.
- Silverman, D. (1993) *Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction*. London: Sage.
- Stern, M. J., & Messer, C. (2009). How family members stay in touch: A quantitative investigation of core family networks. *Marriage & Family Review, 45*, 654-676.
- Thomas, R. (1996). Reflective dialogue parent education practice: Focus on parent development. *Family Relations, 45 (2)*, 189-200.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Walker, S. (2015) Social dynamics of media use on parenting: A conceptual framework. In *Family Communication in the Age of Digital and Social Media, Carol Breuss, Ed. New York: Peter Lang*.
- Wang, F., and Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational technology research and development, 53*, 4 5-23.
- Wartella, E., Rideout, V., Lauriella, A & Connell, S. (2013). *Parenting in the Age of Digital Technology: A National Survey*. Northwestern University, Center on Media and Human Development. [http://web5.soc.northwestern.edu/cmhd/wp-content/uploads/2013/05/Parenting-Report\\_FINAL.pdf](http://web5.soc.northwestern.edu/cmhd/wp-content/uploads/2013/05/Parenting-Report_FINAL.pdf)
- Wenger, E. (1998). *Communities of Practice Learning, Meaning and Identity*, Cambridge University Press.
- Wenger, E., White, N., and Smith, J.D. (2009). *Digital habitats: Stewarding technology for communities*. CP Square, Portland, OR.

## Acknowledgements

Primary thanks goes to the parents and staff and administrators at the Early Childhood Family Education program who are the community through which Parentopia is informed and given shape. I also acknowledge the research and development team. The project is supported by a grant from the NSF (#1210863SoCS).

# How Middle School Students Construct and Critique Graphs to Explain Cancer Treatment

Camillia Matuk, New York University, [cmatuk@nyu.edu](mailto:cmatuk@nyu.edu)  
Jiayuan Zhang, New York University, [anna.j.zhang@nyu.edu](mailto:anna.j.zhang@nyu.edu)  
Marcia C. Linn, University of California, Berkeley, [mclinn@berkeley.edu](mailto:mclinn@berkeley.edu)

**Abstract:** Using graphs in science is challenging as it requires both scientific and representational fluency. We examined how different graphing activities during science inquiry helped to develop these interrelated abilities in students. Grade 7 students (N=117) worked in pairs on a web-based cell biology unit to either generate or critique graphs of the effects of proposed cancer treatments on cell numbers. All students gained in their graph and science explanation abilities. While students who critiqued graphs gave better overall explanations within the unit, students who constructed graphs better applied their conceptual understanding of science to explain their graphs, both within the unit and later on the post test. We interpret these findings in terms of the relative affordances and constraints of critique and construction activities, and observe students' common misunderstandings of graphs. This study has implications for designing instruction that supports students' uses of graphs within science contexts.

## Finding the synergy between graph and science understanding

The role of graphs in scientific practice is not often captured in typical K-12 curricula. Within scientific communities, graphs are used to predict, explain, and communicate complex concepts. In mainstream media, graphs persuade and inform citizens of important societal and environmental trends. The prevalence of graphs in both professional and everyday settings is such that people require a degree of graph literacy to make and critique arguments, and to interpret information for making decisions about personal and policy issues (e.g., Wiley et al. 2009).

For these reasons, there have been calls for curricula that develop students' abilities to use graphs to model, reason about, and communicate science ideas (e.g., Association for the Advancement of Science, 1993; National Research Council, 1996; Next Generation Science Standards, 2013). However, research continues to find that students struggle to understand and effectively use graphs in science contexts. This study explores the effects of different graphing activities incorporated into a collaborative web-based science inquiry unit on cancer and cell division.

## Constructing and critiquing graphs as components of science literacy

Using graphs effectively involves coordinating an understanding of the relevant domain with an understanding of the representational language of graphs. That is, students must be able to encode and relate a graph's visual features to the concepts these represent (e.g., Friel et al., 2001; Shah & Hoeffner, 2002). As graphs are tools used to devise and evaluate solutions to complex science problems, the ability to both construct and critique them are critical components of scientific literacy. diSessa & Sherin (2000) refer to the abilities to critique and construct representations as components of meta-representational competence (MRC). To critique a graph is to determine the accuracy and effectiveness with which it conveys a message. To construct a graph, meanwhile, is a generative activity that is distinct from, yet just as important as critique (Leinhardt et al., 1990). By constructing graphs, students can visually demonstrate how they make sense of scientific information.

Reading and constructing graphs are challenging for students who are still developing their understanding of both science and of graphs. Indeed, prior research documents the many ways that student falter in their uses of graphs: They struggle to use graphs as evidence to support arguments (Lovett & Chang, 2007); they read graphs as literal pictures rather than acknowledge its axes (Clement 1985); they focus on individual points rather than on bigger trends; they fail to use content to explain axes and slope (Beichner, 1994); and they read graphs in terms of irrelevant other representational forms (see review by Leinhardt et al., 1990). Consistent with prior research, Lai et al. (2016) found that middle school students struggled to interpret graph features in terms of their related science concepts; and to recognize and interpret relationships depicted in visual features, such as trends, shapes, and noise. They moreover struggled to translate narratives of scientific phenomena into graphs, and instead provided superficial visual descriptions of graphs rather than descriptions grounded in science; and faltered in their efforts to interpret common graphical patterns, such as curve shapes and noise when making sense of global climate change and growth curves.

Constructing graphs is likewise difficult for students who are prone to create one-point graphs, generate a series of graphs to represent a single factor, and graph an increasing linear function regardless of the actual trend (Mevarech & Kramarsky, 1997). Moreover, graph construction activities can become easily mired in mundane tasks, such as plotting data points, which students are unlikely to connect to relevant scientific ideas. On the other hand, depicting relationships qualitatively may permit students to go beyond the data to make inferences based in their conceptual understanding, and to use graphs to engage in such scientific practices as predicting, arguing, and explaining. Interpreting qualitative features of graphs, however, has proven more difficult for middle school students than interpreting quantitative features of graphs (Hattikudur et al., 2012).

### The need to emphasize graphs in science instruction

Studies find that students' understanding of graphs can be enriched through activities that involve explaining their reasoning and challenging the views of their peers (Kramarski, 2004). As well, curriculum designs that effectively integrate graphs into science contexts can enable students to successfully communicate scientific phenomena through graphs (e.g., Vitale et al., 2015). These findings suggest that instruction might go beyond guiding students in the technical tasks of constructing graphs and recognizing its features, to also supporting students in developing the metacognitive skill of critiquing graphs. The importance of graphs, along with students' struggles to use graphs, suggest a need for instruction that effectively incorporates graphs into realistic science inquiry activities (Lai et al., 2016).

In response, we integrated two graphing activities into a collaborative inquiry unit on cancer and cell division. We investigate the relative value added of students' qualitative construction and critique of graphs for supporting and revealing their abilities to explain and evaluate proposed cancer treatments.

## Methods

### The WISE *Mitosis* unit

We integrated a graphing activity into an existing middle school science unit on cancer and cell division. The unit, called *What makes a good cancer medicine?: Observing mitosis and cell processes* (aka, *Mitosis*) was authored in the Web-based Inquiry Science Environment (WISE, wise.berkeley.edu, Slotta & Linn, 2009). WISE is a free, open source learning environment with tools for authoring media-rich content, and for monitoring and guiding students' work. Units are designed according to Knowledge Integration (KI, Linn & Eylon, 2011), a framework that recognizes students' diverse ideas about science, and guides them in distinguishing, organizing, and integrating these into normative scientific understandings.

*Mitosis* introduces students to the process of cell division and the effects of cancer on the body. Central to the unit is an investigation of potential cancer treatments, in which students observe and compare animations of cells dividing normally, and when treated with different medicine options. Students work in pairs on shared computers, and use tools within the environment to develop, share, and refine their explanations for recommending one medicine over the others.

Next, the unit introduces surgery and chemotherapy as typical cancer treatments that each come with trade-offs: Surgery can quickly remove cancerous cells, but risks damaging healthy organs. As well, any remaining cancerous cells will continue to divide, which may lead to tumors to return. Meanwhile, chemotherapy avoids the risks of surgery, but because it targets all fast dividing cells—not just cancerous cells—it introduces side effects such as hair loss and nausea. Students analyze graphs of each treatment on a patient's number of cells over time. They read that because neither treatment is perfect, a goal in designing cancer treatment is to maximize the effects on cancerous cells while minimizing damage to healthy cells. A culminating graphing activity (further described below) has students use a graph to explain the effects of a possible cancer treatment on the number of cells in the body.

### Participants and study design

Participants were 117 grade 7 students of a middle school in the west coast of the United States. They were taught by the same teacher, who had used previous versions of the same unit in the past several years. Students worked in pairs on a shared computer for approximately 10 consecutive days to complete the unit, including a pre and posttest that assessed their application of the science content. The teacher began each day with a whole class opener, in which she highlighted difficulties she noted in students' work from the previous day, and prepared students for the upcoming activities. Otherwise, the teacher circulated the classroom to assist groups as they worked through the unit at their own pace.

The teacher formed student pairs and randomly assigned half within each of her 4 class periods to complete one of two versions of the unit: *Construct* and *Critique*. These differed in the following activity

(Figure 1). *Construct* students were prompted to prescribe and explain a treatment plan for a cancer patient, and to complete a graph of the effects of their treatment on the numbers of cancerous cells over time. Meanwhile, *Critique* students were given a graph of the effects of a doctor's prescribed treatment on the numbers of cancerous cells over time, and prompted to annotate and explain how and why that treatment would or would not be effective. Afterward, students shared their graphs and explanations, and commented on the work of others in their class through an online discussion forum. They were then instructed to use their peers' ideas to revise their work.

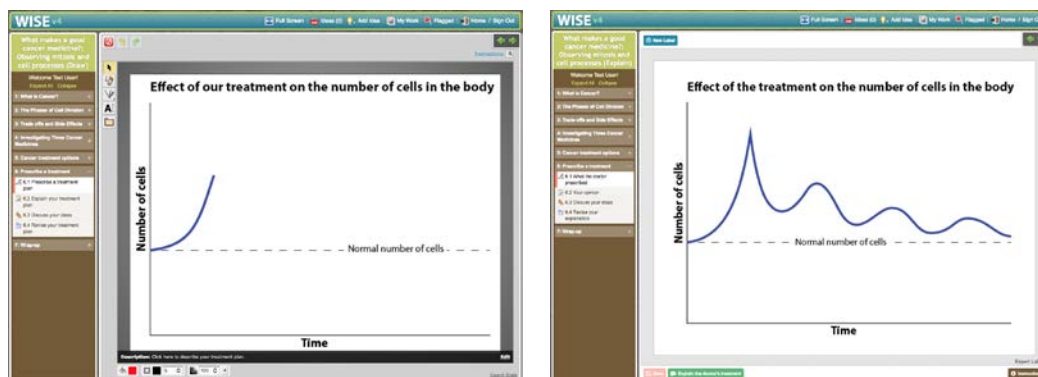


Figure 1. Screenshots of the *Construct* (left) and *Critique* (right) versions of the *Mitosis* unit.

## Data and analysis

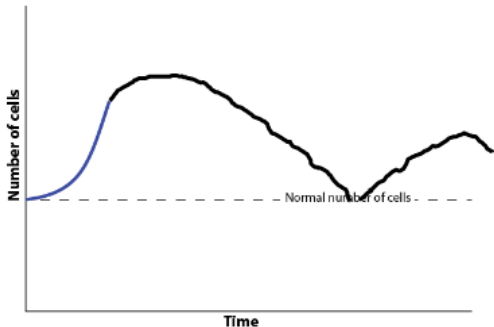
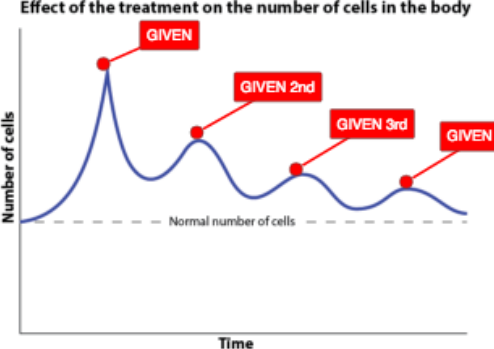
The data considered in this study include students' responses to the graphing activity embedded within the unit, which consisted of their graph artifacts (student-generated graphs from the *Construct* group and annotated graphs from the *Critique* group), and their accompanying written explanations of the cancer treatments. (Data on the ways that students shared and revised their ideas are reserved for future analyses.) We developed a rubric to rate students' abilities to integrate science content understanding with graphing abilities. This rubric identified the presence of several key concepts (Table 1), including: (1) an explanation based in science content learned in the unit (e.g., what cancer is, how cells divide, how chemotherapy works) as opposed to a literal description of the graph; (2) a sense of the imperfections or trade-offs of cancer treatment, as opposed to the belief that treatment is a straightforward and uncomplicated solution; (3) a recognition that to mitigate side-effects, the treatment must be given in multiple brief cycles rather than in a single dose; (4) a description of the graph that conveys the process of treatment by identifying sequences of events, as opposed to overlooking the nuanced changes in cell numbers over time.

The rubric also rated responses for their accuracy. One 4-point scale (0-3) rated the normativeness of students' written explanations, while another captured the accuracy of their graph artifacts (e.g., whether or not students indicated a decreasing number of cells, and whether the graph cohered with the written explanations). We summed the ratings on each aspect to obtain a single score with a possible total of 10 points. This score captures the overall quality of responses in terms of students' abilities to accurately apply relevant science content and graph features in their critiques or constructions of graphs.

The pre and post test consisted of five items that assessed students' abilities to apply concepts learned in the unit (e.g., explaining the importance of the phases of cell division, and describing the mechanism of cancer in terms of its effect on cell division.) For this study, we focused our analysis on one of these items, which had students generate and explain a graph of the number of cancer cells before, during, and after a proposed treatment. We used the rubric described above to code these responses. We also conducted a 30-minute long recorded phone interview with the teacher shortly after she had enacted the unit, and draw on this to help explain our findings.

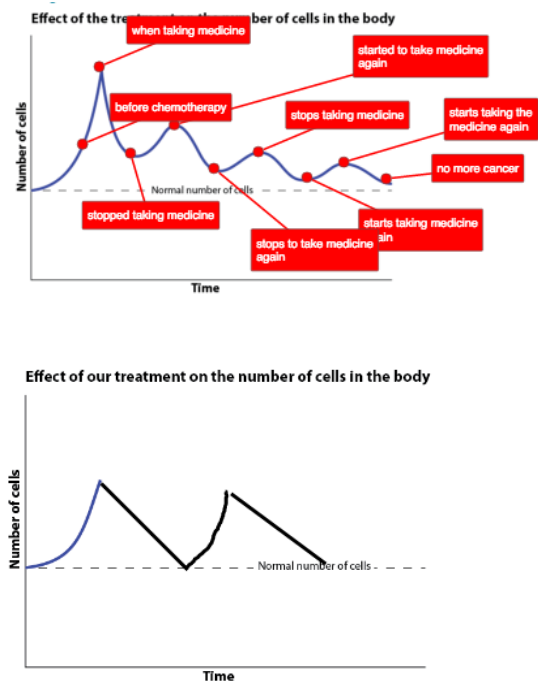
Table 1: Rubric for scoring students' responses, with examples (comparison groups indicated in italics).

Category	Present
1. Content-based explanation	<b><u>Explanation:</u> Uses relevant science content to interpret the graph.</b>
	<u>Example</u>

	<p><i>Critique:</i> The chemotherapy keeps the cells from doing mitosis. <b>When you have cancer your cells divide out of control you can get tumors from too many cells (...)</b> The doctor's treatment plan worked by keeping her cell count near the normal amount but not harming the healthy cells but at the same time stop the cancer cells from dividing.</p>
<p>2. Sense of trade-off</p>	<p><b><u>Explanation:</u></b> Understands that cancer treatment involves a trade-off that can lead to side-effects.</p> <p><u>Example</u>  <i>Construct:</i> If you do chemotherapy for the just the right amount of time, you will get rid of the cancer cells and lose few normal cells.</p>
<p>3. Treatments cycles</p>	<p><b><u>Explanation:</u></b> Uses/recognizes an approach to treatment that mitigates risks and side-effects.</p> <p><u>Examples</u></p> <p><i>Critique:</i> The doctors gave it to her multipul times to kill off the cancer cells...</p> <p><i>Construct:</i> When the number of cells reach the climax she takes not that much medicine. It'll go down but star rising again so she would take the medicine when it starts rising again.</p> <div style="text-align: right;"> <p><b>Effect of our treatment on the number of cells in the body</b></p>  <p><b>Effect of the treatment on the number of cells in the body</b></p>  </div>
<p>4. Narrative description</p>	<p><b><u>Explanation:</u></b> Conveys the process of treatment by identifying sequences of events, and recognizing the change in cell numbers over time.</p>

Examples  
*Critique*

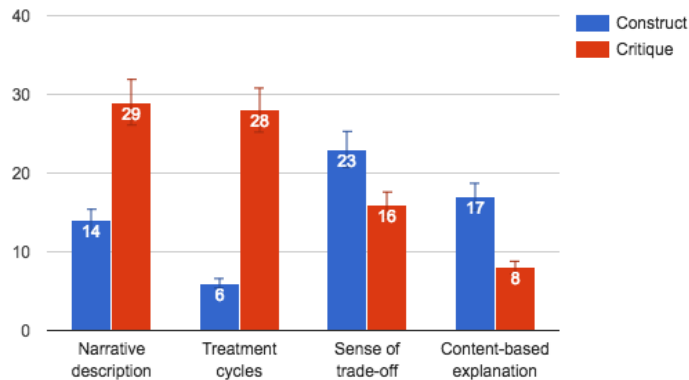
*Construct*: How the treatment would work is that we would apply it until it got back to normal, and then apply it again when the cells went above average. We think this will be effective because it will stop the cancer cells when they come back, but it won't go below normal.



## Findings and discussion

### Group differences on the embedded graphing activity

Students who critiqued the graph had significantly higher overall scores on the embedded graphing activity (N=29; M=9.59; SD=1.05) than students who constructed graphs of the effects of their own treatments (N = 32; M=8.63, SD=1.39),  $t(59) = 3.03, p < .005$ . A 2x4 chi-square test showed significant differences between groups in the occurrences of the specific aspects we coded,  $\chi^2(2) = 21.31, p < .0001$  (Figure 2).



**Figure 2.** *Construct* and *Critique* students differed significantly in the frequencies of *Narrative description*, *Treatment cycles*, and *Content-based explanations* on the graphing item embedded in the unit.

Specifically, *Critique* students (N=29 M=.97, SD=0.19) were more likely than *Construct* students (N=32 M=.19, SD=0.40) to identify the importance of repeated doses of medicine,  $t(59) = 9.64, p < .0001$ . This result might be explained by the fact that *Critique* students had only to identify these features from the given graph, while *Construct* students, who had to generate their own graphs, were left to discover this strategy on their own, which they did not always do successfully. *Critique* students (N=29 M=1.00, SD=0.00) were also more likely than *Construct* students (N=32 M=.44, SD=0.50) to describe the narrative process represented by the graph,  $t(59) = 6.01, p < .0001$ . Being given an already constructed graph may have allowed *Critique* students to focus on developing narrative descriptions of it. Meanwhile, the extra effort of constructing a consensus graph may have led *Construct* students to put less effort into their written explanations. Notably, however,

*Construct* students were more likely to use science content ideas to explain their graphs (N=32, M=0.53, SD=0.51) compared to *Critique* students (N=29, M=0.28, SD=0.45),  $t(59)=2.06$ ,  $p<.05$ . It is possible that constructing a graph encouraged deeper thought into the graph's conceptual meaning.

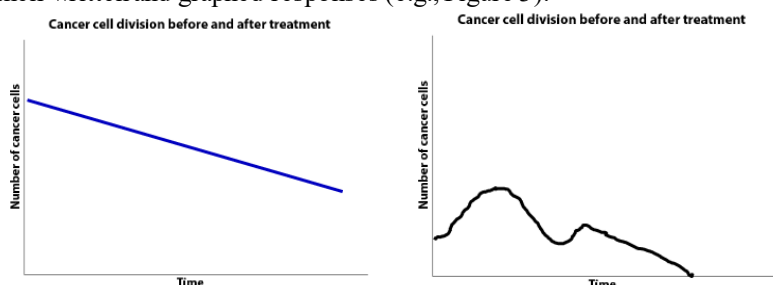
### Gains from pre to post

All students made significant gains on the graphing item from the pre (N=117, M=1.99, SD=2.79) to the posttest (N=117, M=7.42, SD=2.59),  $t(232)=15.41$ ,  $p<.0001$ . *Construct* students gained slightly, but not significantly more (N=62, M=5.48, SD=3.43) than *Critique* students (N=55, M=5.36, SD=4.02). These findings suggest that in spite of differences in performance within the unit, both versions helped students improve their abilities to articulate key ideas, as well as to express these ideas normatively in both written and graphic forms.

Some of the group differences that were apparent within the unit persisted to the post test. Specifically, *Critique* students (M=0.24, SD=0.47) were more able than *Construct* students (M=0.08, SD=0.27) to identify the importance of cycling treatment, as shown by their greater pre to posttest gains on the graphing item,  $t(115)=2.22$ ,  $p<.05$ . It is likely that these students could easily reproduce features of the graph to which they were exposed during the unit. Likewise, the *Construct* students made significantly greater gains (N=62, M=0.45, SD=0.50) in applying content to their explanations compared to *Critique* students (N=55, M=0.22, SD=0.50),  $t(115)=2.52$ ,  $p<.05$ . This finding suggests that *Construct* students were able to translate the abilities gained during the unit to their work beyond the unit.

### Examples of students' challenges with using graphs to explain cancer treatment

In the better cases, students displayed a more nuanced understanding of the cancer treatment process by the end of the unit in both their written and graphed responses (e.g., Figure 3).



**Figure 3.** Left: One student's pretest response, which conveys a simplistic understanding of the effect of the medicine, along with the explanation: "The drug will stop the cells from multiplying..." Right: The same student's posttest response with the accompanying explanation: "My graph... is supposed to show that the cell count was rising before treatment, and after the start of treatment is went down, even past the normal cell count. In the middle of it, the cell count goes back up because there is a break in treatment to make sure there is no overdosage, but once the cell count starts to go back up, the treatment is restarted." (92037)

For the most part, however, we observed problematic features of students' responses that are consistent with other research (e.g., Lai et al., 2016). For example, with the exception of two student pairs, one in each comparison group, students failed to connect their ideas about science to their graphs, and instead restricted their explanations to describing their graph's visual features (e.g., "The treatment started, stopped, started again... This made the number of cells go up and down..."). In another example, one student pair in the *Critique* group mistook the y-axis to represent both the amount of medicine and the number of cells. As they wrote: "The doctor is giving Chemotherapy in smaller and smaller doses until the number of cells is at a normal amount." (218836)

While most students left their pretests blank or else scribbled nonsense and typed "I don't know," a few students created pictorial representations of the effects of cancer treatment (Figure 4); or graphed an increasing linear function (Figure 5). In each of these cases, the students improved their responses by the posttest. In some cases, however, students who displayed one misunderstanding of graphs at the pretest (using it as a pictorial representation of the phenomenon) ended with different misunderstandings by the posttest (drawing a series of graphs as opposed to a single one, and depicting rising linear functions in spite of the actual trend) (Figure 6).



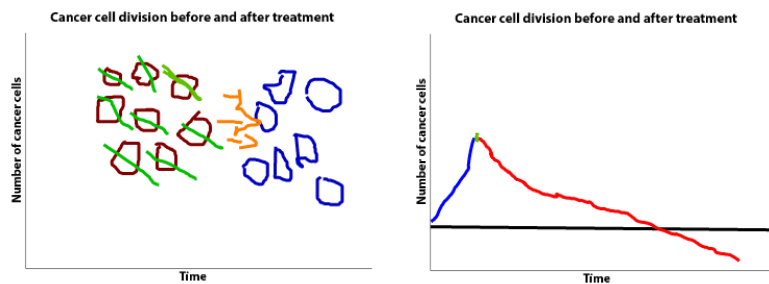


Figure 4. Left: One *Critique* student's pretest response showing a pictorial representation of the cancer treatment process. The graph is accompanied by the explanation: "The red circles are cancer cells. My drug is represented by the green lines crossing them out, then them turning them into regular cells." Right: The same student's posttest response is accompanied by the explanation: "The blue line represents when the person has cancer, and the number of these cells is growing rapidly. The green line represents when the drug is first taken. The red line, represents after the drug has been taken. The number of cells is going down, because the person is losing cells at a normal rate, but their body can't replace the cells." (89152)

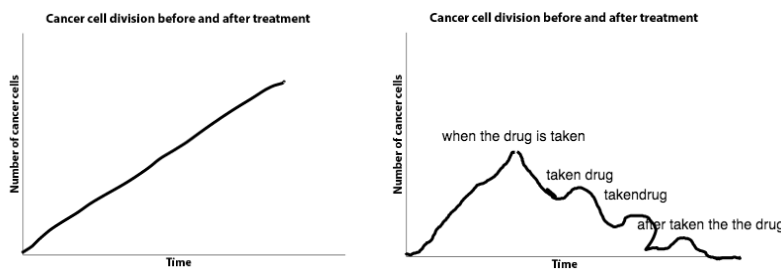


Figure 5. Left: One *Critique* student's pretest response, showing a rising linear function, and the same student's posttest response (right), demonstrating a more accurate narrative of the treatment process and the cycling of doses of medicine. (92399)

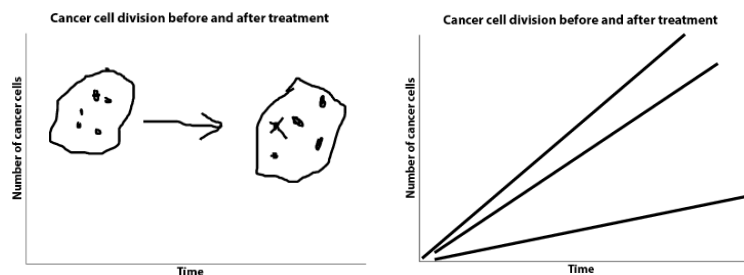


Figure 6. Left: One *Construct* student's pretest response, showing a pictorial representation of cancer cells before and after treatment, and the same student's posttest response, which incorrectly uses a series of rising linear functions to represent the numbers of cancer cells before, during, and after treatment. (91962)

## Conclusions and implications

We compared how students explained cancer treatment when constructing or critiquing qualitative graphs. Our findings suggest that critiquing graphs enabled students to generate better overall explanations, but that these explanations differed in particular ways. Specifically, the explanations of students who critiqued a given graph were more likely to convey a narrative of the underlying process, and to recognize the importance of cycling doses of medicine for mitigating side effects. Meanwhile, students who constructed graphs of their own treatment plans were more likely to use science concepts to explain their graph's meaning. These distinctions might be accounted for by differences in the practices that critique and construction either emphasize or minimize. That is, critique offers students with material upfront (a finished graph and a worked solution), which accomplishes some of the analytic work for students, and allows them to focus on elaborating other aspects of their explanations (e.g., coherent narrative descriptions). Meanwhile, construction requires students to generate material that is not provided. While doing so may prevent students from realizing nuanced solutions (e.g., cycling treatment as a way to mitigate side effects) and from spending as much effort in developing narratives of their explanations, the act of discussing and coming to consensus on a graph to generate may motivate

students to think more deeply about the underlying conceptual meaning, and offer them a context in which to express that understanding.

This study offers an example and classroom trial of a graphing activity embedded within science inquiry instruction. Our interview with the teacher suggests that her opener activities, which modeled responses to other graphing items in the unit, may have prepared students to perform the graph critique and construction activity examined in this study. In spite of overall gains in their science and graphing abilities, the fact that certain students continued to display misunderstandings that are consistent with prior literature reveals an opportunity to investigate how students' graph construction reflects their scientific misunderstandings (cf. Vitale et al., 2015). It also suggests the need to further refine instructional materials that emphasize the synergy between graphing and scientific practices. In continuing our analysis of these data, we will investigate the ways that sharing and discussing graph artifacts with their peers impacted students' revisions, and the ways that technology can partner with teachers to support students' graphing practices in science contexts.

## References

- Beichner R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics*, 62, 750–762
- Clement, J. (1985). Misconceptions in graphing. In: Proceedings of the ninth international group for the psychology of mathematics education. Noordwijkerhout
- DiSessa, A., Sherin, B. L. (2000). Meta-representation: An introduction. *Journal of Mathematical Behavior*, 19, 385–398.
- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in mathematics Education*, 124-158.
- Hattikudur, S., Prather, R. W., Asquith, P., Alibali, M. W., Knuth, E. J., & Nathan, M. (2012). Constructing graphical representations: Middle schoolers' intuitions and developing knowledge about slope and y-intercept. *School Science and Mathematics*, 112(4), 230-240.
- Kramarski, B. (2004). Making sense of graphs: does metacognitive instruction make a difference on students' mathematical conceptions and alternative conceptions?. *Learning and Instruction*, 14(6), 593-619.
- Lai, K., Cabrera, J., Vitale, J. M., Madhok, J., Tinker, R., & Linn, M. C. (2016). Measuring graph comprehension, critique, and construction in science. *Journal of Science Education and Technology*, 1-17.
- Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research*, 60(1), 1-64.
- Linn, M. C., & Eylon, B. S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration*. Routledge.
- Mevarech, Z. R., & Kramarsky, B. (1997). From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. *Educational Studies in Mathematics*, 32(3), 229-263.
- Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14(1), 47-69.
- Slotta, J. D., & Linn, M. C. (2009). *WISE science: Web-based inquiry in the classroom*. Teachers College Press.
- Vitale, J. M., Lai, K., & Linn, M. C. (2015). Taking advantage of automated assessment of student-constructed graphs in science. *Journal of Research in Science Teaching*, 52(10), 1426-1450.

## Acknowledgements

Funding for this research was provided by the National Science Foundation: a DR K-12 under award #1119670.

# The Impact of Peer Tutors' Use of Indirect Feedback and Instructions

Michael Madaio, Justine Cassell, and Amy Ogan,  
mmadaio@cs.cmu.edu, justine@cs.cmu.edu, aeo@cs.cmu.edu  
Carnegie Mellon University

**Abstract:** During collaborative learning, computer-supported or otherwise, students balance task-oriented goals with the interpersonal goals of relationship-building; these goals may conflict, negatively impacting learning. In peer tutoring, for instance, tutors may avoid providing feedback to their partners to avoid the face-threat to their tutee. In this paper, we explore how the interpersonal closeness between tutor and tutee impacts tutors' use of indirectness with feedback and instructions, and the impact those moves have on tutees' problem-solving. We found that stranger tutors use more indirect instructions and provide more positive feedback to their tutee than friend tutors, and that stranger tutees attempted and solved more problems when their tutors used indirect instructions. We found no effect for dyads of friends, suggesting that interpersonal closeness reduces the face-threat of direct instructions. These results demonstrate that designers of CSCL tools should incorporate awareness of students' relationships into their systems, as that relationship impacts students' collaborative learning behaviors.

## Introduction

In collaborative learning interactions, whether computer-mediated or face-to-face, students simultaneously pursue the task-oriented goals of learning and the interpersonal goals of getting along with or building a relationship with other students, much like in other social interactions (Burgoon et al., 1995). In some forms of collaborative learning, such as peer tutoring, students may offer each other advice, instructions, or feedback. Such pedagogical behaviors, while supporting the interactional goal of helping their partner learn, may conflict with the interpersonal goal of relationship-building with their partner due to the potential for such moves to threaten their partner's "positive face", or desire to be approved of by others (Brown and Levinson, 1987). Similarly, other productive learning behaviors, such as engaging in cognitive conflict, have been found to occur more often in groups of friends than strangers, perhaps because of a greater importance for strangers in avoiding face-threat while pursuing interpersonal goals (Azmitia & Montgomery, 1993). To mitigate the interpersonal consequences of pedagogical behaviors that are likely to threaten tutees' face needs, such as giving direct instructions and feedback, peer tutors without sufficient interpersonal closeness with their tutee might avoid providing the necessary feedback altogether. If they are more skilled at attending to interpersonal needs, they might phrase their words in a polite or indirect manner, to reduce the implicit face-threat (Person et al., 1995). Some computer supports for learning, such as some forms of intelligent tutoring systems, apply a polite style to the feedback and instructions provided to students, such as "Shall we calculate the result now?", to mitigate the face threat of direct feedback and instructional directives (Johnson and Rizzo, 2004). An overuse of such polite or indirect instructional moves, however, may have a negative impact on student learning, due to the ambiguity of indirectness (Person et al., 1995). Moreover, indirectness may not be the appropriate way for tutors to deliver feedback in all situations. In fact, previous studies on collaborative learning have shown that the frequency of politeness and the negative impact of face-threat diminishes over the course of a relationship (Ogan et al 2012), following prior work on the decreasing importance of positivity as the interpersonal distance between interlocutors decreases (Tickle-Degnen and Rosenthal, 1990). Further, while expert teachers may be communicatively competent enough to modulate their feedback and instructions to mitigate the face-threat of directness (Kerssen-Griep, 2008), it is not clear that untrained peer tutors are able to do so.

In order to design computational supports for collaborative learning that both effectively manage face needs and support cognitive conflict, we should thus first understand how untrained peer tutors mitigate face-threat to their tutees, either through indirect delivery of instructions and feedback, or by avoiding that feedback altogether. In this paper, we first investigate how reciprocal peer tutors' interpersonal closeness with their tutees, as well as their prior domain knowledge and tutoring self-efficacy, impacts their use and delivery style of potentially face-threatening tutoring moves. We then investigate the relationship between peer tutors' delivery style and their tutees' problem-solving. We found that stranger tutors provide more positive feedback and more indirect instructions than friend tutors do, and found that tutors with high self-efficacy used significantly more indirect instructions than low self-efficacy tutors, particularly for tutors with lower rapport with their tutee. For stranger dyads, independent of their self-efficacy, tutors' use of indirect instructions was positively predictive of their tutees' number of problems attempted and solved, but for friend dyads, there was no effect. These findings can help inform the design of CSCL tools that can detect students' interpersonal closeness and suggest different

ways for students to deliver instructions and feedback to mitigate face-threat, or to intervene to provide that direct support itself when necessary.

## Related work

Reciprocal peer tutoring is a form of collaborative learning where same-age students work together by teaching each other, despite neither of them being an expert (Palincsar and Brown, 1984). Prior work has shown it to be an improvement over individuals learning alone, but the differences in both content knowledge and pedagogical knowledge between novice peer tutors and expert tutors may have significant consequences for both the process and outcomes of tutoring (Palincsar and Brown, 1984). Specifically, prior work has argued that providing instructional feedback, directives, or advice can be highly face-threatening and may be avoided by peer tutors (Person et al., 1995). However, tutors with greater instructional or interpersonal ability may be better able to mitigate face threat to their students, perhaps by presenting feedback in an indirect or polite way (Person et al., 1995), or by using nonverbal immediacy cues as described by Kerssen-Griep (2008). As we have described above, indirectness is one of the conversational strategies that plays a role in face management (others are praise and acknowledgement). To understand the impact of face management on peer tutoring behaviors, in particular the delivery of feedback and instructions, we draw on Brown and Levinson's (1987) theory of face management and politeness. In this theory, students, like any social actors, are motivated by their desire for what is referred to as positive face, or, to be approved of by others, and negative face, which is the desire to be autonomous and unimpeded by others (Brown & Levinson, 1987). As a result, tutors' instructional directives, if they take the form of demands, may threaten students' negative face, and tutors' instructional feedback, if given in a blunt manner, may threaten their tutees' positive face (Brown & Levinson, 1987; Johnson and Rizzo, 2004). However, the severity of that face-threat, and thus, its negative impact on the tutees, can be mitigated by the use of politeness or indirectness, which may be mediated by the interpersonal closeness between students. This follows Tickle-Degnen and Rosenthal's theory that the importance of positivity decreases as the interpersonal closeness between interlocutors grows (Tickle-Degnen and Rosenthal, 1990). However, it is unclear to what extent peer tutors minimize the face-threat of feedback and instructions by being indirect in their delivery, or how that indirectness impacts their tutees' problem-solving. Accordingly, some intelligent tutoring systems have attempted to mitigate the face-threat of providing feedback and instructions by phrasing tutor moves in a polite or indirect way, and they found benefits for some students on some types of questions (Johnson and Rizzo, 2004). However, their model operationalized Brown and Levinson's social distance as simply decreasing over subsequent sessions with a tutee, which does not account for the dynamic nature of relationship development, as even students who work together for multiple sessions may never develop the interpersonal closeness necessary for reducing the face threat of direct feedback.

To better operationalize the dynamic development of interpersonal closeness and its impact on face-management, we draw on the theory of interpersonal rapport, as described by Tickle-Degnen and Rosenthal (1990), as well as Spencer-Oatey's (2008) model of face-management. In these theories, we would expect positive face management strategies to decrease as the relationship between tutor and tutee develops over time (Spencer-Oatey, 2008; Tickle-Degnen and Rosenthal, 1990). Prior work has found, using friendship as a proxy for long-term rapport, that tutoring dyads of friends do in fact engage in more violations of social norms, such as playful teasing and social challenges, and that these are correlated with learning gains in friends. Those same behaviors lead to decreased learning among strangers (Ogan et al., 2012). This suggests that the relationship between tutor and tutee allows each to playfully challenge the other while tutoring, in what might be seen as face-threatening acts, when done between friends. In the same vein, the indirectness that may be appropriate or necessary between strangers, to mitigate face threat from direct feedback or instructions, might be unnecessary or even detrimental between friends. Here, to investigate how a dynamic measure of interpersonal closeness can capture changing face-management needs, we use two measures of interpersonal closeness, 1) externally-rated rapport between participants as a measure of their short-term interpersonal closeness, and 2) the friendship between participants as a proxy for their long-term interpersonal closeness.

Although expert tutors and teachers are able to effectively mitigate the face threat of a tutoring move through indirectness, politeness, or a self-effacing disclosure (Kerssen-Griep 2008), untrained peer tutors may not be as deft in their facework. Although we did not measure tutoring ability, we used the tutors' score on a pre-test as a proxy for their prior domain knowledge, following Rowan et al.'s (1997) findings that teacher prior domain knowledge was predictive of their students' performance. In addition to their content knowledge of the domain, prior work has shown the importance of teachers' instructional self-efficacy for their attention to the interpersonal goals of teaching as well as the instructional goals (Mojavezi, 2012; Saklofske et al. 1988). Teachers' self-efficacy beliefs, or the belief about their ability to impact student outcomes and the confidence that they can do so, have

been shown to impact teachers' use of different types of feedback (Saklofske et al., 1988) as well as students' motivation and achievement (Mojavezi, 2012).

## Methods

We seek to investigate how the interpersonal closeness between peer tutors and their tutees impacts the tutors' use of indirectness with instructions and feedback and how those moves in turn impact their tutees' learning.

## Research questions

- RQ1: Do tutors with lower interpersonal closeness with their tutees provide feedback about their tutees' correctness at different rates than dyads with greater interpersonal closeness? Due to the potential face-threat in giving negative feedback to their partners, we hypothesize that both tutors in dyads of strangers and tutors in low-rapport dyads will provide less negative feedback (Person et al., 1995). Friends and high-rapport dyads would have less need for face-threat mitigation, and will thus use more negative feedback (Spencer-Oatey, 2008).
- RQ2a: Do tutors with a lower interpersonal closeness with their tutee use more indirect language while tutoring than tutors with a greater interpersonal closeness? Based on prior literature on the use of indirect language to mitigate the face-threat of various tutoring moves, we hypothesize that tutors with a lower interpersonal closeness with their tutee will use more indirect language while tutoring than tutors in dyads with greater closeness (Brown and Levinson, 1987; Johnson and Rizzo, 2004).
- RQ2b: Do tutors with a lower tutoring self-efficacy use more indirect language while tutoring than tutors with greater self-efficacy? From prior literature on the use of hedges and subjectivizers to convey uncertainty (Rowland, 2007), we hypothesized that tutors with lower tutoring self-efficacy would use more indirect language due to their uncertainty. However, an alternative hypothesis is that tutors with greater tutoring self-efficacy would attend more to their tutees' needs for face-management and would use more indirect language with their tutoring moves to mitigate the face-threat of feedback and direct instructions (Brown and Levinson, 1987; Kerssen-Griep, 2008; Saklofske, 1988)
- RQ3: Does a tutor's use of indirect feedback lead to improved problem-solving for their tutee? From prior literature on the motivational benefits of face-threat mitigation, we hypothesize that there will be an interaction between a dyad's interpersonal closeness and their use of indirect tutoring language, such that in stranger dyads and low-rapport dyads (both friend and stranger), when tutors use more indirect tutoring moves, their tutees will attempt and solve more problems (Kerssen-Griep 2008). An alternative hypothesis is that direct feedback and instruction leads to more problem solving.

## Dialogue corpus and annotation

The dialogue corpus described here was collected as part of a larger study on the effects of rapport-building on reciprocal peer tutoring. The participants were assigned to 12 dyads that alternated tutoring each other in linear Algebra equation solving for 5 weekly hour-long sessions, for a total of ~60 hours. Each session was structured such that the students engaged in brief social chitchat in the beginning, then had one tutoring period of 20 minutes with one of the students randomly assigned to the role of tutor, then engaged in another social period, then had the second tutoring period where the other student was assigned the role of tutor. This process was repeated for five sessions. As each student was randomly assigned to be the tutor for half of the tutoring periods, they were not expected to have any greater prior knowledge than their partner for the problems they were tutoring them on. All students were supported with a set of instructions on how to teach the particular problems for which they were assigned the role of tutor. The students took a pre-test before the first session and a post-test after the fifth session to assess their learning gains. However, because of a ceiling effect on the pre-test, we used the problems the tutees solved during the sessions as a proxy for a learning outcome measure for each session. The participants (mean age = 13.5, min=12, max=15) came to a lab on an American university campus in a mid-sized city for the study. Half were male and half were female, assigned to same-gender dyads, so that, in other work with this corpus, gendered differences in the social, rapport-building behaviors of the participants could be identified. To investigate how the use of various tutoring behaviors differs between dyads with varying degrees of interpersonal closeness, we used friendship as a proxy for long-term closeness and asked half of the participants to bring a same-age, same-gender friend to the session with them, and for the other half of the dyads, we paired them with a stranger. Audio and video data were recorded, transcribed, and segmented for clause-level dialogue annotation.

The rapport, or short-term interpersonal closeness between the participants, was evaluated using a "thin-slice" approach (Ambady and Rosenthal, 1992), where the corpus was divided into 30-second video slices provided to naive, third-party raters in a randomized order, so they would each slice's rapport, and not the delta

across multiple slices. Three raters rated the rapport present in each slice on a Likert scale from 1-7, and a single rating was chosen for each slice using a bias-corrected weighted majority vote approach, as in Sinha and Cassell (2015). From the 120 thirty-second slices in each session, we calculated a summary rapport score for each session, or the “utopy”, which we use in this paper as our rapport measure. Prior work has shown that statistical summaries such as a measure of central tendency or proportion of high and low ratings of rapport collapse the temporal dimension and are not as robust as more stochastic-based models which capture the evolution of rapport over time (Sinha, 2016). We thus fit a Markov chain of order 1 to the sequence of 120 rapport ratings for each session, and used the resulting transition probability matrix to generate a measure of the entropy (or, uncertainty) and utopy (Dillenbourg, 2015; Sinha, 2016), or likelihood of the dyad being in a high-rapport state.

As part of a larger study on the relationship between rapport-building and peer tutoring, this corpus was annotated for a set of social, rapport-building verbal and nonverbal behaviors, as well as pedagogical, tutoring-related behaviors from both the tutor and tutee. Social conversational strategies, as in Sinha and Cassell (2015), include such phenomena as students’ self-disclosure, reference to shared experiences, praise, and social norm violations, each of which serves a rapport-building function, as described in Zhao et al. (2014), with Krippendorff’s alpha for all codes  $> .75$ . We annotated the dialogue corpus for a set of tutoring and learning behaviors, such as question-asking, “knowledge-telling”, or step-level instructions and procedures, and feedback from tutors on their partner’s correctness (Madaio et al., 2016), with Krippendorff’s alpha for all codes  $> .7$ . In Table 1, we describe two tutoring strategies relevant in this paper for their potential to threaten the tutees’ face: tutors’ feedback, and their instructions, or “knowledge-telling”. Then, to understand the ways that dyads with tutors of differing levels of interpersonal closeness with their tutees (both friendship status and rapport utopy) mitigated the face-threat of their tutoring instructions and feedback, we coded our corpus for four types of indirectness, which are often used to soften or mitigate the face threat of being “bald on record” with direct instructions, advice, or feedback (Brown and Levinson, 1987). We annotated for: apologizing, hedging language, the use of vague category extenders, and “subjectivizing” (Zhang, 2013; Neary-Sundquist, 2013; Fraser, 2010), as described in more detail in Table 2. The Krippendorff’s alpha for five trained raters rating all four codes was  $> .7$ . Then, we labeled a clause as *indirect feedback* if the same clause had an annotation for indirectness and an annotation for feedback, either positive or negative. Typical examples of indirect feedback include: “I think you got it.” “I guess that’s what it is.” “Oh, no, actually, it’s not.” “No, it’s just nineteen.” We labeled each clause as *indirect instructions* if it had an annotation for indirectness and an annotation for “knowledge-telling” (instructions and directives used by the tutor). Typical examples of indirect instructions are: “Actually, just add five here.” “I think you’re gonna divide it by a fraction or something.” “I would probably subtract the sixteen.” Finally, we provided the participants with a questionnaire following the study with a set of items for constructs relevant to rapport-building. We used items about tutors’ self-efficacy for tutoring, a 7-item scale indexing whether the participants thought they were good tutors and were concerned about tutoring quality, motivation and impact on their tutee.

**Table 1: Annotations of Tutoring Moves and Indirect Language**

<b>Code</b>	<b>Definition</b>	<b>Example</b>
Apology	Apologies used to soften direct speech acts.	“Sorry, it’s negative 2.”
Hedges	Qualifying words to reduce the intensity or certainty of utterances.	“You just add 5 to both sides.”
Extenders	Words used to indicate uncertainty by referring to vague categories.	“You have to multiply or something.”
Subjectivizer	Words that reduce intensity or certainty.	“I guess you divide by 3 here.”
Positive Feedback	Explicit evaluations of correct answers or steps.	“Yep, that’s it!”
Negative Feedback	Explicit evaluations of incorrect answers or steps.	“No, that’s wrong.”
Knowledge-telling	Explicitly stating procedures or steps for the tutee.	“Multiply it by the fraction.”

## Results

First, because friends are generally more talkative than strangers in their peer tutoring interactions (as described in more detail in Madaio et al., 2016), we normalized the amount of indirect utterances by the total number of utterances used by that participant in each session. By far the most frequently used marker of indirectness in our dataset was the use of hedges (e.g. “just”, “actually”, etc.) with normalized mean = .05 and standard deviation = .04, followed by subjectivizers (e.g. “I think”, “I guess”, etc) (m=.02, sd=.03), and apologies (m=.01, sd=.01), and with vague category extenders by far the most infrequent (e.g. “and stuff”, “or something”, etc.) (m=.002,

sd=.01). This distribution aligns with findings from Rowland’s (2007) work studying the use of hedges, subjectivizers (what he calls “shields”), and extenders in student-teacher mathematics lessons. To analyze tutors’ use of indirect delivery with their instructions and feedback, we first created an aggregate count of all indirect language, from each of the 4 indirectness codes described above for each participant. Then, after finding the co-occurrence of indirect language with tutors’ feedback and instructions, we normalized those frequency counts by the total number of “on-task” utterances, or the total number of annotated task-related questions, explanations, feedback, and metacognitive utterances, following Madaio et al. (2016).

Table 2: Means and Standard Deviations of Normalized Frequency of Indirectness and Indirect Tutoring

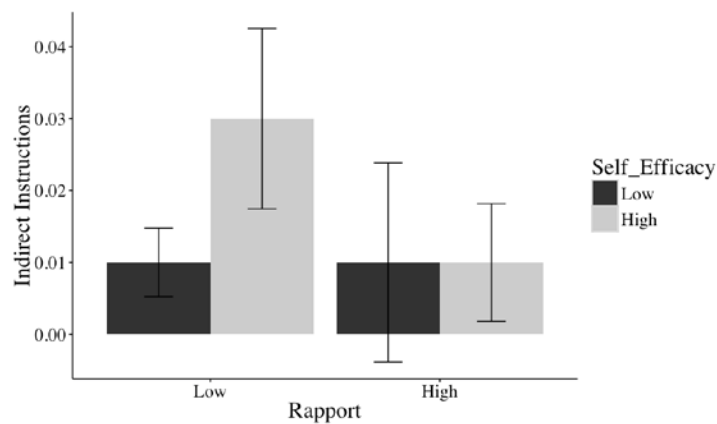
<b>Dyad Type</b>	<b>Indirectness</b>	<b>Indirect Feedback</b>	<b>Indirect Instructions</b>
All Tutors	.09 (.06)	.002 (.005)	.01 (.02)
Friend Tutors	.06 (.05)	.002 (.005)	.01 (.02)
Stranger Tutors	.11 (.06)	.002 (.01)	.02 (.03)

### Tutors’ use of positive and negative feedback (RQ1)

First, we investigated our hypothesis that tutors with lower interpersonal closeness with their tutees would use fewer instances of negative feedback (Person et al., 1995). To test this hypothesis, we ran a repeated measures ANOVA on the normalized frequency of tutors’ negative feedback, crossing the between-subjects factors of Relationship (Friend/Stranger), Rapport (High/Low), Self-Efficacy (High/Low), and the median split of the tutors’ Pre-Test performance (High/Low) with the within-subject factor of Session and Tutoring Period, and using each dyad’s tutoring period and session number as error terms. There was no statistically significant effect of any of the factors on tutors’ use of negative feedback. We then investigated whether stranger tutors were using more *positive feedback* to boost their tutees’ face, rather than avoiding negative feedback. We thus ran the same model on tutors’ use of positive feedback, finding a highly significant effect of relationship on the amount that tutors used positive feedback ( $F(1,64) = 12.8, p < .001$ ). To find the direction of that difference, we ran a t-test, and found that stranger tutors were significantly more likely ( $t(71) = 3.77, p < .001$ ) to use positive feedback ( $m = .17, sd = .12$ ) than friend tutors ( $m = .08, sd = .08$ ). We hypothesized that stranger tutors may be using more positive feedback because their tutees were solving more problems correctly. Therefore, we conducted a t-test, which showed that stranger tutees did not solve significantly more problems than friend tutees.

### Tutors’ use of indirect feedback and instructions (RQ2)

We then investigated the hypothesis that stranger tutors, and tutors in low-rapport dyads, would use a more indirect style than friend tutors and tutors with high-rapport with their tutees. We ran a repeated measures ANOVA on the tutors’ *indirect feedback*, using the same within-subject and between-subject factors and the same error terms as in RQ1. To investigate our alternative hypothesis about tutors’ self-efficacy interacting with their relationship with their tutee (RQ2b), we included interaction terms for Relationship, Rapport, and Self-Efficacy. There was no statistically significant effect of any factor on tutors’ use of indirect feedback. We then ran the same repeated measures ANOVA model on tutors’ use of *indirect instructions*, finding a highly significant main effect of relationship on the tutors’ use of indirect instructions ( $F(1,70) = 7.4, p < .01$ ). To find the direction of that difference, we ran a Mann-Whitney-Wilcoxon test, for comparing means of non-normal distributions, which showed that stranger tutors were significantly more likely ( $U = 3749, p < .05$ ) to use indirect instructions ( $m = .02, sd = .03$ ) than friend tutors ( $m = .007, sd = .02$ ). There was a significant interaction effect of Rapport and Self-Efficacy on all tutors’ use of indirect instructions ( $F(1,70) = 4.5, p < .05$ ), as seen in Figure 1. High-self-efficacy tutors with low rapport with their tutees were significantly more likely ( $U = 4052, p < .001$ ) to use more indirect instructions ( $m = .03, sd = .02$ ) than high self-efficacy tutors with *high* rapport with their tutees ( $m = .01, sd = .01$ ). This supports our hypothesis that tutors with lower interpersonal closeness (i.e. rapport) with their tutees would use more indirect instructions than those with greater interpersonal closeness. However, it is primarily the high self-efficacy tutors with low rapport with their tutees that appear to use this strategy, as they were (marginally) significantly more likely ( $U = 928, p = .07$ ) to use more indirect instructions ( $m = .03, sd = .02$ ) than low self-efficacy tutors with low rapport with their tutees ( $m = .01, sd = .01$ ).



**Figure 1.** Interaction Effect of Rapport and Self-Efficacy on Tutors' use of Indirect Instructions.

### Impact of tutors' indirect tutoring on tutees' problem-solving (RQ3)

Finally, we investigated our hypothesis that tutors' greater use of indirect feedback and instructions is associated with greater problem-solving for the tutee, particularly for dyads of lower interpersonal closeness. We ran a linear mixed effect model using the tutees' percent of problems solved in each tutoring session as the dependent variable, and using the normalized frequency of tutors' indirect feedback and instructions, tutors' *direct* feedback and instructions, and tutees' own self-explanations as fixed effects, along with interaction terms for each of the above with Relationship and Rapport, with random effects of Dyad and Session. We included the tutees' self-explanations, following the findings of Madaio et al. (2016) that tutees' self-explanations (i.e. annotations of knowledge-telling) were a more significant predictor of their learning than their tutors' instructions. We included the *direct* feedback instructions (in addition to the indirect) to see whether that directness impacted tutee learning, due to the potential for negative face-threat. While we found in response to RQ1 that strangers and friends solved the same number of problems overall, in this model, stranger tutors' use of indirect instructions was positively predictive ( $\beta = .64, p = .05$ ) of their tutees' problem-solving. We further investigated whether the use of indirectness with instructions and feedback might serve a motivational role, leading to an increased amount of *problems attempted* for the tutee. We thus ran the same mixed effects model, but with the tutees' number of problems attempted as the dependent variable. In this model, stranger tutors' use of *indirect instructions* was significantly positively predictive ( $\beta = .84, p < .001$ ) of their tutees' amount of problems attempted. No other factors were significant.

### Discussion

In this paper, we investigated the extent to which the interpersonal closeness (both short- and long-term) between peer tutors and their tutees impacts their learning, by way of the face-management that tutors engage in when providing instructions and feedback. A CSCL system that is not aware of the interpersonal closeness between collaborating students may mischaracterize the impact of face-threat mitigation behaviors and respond inappropriately. Here, the differences we found in tutors' use of indirectness with various tutoring behaviors suggest that interpersonal, relational aspects of the tutoring interaction impact the pedagogical task goals and vice-versa. Although stranger tutors used more positive feedback than friends, their tutees did not solve significantly more problems than friend-only dyads. This suggests that they may be using that positive feedback to boost their partners' face rather than accurately diagnosing the correctness of their partners' problem-solving. Further, the lack of significant effects of the positive feedback on tutee learning indicate that, while it may serve a relational, face-boosting role, it may not serve a pedagogically useful role, and may even, as Person et al. (1995) pointed out, lead to ambiguity about the correct procedures or answers, possibly eroding the tutee's trust in their tutor over time. Further, the stranger tutors' use of more indirect instructions than friend tutors suggests that stranger tutors may be hedging or qualifying their instructions to avoid the face-threat of directness. The benefits of this face-threat mitigation for dyads with low interpersonal closeness can be seen in the positive effect of indirect instructions on the amount of problems stranger tutees attempted and solved correctly. However, the benefits of indirect feedback for tutees disappear as the relationship strengthens between tutor and tutee. While interpersonal closeness affects tutors' use of face-threat mitigation strategies, not all tutors engage in equal amounts of face-management. Tutors' self-efficacy for tutoring also impacted their use of indirect instructions, as seen in the interaction between rapport and self-efficacy. As expected, tutors with low rapport with their tutee were more



likely than those with high rapport to give instructions indirectly. However, this effect was present only for highly self-efficacious tutors. Unexpectedly, there was no difference between low and high rapport dyads for tutors with low self-efficacy. This suggests that more self-efficacious tutors may be better able to modulate the directness of their instructions to attend more to the need for face-management when their interpersonal closeness with their tutee is lower, irrespective of their actual domain knowledge.

For the computer-supported collaborative learning community, researchers who design tools to support peer tutoring should be aware of the ways in which the interactional goals of tutoring may be impacted by the interpersonal goals of face-threat mitigation in the context of relationship-building. That is, the collaborative learning behaviors exhibited by students may differ depending on whether they are friends or strangers, or have a greater or lower rapport between them, and those same behaviors may have different impacts on student learning, depending on that closeness. Designers of collaborative intelligent tutoring systems, as in Olsen et al.'s (2014) work, might build in awareness of the interpersonal closeness between students, and, in addition to cognitive instructional supports, might recommend students phrase their instructions to each other more indirectly when interpersonal closeness is lower. These findings can also inform the design of educational dialogue systems, or conversational agents which could support peer tutoring by detecting and responding appropriately to the interpersonal dynamics between the human students and the conversational agent, *and* by building a deeper rapport with the students over time, as seen in (Zhao et al. 2014; Sinha et al., 2015) so as to be able to be more direct and blunt with instructions and feedback. To detect the interpersonal closeness described here, Yu et al. (2013) developed a method for the automatic prediction of friendship, which (the lack thereof) was the strongest predictor of the use of indirect language with instructions. Zhao et al., (2016) have developed a method for the automatic detection of rapport based on temporal association rules between multimodal data such as students' conversational moves and nonverbal behaviors and the subsequent change in their rapport. More broadly, this work contributes to a more robust understanding of the ways in which interpersonal closeness impacts the collaborative learning process by way of peer tutors' face-management. While prior work has hypothesized that social distance impacts the use of politeness and indirectness, our work provides two operationalizations of that social distance, friendship and rapport, and shows how each type of interpersonal closeness differently impacts peer tutors' face-threat mitigation while tutoring, showing the impact of those face-management tactics for different tutees' problem-solving.

## Future work and limitations

This work is part of a larger research program to understand the ways in which interpersonal rapport impacts teaching and learning, and will be used to inform the design of a conversational agent that simulates a peer tutor as the front end of an intelligent tutoring system. Our goal for this agent is to build rapport with its tutee, reducing the face-threat of particular instructional moves when necessary by using instructional moves in socially appropriate ways. One of the limitations of this work for investigating differences in the effects of rapport that can inform such an agent is the small sample size, particularly for dyads of strangers. We are currently conducting a similar study with 16 dyads of strangers to better understand how the rapport-building process develops within dyads starting from the same interpersonal baseline, and how that rapport impacts their teaching and learning process and outcomes. Another limitation of this work is the culturally dependent nature of what may be perceived as face-threatening or indirect by the interlocutors, and future work should take into account the culture of the participants in understanding face-threat. In addition, we are currently involved in investigating other potentially face-threatening pedagogical behaviors, to understand whether and how high-rapport dyads engage in for instance, cognitive conflict, help-seeking, help-offering, and others. While in this paper, using the normalized aggregate frequency of a particular set of annotated behaviors revealed interesting differences between groups of dyads, some of the most beneficial tutoring behaviors may occur infrequently or may have their benefits mediated by contingent patterns of use and responses from their partner (Ohlsson et al., 2007). Therefore, an analysis that does not take this contingent, temporal pattern of use into account may miss important effects. To better understand the contingent patterns of use for particular tutoring and learning behaviors, we plan to use temporal association and sequence mining to discover the sequences of pedagogical and social behaviors that contribute to greater rapport and learning, to predict rapport from multimodal features. We intend for this work to contribute to the design of socially-aware computer-supported collaborative learning systems, which can more appropriately respond to learners' social states in pedagogically beneficial ways.

## References

Ambady, N. and Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, 111(2):256, 1992.

- Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social development*, 2(3), 202-221.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language use*. Cambridge, England: Cambridge University Press.
- Burgoon, J. K., Stern, L. A., & Dillman, L. (1995). *Interpersonal Adaptation: Dyadic Interaction Patterns*. Cambridge University Press. <http://doi.org/10.1007/s13398-014-0173-7.2>
- Dillenbourg, P. (2015). *Orchestration graphs: Modeling scalable education*. EPFL Press.
- Fraser, B. (2010). Pragmatic competence: The case of hedging. *New Approaches to Hedging*, 15–34. Retrieved from <http://www.bu.edu/sed/files/2010/10/2010-Pragmatic-Competence-The-Case-of-Hedging.pdf>
- Johnson, W. L., & Rizzo, P. (2004). Politeness in tutoring dialogs: “Run the factory, that’s what I’d do.” *Intelligent Tutoring Systems, Proceedings*, 3220, 67–76. [http://doi.org/10.1007/978-3-540-30139-4\\_7](http://doi.org/10.1007/978-3-540-30139-4_7)
- Kerssen-Griep, J., Trees, A. R., & Hess, J. A. (2008). Attentive Facework During Instructional Feedback: Key to Perceiving Mentorship and an Optimal Learning Environment. *Communication Education*.
- Madaio, M. A., Ogan, A., & Cassell, J. (2016). The Effect of Friendship and Tutoring Roles on Reciprocal Peer Tutoring Strategies. *Intelligent Tutoring Systems*.
- Mojavezi, A., & Tamiz, M. P. (2012). The impact of teacher self-efficacy on the students’ motivation and achievement. *Theory and Practice in Language Studies*, 2(3), 483–491.
- Neary-Sundquist, C. (2013). The use of hedges in the speech of ESL learners. *Elia: Estudios de Lingüística Inglesa Aplicada*, 13, 149–174.
- Ogan, A., Finkelstein, S., Walker, E., Carlson, R., & Cassell, J. (2012). Rudeness and rapport: Insults and learning gains in peer tutoring. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7315 LNCS, 11–21.
- Ohlsson, S., Di Eugenio, B., Chow, B., Fossati, D., Lu, X., & Kershaw, T.C. (2007). Beyond the code-and-count analysis of tutoring dialogues. In *Proceedings of the 2007 conference on AI in Education*.
- Olsen, J. K., Belenky, D. M., Aleven, V., & Rummel, N. (2014, June). Using an intelligent tutoring system to support collaborative as well as individual learning. In *International Conference on Intelligent Tutoring Systems* (pp. 134-143). Springer International Publishing.
- Palinscar, A. S., & Brown, A. L. (1984). Reciprocal Teaching of Comprehension-Fostering and Comprehension-Monitoring Activities. *Cognition and Instruction*, 1(2), 117–175.
- Person, N. K., Kreuz, R. J., Zwaan, R. A., & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and instruction*, 13(2).
- Rowan, B., Chiang, F., & Miller, R. (1997). Using Research on Employees' Performance to Study the Effects of Teachers on Students' Achievement. *Sociology of Education*, 70(4), 256-284. doi:10.2307/2673267
- Rowland, T. (2007). ‘Well maybe not exactly, but it’s around fifty basically?’: Vague language in mathematics classrooms. In *Vague language explored* (pp. 79-96). Palgrave Macmillan UK.
- Saklofske, D., Michayluk, J., & Randhawa, B. (1988). Teachers’ efficacy and Teaching Behaviors, 407–414.
- Sinha, T., & Cassell, J. (2015, November). We click, we align, we learn: Impact of influence and convergence processes on student learning and rapport building. In *Proceedings of the 1st Workshop on Modeling INTERPERSONAL Synchrony And influence* (pp. 13-20). ACM.
- Sinha, T. (2016). Cognitive Correlates of Rapport Dynamics in Longitudinal Peer Tutoring. Available at <http://tinyurl.com/RapportDynamicsSinha2016>
- Spencer-Oatey, H. (2005). (Im)Politeness, face and perceptions of rapport: unpacking their bases and interrelationships. *Politeness Research*, 1(1)(Im), 95–119. <http://doi.org/10.1515/jplr.2005.1.1.95>
- Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological inquiry*, 1(4), 285-293.
- Yu, Z., Gerritsen, D., Ogan, A., Black, A., & Cassell, J. (2013). Automatic Prediction of Friendship via Multi-model Dyadic Features. *Proc. of the SIGDIAL 2013 Conf: Association for Computational Linguistics*.
- Zhao, R., Papangelis, A., & Cassell, J. (2014). Towards a Dyadic Computational Model of Rapport Management for Human-Virtual Agent Interaction, 514–527.
- Zhao, R., Sinha, T., & Cassell, J. (2016). Socially-Aware Virtual Agents: Automatically Assessing Dyadic Rapport from Temporal Patterns of Behavior. In *Intelligent Virtual Agents*.

## Acknowledgments

The research reported here was supported, in whole or in part, by the National Science Foundation Cyberlearning Award No.1523162, and the Institute of Education Sciences, U.S. Department of Education, through Grant R305B150008 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

# Mobile City Science: Technology-Supported Collaborative Learning at Community Scale

Katie Headrick Taylor, University of Washington, kht126@uw.edu  
Deborah Silvis, University of Washington, dsilvis@uw.edu

**Abstract:** In a new era of digital media and democracy, there is widespread concern that technologies have incapacitated us from learning and teaching across diverse communities and perspectives. While this notion may ring true in certain contexts, this paper describes a study, “Mobile City Science,” that designed a novel learning experience in which educators and young people used mobile and place-based technologies to document and analyze the diverse perspectives of community members living in rapidly changing urban areas. The objective of this work was to teach young people digital literacies associated with “city science,” an emerging interdisciplinary field that creates data-driven approaches to complicated community issues. Participants were videotaped as they collected and analyzed information about a specific neighborhood using wearable cameras, GPS devices, heart rate monitors, and a GIS software. Early findings show that Mobile City Science uses technology to engage people with diverse perspectives around a community scale problem.

## Major issues addressed

Technology has changed the nature of political engagement. For every “success” story of increased government transparency and youth mobilization, there is an instance of political balkanization and divisiveness (e.g., Manjoo, 2016). In this new era of digital media and technology, many have argued that the sheer ubiquity of our technological interactions have incapacitated us from learning and teaching across diverse communities and perspectives. In an article titled, “How We Broke Democracy,” Rose-Stockwell wrote, “If we cannot build the tools of our media to encourage empathy and consensus, we will retract further into toxic divisions that have come to define us today” (2016). This study, “Mobile City Science,” represents one such attempt to use technology and digital media as a means of encouraging empathy and building consensus around a “live” community problem. Importantly (and in contrast to Facebook and Twitter), technologies in Mobile City Science engage young people and youth educators *in neighborhoods*, in face-to-face interactions, to generate new information and representations of diverse perspectives.

Before the most recent presidential election, social science was promoting the idea that “big data,” produced by our multiple devices and technologies, would make cities and their citizens “smarter,” or work better together. After the results of the election came through on November 8, 2016, several fundamental questions are now being asked about the promise of big data. Whose lives do these data actually represent? Who learns what from these data? What is the origin of data, and who has legitimacy to make interpretations and arguments from it? How did communities become so balkanized and bifurcated, bolstered by “data?”

While these questions remain problematic issues for political engagement at large, *they also open-up novel teaching and learning opportunities for underrepresented young people to create and engage with vast amounts of data across stakeholders that may have divergent perspectives on community issues.* Creating insights and data-driven approaches to community issues, or “city science,” is an interdisciplinary field that is emerging alongside ubiquitous computing (MIT Media Lab, n.d.). Geospatial applications and mobile devices are especially conducive to this kind of data-driven inquiry process; these tools support and promote moving around the community and interfacing with shopkeepers, residents, and visitors *in place*. The mapping capabilities of geospatial apps and tools support a spatio-temporal way of recording, interpreting, arguing from the data collected around the neighborhood. These technological affordances have been shown to differently engage people in community-based issues, in ways that leverage physical mobility, be it walking, bussing, or bicycling around a geographic area (e.g., Nold, 2009; Taylor, 2013).

## Potential significance of the work

As a whole, this research takes seriously the role of underrepresented young people using technology in Jane Jacobs’ (1961) provocation that “...cities have the capability of providing something for everybody, only because, and only when, they are created by everybody” (p. 238). Mobile City Science (MCS) provides four key contributions to the learning sciences and other fields concerned with new ways of engaging underrepresented youth in community-level issues and dialogue through technology. First, MCS informs and contributes to theories of embodied learning (e.g., Alibali & Nathan, 2012; Farnell, 1999; Glenberg, Gutierrez, Levin,

Japuntich, & Kaschak, 2004; Goldin-Meadow, Cook, & Mitchell, 2009) by analyzing how and what young people, and the people that educate them, learn about complex community issues from being on-the-move through their neighborhoods with mobile and location aware technologies. Second, MCS formalizes innovative ways of teaching community engagement to young people living in underserved areas of the city; MCS is a set of on-the-move teaching and learning experiences that support young people in collecting, analyzing, and arguing from spatial and other forms of data (e.g., GPS tracks, geo-referenced video files, density plots, place-elicited interviews). Again, these data represent the diversity of lived experiences within the geographic area. Third, MCS provides accessible, technologically enhanced ways for youth-serving organizations, community developers, urban planners, and/or social science educators to engage young people in civic processes and conversations happening at the scale of the city. Finally, MCS will contribute to a new theory of social change where technologies potentially democratize (rather than balkanize) learning and participation (e.g., Bilkstein, 2013; Papert, 1991; Resnick, et al., 2000; Wilensky & Papert, 2010) in processes of community development to include young people's data-driven perspectives in planning and policy implementation.

## Approaches

The past decade has seen rapid growth in development and distribution of location-aware technologies (e.g., GPS in mobile devices), digital mapping tools (e.g., Google Maps, Open Maps), and tools for spatial analysis and modeling (e.g., QGIS, MapBox, mobile augmented reality). These technologies create new opportunities for linking data on personal mobility with a growing variety of spatially organized, open, and large-scale data sets (Busch, 2014; Kahn & Hall, 2016). In the hands of consumers, including school-aged youth, these tools and diverse sources of information are increasingly taken up in emerging practices of "neo-geography" (Goodchild, 2009; Graham, 2009; Liu & Palen, 2010). Some of these practices involve substantive learning opportunities in overlapping conceptual domains for STEM disciplines (e.g., Enyedy & Mukhopadhyay, 2007; Radinsky, 2008) and the humanities (e.g., Farman, 2013, 2015; Van Wart, Tsai, & Parikh, 2010).

## Theoretical approaches

Engaging young people and adult stakeholders in mobile experiences through the city exemplifies a blurring of physical experience and representation that is so central to theories of embodiment. Blurring the disconnect between the physical and the represented is the basis for embodied learning, and is arguably a more authentic means for teaching and learning irrespective of content domain (e.g., Alibali & Nathan, 2012; Farnell, 1999; Glenberg et al., 2004; Goldin-Meadow, Cook, & Mitchell, 2009). Embodied learning theorizes learning as a process that is distributed across all of the sense-making modes of the body. This multimodality of sense-making is fundamental to not just learning *in situ*, but to how we organize "perceptual symbol systems" (Barsalou, 1999) individually and socially. For instance, Lakoff & Johnson (1980) demonstrated how seemingly abstracted metaphors are grounded in, or harken back to, some fundamental embodied experience that makes them almost universally understood in interactions.

Importantly for this work, researchers pursuing questions with an embodied learning lens carefully consider how all of these meaning-making modes are mediated by the various technologies that permeate daily life (e.g., Hall, Ma & Nemirovsky, 2015; Lee & Thomas, 2011; Keating & Sunakawa, 2011). An aim of MCS is to contribute to a more robust theory of embodied learning by providing descriptive and comparative analyses of how young people and adult professionals make "body sense" (Cajete, 2000) of the urban spaces in which they are engaging with technology. In a special issue of the *Journal of the Learning Sciences*, researchers provided arguments for the centrality of the body for learning mathematics (Hall & Nemirovsky, 2012). While that work has most recently promoted embodied cognition as a viable theory of learning, educators and educational researchers still struggle against the tendency to fetishize abstracted, "pure knowledge" over the ways in which our moving, feeling bodies make sense of the world. Additionally, and as Stevens (2012) pointed out in his commentary to the "Modalities of Body Engagement in Mathematical Activity and Learning" *JLS* Special Issue, attempting to build a broad understanding of how the body and learning relate via solely classroom-based studies is a major limitation to robust theory-building. Thus, MCS provides a (literally) grounded account of embodiment that problematizes the absence of space and place in previous accounts of learning.

Cognitive and learning scientists agree that "the original function of any brain is to control motion – only mobile organisms have brains" (Streek, Goodwin, & LeBaron, 2011, p. 7). However, leveraging physical mobility as a learning resource is still undervalued and underutilized in more traditional learning arrangements. Some researchers have pushed the possibilities of learning on-the-move with location aware and wearable technologies for creating place-based, mobile experiences (e.g., Hall, Ma, & Nemirovsky, 2014; Rosner, et al., 2015; Ryokai & Agogino, 2013; Taylor & Hall, 2013; Taylor, accepted). One early example of using mobility

as a means for learning about and showing an experience of a place is the art of Jeremy Wood, an artist working at the intersection of cartography, visual arts, and geometry (Lauriault & Wood, 2009). Wood made visible his own embodied experiences through the landscape by re-appropriating GPS, a tool historically used by the US Military to precisely locate and track its missile submarines (Sample, 2014). Wood's "GPS drawings" represented a coordinating effort of walking (or driving or biking or swimming) and technology to tell a story about a place. These stories, in the form of geospatial representations, taught viewers about a practiced, lived landscape that an overhead satellite map makes invisible. In this way, physical mobility (mediated by geospatial technology) not only allowed the maker to see a place differently through first walking it and then viewing the GPS tracks, but also taught an audience about a different experience of that place.

This novel way of using mobile mapping technologies highlights how walking especially remains an act of civic and political agency (Rosner, 2015; Ingold, 2007; Suchman, 1995). Silent marches, migrations, and "memory walks" (Bonilla, 2011) are acts of dissatisfaction with the status quo; these walks also represent a process of teaching and learning where diverse publics inform more powerful others of inequitable lived experiences. This notion of mobility was a launching point for Mobile City Science; this designed experience supports youth to produce maps that are representative of their lived experiences of being a young person in a rapidly shifting urban setting, and the experiences of others. Moreover, MCS is intended to bridge the world of mapping and quantification familiar to urban planners and cartographers to the world of place-based storytelling and affect to produce a new opportunity space for teaching and learning at the interface of diverse publics (e.g., young people, shopkeepers, residents, urban planners, neighborhood school educators, parents).

## Methodological approaches

Mobile City Science is a community-based design research project (e.g., Bang, Faber, Gurneau, Marin, & Soto, 2016) to understand and create new forms of techno-civic engagement for young people with mobile and location-based tools. We are asking these top-level research questions that guide the design study: (1) How do participants (i.e., young people and youth educators) use the designed MCS activities (e.g., historic neighborhood geocache, GPS drawing) to engage diverse perspectives on a community issue? (2) What kinds of data do participants collect and produce about their communities using these tools? (3) Whose lived experiences and perspectives do these data represent? (4) How are young people using data, and these tools, to make informed arguments and representations about changes to their communities?

We are working closely with three youth-serving organizations, Oasis Center, Digital Youth Network, and New York Hall of Science. These three organizations are highly respected for their technologically rich versions of "citizen science" they already do with young people and adult community members in Nashville, Chicago, and Queens, New York. We held, and are still holding, co-design meetings that are audio recorded and analyzed thematically. We held a training for youth educators from these organizations to implement the Mobile City Science curriculum in their respective communities. The four-day training was audio and video recorded.

The MCSC we create builds upon the current on-the-move citizen science programs that these youth-serving institutions already deliver. Importantly, MCS activities leverage mobile and location aware technologies for youth to collect, analyze, and argue from spatial and other forms of data. Some of the following teaching activities, and corresponding learning objectives, are included in past and ongoing MCS implementations (see Figures a, b, c, and d below for examples):

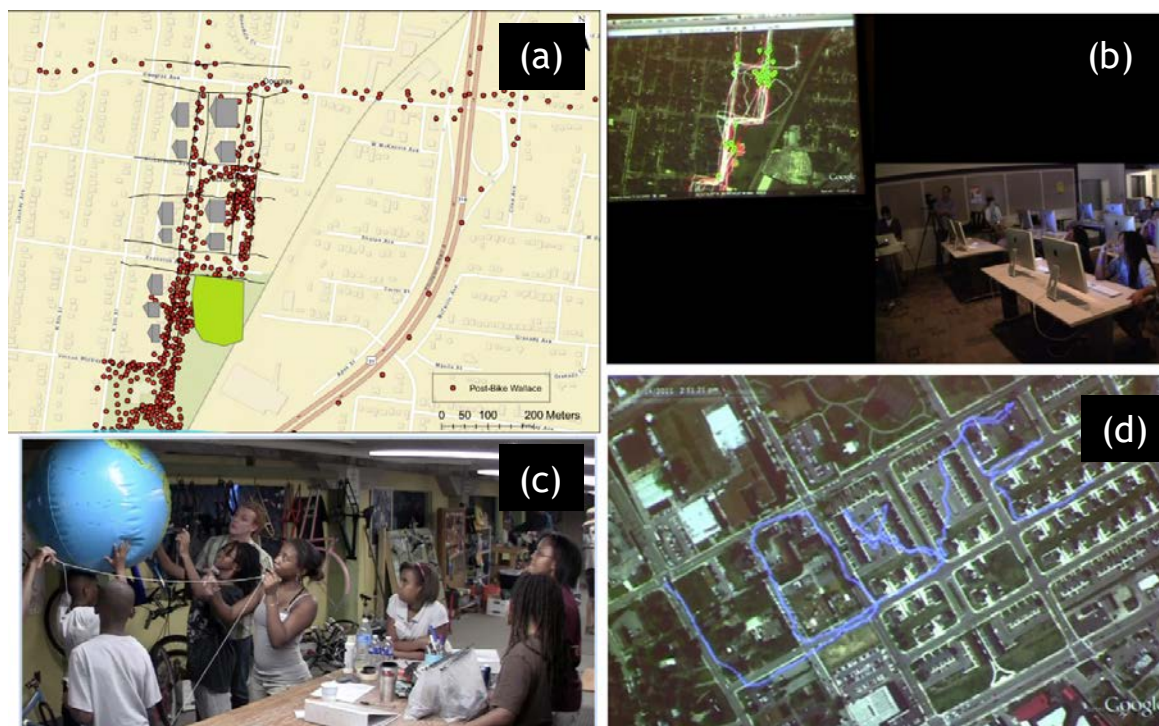
### MCS designed activities

- *Pre/Post Assessment*: On the first and last days of the MCSC, young people are asked to draw their neighborhood on paper (i.e., "free recall maps," Hart, 1977) and asked open protocol questions as they draw. The changes between the free recall maps drawn on the first day of the study and the final day help us understand how young people's perceptions of their neighborhoods, and maps as a particular way of representing their neighborhoods, have changed over the course of the curriculum.
- *Youth Neighborhood Data Collection* supports young people to understand the *current* state of their neighborhood. Two examples of collecting neighborhood data are below:
  - Youth conduct **walking and biking audits** of the neighborhood with maps in hand. As they move through the neighborhood, they annotate the map with the routes they take, things that are missing on the map, map inaccuracies, and conventions that would be helpful for this way-finding exercise. Young people also take geo-tagged digital photos, and collect heart rate data (dramatic elevation changes are a real barrier for pedestrians and cyclists, especially in Seattle). They add this data, including the photos and heart rate information, into an open source GIS software (i.e., Google Earth). There are two important learning objectives

associated with this activity. The first learning objective is to learn about maps and spatial data through different representations (i.e., a paper map and a digital map) of a familiar place. The second related objective is for young people to learn about what different types of maps show, ignore, and highlight and how these different portrayals achieve a different purpose and can convey very different messages regarding the same place.

- Youth conduct an **historic geocache** of the neighborhood with GPS devices in hand to collect historic data about the place in which they live (e.g., famous residents, important institutions, urban development decisions). At each cache, they have to produce some form of data (e.g., geo-tagged photo, geo-referenced video, GPS tracks, interview). Participating youth add this data as another map layer into an open source GIS software. The learning objectives for this activity are threefold. First, young people learn how different spatial phenomena are connected and impact people differently, over time. Secondly, young people gain facility with GPS devices and open source GIS software through recording and representing their own data. Related to this second objective, youth will begin to understand the affordances and limitations of these technologies based on one's purpose for using them. Third, young people are encountering notions of difference and multiple experiences of the same place.
- *Youth Data Analysis* supports young people to interrogate data as a different way of *seeing* one's neighborhood. Two examples of analyzing data are below:
  - Participating youth collect and analyze their **personal time geography** with GPS track data that show "typical" routes and locations that young people use and frequent (Hagerstrand, 1979). They analyze this data in a GIS to learn how to look for patterns, problems, and possibilities from spatial data. For example, patterns might be in relation to the grid network. Problems might be about the interstate system cutting through the neighborhood. And possibilities could be about additional bus routes for youth to access a library.
  - Thinking closely about the data they have collected, young people create **asset maps** of the neighborhood and greater community. Asset maps show the things that are working well about the geographic area, the deficits that exist, and the assets young people are imaging for the future of this place. As an instructional activity, asset mapping will support youth to think about places as developing over time and that this development is, in part, planned and mapped first.
- *New Neighborhood Maps* support young people to show and say something new about their neighborhoods from the multiple perspectives they have encountered and considered. Two examples of creating new neighborhood maps are below:
  - In the neighborhood, participating youth produce **GPS drawings** with GPS devices in hand (e.g., Lauriault & Wood, 2009). GPS drawings are novel messages or images written or drawn at the scale of the neighborhood chosen and produced by youth. After walking or biking these new messages or images, the GPS tracks are loaded into a GIS and shared with the group. As a neo-geography practice, young people learn about the emerging affordances of these mapping tools for showing affective responses to the places in which they live.
  - Young people build geospatial simulations, or **mobile augmented reality** (MAR; Heinrich et al., 2008; Rothfarb, 2011; Ryokai & Agogino, 2013) experiences using *Aris* (an open source platform for building map-based, mobile tours and interactive stories). These MAR experiences allow others to experience potential changes to the neighborhood and community from the perspective of the young person making it. In other words, as a person walks through a young person's community with *Aris* open on her phone, that young person's suggested changes to the built environment *in that place* will populate the smartphone screen (think Pokemon Go without the monsters). Building MAR experiences is a chance for young people to consider ways of building persuasive spatial arguments around *their own experience of living in an underserved area* of the city.
- *Youth Data-Driven Arguments* support young people to recommend evidence-based changes for their neighborhoods through maps. Participating youth create **counter-maps** of their neighborhoods, using all of the data collection and analysis they have done over the course of the study, and share these with their peers, adult educators, and researchers. This part of the curriculum allows young people to practice making claims to space and resources from spatial data. Young people share counter-maps with the community stakeholders with whom they came into contact during the above on-the-move

activities, and also interested professionals (e.g., planners, transportation engineers, public school administrators and educators).



Figures 1a, 1b, 1c, 1d. These images show snapshots of some of the teaching activities involved in MCS. Image (a) shows a participant’s free recall map geo-referenced to track data he collected and a base layer map of his neighborhood using a GIS. Image (b) shows participants examining and talking about their GPS track data during the personal time geography activity. Image (c) shows participants learning about GPS technology before doing the historic neighborhood geocache. Image (d) shows what one group wrote/walked through their neighborhood during the GPS drawing activity.

### Data collection and analysis

For each MCS implementation, we are collecting video recordings of youth engagement, including their own wearable camera footage. We also collect artifacts that young people (and educators in the training) create as analyzable materials. These artifacts include free recall maps, GPS tracks, geo-tagged photos, digital maps, mobile augmented reality (MAR) experiences, and interviews with community members. These artifacts are treated as evidence of learning, or youth changing their perspectives over time, since their technical and community engagement practices become more sophisticated over the course of their participation in MCS. We also conducted ongoing and retrospective analyses of the design experiment as a “paradigm case” (Cobb et al., 2003, p. 13; Brown, 1992) of on-the-move embodied learning throughout and about the city. As we run and analyze implementations, we refine the design of MCS to address youth-relevant issues that are specific to each community, and to better address our design objectives. Implementations also provide comparative material from which to see differences in youth participation and learning based on the specific context and the refined design of the “experience.”

### **Major findings**

While iterations of Mobile City Science are still underway in Chicago and New York, we can report some major findings from the first rounds of youth participation (Nashville) and educator trainings (Seattle). Also, some findings come from facilitated design charrettes where youth participants shared their findings and arguments with urban planners, directors of the non-profits serving the youth of that community, cartographers, and parents.

The first major finding of this work is that moving between daily, lived experience (at the embodied, mobile, and practiced scale) and the representation of that experience (at the abstracted, static, and mapped scale) provides young people, and adult educators, with new insights about the technological and data

infrastructures that support community level decisions. Because MCS participants are generating *original* data of a familiar neighborhood or community, they see how and what gets distilled from lived experience to a “data point” so that it can be mapped. Participants often find this distillation process (e.g., Bowker & Star, 2000) unnerving or frustrating. In the GPS drawing activity, for instance, a group of young women expressed concern about their track data not showing how difficult it was to walk up and down steep hills and the overall physical effort that task required. After talking with a local merchant during the geocache, a participant struggled to find a way to *show*, through maps, the shopkeeper’s anxieties about a quickly gentrifying neighborhood

The second major finding is that re-mediating the daily mobility of young people with old and new technologies (e.g., bicycles, GPS devices) provided youth with powerful comparative leverage when ground-truthing maps and arguing for “on the ground” changes. On-the-move activities leveraged all of the sense-making modes of the body so that learning/problem-solving experiences were vividly recounted and used as evidence in conversations with (adult) stakeholders. During a final community design charrette, several youth argued for a protected bicycle path along a high traffic corridor. To make their point, several young people animatedly recounted, and then showed their maps of, being “buzzed” by a city bus while they were on bicycles along a designated bike route. The MCS facilitators also showed the head camera footage of that event. The adult stakeholders that attended were visibly moved by this data-driven argument.

Finally, the familiar context of the neighborhood and/or school community provided participants with a biographically relevant setting and authentic motivation for imagining a better future for that place. The familiar context of the neighborhood and/or school community allowed young people to leverage a variety of learning resources – neighbors, favorite locations, memories – that are often absent in de-contextualized, stationary mapping exercises or citizen science more generally. The adult residents, professionals, and stakeholders with whom youth came into contact during the on-the-move activities often fueled this interest and motivation for youth as they provided previously unknown stories and concerns about the community. In another example from the historic neighborhood geocache activity, a group of MCS participants spoke with a local shopkeeper who had been operating her store in the same location for more than twenty years. The shopkeeper alerted the MCS participants that the neighboring university was advocating for a dramatic change to the “character” of the neighborhood to include high-rise buildings; she was afraid that this shift would force her out of her building, unable to afford the higher rents. While the MCS participants were familiar with the argument to increase density in urban areas for several reasons, they were not aware of the consequences this could have on longtime residents and shopkeepers in older, rent stabilized buildings.

## Conclusions and implications

While technologies continue to emerge, the demands they make on us change, but so do the opportunities they give us. Rather than thinking only of digital media and technology as keeping us apart (cf., Turkle, 2011), isolating us from the range of lived experiences and perspectives that exist within even a single community (or home), Mobile City Science follows the tradition of computer-supported collaborative learning. Mobile City Science represents an attempt to use technology and digital media as a means of learning through empathy and consensus building around a “live” community problem. The technologies in Mobile City Science engage young people and youth educators *in neighborhoods*, in face-to-face interactions, to generate new information (data) and representations of diverse perspectives. MCS as a learning experience supports young people to identify important local issues, understand multiple perspectives on those issues, and synthesize across these data to produce a cogent recommendation for neighborhood change. Similar to the intended outcomes of work by Gordon, Elwood, and Mitchell, (2016), MCS “expands adult-centric notions of civic agency and develops participatory mapping practices that elicit young people’s knowledge on their own terms” (p. 2). Creating and arguing from data produced with geospatial technologies is an emerging techno-civic literacy that young people must learn for influencing change in their communities.

After the latest presidential election, President Obama stated that our current media ecosystem “means everything is true and nothing is true” (Remnick, 2016). By this statement, he meant that “non-truths” are easily circulated through digital media and people cannot only become misinformed but also bolstered by misinformation that agrees with their existing worldview. This issue is a national teaching and learning crisis; people can use technology to roadblock consensus building from a shared set of data treated as “truth.” In such a context, we argue that the learning sciences must focus its efforts on creating not just tools, but complete teaching and learning experiences, that encourage empathy and consensus across a *diverse* range of stakeholders and community members. Otherwise, we may continue to “retract further into toxic divisions that have come to define us today.” We hope Mobile City Science is one such teaching and learning experience for young people and the adults that educate them in youth-serving organizations.



## References

- Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, 21(2), 247-286.
- Bang, M., Faber, L., Gunneau, J., Marin, A., & Soto, C. (2016). Community-based design research: learning across generations and strategic transformations of institutional relations toward axiological innovations. *Mind, Culture, and Activity*, 23(1), 28-41.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences*, 22(04), 577-660.
- Blikstein, P. (2013). Digital fabrication and 'making' in education: The democratization of invention. *FabLabs: Of machines, makers and inventors*, 1-21.
- Blikstein, P., & Wilensky, U. (2009). An atom is known by the company it keeps: A constructionist learning environment for materials science using agent-based modeling. *International Journal of Computers for Mathematical Learning*, 14(2), 81-119.
- Bonilla, Y. (2011). The past is made by walking: Labor activism and historical production in postcolonial Guadeloupe. *Cultural Anthropology*, 26(3), 313-339.
- Bowker, G. C., & Star, S. L. (2000). *Sorting things out: classification and its consequences*. Cambridge: MIT Press.
- Brown, A.L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2(2), 141-178.
- Cajete, G. (2000). *Native science: Natural laws of interdependence*. Santa Fe, NM: Clear Light Publishers.
- Cobb, P., Confrey, J., Disessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9.
- Enyedy, N., & Mukhopadhyay, S. (2007). They don't show nothing I didn't know: Emergent tensions between culturally relevant pedagogy and mathematics pedagogy. *The Journal of the Learning Sciences*, 16(2), 139-174.
- Farman, J. (2013). *The mobile story: Narrative practices with locative technologies*. Routledge.
- Farnell, B. (1999). Moving bodies, acting selves. *Annual Review of Anthropology*, 341-373.
- Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and Imagined Activity Can Enhance Young Children's Reading Comprehension. *Journal of Educational Psychology*, 96(3), 424.
- Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, 20(3), 267-272.
- Goodchild, M. F., & Janelle, D. G. (2010). Toward critical spatial thinking in the social sciences and humanities. *GeoJournal*, 75(1), 3-13.
- Gutiérrez, K. D., & Vossoughi, S. (2009). Lifting off the ground to return anew: Mediated praxis, transformative learning, and social design experiments. *Journal of Teacher Education*.
- Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science*, 24(1), 6-21.
- Hall, R., Ma, J. Y., & Nemirovsky, R. (2014). Rescaling Bodies in/as Representational Instruments in GPS Drawing. In V.R. Lee (Ed.), *Learning Technologies and the Body: Integration and Implementation In Formal and Informal Learning Environments* (112-135). Routledge.
- Hall, R., & Nemirovsky, R. (2012). Introduction to the special issue: Modalities of body engagement in mathematical activity and learning. *Journal of the Learning Sciences*, 21(2), 207-215.
- Halverson, E. R. (2013). Digital art making as a representational process. *Journal of the Learning Sciences*, 22(1), 121-162.
- Hart, R. (1979). *Children's experience of place*. Irvington.
- Heinrich, M., Thomas, B.H., & Mueller, S. (2008). ARWeather: An augmented reality weather system. In *Proceedings of the 7<sup>th</sup> IEEE/AM International Symposium on Mixed and Augmented Reality* (2008 March). Washington, D.C.: IEEE Computer Society.
- Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive science*, 19(3), 265-288.
- Ingold, T. (2007). *Lines: a brief history*. Routledge.
- Jacobs, J. (1961). *The Death and Life of Great American Cities*. Vintage.
- Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The journal of the learning sciences*, 4(1), 39-103.
- Kahn, J. & Hall, R. (2016). Getting personal with big data: Stories with <sup>[11]</sup>multivariable models about global health and wealth. Paper presented at the annual meetings of the *American Educational Research Association*, Washington, D.C. Best Paper award, SIG for Learning Sciences and Advanced Technologies.

- Keating E, Sunakawa C. (2011). 'A full inspiration tray': multimodality across real and virtual spaces. In Goodwin, C. & LeBaron, C. (Eds.), *Embodied Interaction in the Material World* (194–206). Cambridge: University Press.
- Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, 4(2), 195-208.
- Lauriault, T. P., & Wood, J. (2009). GPS tracings—personal cartographies. *The Cartographic Journal*, 46(4), 360-365.
- Lee, V. R., & Thomas, J. M. (2011). Integrating physical activity data technologies into elementary school classrooms. *Educational Technology Research and Development*, 59(6), 865-884.
- Manjoo, F. (2016, November 16). Social media's globe-shaking power. *The New York Times*. Retrieved from [http://www.nytimes.com/2016/11/17/technology/social-medias-globe-shaking-power.html?\\_r=0](http://www.nytimes.com/2016/11/17/technology/social-medias-globe-shaking-power.html?_r=0)
- MIT Media Lab. (n.d.). Retrieved from <https://www.media.mit.edu/special/groups/city-science>
- Nold, C. (2009). Emotional Cartography. *Bio Mapping website*.
- Papert, S. (1991). Situating Constructionism. In S. Papert & I. Harel (Eds.), *Constructionism*. Cambridge, MA: MIT Press.
- Radinsky, J. (2008). Students' roles in group-work with visual data: A site of science learning. *Cognition and Instruction*, 26(2), 145-194.
- Remnick, D. (2016, November 28). Obama reckons with a Trump presidency. *The New Yorker*. Retrieved from <http://www.newyorker.com/magazine/2016/11/28/obama-reckons-with-a-trump-presidency>
- Resnick, M., Berg, R., & Eisenberg, M. (2000). Beyond black boxes: Bringing transparency and aesthetics back to scientific investigation. *Journal of the Learning Sciences*, 9(1), 7-30.
- Rose-Stockwell, T. (2016, November 11). How we broke democracy: Our technology has changed this election, and is now undermining our ability to empathize with each other. *Medium*. Retrieved from <https://medium.com/@tobiasrose/empathy-to-democracy-b7f04ab57eee#.m7ru89641>
- Rosner, D. K., Saegusa, H., Friedland, J., & Chambliss, A. (2015, April). Walking by Drawing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 397-406). ACM.
- Rothfarb, R.J., (2011). Science in the city AR: Using mobile augmented reality for science inquiry activities. In *Proceedings of the ACM SIGGRAPH 2011 Posters*. ACE, New York, NY.
- Ryokai, K., & Agogino, A. (2013). Off the paved paths: Exploring nature with a mobile augmented reality learning tool. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 5(2), 21-49.
- Sample, M. (2014). Location is Not Compelling (Until It is Haunted). In J. Farman (Ed.), *The mobile story: Narrative practices with locative technologies*. Routledge.
- Stevens, R. (2012). The missing bodies of mathematical thinking and learning have been found. *Journal of the Learning Sciences*, 21(2), 337-346.
- Suchman, L. (1995). Making work visible. *Communications of the ACM*, 38(9), 56-ff.
- Taylor, K. H. (2013). Counter-mapping the neighborhood: A social design experiment for spatial justice. Ph.D. Nashville, TN: Vanderbilt University.
- Taylor, K.H. (accepted). Learning along lines: Locative literacies for reading and writing the city. *Journal of the Learning Sciences*, (forthcoming).
- Taylor, K. H., & Hall, R. (2013). Counter-mapping the neighborhood on bicycles: Mobilizing youth to reimagine the city. *Technology, Knowledge and Learning*, 18(1-2), 65-93.
- Townsend, Anthony M. (2013) *Smart cities: big data, civic hackers, and the quest for a new utopia*. WW Norton & Company.
- Van Wart, S., Tsai, K. J., & Parikh, T. (2010, December). Local ground: A paper-based toolkit for documenting local geo-spatial knowledge. In *Proceedings of the First ACM Symposium on Computing for Development* (p. 11). ACM.
- Wilensky, U., & Papert, S. (2010). Restructurations: Reformulating knowledge disciplines through new representational forms. Paper presented at *Constructionism 2010*, Paris, France.

## Acknowledgments

The work is funded by NSF #1645102. We would like to acknowledge the loving and empathetic work the adult educators do with youth at Oasis Center, Digital Youth Network, and New York Hall of Science. We are grateful to have the chance to work with you and with the communities of young people you serve.

# Scientific Discourse of Citizen Scientists: A Collaborative Modeling as a Boundary Object

Joey Huang, Indiana University, [huang220@indiana.edu](mailto:huang220@indiana.edu)  
Cindy Hmelo-Silver, Indiana University, [chmelosi@indiana.edu](mailto:chmelosi@indiana.edu)  
Rebecca Jordan, Rutgers University, [rebecca.jordan@rutgers.edu](mailto:rebecca.jordan@rutgers.edu)  
Troy Frensley, Virginia Tech, [btfren@vt.edu](mailto:btfren@vt.edu)  
Steven Gray, Michigan State University, [stevenallangray@gmail.com](mailto:stevenallangray@gmail.com)  
Greg Newman, Colorado State University, [Gregory.Newman@colostate.edu](mailto:Gregory.Newman@colostate.edu)

**Abstract:** In this study, we examine participants' practices in two citizen science projects in order to explore the use of scientific knowledge and practices as they engage in collaborative modeling. They use the Mental Modeler, an online resource to facilitate science engagement and collaboration. This paper applies an analytical approach that uses visual representations to understand the shifts in scientific discourse and interpret complex interaction patterns between participants and facilitators in two citizen science projects. The findings suggest that the Mental Modeler serves as a boundary object that allows participants and facilitators to collaboratively engage with scientific topics and practices through the development of scientific discourse and learning.

**Keywords:** citizen science, collaborative learning, scientific practices, engagement

## Introduction

Citizen science refers to partnerships between scientists and the public in scientific research in which real data are collected and analyzed (Jordan, Ballard, & Phillips, 2012). Citizen science can provide opportunities for informal learning as citizen volunteers are engaged in scientific practices, including modeling, gathering evidence, and evaluating arguments. Most citizen science projects are contributory in which scientists design the project and include the public in data collection (e.g., Nicosia et al., 2014). Participation tends to only involve data collection rather than engaging in a full range of science practices. However, collaborative and co-created projects have great potential benefits to maintain participants' engagement and closely address stakeholders' interests and concerns (Jordan et al., 2016). In addition, collaborative science projects can enhance public engagement where researchers and citizen scientists collect new information and learn from each other in relation to the local environmental issues. The processes of collaborative decision-making and project planning may increase the capacity of scientific discourse and practices of both researchers and scientists. This study included two collaborative projects in order to examine the processes of project planning, modeling, and collaborative learning of citizen scientists.

Although collaborative learning has been emphasized in science learning (e.g., Cornelius et al., 2013), it is only recently that informal online collaborative tools have been designed and used to facilitate learning and engagement in scientific practices among citizen scientists. The purpose of this study is to apply an analytical approach to investigate participants' scientific knowledge, practices, and engagement with the use of online collaborative tools. In addition, the analytical approach provides visual representations to better understand the shifts in scientific discourse and aid in interpretation of complex patterns between participants and facilitators of citizen science projects.

## Theoretical framework

Public engagement with science (PES) refers to an avenue for collaborative discourse in informal settings in which individuals with varied life experiences and scientific expertise participate and engage with scientific activities or events (McCallie et al., 2009). During the process, they are able to share and articulate their perspective, ideas, and knowledge within the scientific community of practice. However, although practices of scientific argumentation have been considered a necessity in terms of learning and doing science, they are rarely taught in formal education settings (Duschl et al. 2007). This study aims to illustrate and examine a designated online resource to better understand scientific discourse and engagement among citizen scientists.

Collaborative Science (<http://www.collaborativescience.org>) includes a set of online resources such as videos, scenario modeling tools, and supporting education materials on ecosystem functions, which are designed to scaffold the process of collaboration between researchers and citizen scientists. The resources provide a framework for engaging in environmental management projects through adaptive management and a modeling

tool, Mental Modeler ([www.mentalmodeler.org](http://www.mentalmodeler.org)) (Gray et al., 2013, 2013). Mental Modeler enables analysis of representative shifts in participants' individual and collective knowledge of their management problem. Mental Modeler, shown in Figure 1 is based on Fuzzy Cognitive Mapping (FCM) which provides functionality for users to develop concept mapping in terms of the variety and the impact strength of environmental and social factors influencing their driving environmental problems (Kosko, 1986; Özdesmi & Özdesmi, 2004; Gray et al., 201). MentalModeler (Gray et al., 2013) can help participants track, monitor, and manage the processes of addressing ecological concerns or issues related to their local environments. The practice of modeling continually challenges participants to reflect on and revise their ideas based on observations, serving as a working artifact to collect and drive individual and collaborative knowledge development within a community (Wartofsky, 2012). In addition, boundary objects refer to artifacts or media, which serve as a bridging function for citizen scientists and facilitators to communicate with each other across contexts and environments. These mental models can also serve as a boundary object facilitating communication between the community and partners, particularly those in scientific communities (i.e. land managers, professional researchers), by providing opportunities for more specific presentation of their concept development and feedback about their planning (Akkerman & Bakker, 2011; Hmelo-Silver, 2003). Although the Mental Modeler serves as a boundary object, we argue that the interactions between citizen scientists and facilitators across sites affect not only individuals but also the different social and scientific practices at large. In this study, we aim to examine participants' uses of scientific practices as well as scientific knowledge through the development of concept mapping with the use of Mental Modeler. In this study, we address the following research questions: (1) How does the opportunity to engage in a collaboratively created citizen science project afford engagement in scientific practices?; (2) How does a collaborative modeling tool serve as a boundary object that supports social practices of science that involve generating conjectures, constructing and evaluating models, and application of science knowledge?; (3) What is the role of a more knowledgeable facilitator in supporting these practices?

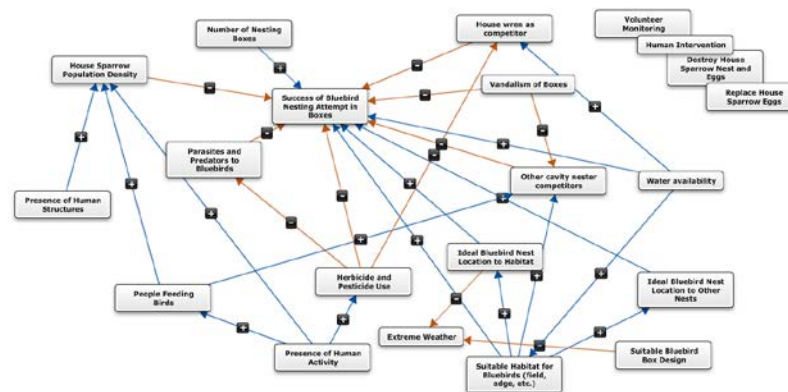


Figure 1. The Sparrow Swap Group Model.

## Methods

Two groups of participants were included in the study. Participants from each group were engaged in two one-hour webinar discussions. Participants had developed individual mental models before the webinars, and the goal of the webinar discussions is to develop consensus based group models. The first webinar section was about model construction, and the second one focused on model refinement. The first group was called Sparrow Swap, which had ten participants. A citizen scientist contacted the research team to share this project idea, which was then shared with the Virginia Master Naturalist (VMN) volunteers via email. Participants joined the project voluntarily based on their interest. The primary goal of the Sparrow Swap project was to test the effectiveness of replacing house sparrow eggs with artificial eggs as a technique to decrease competition with between native bluebirds and non-native house sparrows at nesting sites. The house sparrow eggs that were removed were sent to researchers for curation at the North Carolina Museum of Natural History and for testing of contaminants in the eggs to determine levels of certain pollutants in the environment. Participants used the Mental Modeler to collectively develop their understanding of the key factors contributing to the complex issue of house sparrow competition on native songbirds.

The second group was the Booker T. Washington Native Plantings Experiment (BOWA). Participants were recruited via the Virginia Master Naturalists email list serve and during a presentation to the VMN chapter closest to the Booker T. Washington National Monument. The group had 12 members and the purpose of this

project was to measure the success of native grass restoration following the removal of an invasive grass species (Johnson grass). Mental Modeler was used to help participants identify the relationships between various native species and exotic Johnson grass in order to promote the success of native seed planting.

During the discussion, participants from both groups developed and refined their group models. Facilitators from the research team helped guide the group discussion with the use of Mental Modeler. In this way, participants and the facilitator could reason collaboratively and make decisions about what research questions they were addressing, add, changing, or revising variables, and developing a data collection plan.

In order to create the initial consensus-based group models, it required two discussion sessions per group. Thus, the data sources were drawn from four webinars of the two groups, Sparrow Swap and BOWA, which were recorded and transcribed. Each webinar was about 72 minutes long, totaling 288 minutes, and with 7-12 participants and 1-5 facilitators. Two coders scored a shared 20% of the data set, achieving a substantial level of interrater reliability, kappa = 0.84. One researcher coded the rest of the data. The videos were coded through the qualitative coding scheme presented in Table 1. The coding scheme was modified from Hmelo-Silver's (2003) study, originally developed for an analysis of cognitive and social processes involved in the construction of a joint problem space in collaborative inquiry with a simulation. The coding scheme was adapted in order to capture the representations of participants' use of scientific knowledge and practices and group dynamics and interactions and differentiate facilitators' actions in terms of monitoring, explaining concepts, and providing research instructions during the discussions (Hmelo et al., 2000, Hmelo-Silver et al., 2002). The coding scheme included 6 major categories and 24 subcategories. The first part of the coding scheme focused on the individual level, including Knowledge used in the discussion (K1-K4), Scientific Practices related to modeling (S1-S2), and Metacognition (M1-M4). The second part of coding scheme focused on the social interaction, including Questioning (Q1-Q5), Responses (R1-R6), and Facilitator(s) Input (F1-F5). Data were coded at the unit of the conversational turn by speakers or were parsed when ideas or topics changed. Turns could be coded on multiple dimensions. The total numbers of turns for the first webinar of the Sparrow Swap group was 325, and the second one was 208. In addition, the total turn numbers of initial discussion for the BOWA group was 211, and 264 for the second group discussion.

To study how the conversation unfolded, we conducted a temporal analysis. The data were analyzed and represented via Chronologically-Ordered Representation of Discourse and Tool-Related Activity (CORDTRA) diagrams. CORDTRA diagrams provide visual representations, which can help in interpreting complex patterns and analyzing collaborative learning processes in CSCL (Computer-Supported Collaborative Learning) environments. In addition, CORDTRA analysis includes the analysis of coding schemes that quantify different types of discourse moves which occurred in the discussion and a chronological picture in which multiple process are represented in parallel on one timeline (Hmelo-Silver et al., 2011).

Table 1: Qualitative Coding Scheme

Category	Code	Subcategory	Definition
<b>Knowledge</b>	<b>K1</b>	Conceptual Knowledge	Knowledge expressed with justification/explanation (based on scientific practices).
	<b>K2</b>	Conceptual Conjecture	Knowledge expressed without justification/explanation.
	<b>K3</b>	Anecdotal/Pattern of Experience	Experience related as a one-time occurrence or story from another person/Experience based on a regular observation or occurrence.
	<b>K4</b>	Research Experience	Experience related as part of previous regular field work.
<b>Scientific practice</b>	<b>S1</b>	Top-down modeling/planning	Representing conceptual knowledge or learning more about their management problem/ Practical concerns.
	<b>S2</b>	Bottom-up Modeling/planning	Adding removing components, low-level modeling and planning
<b>Metacognition</b>	<b>M1</b>	Monitoring	Checking group progress, model components, planning concepts, or asking for other explanations.
	<b>M2</b>	Evaluation /Reflection	Thinking about specific actions and their outcomes.
	<b>M3</b>	External resources	Seeking expertise/resources outside of the group.
	<b>M4</b>	Stakeholder Concerns	Consideration of how external social factors impact participants' planning.
<b>Questioning</b>	<b>Q1</b>	Tool-related	Questions or issues pertaining to tool-use
	<b>Q2</b>	Explanation related	Questions about causal antecedents, consequents, enabling conditions; tend to get

			at how and why.
	<b>Q3</b>	Definitional	Participant asks definition for their ideas, or specific values related to the project.
	<b>Q4</b>	Clarification	Participant seeks verification for their ideas, or specific values related to the project.
	<b>Q5</b>	Meta question	Range of metacognitive questions to elicit meta responses, support group dynamics or progress, self-regulated learning.
<b>Responses</b>	<b>R1</b>	Agreement with facilitator	When participants show agreement to the views of the facilitator(s), coded in context of facilitator(s) statement.
	<b>R2</b>	Agreement with group member	Participant agrees with view of their group member.
	<b>R3</b>	Brief answer	Answers to general questions that do not include any explanation
	<b>R4</b>	Minimal justification	Answers that include a reason or justification.
	<b>R5</b>	Elaborate justification	Answers that include a detailed explanation to justify one's beliefs or share one's knowledge.
	<b>R6</b>	Conceptual conflicts	Acknowledge or express different opinions through interaction/ over a component of model or broader concepts involved in the project.
<b>Facilitator(s) input</b>	<b>F1</b>	Monitoring	Facilitator check-in/monitor progress, and encourages participation.
	<b>F2</b>	Explaining concepts	Addresses higher-level concepts that might help build the model.
	<b>F3</b>	Research Instructions	Facilitators giving explicit guidance about research interventions.

## Results

For the CORDTRA diagrams (Figure 2), discourse codes were arranged in chronological order on the horizontal axis (turn numbers). The vertical axis shows the categories of the qualitative coding scheme, from the top to the bottom, K1 to F5. Each point refers to what code(s) was/were coded at specific turn by speakers. CORDTRA diagrams help distinguish the group dynamics and collaborative activity between two webinar sections. The results were also interpreted based on the percentages and frequencies of each code (Table 2). The percentage and frequency can help us to explain how cognitive engagement and communications may be different across the two groups.

First, for the Sparrow Swap group, participants shared more anecdotal (K3) and research experiences (K4) in the second group discussion than in the first meeting (Table 2), whereas there were slightly fewer knowledge practices (K1 & K2) shared in the second discussion. Since the first group discussion focused on identifying the variables and defining the relationships between variables, bottom-up modeling (S2) was more prominent in the first discussion. The second group meeting focused more on conceptual knowledge (S1) and practical concerns (M4) for conducting the projects. Top-down modeling/planning was more prominent in the second discussion. The second group discussion also involved more group interactions (Questioning and Responses). The facilitators provided more monitoring for checking group progress and encouraging participants' contributions in the first group meeting. In addition, the CORDTRA diagrams (Figure 2) helps us to zoom in certain time period to examine the interactions between participants and facilitator as well as scientific discourse more closely.

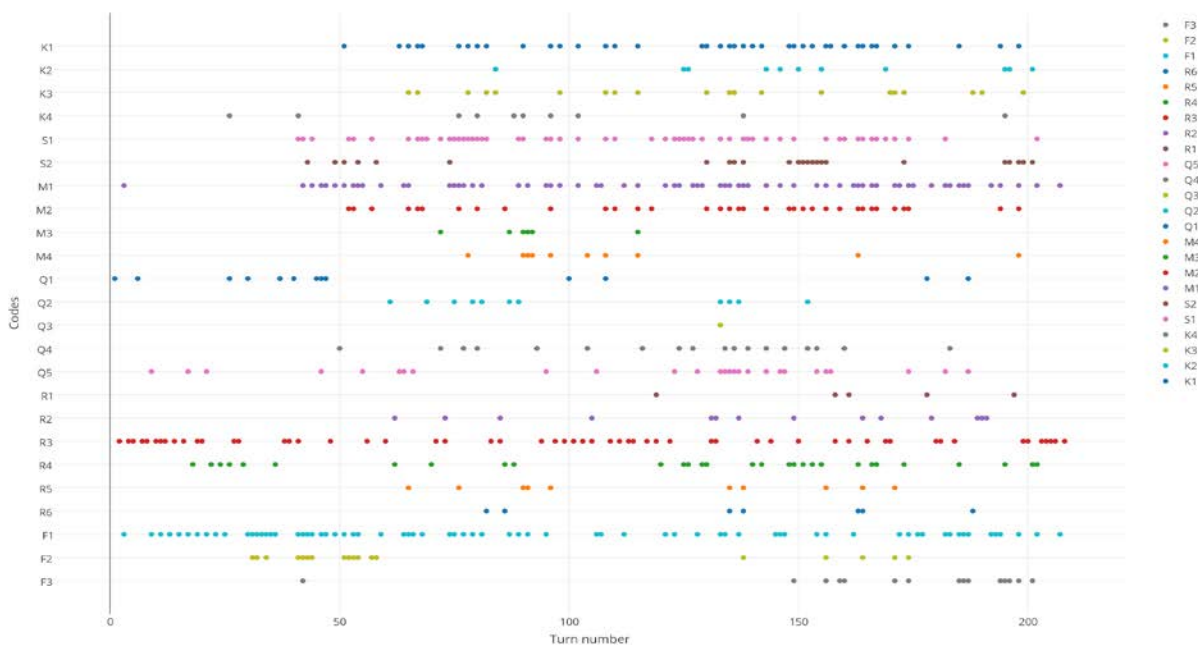


Figure 2. CORDTRA Sparrow Swap Group (Section I).

To better understand the interaction, we provided an excerpt from Sparrow Swap below to illuminate how scientific knowledge and practices are engaged while the group uses Mental Modeler. Additionally, the data supported that the Mental Modeler tool served as a boundary object to help participants collaboratively build the model. In this group, Rena and Sam are the facilitators and the others are citizen scientists. In particular, this excerpt was selected to illustrate how the tool is a dialogical learning resource to help collective knowledge building. This occurred after Amy raised a question about bluebird boxes compared with natural habitats and the facilitator asked the group a meta-level question about what they think they should be talking about:

- 01 Rena: Is there a reason to specifically focus on our boxes (*bluebird boxes*)?
- 02 Amy: I think so because I mean, maybe this is jumping the gun, but we are talking about  
 03 swapping eggs eventually so we are only going to be doing that on a bluebird trail with  
 04 bluebird boxes... because bluebird population density. I mean it is going to be different if  
 05 we have a natural habitat or just boxes. When we do this diagram. Does that make sense?
- 06 Rena: It does to me, others?
- 07 Anna: I changed it to success of bluebird nesting attempts in boxes specifically.
- 08 Rena: And so bluebird population density, how do we want to treat that in this model?
- 09 Sam: I have questions about whether or not density is the interesting thing or is it the population  
 10 in the area, and I do not know if those are two different things... if these are two different  
 11 variables...or just overall population.
- 12 Donna: I think the population in the area...and I am thinking that the population density, well the  
 13 more bluebirds there are, the more likely one of them is going to choose a box, so...Unless  
 14 there is an abundance in the natural habitat...I can't see how there would be a negative  
 15 impact on nesting attempts if you have...high population density.
- 16 Rena: That makes sense. Do we want to keep both of these in the model (*see Figure 3 below*)?
- 17 Sam: I wonder as I hear Donna talking about it is there a feedback relationship between the  
 18 density and success from year to year? So the more successful they are the higher the  
 19 density, the higher the density, potentially the more success? Or is there kind of an arrow  
 20 going one way, and an arrow going the other between these two?

- 21 Amy: ...I think that the more successful attempts you have the greater the population density.  
 22 Rena: Ok great.  
 23 Lee: I think it depends more on the population density of house sparrows.  
 24 Rena: Ok, let's look at our house sparrow population density then. What do we think an  
 25 arrow...right there? Is that what you are thinking Lee?...

During this five-minute conversation, participants and facilitators applied the tool as a boundary object to discuss the components and the relationships between the components (Figure 3) as they engage in bottom-up and top-down modeling/planning (Line 02-05). This suggests that they were reacting to issues raised in the discussion. Donna applied conceptual knowledge based on her pattern of experiences in Lines 12-15 with justification based on scientific practices to justify the connection between population density and bluebirds' nesting attempts. The process of collaborative decision-making was dynamic and dialogical based on the use of the Mental Modeler (Line 16-25).

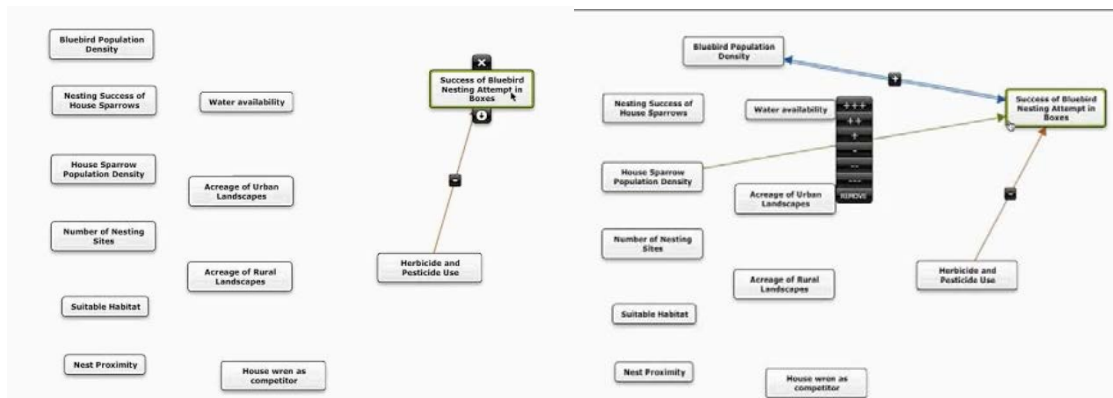


Figure 3. Sparrow Swap model excerpt (Turns 91-105) that focuses on components that participants changed.

For the BOWA group, there are three major findings. First, more frequent and dense conceptual knowledge/conjecture (K1 and K2) was found during the second group meeting (Table 2). In addition, more of the individuals' anecdotal and patterns of experiences (K3) were shared among participants in the second group discussion, which means participants applied the experiences related as an one-time occurrence or based on a regular observation rather than an research experience which involves regular field work. Also, top-down modeling and planning (S1), monitoring (M1), seeking external resources (M3), and stakeholder concerns (M4) also appeared more frequently in the second group discussion. However, there was less facilitator monitoring (F1), explaining concepts (F2), and research instructions (F3) during the second group discussion.

More bottom-up modeling/ planning (S2) tasks were observed during the first group meeting for both projects. This finding suggests that participants began with a stage of defining system components and adding or removing variables from the model. Once the variables and their relationships were determined after the first discussion, participants shifted their discussions to top-down modeling/planning (S1), representing conceptual knowledge, and showing practical concerns and management problems. Although space precludes including the CORDTRA diagrams here, our visual inspection shows that how top-down versus bottom-up planning appeared differently throughout the two sections of webinar. Furthermore, the monitoring (M1) and facilitator monitoring (F5) codes were shared among participants and facilitators for most of time across all four webinar sections. In addition, facilitators' inputs (F1-F3) were more frequent in the beginning and close to the end of discussions. Because there was more explaining and guiding during the first group discussion, facilitators' inputs were coded much more frequently in the first than the second group discussion.

Table 2: Frequency and Percentage Results of Sparrow Swap & BOWA Groups

		Sparrow Swap		BOWA	
Category	Codes	I (325 Turns)	II (208 Turns)	I (211 Turns)	II (264 Turns)



		Frequency	%	Frequency	%	Frequency	%	Frequency	%
<b>Knowledge</b>	<b>K1</b>	63	19.4	40	19.2	57	27.0	72	27.3
	<b>K2</b>	23	7.1	11	5.3	17	8.1	30	11.4
	<b>K3</b>	27	8.3	20	9.6	10	4.7	19	7.2
	<b>K4</b>	6	1.8	10	4.8	3	1.4	3	1.1
<b>Scientific Practices</b>	<b>S1</b>	22	6.8	56	26.9	41	19.4	77	29.2
	<b>S2</b>	143	44.0	24	11.5	52	24.6	53	20.1
<b>Metacognition</b>	<b>M1</b>	172	52.9	67	32.2	93	44.1	134	50.8
	<b>M2</b>	16	4.9	35	16.8	38	18.0	33	12.5
	<b>M3</b>	2	0.6	6	2.9	9	4.3	25	9.5
	<b>M4</b>	0	0.0	10	4.8	9	4.3	28	10.6
<b>Questioning</b>	<b>Q1</b>	14	4.3	13	6.3	6	2.8	2	0.8
	<b>Q2</b>	17	5.2	11	5.3	11	5.2	16	6.1
	<b>Q3</b>	3	0.9	1	0.5	3	1.4	0	0.0
	<b>Q4</b>	21	6.5	18	8.7	12	5.7	16	6.1
	<b>Q5</b>	52	16.0	27	13.0	46	21.8	38	14.4
<b>Responses</b>	<b>R1</b>	14	4.3	5	2.4	13	6.2	17	6.4
	<b>R2</b>	18	5.5	14	6.7	24	11.4	33	12.5
	<b>R3</b>	74	22.8	57	27.4	49	23.2	75	28.4
	<b>R4</b>	27	8.3	30	14.4	29	13.7	26	9.8
	<b>R5</b>	4	1.2	10	4.8	8	3.8	12	4.5
	<b>R6</b>	13	4.0	7	3.4	6	2.8	17	6.4
<b>Facilitator(s) input</b>	<b>F1</b>	134	41.2	71	34.1	92	43.6	76	28.8
	<b>F2</b>	10	3.1	18	8.7	18	8.5	11	4.2
	<b>F3</b>	7	2.2	15	7.2	8	3.8	5	1.9

## Conclusions and significance

This study investigated changes over time in terms of scientific practices, monitoring, and interactions for two different citizen science projects mediated by group modeling practices. The first webinar section focused on model construction, and the second one was about model refinement. Applying qualitative coding schemes helped us to examine the process of negotiation and group interactions and identify different phases of activity and patterns of action. In particular, the way in which participants applied scientific knowledge and practices in modeling and planning in different phases of group discussion, the overall relation of the discourse between facilitators and participants, and the timing of facilitator input during the discussion suggest that collaborative modeling provides a context for rich discussions with respect to collaborative problem-solving and decision-

making. The results suggested that these co-created citizen science projects provided participants with opportunities to work collaboratively and facilitated engagement in scientific practices. Additionally, this supports our conjecture that the models serve as boundary objects for engaging in science practices such as developing and using modeling, clarifying and identifying system components, and constructing solutions. We found there were patterns for group modeling in terms of shifting from bottom-up level to top-down level of discussions. The findings related to the timing of facilitator input suggest that facilitator engagement was strongest in initiating and framing discussions and helping to wrap up the sessions with implications for engagement and ownership of the model as a shared object for negotiation of ideas and knowledge related to the environmental issue that the group was addressing. The facilitators' inputs were less involved in the second sessions for model refinement, which indicated the growth of engagement among participants. This study demonstrates how a collaborative modeling tool can serve as a boundary object that allows citizen scientists and facilitators can engage in meaning making around scientific practices. This suggests that CSCL research and practice can contribute to public engagement in science accessible to a broader citizenry.

## References

- Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of educational research*, 81(2), 132-169.
- Cornelius, L. L., Herrenkohl, L. R., & Wolfstone-Hsy, J. (2013). Organizing collaborative learning experiences around subject matter domains. In C. E. Hmelo-Silver, C. A. Chinn, C. K. K. Chan & A. M. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 333- 350). New York: Routledge.
- Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. National Academies Press.
- Gray, S. A., Gray, S., Cox, L. J., & Henly-Shepard, S. (2013, January). Mental modeler: a fuzzy-logic cognitive mapping modeling tool for adaptive environmental management. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp.965-973). IEEE.
- Gray, S. A., Zanre, E., & Gray, S. R. J. (2014). Fuzzy cognitive maps as representations of mental models and group beliefs. In *Fuzzy Cognitive Maps for Applied Sciences and Engineering* (pp. 29-48). Springer Berlin Heidelberg.
- Hmelo-Silver, C. E., Jordan, R., Liu, L., & Chernobilsky, E. (2011). Representational tools for understanding complex computer-supported collaborative learning environments. In *Analyzing interactions in CSCL* (pp. 83-106). Springer US.
- Hmelo-Silver, C. E., Nagarajan, A., & Day, R. S. (2002). "It's harder than we thought it would be": A comparative case study of expert–novice experimentation strategies. *Science Education*, 86(2), 219-243.
- Hmelo, C. E., Nagarajan, A., & Day, R. S. (2000). Effects of high and low prior knowledge on construction of a joint problem space. *The Journal of Experimental Education*, 69(1), 36-56.
- Hmelo-Silver, C. E. (2003). Analyzing collaborative knowledge construction: Multiple methods for integrated understanding. *Computers & Education*, 41(4), 397-420.
- Jordan, R. C., Ballard, H. L., & Phillips, T. B. (2012). Key issues and new approaches for evaluating citizen-science learning outcomes. *Frontiers in Ecology and the Environment*, 10(6), 307-309.
- Jordan, R., Gray, S., Sorensen, A., Newman, G., Mellor, D., Hmelo-Silver, C., ... & Crall, A. (2016). Studying citizen science through adaptive management and learning feedbacks as mechanisms for improving conservation. *Conservation Biology*, 30(3), 487-495.
- Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of man-machine studies*, 24(1), 65-75.
- McCallie, E., Bell, L., Lohwater, T., Falk, J. H., Lehr, J. L., Lewenstein, B. V., ... & Wiehe, B. (2009). Many experts, many audiences: Public engagement with science and informal science education. *A CAISE Inquiry Group Report*, 1-83.
- Özesmi, U., & Özesmi, S. L. (2004). Ecological models based on people's knowledge: a multi-step fuzzy cognitive mapping approach. *Ecological modelling*, 176(1), 43-64.
- Wartofsky, M. W. (2012). *Models: Representation and the scientific understanding* (Vol. 48). Springer Science & Business Media.

# Cross-Community Interaction for Knowledge Building in Two Grade 5/6 Classrooms

Jianwei Zhang, University at Albany, SUNY, jzhang1@albany.edu  
Maria Bogouslavsky, University of Toronto, maria.bogouslavsky@gmail.com  
Guangji Yuan University at Albany, SUNY, gyuan@albany.edu

**Abstract:** This study explores cross-community interaction in two Grade 5/6 knowledge building communities. The two classrooms studied human body systems with the support of Knowledge Forum over a 10-week period. As the students conducted focused inquiry and discourse within their own community, they reviewed productive threads of ideas and posted syntheses in a cross-community space, as synthetic boundary objects. A set of idea thread syntheses from previous classrooms studying human body systems was also posted in the cross-community space. Qualitative analyses of classroom videos, online discourse, and interviews provide a rich description of how the students conceived, generated, and interacted around the synthetic boundary objects for knowledge building across communities.

## Introduction

Schools need to engage students in sustained inquiry and knowledge building discourse by which ideas are continually developed, refined, and built upon, giving rise to higher-level goals (Scardamalia & Bereiter, 2006). In real-world knowledge creation, the trajectory of sustained inquiry and discourse in each community is further supported by interactions across communities that work as an interconnected field. A creative field leverages the work of all communities and their members by accumulating a shared, easily accessible knowledge base, represented using various inscription systems, facilitating dynamic idea contact and cross-fertilization (Csikszentmihalyi, 1999). Such cross-community interactions help to sustain productive knowledge building over time across generations, with newcomers learning from the existing ideas, practices, and role models and further making novel contributions. Fostering cross-community interactions for sustained knowledge building is a new challenge and opportunity for collaborative learning research.

Existing designs and research of collaborative learning focus on micro-level discourse; new advances are needed to support and understand emergent interactions at the higher social levels (Stahl, 2013). With online systems automatically preserving student discussions and supporting virtual sharing (Reil, 1994), it becomes feasible and important to use “the persistent record of interaction and collaboration as a resource” (Stahl, Koschmann, & Suthers, 2006, p. 419) for sustained knowledge building across the boundaries of different communities. Several researchers have made initial explorations to engage students in interactions across knowledge building communities. A common strategy is to have each community directly share its online discourse space with other communities. In a study conducted by Lai and Law (2006), two classrooms from Hong Kong and Toronto, respectively, engaged in collaborative knowledge building supported by Knowledge Forum, a collaborative online environment (Scardamalia & Bereiter, 2006). Students in each classroom had access to the online discussions of their partner classroom. Through the cross-classroom sharing, students learned from the practices (e.g. questioning) of the partner classroom to improve their own work. Laferrière and colleagues (2012) conducted a study that engaged three international sites. Each classroom gave other classrooms access to their online discourse space so they could read their notes and respond. The cross-community collaboration led to productive classroom changes. Meanwhile, difficulties arose for students to understand and build on other communities’ extended discourse without a clear sense of the contexts. New designs of boundary-crossing support are needed to make knowledge progress accessible across communities.

This research designs cross-community interaction using a multilevel emergence approach, focusing on interactions mediated through boundary objects. “Boundary objects” are artifacts (e.g. reports, tools, models) used to bridge the boundaries (discontinuities) between different social worlds (Star & Griesemer, 1989). Wenger defines them as “forms of reification around which communities of practice can organize their interconnections.” (Wenger, 1998, p. 105) Objects from a community often have contextual meanings not accessible to other communities. What makes boundary objects effective for bridging different communities of practice is their interpretative flexibility as a “means of translation” (Star & Griesemer, 1989): they have a structure that is common enough to make them recognizable across the different social worlds and allow different communities to interact and work together. As Akkerman and Bakker (2011) suggest, interactions with shared boundary objects help members of different communities to identify, understand, and reflect on their

different practices, leading to an enriched view within each community and potentially the creation of new, in-between practices.

As noted above, raw distributed online discourse records are hard to be used as boundary objects to bridge the boundaries between different knowledge building communities. In this study, following a multi-level design, students generate synthetic boundary objects for cross-community sharing on the basis of the extended knowledge building discourse within their own community's space. The synthetic boundary objects take the form of idea thread syntheses framed using shared structures of inquiry. Students in each community engage in focused inquiry and interactive discourse within their own community's space, with small-groups formed and reformed to address emergent problems of inquiry. As progress is made, students working on the various areas selectively review and synthesize fruitful threads of inquiry emerged from their discourse. The selective reviews and syntheses of inquiry threads can facilitate peer learning and build-on across inquiry topics within each classroom (Zhang et al., 2015); they may further be shared as boundary objects to enable cross-community interaction. Students from another classroom (or a subsequent student cohort) can use the syntheses of idea threads to view into the discussions and understand the extended journey and progress of inquiry.

To support students' reflective review and structuring of distributed online discourse, our team developed Idea Thread Mapper (ITM) (Zhang et al., 2015). ITM interoperates with Knowledge Forum (Scardamalia & Bereiter, 2006) and potentially other platforms that support online knowledge building discourse. Using ITM, students identify focal objects of inquiry addressed by their collective discourse and select important discourse contributions related to each focus. The discourse entries with a shared focus are displayed on a timeline, as an idea thread, extending from the first to the last entry. Each idea thread has a "Journey of Thinking" synthesis. Students review their idea progress in the thread and co-author/update the "Journey of Thinking" synthesis. In line with the focus of knowledge building on continual idea improvement through progressive problem solving (Scardamalia & Bereiter, 2006), we designed a set of scaffolds for Journey of Thinking synthesis, including (a) overarching topic and problems, (b) we used to think...now we understand... and (c) deeper research needed. ITM has been used by a set of classrooms that studied various science topics, with a rich set of idea threads and Journey of Thinking syntheses archived.

This research explores cross-community interactions mediated through idea thread syntheses, as synthetic boundary objects. Given the exploratory nature of this research topic, we used qualitative methods to provide a rich description of a purposefully designed case of cross-community interaction in two partnering classrooms. Students had access to archived idea thread syntheses from previous cohorts, generated idea thread syntheses based on their ongoing work, and shared the syntheses with their partner classroom. Our research questions ask: (a) How did the teachers and students conceive the nature of synthetic boundary objects in the form of idea thread syntheses? (b) How did the students generate synthetic boundary objects based on their community's knowledge building discourse and inquiry work? (c) How did the students interact with the synthetic boundary objects from other communities, with what support from the teacher?

## Method

### Classroom contexts and designs

This study tested cross-community interaction in two Grade 5/6 knowledge building communities. The two classrooms were taught by two teachers: Mr. B and Mr. M. Both teachers had multiple years of teaching experience, and Mr. B was more experienced with teaching grade 5/6 science using knowledge building pedagogy and technology. There were a total of 24 students in Mr. M's room and 23 students in Mr. B's room, with a total of 39 students consented to participate in this research. The two classrooms studied human body systems with the support of Knowledge Forum over a 10-week period. On an ongoing basis, students in each classroom contributed and built on one another's ideas in their own classroom's Knowledge Forum views (workspaces). Cross-community interaction was supported through a "Super View" on Knowledge Forum where students accessed and posted idea thread syntheses. A visual was added to the "Super View" to facilitate the sharing process: two trees with a number of branches where Super Notes about various inquiry topics could be placed (see Figure 1). Each classroom had its own "tree of knowledge," and students could take a look at their peer classroom's knowledge at any point of the knowledge building process for mutual learning and idea connection. Online posts to share idea thread syntheses in the Super View were called "Super Notes" by the teachers and students. Each Super Note was organized using the Journey of Thinking scaffolds of Idea Thread Mapper that interoperates with Knowledge Forum. Prior to this study, a set of classrooms from two schools had used Idea Thread Mapper to organize their knowledge building discourse about human body systems and created idea thread syntheses. Based on these syntheses, an initial set of Super Notes (idea thread syntheses) was posted to the Super View, each framed using the scaffolds: *Our research topic and problems, We used to*

*think...Now we understand..., We need deeper research.* The teacher in each classroom first introduced the Super View in the third week of the inquiry when their students had generated their own questions and conducted initial research about the various topics related to the human body. Students read the Super Notes from the previous classes and reflected on what they could learn from the questions and ideas. With deeper research conducted in each classroom in the next two to three weeks, students working on various themes started to create Super Notes to summarize their progress for sharing with their own classmates as well as with the other Grade 5/6 classroom. Students from the two classrooms read each other's Super Notes and discussed the insights gained. A whole class meeting was organized in each room for students to reflect on what they had learned from their peer classroom and the prior classes and planned for possible deeper research.

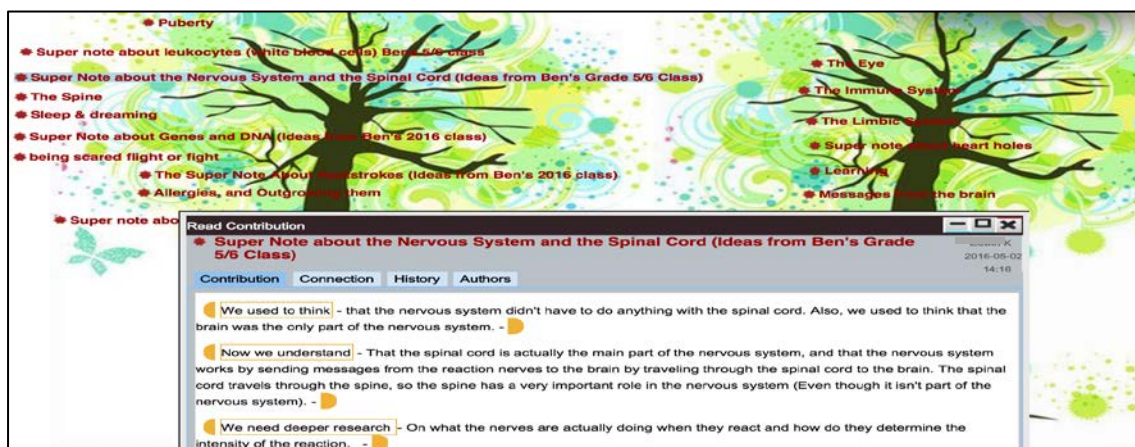


Figure 1. The “Super View” for Sharing Journey of Thinking Synthesized by Different Classrooms.

## Data sources and analyses

The data sources included observations and video recordings of classroom discussions, online discourse in each classroom's regular views and the shared Super View, and student and teacher interviews. The video records and related observation notes captured major classroom episodes (seven lessons of Mr. B and six lessons of Mr. M) in which students were introduced to the Super View, discussed information learned from the Super Notes of the other classroom, and created Super Notes based on their own work for cross-classroom sharing. We interviewed the two teachers and 13 students at the end of the unit. The students were asked how a “Super Note” was different from other notes, how they decided what specific ideas should be included in their Super Notes, and how was reading the Super Notes of other classrooms helpful for their knowledge building. The teachers were asked how the cross-community interaction helped their students and how they facilitated such interactions. The interviews were video-recorded and transcribed for analysis.

Guided by each of the three research questions, we conducted detailed analysis of the classroom videos and interviews in connection with student discourse in the shared Super View and their own regular Knowledge Forum views. Specifically, to develop an overall sense of the student-generated Super Notes in connection with their knowledge building discourse, we conducted content analysis (Chi, 1997) to code student notes in their regular Knowledge Forum views based on the inquiry topics reviewed by their Super Notes. Two coders independently coded 81 notes (22% of the total notes) resulting in an inter-rater agreement of 98%. Following procedures of grounded theory analysis (Strauss & Corbin, 1998), the researchers read and re-read the transcriptions of the classroom discussions and interviews, created open codes, which were then clustered into primary themes to capture prominent patterns addressing each of the three research questions. The authors then co-reviewed the open codes and initial themes and discussed any disagreements. The themes were further validated through checking data against the themes, relating and comparing the themes identified from student and teacher data, and triangulating the identified themes across the data sources. The refined coding themes are elaborated in Results under the three research questions.

## Results

How did the students and teachers conceive the nature and role of synthetic boundary objects from other communities?

Qualitative analysis of student interviews revealed interrelated themes about how the students conceived the nature and role of the Super Notes (see Table 1).

Table 1: Students' conceptions of the Super Notes captured in the interviews

Themes of conceptions	Class	Examples from the individual interviews
Super Notes as a summary of "big idea" and knowledge basis.	Both classes	"I think it is to focus on the entire idea of the topic you are focusing on and not just the tiny details you wanna share with the whole class." [...] "So it's just I think the basic basic idea."
Super Notes as refined and "verified" knowledge	Both classes	"Well, we definitely did not include like the information that we did not know much about, because that would mean that.. like if you were not sure if that was right or not, then it would not be good to include it because that would mean that you are technically making a Super Note, which is partially not true..."
Super Notes as a journey of thinking	Mr. M's class	"It's like one huge note that reflects on all of your ideas, and what you used to think and what you now know. So I thought it was a great idea rather than making a bunch of notes on your progress of learning a topic."
Super Notes as knowledge for others	Mr. B's class	"...it's kind of to let everybody know what you are researching, but without having to read like all the stuff that you've read and having the basic knowledge of that topic."
Super Notes as well-phrased and polished ideas	Mr. B's class	"When you look at normal notes, it's kind of hard to understand: there maybe some spelling errors, and maybe some like grammar errors... If you look at Super Notes that are amazingly written and they are really simple and they help people understand what is the main focus of this Super Note."

Both teachers conceived and presented the "Super View" as a higher-level space of discourse that required students to formulate and summarize "big ideas" investigated so far. Unlike regular KF notes, the Super Notes needed an additional level of reflection and refinement: both teachers asked students to show them the final draft of their Super Notes before posting to the Super View. While sharing this common conception, the two teachers gave slightly different emphasis. Mr. B, who were more experienced with knowledge building, explicitly emphasized that the Super Notes were about sharing the "journey of thinking" rather than specific information. The goal was to show how their understanding and thinking had evolved during the course of the inquiry. Therefore, the Super Notes had a metacognitive layer that was weak in regular notes. The teachers commented that the Super Note scaffolds played a crucial role in making this metacognitive layer visible by framing the process of thinking from "what you used to think" to "what you understand right now" with "an eye on helping someone." While both teachers emphasized using the scaffolds as a way to structure the Super Note, Mr. B underlined the importance of clarity, so the Super Notes could be understood by students from a different classroom, who lacked the knowledge about the classroom contexts. For Mr. M, the Super Note was about pulling out the important ideas from their regular KF views and bringing together "important points or discoveries that everybody should know about." Unlike Mr. B, he did not explicitly emphasize that the Super View was intended to be a communal place shared with the other classroom. During the interview, Mr. M commented that some of his students were somehow afraid or hesitant to take the leap into the "super level."

### How did the students generate synthetic boundary objects?

This question was investigated through tracing student online discourse in relation to their Super Note topics as well as qualitative analysis of the video-recorded classroom interactions and student reflections captured in the interviews. Figure 2 reports the Super Note topics covered by the two classrooms and the number of regular Knowledge Forum notes/posts related to each topic. Mr. B's students created a total of ten and Mr. M's students created six Super Notes. Their Super Notes addressed both shared and unique topics of inquiry. While some of the Super Note topics had intensive discussions on Knowledge Forum, a few other topics had only been addressed by very few regular notes. These topics were very specialized (e.g. allergies, heart stroke, and scariness) but were identified as interesting and helpful for other students. Another related factor, according to the teachers, was that some of the inquiry work was documented in students' personal notebooks and shared face-to-face, therefore, not reflected on Knowledge Forum.

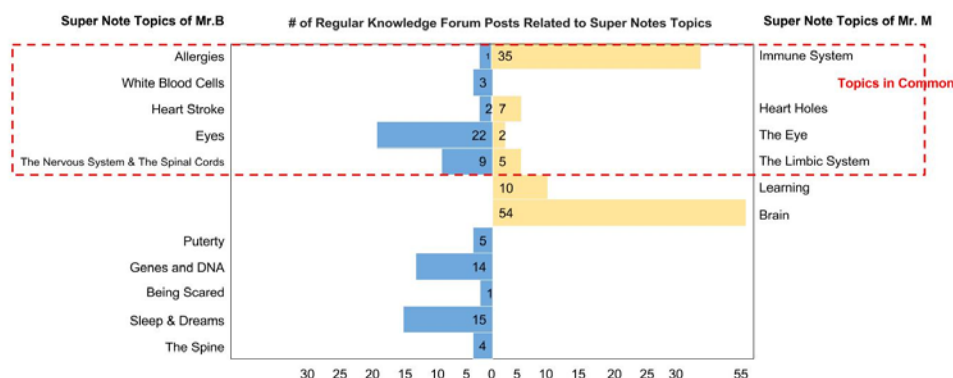


Figure 2. Super Notes Created by the Two Classrooms in Relation to Regular Knowledge Forum Notes.

The qualitative analyses of videos and interviews revealed essential processes that went beyond simple summarization of information to including high-level reflection on progress and gaps, rising above distributed ideas and information sources for coherent understandings, and selective integration of “juicy” ideas for cross-classroom sharing. The specific processes are elaborated below.

(a) Whole class discussion to identify productive areas of inquiry and form specialized groups. In each classroom, the teacher asked his students to reflect on their inquiry work in various areas about the human body and form into small group based on specialized interest to generate Super Notes.

(b) Reviewing previous Knowledge Forum posts and personal notes to identify knowledge advances, in small groups. Prior to the Super Note intervention, students wrote in their regular Knowledge Forum views to share specific questions and ideas, explore information from authoritative sources (e.g. books, videos), and discuss findings from experiments. Students also took personal notes about their research. The teachers encouraged their students to review their online posts and personal notes as a starter for the Super Notes. For instance, in the group focusing on DNA, four students first updated their personal notes taken in MS Word and shared their documents by passing around each other’s computer. Summarizing four separate long documents was challenging, so Mr. B, noticing the challenge, approached the group and suggested to choose one person to type the big Super Note. The small group discussed what should be included in the Super Note.

(c) Deepening research using authoritative sources, as groups and individuals. Reviewing and analyzing existing work and ideas, students noticed questions and issues that needed to be clarified. This pushed the students to conduct further research using information from books, videos, and websites. Some of the sources were beyond the students’ level of reading. The teachers worked as a co-learner and helper to interpret the information, explain scientific terms, and model rephrasing ideas using simpler terms.

(d) Combining each other’s ideas and expertise for coherent group understanding. To understand the complex mechanisms underlying the topics (e.g., DNA, immune system, heart holes), students worked collaboratively to develop specialized understandings and further combined the information to elaborate the full picture. As a student from Mr. M’s class reflected: “I actually worked with a friend on this [topic about heart holes], and she was mostly working on where heart holes are, like I told you they are on septum, and I was working on how they heal. And she asked a question on the regular view on “how the heart holes heal?” so I’ve researched that and we kind of combined our ideas, and put [them] in a Super Note.” The Super Note created by these two students is shown in Figure 3.

**Super note about heart holes**

**Our topic and problems** Why do hearts get holes in it? And how does it heal? 📄

**We used to think** that it would heal by itself overtime. We also thought that the hole(s) was on the outside of the heart 📄

**Now we understand** The hole is on the septum which is between the two chambers of the heart. One chamber sends lots of oxygen rich blood to the body and the other chamber sends not oxygen rich blood to the lungs. It can be dangerous when the blood mixes because it’s like breathing carbon dioxide instead of oxygen. 📄

**Now we understand** that some people who are born with heart holes in their septum will heal overtime. Whereas some people will have to get an open heart surgery to close the holes where a machine takes over the heart's pumping action and moves blood away from the heart. 📄

Figure 3. A Super Note about Heart Holes from Mr. M’s Class.

(e) Selecting ideas for sharing through group and individual reflection. In light of the diverse ideas reviewed, students in each group decided on what should be included in their Super Note. The analysis of the

interviews identified a number of criteria that the students had in mind when selecting information: consistency, importance, depth of understanding, and relevancy. For example, a student said: *“Well, we definitely did not include like the information that we did not know much about... So we’ve looked at things that we’ve seen consistent... that we all knew.”*

(f) Summarizing “big ideas” using accessible language with the Super Note scaffolds. Both teachers encouraged students to use the Super Note scaffolds to summarize ideas: start with “we used to think,” continue with “now we understand,” and finish with “we need deeper research.” Both classrooms analyzed Super Notes from previous classrooms to illustrate the use of the scaffolds. Students, reminded by their teacher, took conscious efforts to present the information in a simple and clear way to make it accessible to students from other classrooms.

(g) Sharing the Super Notes with the teacher before posting. Both teachers asked students to show their final drafts of Super Notes before posting them to the Super View. Doing so encouraged students to write careful Super Notes and ask their teacher for advisory input.

### How did the students interact with synthetic boundary objects?

On average, each of the Super Notes from archived idea threads of prior classrooms was read by 19.83 students. Students read their peer classroom’s Super Notes more actively (34.6 users per note) than those from the prior classrooms. Beyond individual reading, each classroom had a whole class discussion about the information from the Super Notes, followed by further small group discussions. Through qualitative analysis of the video records of the whole classroom discussions and student interviews about how they approached the Super Notes, we identified specific patterns of interactions, which clustered around three themes, as elaborated below.

(a) Identifying knowledge from other classrooms for possible connection. In whole classroom discussions, students identified Super Notes from other classrooms that were relevant and interesting and contained new information, such as by saying *“I am interested in K’s note about allergy.”* They related the Super Notes to their own understanding in order to comprehend the topics and enrich their learning.

(b) Comparing knowledge work between communities triggering student reflection. Students discussed how new and unique ideas from the Super Notes helped them to go beyond the limitation of their own knowledge. The teachers facilitated the discussion by raising deeper questions for reflection. For example, Mr. B asked: *“What was the idea that came from the Super Notes that you hadn’t thought before and that pushed your thinking further?”* A student responded: *“Well I never really thought about what side of the brain controls what side of the body, while I already know, but it turns out that, your left side of the brain controls the right side of your body”.* Similarly, Mr. M asked his students: *“What topic either strikes you as new information or something that you’d like to pick up as a thread and go deeper into?”* Two students responded that they learned from the Super Note about allergies, a topic investigated by Mr. B’s students but not covered by the students of Mr. M. Another student pointed out a deep concept learned from a Super Note of Mr. B’s class related to their own work: *“Me and J are doing immune system, ... and we saw these notes about white blood cells, and that was really cool ‘cause white blood cells were part of your immune system. We don’t really know about them individually... Yeah, it was really helpful for us...”*

In addition to reflecting on new knowledge and information gained from the other communities, students further adopted the epistemic form of reflective thinking: “we use to think...now we understand...” and talked about their journey of inquiry accordingly. In Mr. B’s class, student EL reflected: *“I used to think, there was a person and then they had a brain, and then the brain told the body what to do and that was the end of it, and now I understand that like each part of the body has its own little system.”* Student L responded: *“Everything is ...part of like a system, like everything is like I can say work together.”*

(c) Integrating knowledge across communities to develop complementary perspectives and deep understanding. The Super Notes from each classroom were written by students specialized in the related inquiry topics to selectively synthesize key problems and ideas using simple language. When reading the deep questions and ideas from their partner classroom’s Super Notes, the students needed to unpack the information to understand the journey of inquiry presented, detect gaps of understanding, and bring together the knowledge from their own and from the other community to address the gaps and problems. With their teacher’s facilitation, students engaged in extended discourse to collaboratively solve problems and develop explanations. For example, students in Mr. B’s room discussed the Super Note about heart holes written by Mr. M’s students (see Figure 3), which highlighted why heart holes can be dangerous. The students in Mr. B’s room indicated interests in this topic and discussed the specific reasons and mechanisms.

[16] K: like the heart hole, I heard of them, but I didn’t know how that really works.



- [17] S: if you can have a hole in your heart, without it, like, immediately, you exploded.
- [18] Teacher: Well, but what was the problem if you have a hole in your heart?
- [19] D: It's like really dangerous if the blood mixes.
- [20] Teacher: Right, the blood mixes, but why is it bad if the blood mixes?
- [21] B: Because if they mix together, if they mix, they will be as bad as like breathing carbon dioxide.
- [22] Teacher: A, do you want to build on?
- [23] A: Because the blue side like that has no oxygen.
- [24] Teacher: This side, no oxygen (writes "no oxygen" on the blue part of the figure on the Smart Board).
- [25] A: And other part has oxygen.
- [26] Teacher: This does have oxygen. So if they mix, it's like you are breathing air with no oxygen in it, it will be like suffocating.
- [27] S: (reads the Super Note) It says the hole is on the septum, which is between the two chambers of the heart. One chamber sends lots of oxygen rich blood to the body and the other chamber sends not oxygen rich blood to the lungs...
- [28] Teacher: I think a lot of people might have thought the heart pumps blood to the body, but it's more complicated than that. What does it actually do?
- [29] S: I am pretty sure that the blood comes through without oxygen can go around the body, and then it goes through and then it collects oxygen, gives it to the body, it comes out the other way, it keeps going around in the cycle.
- [30] Teacher: Do you want to build on that, M?
- [31] M: Well, it goes through all four chambers, well in the right chambers, its deoxygenated the blood in there, and its goes through of the heart, which pumps oxygen inside the blood and then it gets sent out through the body.
- [32] Teacher: So it's working with oxygenated blood, and blood with no oxygen. C?
- [33] C: While, pretty much blood with no oxygen goes to the lungs, and the lungs give it oxygen, and then it circles back to the heart, and the heart pumps out.

In line 18, the teacher rephrased student K's question of "how that really works" as "what was the problem if you have a hole in your heart?" He facilitated interactive input from his students, who brought knowledge about the respiratory system and circulatory system to analyzing the impact of heart holes. Building on student input, in line 28, the teacher highlighted that the function of the heart is more than pumping blood to the body and invited students for full explanations. In lines 29-33, students S, M, and C built on to one another to elaborate the explanations. Following the above discussion, the teacher and his students improvised a participatory simulation to demonstrate how the blood travels to collect and transmit oxygen. The teacher played the heart, and three students played the blood cell, lungs, and the rest of the body, respectively, with the whole class involved in discussing where the blood cell should go, and with what changes in the process of traveling.

## Discussion

As the findings suggest, the fifth- and sixth-graders showed complicated conceptions of the Super Notes for cross-community interaction, which are in parallel with features of boundary objects as described in the literature (Star & Griesemer, 1989; Wenger, 1998). They conceived that Super Notes should present "big ideas" and refined (verified) knowledge potentially relevant and interesting to other communities, and that knowledge and ideas should be structured consistently as a journey of thinking (in line with the Super Note scaffolds) and well-phrased and polished, so students from other classrooms could understand. Such conceptions of Super Notes were demonstrated in students' practices to generate their Super Notes. They reflected on productive areas of inquiry emerged from their community's discourse and inquiry work; reviewed diverse ideas from their online posts, personal notes, peers' input, and authoritative sources; selected information based on importance, depth, consistency, and relevance to others; and summarized and phrased their Super Note content using the scaffolds to make their knowledge readable and accessible for students from the other classroom. As the students commented, their peers' Super Notes "are amazingly written, and they are really simple and they help people understand what is the main focus of this Super Note." The analyses of the classroom discussions and

student interviews suggest that the students engaged in active and substantial interactions with the Super Notes from other classrooms, with more attention paid to the Super Notes created by their partner classroom than those from prior classrooms. The patterns of interactions support productive mechanisms of boundary crossing suggested by Akkerman and Bakker (2011). Students identified relevant and interesting Super Note topics from other classrooms, compared the different perspectives and inquiries, triggering deep reflection on their own inquiry to integrate and build new knowledge across communities. The cross-community interaction was far beyond writing and reading summaries but served to foster deeper intentionality in each community and expanded/integrated understandings in-between. In the processes of creating their own Super Notes as well as the processes of interacting with others' Super Notes, students engaged in deep reflection as individuals, groups, and a whole community to review, reframe, and "rise above" (Scardamalia & Bereiter, 2006) their collective knowledge progress and carried out personal and group efforts to address gaps and limitations. The teachers framed the Super View as a higher-level discourse space that required high-level thinking and reflection; supported group processes to identify, select, and summarize "big ideas;" and facilitated knowledge building discourse to identify, compare, and connect knowledge across communities for deep understanding.

Taken as a whole, the results suggest productive patterns of cross-community interaction around synthetic knowledge objects. Implementing cross-community interaction is challenging for both students and teachers. To better support students and teachers, we recently designed a cross-community collaboration platform based on ITM. A multi-year design-based research is underway to examine the processes and impacts of cross-community knowledge building in an international network of classrooms.

## References

- Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81* (2), 132-169.
- Chi, M. T. H. (1997). Quantifying qualitative analysis of verbal data: A practical guide. *Journal of the Learning Sciences, 6*, 271-315.
- Csikszentmihalyi, M. (1999). Implications of a systems perspective for the study of creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 313-335). Cambridge, UK: Cambridge University Press.
- Laferrriere, T., Law, N., & Montané, M. (2012). An international knowledge building network for sustainable curriculum and pedagogical innovation. *International Education Studies, 5*, 148-160.
- Lai, M., & Law, N. (2006). Peer scaffolding of knowledge building through collaborative groups with differential learning experiences. *Journal of Educational Computing Research, 35* (2), 123-144.
- Riel, M. (1994). Cross-classroom collaboration in global Learning Circles. *The Sociological Review, 42*(S1), 219-242.
- Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 97-115). New York: Cambridge University Press.
- Stahl, G. (2013). Learning across levels. *International Journal of Computer-Supported Collaborative Learning, 8*(1), 1-12.
- Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409-426). New York: Cambridge University Press.
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, "translations" and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*, 387-420.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2<sup>nd</sup> ed). Thousand Oaks, CA: Sage.
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, England: Cambridge University Press.
- Zhang, J., Chen, M.-H., Tao, D., Lee, J. Sun, Y., & Judson, D. (2015). Fostering sustained knowledge building through metadiscourse aided by the Idea Thread Mapper. In O. Lindwall & S. Ludvigsen (Eds.), *Proceedings of the 11th International Conference on Computer Supported Collaborative Learning* (Vol. 1, pp. 166-173). Gothenburg, Sweden: International Society of the Learning Sciences.

## Acknowledgments

This research was sponsored by National Science Foundation (IIS#1441479). We thank the teachers and students at Dr. Eric Jackman Institute of Child Study in Toronto for their creative work enabling this research.

# Inclusive Collaborative Learning With Multi-Interface Design: Implications for Diverse and Equitable Makerspace Education

Gabriela T. Richard, The Pennsylvania State University, grichard@psu.edu  
Sagun Giri, The Pennsylvania State University, sqg5512@psu.edu

**Abstract:** While the maker movement and its associated affordable and accessible practices and toolkits have reinvigorated interest in pre-collegiate STEM, invention and creativity, many have critiqued makerspaces as implicitly exclusionary, particularly across gender, race and ethnicity. In an effort to rectify past participatory inequities, we designed a maker workshop for high school youth that capitalized upon multiple digital and physical interfaces to create simultaneously digitally and physically responsive projects, which encouraged team-based distributed creativity and development. We explore how the tools and the curricular design encouraged and fostered collaboration and inclusivity, as well as disrupted previous implicit associations around computing and creativity. We discuss the teams, the projects created and how the learning activities provided opportunities for inclusion and equity.

## Introduction

The maker movement and its associated affordable and accessible practices and toolkits have reinvigorated interest in pre-collegiate STEM, invention and creativity (e.g., Blikstein, 2013; Halverson & Sheridan, 2014; Honey & Kanter, 2013; Peppler, Halverson & Kafai, 2016). Most notably, scholars have proposed the benefits of maker activities to foster collaborative learning, a growth mindset, interest-driven knowledge building, and resourcefulness (e.g., Dougherty, 2013; Halverson & Sheridan, 2014; Peppler, Halverson & Kafai, 2016). In particular, makerspaces are touted as increasing transparency in learning computation, both with the tools and activities themselves (Resnick, et. al, 2000; Resnick & Rosenbaum, 2013) and in the spaces designed around them (Peppler, Halverson & Kafai, 2016). Further, the general spirit of constructionist learning for all, as fostered through makerspace learning, has been celebrated widely for opening up previous participatory divides.

However, many have critiqued makerspaces and associated interest-driven STEM and DIY learning activities as implicitly exclusionary in different ways (e.g., Buchholz, et. al, 2014; Kafai, Fields & Searle, 2014; Richard, et. al., 2015). In other words, while maker tools and materials have opened up previous “black boxes” (Resnick, Berg & Eisenberg, 2000), the tacitly exclusionary assumptions and practices still remain. Some have engaged in efforts to increase diversity in access to makerspace literacies directly, such as through recruitment, while others have focused on ways that learning activities or spatial designs can be more supportive or inclusive to previously underrepresented groups (e.g., Kafai, Fields & Searle, 2014; Kafai, et. al, 2014; Kafai, Searle, Martinez & Brayboy, 2014; Richard, et. al., 2015). Still others have explored and capitalized on the tools themselves (particularly e-textiles) as avenues to inclusive participation (e.g., Buechley, et. al., 2008) or as disruptive to gender exclusive patterns associated with making (e.g., Buchholz, et. al, 2014).

Nevertheless, how the tools and arrangements of learning activities can support diverse participation is still underspecified. This is not an immaterial question, as the lack of equitable participation in computing and technology is still a widespread issue from pre-collegiate education and beyond, particularly for African Americans, Latinos and young women, across race and ethnicity (e.g., Heitin, 2014; NSF, 2015). Herein, we explore how a youth-oriented workshop that combined multiple maker digital and physical toolkits fostered inclusive participation. For example, most past efforts have focused on one specific tool or activity, whereas the act of creating an e-textile-based multi-interface design that is both digitally and physically reactive and responsive (i.e., bidirectional) purposefully incorporates several toolkits with differing affordances and implicit associations. In an effort to rectify past participatory inequities, the first author designed two instantiations of a maker workshop for high school youth that integrated e-textile design, coding, digital media design and simplified physical computing, each of which would elicit different interests and biases (Richard & Kafai, 2015). We explore how the curriculum encouraged and fostered collaboration and inclusivity, as well as disrupted preconceptions around computing and creativity, through our guiding research question: How does designing multi-interface bidirectionally responsive projects foster collaboration and inclusivity?

## Background

In recent years, there has been increasing attention paid to who participates in interest-driven digital learning activities (e.g., Ito, et.al, 2013). While makerspaces can cultivate a variety of learning processes and practices, including low tech ones, many efforts have focused on computational thinking skills and creative practices.

## Equity and inclusivity in computing and making

In line with efforts to diversify toolkits and modalities, many have explored how makerspaces can promote equitable and inclusive participation. For example, research has found that integrating crafting with engineering design, in the ways enabled through the Lilypad Arduino, can change the ways that gendered interests are expressed, often resulting in sewing being seen as valuable to computation, and young men being freer in expressing these interests (Kafai, Fields & Searle, 2014; Kafai, et. al, 2014; Kafai, Searle, Martinez & Brayboy, 2014; Richard et al, 2015). Others have found that e-textiles can disrupt past preconceptions and patterns around gendered ability, access and authority in making (Buchholz, et. al, 2014). Further, researchers have explored how utilizing the Lilypad Arduino and e-textile design could create cultural connections between Native American indigenous practices and computational thinking (Kafai, Searle, Martinez & Brayboy, 2014). However, some have noted that utilizing the Lilypad and e-textile design, as a standalone activity, does not necessarily tap into a variety of interests and practices alone (e.g., Richard & Kafai, 2015).

Moreover, research has shown that it is equally important to foster culturally supportive structures in the design of technology-oriented learning spaces, in order to encourage and sustain diverse hobbyist activities and workforces (e.g., Scott, Sheridan & Clark, 2015). For example, scholars have noted that, while past efforts to engage youth in computing often included diverse participation, it did not necessarily result in more equitable computational participation (Scott, Sheridan & Clark, 2015). This lack of equitable participation can partially be explained by the design of technologies that are theoretically value free, but fail to recognize the ways they may be implicitly exclusionary (Lachney, Babbitt & English, 2016). In other words, inclusive participation needs to make a concerted effort to recruit diverse participation, have diverse mentors, allow learners to explore their intersecting and diverging experiences, critique and redesign media representations with accessible tools (such as Scratch) and promote an overall supportive and inclusive environment, whether this is through tapping into culturally relevant pedagogies and practices (Lachney, Babbitt & English, 2016; Scott, Sheridan & Clark, 2015) or through explicit efforts for inclusion (Richard, et.al., 2015).

## Collaboration and inclusivity

In this paper, we subscribe to Roschelle's (1992) definition of collaboration, which is "building and maintaining a shared understanding of a problem or task, distributing responsibility across members, sharing expertise, mutually constructing, and negotiating cognition" (Van den Bossche, Gijsselaers, Segers & Kirschner, 2006, p. 494). Similarly to Peppler and colleagues (2015), we see collaboration as involving effective articulation of ideas and sharing responsibility and contributions. Making involves projects that are big and ambitious, which can necessitate collaboration and division of labor between participants, skills that are seen as essential aspects of 21<sup>st</sup> century skills. We additionally contend that providing an intentional diversity of physical and digital toolkits in maker activities further encourages collaboration because of both the complexity involved in the design and integration process and also the capitalization on varying participant interests and skills. We contend that both aspects are important in meeting the goals of purposefully inclusive collaborative learning in making.

We also challenge that collaboration has to be open, equitable, interdependent, and cohesive to foster effective task performance. Effective collaboration results in deeper learning, critical thinking, collective understanding and long-term retention, and can help build group cohesion (e.g., Johnson & Johnson, 1989). However, simply placing students in groups is not sufficient for collaboration. There are social and interpersonal factors such as social relationships amongst team members, collective efficacy of the group, team cohesion on tasks and feeling of safety within the group that are equally important to collaboration, inclusivity, and team performance. Miyake and Kirschner's (2014) model of team learning beliefs and behaviors describes four factors that have positive effects on team learning and collaboration. We argue that these factors also can have positive impact on inclusivity: making team members feel that they are a valuable part of the team while promoting a safe environment to share their ideas, knowledge and opinions, and take risks. These four factors are psychological safety, cohesion, interdependence and group potency.

Social conditions and team-level beliefs about interpersonal relations with team members, collective belief on the efficacy of the group, team cohesion on tasks at hand and a feeling of safety within the group are equally important to collaboration and team performance. Effective risk taking (Kreijns, Kirschner & Jochems, 2002) requires a sense of safety with team dynamics. Team psychological safety is defined as a "shared belief that the team is safe for interpersonal risk taking" (Edmondson, 1999, pp. 350). For example, team members may not feel comfortable with each other, or, conversely, close-knit team members may isolate others. Promoting a sense of psychological safety has been found to improve team performance and innovation. Akin to psychological safety is a sense of assurance in the group's ability – or group potency (i.e., collective self-efficacy) – which can lead to perseverance during predicaments and conflict (Shea & Guzzo, 1987). Additionally, Miyake and Kirschner (2014) suggested a shared commitment to a task leads to better learning

and performance. Team cohesion has two dimensions: task and social cohesion. Task cohesion emphasizes the commitment to and enjoyment from the collective effort by all members to work collaboratively towards completing a task. Social cohesion, on the other hand, refers to group cohesiveness resulting from the emotional bonds and friendship, which tends to be motivated towards completing tasks just to appease others. Tasks themselves can be viewed differently by team members. For example, task interdependence can lead to more communication since groups see interconnections between sub-tasks that contribute to overall group performance, whereas, positive outcome interdependence can lead to more constructive conflict since individual members' benefits are associated with collective successful task completion (e.g., Van den Bossche, et. al, 2006).

## Methods

### Setting and participants

The setting was a hybrid formal/informal learning experience for 9<sup>th</sup> graders in a large city in the Northeastern United States. Participants (10 girls, 8 boys from diverse ethnic, racial and socioeconomic backgrounds) were students from a science magnet high school. As part of the school's required out-of-school learning experience, students could choose from an assortment of informal workshops coordinated through a partnering science museum, covering a variety of science and technology topics and offering various forms of instruction. The workshop ran for eight weeks, once a week, for about 2 hours (time and duration limitations imposed by the school). Eighteen students enrolled in the workshop and 17 completed all course activities (one left the school), working on teams of 2-3 students to create final projects (artifacts).

### Curriculum

The workshop had been taught previously but redesigned to respond to limitations during the previous instantiation of its experimental curricula (Richard & Kafai, 2015). The novel curriculum of both instantiations was designed and taught by the first author, and focused on utilizing novice friendly digital and physical toolkits to teach youth how to create bidirectionally responsive projects. In other words, students learned how to create projects that were responsive in multiple physical and digital interfaces. While past efforts have taught students how to create projects with digital and physical interaction, this curriculum was unique because they had to engage in designs that were simultaneously responsive in both environments (Richard & Kafai, 2015).

Learners worked with Scratch, a block-based coding and media creation platform, the MakeyMakey, a physical computing plug-and-play device that can make anything conductive control the computer, the Lilypad Arduino, a microcontroller to create e-textile-based wearable physical interfaces, and ModKit, a visual, block-based coding environment, to code the Lilypad Arduino. The purpose was to encourage learners to create digitally and physically responsive wearable games, which would mirror many of the kinds of design and development practices utilized in creating current high tech wearable products. Further, we wanted to foster the kinds of computational literacies involved in both designing these kinds of technologies and also in working collaboratively on a team to do so. However, they could also interrogate and redesign the systems behind them. Moreover, we felt they would learn that design involves the distribution of expertise, and how to collectively negotiate ideas and divide responsibilities. We anticipated that the process of team-based design would integrate their unique interests and skills, making each part essential to the whole, so they begin to see the value of crafting, engineering design, art and media making, and coding as equally valuable to the final product.

In the previous curriculum, youth learned how to use each of the toolkits during the first half of the workshop, where they would engage in both didactic instruction and project-based work. However, based on the novelty of the curriculum, it was more akin to problem-based than project-based learning. Between the first and second version of the workshop, the first author trained and worked with a former graduate student to create buildable models of each aspect of the curriculum. During the second iteration, discussed here, students learned the material primarily by engaging with project-based work from model designs, such as a simple textile glove that could connect to Scratch with the MakeyMakey and conductive fabric, and a standalone e-textile project with the Lilypad Arduino that would control LEDs with conductive fabric. They also learned how to remix in Scratch, how to create basic interactions, such as movement, and how to create customized sprites. Finally, they learned how to connect all of the pieces together with a more complex e-textile glove that produced lights, sound and vibration while controlling a Scratch game. They spent half of the fifth class learning design-thinking principles and creating storyboards of their design ideas. Afterwards, they chose their own teams, and negotiated their designs in order to start the final project creation process over the final three classes.

### Methodology

Data utilized here is in the form of video recorded class observations and interviews with 16 students who consented (pseudonyms used to protect identity). A total of 3 fully bidirectional and 3 unidirectional projects

were created (unidirectional projects only utilized Scratch and the MakeyMakey to create wearable game controllers with no physical responsiveness). This data analysis focuses on the bidirectionally responsive games, of differing complexity, and one unidirectional game, as case studies for the kind of affordances that can be fostered through *bidirectionally responsive making*.

In line with our research question, we explored observations and interviews to understand whether and how collaboration and inclusivity were encouraged through the workshop and toolkits. We utilized Miyake and Kirschner's (2014) team learning model as a frame for understanding these interactions. Further, we explored whether team members felt safe to explore their own individuality within the group (psychological safety) and if the team helped cultivate an environment for open communication, diverse ideas, and supportive and constructive feedback. Further, we analyzed whether teams valued task cohesion or social cohesion and how their choices may have affected overall team performance. To explore inclusivity, we looked at evidence of a change in computational perspectives (Brennan & Resnick, 2012) and perceptions of computing as a field (Kafai, et. al., 2014). We utilize perceptions to understand how the workshop fostered digital identity formation (i.e., technology self-efficacy), which is instrumental to equitable participation in computing (Goode, 2010).

## Findings and discussion

### Bidirectional projects

The “Dino Party” (see figure 1) project is a pet care game where players utilize a carefully crafted, visually-responsive, wearable bracelet to command a dinosaur pet to play and eat in a digital Scratch game of their design. The team designed the physical controller so that both the MakeyMakey and Lilypad were connected. When different conductive fabric buttons were touched, the pet sprite would perform a sequence of actions in the digital environment, which they had coded in Scratch and connected to the fabric with the MakeyMakey, and, simultaneously, the tricolor LED would change color in the physical environment, which they had designed the e-textile circuit to do and coded in the Lilypad Arduino through the ModKit coding environment. Project members consisted of 2 girls (Sadia and Elisa) and 1 boy (Andre). The “Dino Party” team consisted of team members who were mostly inexperienced with programming, digital design, e-textiles and circuitry. Andre and Sadia, for example, did not have experience in any of the content areas, while Elisa had some background and experience with Scratch and circuitry. Sadia explained that in the beginning “just Andre sewed, I made the glove and then Elisa programmed, but we alternated” so that everyone “got a fair chance of doing something.” Sadia, Elisa and Andre were initially “clueless” about their final project because all their ideas had “pros and cons.” They “listed the pros and cons and...the challenges,” divided up the tasks and also switched tasks. The division of labor and the opportunity to perform other sub-tasks such as sewing, programming and designing shows that the team promoted a safe and collaborative group learning environment. The team created a practice of collaboration where decision making was structured into a cost benefit analysis. Sadia described that the team had “different ideas” regarding the sprite design though they “didn’t really argue too much.” The team dynamic was such that during instances of constructive conflict the team “compromised” and “talked it out.” The team also was not punitive when somebody “didn’t code right” and instead were “supportive of each other.”

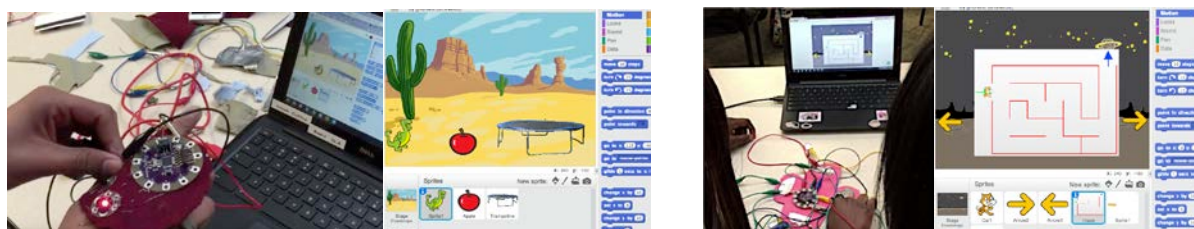


Figure 1 (left). “Dino Party” project; Figure 2 (right). “Scratch Cat...” project.

Overall, the team seemed to be motivated more by task cohesion. The group had devised a strategy around the division of labor such that “Andre sewed, [Sadia] made the glove and...Elisa programmed” but they “alternated” so that everyone had a “fair chance of doing something.” Sadia further described that the team “took a couple of weeks to get [the project] programmed, coded, [and] sewed [with] the conductive fabric” but they felt “accomplished.” Andre mentioned that the team worked on “three different things going on at the same time” and Elisa said that it was “fun to associate [all of the components] together.” All team members described feeling proud of what they had accomplished together, with Sadia stating that she felt “successful of what [she] made.” The Dino Party team had also high levels of group potency. While, according to Sadia, combining the LilyPad, the Makey Makey, and Scratch was “sometimes overwhelming” because of “all these wires and coding” and the constant “plugging it up and unplugging,” the team pushed through it. Further, all members of

the Dino Party team expressed positive shifts in perceptions of computing. For example, Sadia, who originally thought computing was “boring” or something “computer geeks do,” realized that working with computers is “actually a lot of interactive work with your people.” While Elisa and Andre wanted to pursue careers in medicine, they saw new diversity in how computing could be applied in their lives: Andre saw computing as including “tangible real world objects” and Elisa felt the workshop “helped [her] in engineering” so she can now say “you can do it this way instead of that way.”

“Scratch Cat Screeches at You” (figure 2) is a maze game, where players utilize a petal-like controller, with conductive fabric that serves to move the Scratch cat sprite up, down, left or right (figure 5). The project is bidirectional in that the e-textile controller is connected both with the MakeyMakey and Lilypad, and both Scratch and the Lilypad are coded for responsiveness. For example, pressing the upward facing petal on the textile controller would trigger a “when [up arrow] pressed” sequence of events in Scratch, as well as turn on an LED connected via the Lilypad. Project members were two girls (Evelyn and Jackie). The “Scratch Cat...” team was a unique case, consisting of a member with lots of experience in coding and media making in Scratch but no experience in circuitry and e-textiles (Evelyn) and another member with no background in any of the domains (Jackie). Jackie described how Evelyn set up the Lilypad and did most of the Scratch coding while she “designed the controller to be a flower.” Jackie also connected the “trilight [tricolor LED] to the thing that made everything light up [the Lilypad Arduino]” with Evelyn’s help. Evelyn stated that the project was “very collaborative”: they “split off, the Makey Makey and the programming because Jackie could sew better than [Evelyn] could and [Evelyn] could program better than [Jackie] could.” Similarly, Jackie wrote a “few Scratch commands” because she felt it was not “fair to have [Evelyn] do all the Scratch work.” Overall, they appeared more motivated by task cohesion, partially because they didn’t have social bonds before the workshop. However, there was evidence of a lack of psychological safety in the team dynamic. Due to Evelyn’s prior experience with Scratch, Jackie did not feel confident working with it because she feared that she might “mess” things up. Evelyn was highly invested in her Scratch coding expertise and this unspoken hierarchy was enforced in the team. While she did ask Evelyn to let her write a “few commands,” she gave up because she could not code one of the sprites. Jackie also complained that Evelyn “sometimes...didn’t listen to [her] ideas” and vice versa, and that Evelyn “moved too fast” or they would “moved on” to other tasks without communicating.

The team experienced instances of high and low group potency throughout the design process. Jackie felt that “it was good” working with her partner because she helped her “deal with challenges.” The team was effective at “keep [her] on task” and “helped her learn...about Scratch and Modkit.” Evelyn commented that “it was frustrating” working with ModKit and Lilypad Arduino “because [they] never worked” and they were “angry at times” when combining the environments. However, the team pushed through their difficulties, finished the project, and even thought their partnership was strong enough to team up for a science fair. Moreover, despite some of the limitations in team dynamics, both team members expressed shifts in their perceptions of computing. Jackie originally thought computing “would be boring or just dry” but afterward realized that it could be “fun,” “challenging,” and something she could “learn” from. She even discussed creating gloves on her own that could keep someone warm in the winter. Evelyn felt the workshop helped her “work through” computational creation, and “felt better” about computing, which would “definitely” play an important role in her future aspiration to be an astrophysicist.



Figure 3. “Whack-A-Dragon” project.

“Whack-A-Dragon” (figure 3) was a game akin to whack-a-mole that utilized dragons and ghosts instead of moles. Sprites would randomly appear on the screen in Scratch: hitting a dragon sprite would reward you in the game, whereas hitting a ghost sprite would penalize you. A colorful felt-based board with conductive fabric served as the physical interface, where you could hit different parts of the board with a mallet created from a recycled water bottle. The conductive fabric corresponded to screen coordinates in Scratch, and, when hit, would trigger a musical sound in the physical environment (through the Lilypad Buzzer) and sounds and actions in the game. Three boys (Jeremy, Hayden and Eric) made up the project team. Most team members had no prior experience with any of the domains or competencies associated, except for Jeremy, who had some minor circuitry experience and also had previously designed a car out of recycled materials. Hayden described

that “the work was shared [amongst the] group,” such that he was “the one who did Scratch” and the team “wanted another person mostly to do the sewing.” Observations showed that Eric coded the Lilypad in ModKit with Jeremy’s help and Jeremy did all of the sewing exclusively. As Hayden explained, team members “would contribute equally [to] each [of] their roles, and then other people would help them with their roles sometimes.”

While class observations seemed to indicate a lack of psychological safety and cohesion, all members indicated that they felt safe to communicate and negotiate. For example, Eric felt confident within the group to “ask questions” to team members and share “creative ideas.” He also described that one of the benefits of teamwork was that he could “count on them to figure it out” if he “didn’t understand something.” Jeremy recalled some disagreement when it came to the project but, while “com[ing] to an agreement... was pretty hard,” they “voted” and were able to resolve it fairly easily. While they demonstrated levels of task cohesion, such as when Hayden expressed feeling “good” about “figuring out” how to create the final project, they overall evidenced social cohesion. For example, Eric expressed that it was good to “collaborate with [his] friends...because you can ask your partner’s questions and have creative ideas,” as well as work on different components of the project in order to make “it into 1 big thing to show what we’ve learned and what we can now do.” Jeremy felt the workshop “changed the design of the group” regarding “how [they all] worked together” in a “positive” way. While the literature asserts that task cohesion leads to better learning and performance behavior (and the team did struggle with staying on task), by the end of the workshop, they all expressed significant positive changes in their computing self-efficacy and identification. For example, Jeremy felt the workshop “definitely” changed his perceptions of computing in that he “never really knew it was that simple to program” and was now “going to expand [his] experience.” Similarly, Hayden felt programming was “something [he had] wanted to get more into” and the workshop “helped reinforce [his] feelings about it.” Eric felt Scratch “was fun” and enabled him to “create anything.” He also enjoyed working with the Lilypad and e-textiles because “it brought more [of] the real world [where] you [can] create an actual game you can play using physical objects not just in the computer.” While he contended that he now thought “computing [was] more fun,” he also felt “it was more difficult” because “there’s so many possibilities,” though he could “take advantage of it.” He was especially motivated by design “because...you can have so many options and just be creative,” particularly with wearables because they “make it more realistic.”



Figure 4. “Bad Hair Day” project.

### Unidirectional project

“Bad Hair Day” (Figure 4) was a creatively designed wearable game that utilized the MakeyMakey to control Scratch. The textile controller repurposed a hat to represent the kinds of embodied interactions involved in getting your hair done. Touching different parts of the hat would result in having a character on screen get her hair shampooed, blow dried and styled. Three girls (Aminata, Tyra and Mia) were project team members. The “Bad Hair Day” team had some past experience with some of the domain competencies and practices. While Mia had no background, Aminata had had some limited experience with Scratch and Tyra had some experience with circuits. Surprisingly, Mia took on the group leadership role, encouraging the other members to collaborate and distribute aspects of project creation, while the other two members were mostly otherwise directed when she was not in attendance (due to her high school athletic obligations, she often missed classes). While Aminata was instrumental in coming up with the design concept during the design thinking aspect of the workshop, she found Scratch to be challenging, describing it as “hard for [her].” She also knew that she “would give up easily” if she worked alone, but her teammates encouraged her and - in her words - said “let’s try it out.”

There were some arguments between Tyra and Aminata: according to Tyra, they were not “compromising,” and, according to Aminata, they had “different ideas.” Although there were arguments, constructive conflict and negotiation between Tyra and Aminata’s differing ideas meant that team members felt secure to voice their concerns. For the most part, this group demonstrated more social cohesion than task cohesion, which may partially explain why their project ended up being unidirectional. For example, when asked about challenges with working the team, Tyra responded that they had to “sew everything on” and that “somebody had to take it home” and remember to bring it the next week. They “had to text each other” to remind people to bring the hat to class. Similarly, when asked about troubleshooting practices, Mia mentioned



that they “had to” to code with a forever loop to create sounds that mimicked shampooing, washing, and blow-drying when different parts of the hat controller were touched. This frequent emphasis on having to do something implied that the team was more driven by social cohesion than task cohesion. However, they started to evidence task interdependence. Tyra expressed that her group “took turns sewing everything together,” “helped each other” and did “everything together.” Aminata explained this was because if a team member “did not understand something...maybe [their] friend will” and they would “correct it.” This task interdependence meant that team members “taught each other” when some team members “were stuck” or “didn’t understand something.” Over the course of time, they expressed more task cohesion, coupled with group potency. Determined to finish the project, Aminata “slept over” at Tyra’s house “every Saturday...to work on [their] project” outside of the workshop. Class observations showed that Mia and Tyra were initially more active workshop members while Aminata had checked out early on. However, after the team adopted her hair salon idea, she was more engaged and rallied the team to complete the project. Tyra described that, initially, they could not connect the MakeyMakey to the computer because “every time we would almost have it, it would be a problem so we had to start over again.” However, they kept “com[ing] back” to the task and finally “made it work.” According to Aminata, the team had nearly “[given] up” but decided that it was worth completing and when “it all came together” they were “proud.”

While the team initially struggled, by the end of the workshop, each member expressed newfound positive changes in their perceptions of computing to varying degrees. For example, while Aminata wants to be a doctor and feels that she has not been “good at computers,” after participating in the workshop she “like[s] them because [she] was really into [them] here.” Similarly, Mia, who still has reservations about a career in computing, expressed wonderment in what she learned during the workshop: “Until, like we came to the workshop, I didn’t know... I could be standing over here, and... doing something on my computer from right there.” She especially enjoyed creating wearable technologies, which she thought were “fun.” Tyra who “didn’t have anything really to do with computers or technology and stuff,” apart from schoolwork and entertainment, felt better about a career in computing. She previously thought that programming was a “basic job” where people “sat at a desk and typed some stuff up and that was it.” After participating in the workshop, Tyra’s perspective is of admiration and appreciation, stating that programmers are “doing a really good job” because it is “hard.” Her perspective shifted from “know[ing] nothing about computers” to thinking “wearable technology is fun,” and enjoying seeing “the wires and everything...[and] really play[ing] around with [them].”

## Conclusion

Findings indicate that bidirectionally responsive making shows significant promise for fostering inclusive collaborative practices in makerspaces by implicitly scaffolding and distributing making activities, and helping learners better appreciate the creative, diverse and meaningful ways computing could be used in their lives. For example, the tools themselves encouraged the distribution of activities, while also being attuned to the whole, enforcing task interdependence and often encouraging individual and group self-efficacy. Further, for most teams, there appeared to be changes in the ways that gendered interests were expressed, as other studies discussed previously have found. While the all boy team still expressed interest in the technical aspects (i.e., “real world” and “tangible” connections over crafting), the mixed gender teams were more purposeful in distributing tasks and expressing diverse interests. However, unlike other interventions that integrate one toolkit or maker activity (such as standalone e-textile work) as a means of diversifying access or changing perceptions, bidirectionally responsive making encourages learners to distribute activities, such that collaboration is an essential aspect of design work. This process also helps integrate learners from their areas of interest or expertise. While this can sometimes create friction, it can also produce opportunities for meaningful peripheral learning, both of which were evidenced on the “Scratch Cat...” team. Furthermore, learners from groups underrepresented in computing (on teams such as “Dino Party” and “Bad Hair Day”) expressed stronger computing identity and self-efficacy. We contend that having a shared artifact of their efforts can be a powerful motivator and serve as evidence of mastery, regardless of their design team roles. We see this across teams whose members expressed not only a greater understanding of computing and the diverse application it had in their lives, but also awareness of aspects of the development process that were appealing to them, whether it be crafting, design, engineering or coding. During both instantiations of the workshop, allowing members to choose their own teams mostly ended up being an added benefit, which helped them finish their projects. However, there is always the concern that some students will be left behind. While not all groups were focused on task cohesion, most eventually demonstrated a shift from group cohesion (or friendship-oriented goals) to task goals, such as with the “Bad Hair Day” team.

An important limitation in our findings is the cost associated once multiple maker toolkits are integrated in the curriculum. This is an important consideration, despite the increased cost accessibility of maker

and DIY materials (Blikstein, 2013). Once each toolkit is factored in, the total cost of materials per team is estimated between \$100-150 (or \$30-50 per student); on the other hand, having a purposeful team-based activity allows learners to distribute the costs as well as the benefits of learning a variety of tools. Another limitation happened when the school stopped providing PC laptops in favor of Chromebooks. However, Chromebooks cannot download software, such as ModKit. Incorporating the Raspberry Pi in the curriculum would add a layer of complexity but also benefits such as learning computer hardware and providing a more affordable and extensible Chromebook alternative. A future direction for our work will be to further understand the benefits and limitations of bidirectionally responsive making for inclusive collaborative learning for older and younger learners and with alternative toolkits and materials, such as paper or squishy circuits and Raspberry Pis.

## Acknowledgements

Special thanks to Yasmin Kafai for supporting the project, Lei Radium Yang for helping design the buildable models for the second iteration of the workshop, and R. William Ashley for assisting with data analysis.

## References

- Blikstein, P. (2013). Digital fabrication and 'making' in education: The democratization of invention. *FabLabs*, 1-21.
- Brennan, K & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. *Annual Meeting of AERA*, Vancouver, BC, Canada.
- Buchholz, B., Shively, K., Peppler, K., & Wohlwend, K. (2014). Hands on, hands off: Gendered access in crafting and electronics practices. *Mind, Culture, and Activity*, 21(4), 278-297.
- Buechley, L., Eisenberg, M., Catchen, J., & Crockett, A. (2008). The LilyPad Arduino: using computational textiles to investigate engagement, aesthetics, and diversity in comp sci ed. In *Proc of SIGCHI* (pp.423-432). ACM.
- Dougherty, D. (2013). The maker mindset. In M.Honey & D.Kanter (eds.), *Design, make, play* (pp. 7-11). Routledge.
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Admin sci quarterly*, 44(2), 350-383.
- Goode, J. (2010). The digital identity divide. *New media & society*, 12(3), 497-513.
- Halverson, E. R., & Sheridan, K. (2014). The maker movement in education. *Harvard Ed Review*, 84(4), 495-504.
- Heitin, L. (2014, Jan 10). No girls, blacks or Hispanics take AP computer science exam in some states. *Ed Week*.
- Honey, M., & Kanter, D. (Eds.). (2013). *Design, make, play*. Routledge.
- Ito, M., Gutiérrez, K., Livingstone, S., Penuel, B., Rhodes, J., Salen, K., ... & Watkins, S. C. (2013). *Connected learning: An agenda for research and design*. Digital Media and Learning Research Hub.
- Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Interaction Book Co.
- Kafai, Y., Fields, D., & Searle, K. (2014). Electronic textiles as disruptive designs. *Harvard Ed Rev*, 84(4), 532-556.
- Kafai, Y. B., Lee, E., Searle, K., Fields, D., Kaplan, E., & Lui, D. (2014). A crafts-oriented approach to computing in high school. *ACM Transactions on Computing Education (TOCE)*, 14(1), 1-20.
- Kafai, Y., Searle, K., Martinez, C., & Brayboy, B. (2014). Ethnocomputing with electronic textiles. In *Proc of the 45th ACM SIGSCE* (pp. 241-246). ACM.
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2002). The sociability of computer-supported collaborative learning environments. *Educational Technology & Society*, 5(1), 8-22.
- Lachney, M., Babbitt, W. & English, R. (2016). Software Design in the 'Construction Genre' of Learning Technology: Content Aware Versus Content Agonistic. *Computational Culture*, 5.
- Miyake, N., & Kirschner, P. A. (2014). The social and interactive dimensions of collaborative learning. In R.K. Sawyer (ed.), *The Cambridge handbook of the learning sciences* (pp. 418-438).
- NSF (2015). *Women, Minorities, and Persons with Disabilities in Science and Engineering*. NSF: Arlington, VA.
- Peppler, K., Maltese, A., Keune, A., Chang, S. & Regalla, L. (2015). *Survey of Makerspaces, pt II*. Maker Ed Initiative.
- Peppler, K., Halverson, E., & Kafai, Y. B. (2016). *Makeology (Vol.1)*. Routledge.
- Resnick, M., Berg, R., & Eisenberg, M. (2000). Beyond black boxes: Bringing transparency and aesthetics back to scientific investigation. *The Journal of the Learning Sciences*, 9(1), 7-30.
- Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B. & Kafai, Y.B. (2009). Scratch: Programming for all. *Comm of ACM*, 52(11), 60-67.
- Richard, G. T., & Kafai, Y. B. (2015). Responsive make and play: youth making physically and digitally interactive and wearable game controllers. In A.Nijholt (ed.), *More Playful User Interfaces* (pp. 71-93). Springer.
- Richard, G. T., Kafai, Y. B., Adleberg, B., & Telhan, O. (2015). StitchFest: Diversifying a College Hackathon to broaden participation and perceptions in computing. In *Proc of the 46th ACM SIGSCE* (pp. 114-119). ACM.
- Scott, K. A., Sheridan, K. M., & Clark, K. (2015). Culturally responsive computing: a theory revisited. *Learning, Media and Technology*, 40(4), 412-436.
- Shea, G. P., & Guzzo, R. A. (1987). Group effectiveness: What really matters? *Sloan Management Review*, 3, 25-31
- Van den Bossche, P., Gijssels, W. H., Segers, M., & Kirschner, P. A. (2006). Social and cognitive factors driving teamwork in collaborative learning environments. *Small group research*, 37(5), 490-521.

# Investigating Immersion in Relation to Students' Learning During a Collaborative Location-Based Augmented Reality Activity

Yiannis Georgiou and Eleni A. Kyza

ioannis.georgiou@cut.ac.cy, Eleni.Kyza@cut.ac.cy

Media, Cognition and Learning Research Group, Cyprus University of Technology

**Abstract:** Immersion has been argued to affect students' learning in settings such as virtual worlds and digital games. However, a review of the literature indicates a lack of empirical studies investigating immersion in relation to the learning process. The present case study characterizes students' immersive experiences during a location-based augmented reality science activity. Two pairs of students were purposefully selected from a cohort of eighteen 11<sup>th</sup> graders, due to their diametrically opposing views about their immersive experience. The analysis of students' discourse during the activity, and of post-activity interviews, yielded a coherent indicator of immersion. To investigate whether each pair's immersion affected the learning process, we analyzed the pairs' activity logs, discourse and learning outcomes. Findings show that immersion was related to the learning process, dramatically affecting students' learning behaviors, such as collecting and interpreting the available data, as well as problem-solving patterns.

## Introduction

Immersion is a widely-used construct in the literature on digital technologies, such as computer and video games, virtual worlds or location-aware augmented reality (AR) apps. According to Dede (2009), immersion is "the participant's suspension of disbelief that she or he is 'inside' a digitally enhanced setting" (p.66). Conceptualizing immersion as a graded process of cognitive and emotional involvement, researchers have argued that heightened levels of immersion can facilitate science learning (Cheng, She & Annetta, 2015). Based on the review of the extant literature there is a lack of empirical studies investigating immersion in relation to the learning process or to students' collaboration; this gap is important to be addressed, given that immersion represents a psychological experience unfolding during the learning process (Jennett et al., 2008).

Empirical studies on the topic, mostly using quantitative methodologies, have previously investigated immersion in relation to students' learning gains in the context of game-based virtual worlds, and have often resulted in contradictory findings. Although some of these studies have provided empirical support for the positive effect of immersion on students' learning (e.g. Hickey et al., 2009; Ketelhut et al., 2010), other studies found weak or no relation between immersion and learning outcomes (e.g. Cheng et al., 2015; Hsu & Cheng, 2014). Even though the latter studies have not identified a positive relation between learning outcomes and immersion, they have indicated that immersion is highly related to students' game scores, suggesting that immersion has a significant impact on students' performance during the learning process. On a similar note, Hsu and Cheng (2014) found no relation between higher levels of immersion and 7<sup>th</sup> graders' conceptual understanding, but identified relations between high levels of immersion and students' problem-solving skills. These findings led them to assume that higher levels of immersion may affect students' problem-based patterns during the learning process, which may not be identified by simply looking at the learning outcomes.

The present study investigates the claim that immersion relates to the learning process in the context of a collaborative location-based AR activity. As in other studies of immersion, augmented reality is a context where immersion is assumed to support learning, but this claim has not been empirically investigated (Cheng & Tsai, 2013; Dunleavy, Dede & Mitchel, 2009). Since there is scant research on investigating immersion and its relation to learning in location-based augmented reality settings, the first goal of this study was to characterize immersive experiences as experienced by the students in the field and as reported at the end of the activity. A second goal of this study was to investigate the relationship between students' immersive experiences, their learning process and outcomes. Understanding immersion in location-based augmented reality settings and its relation to learning can help us build more engaging learning environments and support learning in informal and outdoors settings.

## Theoretical framework

Location-based augmented reality (AR) settings for science education are assumed to increase students' immersion and impact learning outcomes, due to set of unique characteristics (Dunleavy et al., 2009). In particular, location-based AR settings differ from other digital immersive environments as they: (a) employ

mobile and location-aware interfaces, (b) combine physical and digital spaces, thus creating blended spaces, (c) extend the activity outside the limits of traditional space (e.g. the screen) into the physical space, and (d) provide students with rich interaction possibilities with the physical world, as well as with the virtual elements augmenting reality (De Souza E Silva & Delacruz, 2006). However, learning in location-based AR settings is often considered as a highly challenging task. Based on existing literature, location-based AR settings for learning in science should be structured around authentic but complex real-world problems; for their solution students are often asked to work collaboratively to collect and synthesize relevant data, as they progress through multiple, virtual or real-world data sources (Dunleavy et al., 2009; O'Shea, Mitchell, Johnston, & Dede, 2009). In addition, collaborating students in location-based AR settings are required to apply a set of complex skills, such as collaborative problem-solving, inquiry-based skills, geo-spatial navigation skills and handheld manipulation (Dunleavy et al., 2009).

Immersion, as a multi-level process of cognitive and emotional involvement, can be crucial in terms of defining students' performance, given the complex nature of collaborative location-based AR activities. Students, who are highly immersed in location-based AR settings, feel surrounded by a blended, yet realistic augmented environment, as being in a unified world (Cheng & Tsai, 2013). When this occurs, "students quickly enter a state of suspended disbelief, accept the blended real and digital environment, give their attention over to it, and engage in the variety of options available to them to access content related to the topic being addressed" (Cabiria, 2011, p. 240). Despite these assertions, Cheng and Tsai (2013) have argued that even though immersion is expected to relate to students' behaviors in AR-related learning, there is still a lack of studies investigating how the learning process unfolds in such contexts. The present case study focused on two pairs of high school students, who reported diametrically opposite views about their immersive experience during a collaborative AR location-based activity, to investigate: (a) How can we characterize immersion in location-based AR activities, and (b) What is the relation between immersion and learning?

## Methodology

### Participants

Eighteen 11<sup>th</sup> grade students, working in pairs, participated in the augmented reality activity using mobile devices; their AR experience lasted for approximately 2 hours. Students were randomly assigned to pairs. This case study purposefully focuses on two pairs: Janet and David (Pair 1) and Susan and Jack (Pair 2) [names are pseudonyms]. These two pairs were selected due to their diametrically opposing views regarding their immersive experience, as expressed by them in interviews, which took place after the activity. This focus provides the opportunity to explore whether and how immersion is related to the learning process during the location-based AR activity.

### Learning intervention

The collaborative location-based AR activity took place at a lake near an environmental science center. During the activity, which took the form of a narrative-driven, inquiry-based investigation, students worked in pairs to investigate the mysterious decline of mallard ducks inhabiting the lake; each pair was provided with a tablet equipped with the TraceReaders AR app (Georgiou & Kyza, 2013). The goal of the activity was to engage students in an evidence-based, explanation-building process, and to expand students' understanding of scientific concepts related to the lake ecosystem. As students moved around in the physical world, a map in the AR app displayed information corresponding to different hotspots. The hotspots were triggered once the students were within a radius of 20 meters; once triggered, the app displayed a variety of multi-modal information (e.g. videos, texts, photographs, and audio), which was relevant to the inquiry-based investigation.

### Data collection

To characterize immersion and investigate its relation to learning, data were collected during and after the pairs' AR activity. The following data were collected during the students' investigation: (a) Log files: Students' actions during the intervention were captured in a log file documenting the history of the students' actions, such as time spent on each activity in the app; (b) Audio-taped discussions: Each pair's discussions were audio-recorded through an integrated recorder from within the AR location-aware app; (c) Pairs' final videos: The overall performance of each pair was evaluated based on whether they had reached an evidence-based conclusion at the end of their investigation. For this purpose, each pair was asked to prepare a 3-minute video at the end of their investigation, in which they presented their final conclusions and arguments. Each student also participated in a group interview which took place after the learning activity and lasted for 90 minutes; two group interviews were held. The nominal group technique (McPhail, 2001) was used for the post-session

interviews. According to this technique, students were initially asked to individually write down and justify their viewpoints regarding the immersive nature of the location-based AR activity. As a second step, students were asked to share their ideas with the group; the interviews concluded with a debriefing discussion. In this way, we received both the individual input from all group members and had access to richer discussion resulting from group interaction on the topic.

## Data analysis

The data were analyzed using mixed methods to answer the questions about the process of immersion during the AR activity and the relation of immersion to student learning. To characterize students' immersion the views of the four students expressed during the post-session interview were qualitatively analyzed to develop an immersion indicator, reflecting students' immersion for each pair. For this purpose, we used a coding scheme by Scoresby and Shelton (2011), which defined immersion as a linear process according to which interest for the activity *content*, and *emotion* evoke *motivation*, which in turn results in *engagement* (see Table 1). Thus, the statements of each pair were categorized per student and according to these four immersive states (content, emotion, motivation, and engagement). Statements per state were also classified as negative or positive, thus providing a more nuanced indication of the ways students experienced each different state. Furthermore, students' statements about each state were grouped using a thematic analysis approach (Attride-Stirling, 2001). The immersion indicators, derived from coding the views of the students in each pair, were supplemented with the analysis of the pairs' discourse during the learning process, which was also coded as positive or negative using the Scoresby and Shelton (2011) coding categories. This process provided a systematic way to characterize students' immersion, addressing both the cognitive and emotional involvement with the location-based AR activity. The inter-rater agreement between two independent researchers, who coded 25% of the data corpus, was estimated using Cohen's kappa and was satisfactory, at  $\kappa = .816$ ,  $p < .001$  for the pairs' statements and  $\kappa = .741$ ,  $p < .001$  for students' discourse.

Table 1: Coding scheme for characterizing students' immersion (based on Scoresby and Shelton, 2011)

Immersive state	Definition
Content	Students indicate their interest about the activity in terms of expressing their likes and their dislikes about the different aspects of the activity e.g. the actions performed during the activity, media design (e.g. graphics and sounds), level of difficulty.
Emotion	Students indicate their feelings about the activity, expressing an emotional connection or disconnect with the activity.
Motivation	Students indicate their motivation expressing whether they were looking forward or not to discovering what happens next and accomplishing the learning mission.
Engagement	Students indicate their engagement, or lack of, with the learning process and activities.

In order to relate students' immersion with each pair's learning process, we analyzed data from: (a) log files, (b) audio-taped discussions and (c) each pair's final videos. Quantitative data derived from the log files of the two selected pairs were analyzed descriptively, in order to outline each pair's learning process. The two pairs were contrasted in terms of (a) the number of hotspots visited, (b) the time allocated at the different hotspots for examining the data sources, and (c) the time allocated for examining the data sources, which included inscriptions such as tables, graphs and diagrams. Students' audio-taped discussions were analyzed according to a slightly modified coding scheme by Nilsson and Svingby (2009), in order to classify students' discourse according to learning actions during the collaborative location-based AR activity (see Table 2). As part of the audio-taped discussion analysis, an inter-rater process was employed during which two independent researchers coded the 25% of the data corpus. Cohen's kappa was run to determine the agreement between the raters, with satisfactory agreement ( $\kappa = .802$ ,  $p < .001$ ). Finally, each pair's final video was qualitatively analyzed to determine if each pair had reached an evidence-based conclusion by the end of the learning intervention.

Table 2: Coding students' discourse

Category	Description
Obtaining information	Identifying information from the learning environment through reading sources (text, tables, diagrams) or watching videos
Capturing data	Taking photos from the field and keeping notes about them as data

Category	Description
Problem solving	Discussing the content and how to solve the problem
Navigating	Discussing navigation issues related to the augmented reality app or to the hotspots augmenting the physical space
Interacting with other pairs	Discussing application-related issues with other pairs
Interacting with teachers	Receiving feedback from teachers during the learning process
Off-task discussions	Discussing issues irrelevant to the learning environment
Technical issues	Discussing problems with the use of the tablet or the AR app

## Findings

### How is immersion experienced in location-based AR investigations?

The examination of students' immersion indicators showed that Pair 1 (Jack and Susan) achieved high levels of immersion (see Table 3 for a summary of the assessment of their immersion experience). As Jack reported, the activity captured his interest due to the user-friendly app, its topic, diversity of data provided, its nature-based location and its location-aware qualities. Even though he did not provide any indications regarding any emotional connection with the activity, Jack also expressed his motivation by explaining how he felt challenged to analyze and reflect on the data collected. He also mentioned how he and Susan were engaged with the learning process, explaining how they were actively involved with collecting and reflecting on their data. Not all of Jack's statements were positive. Jack negatively evaluated the realism of the virtual characters, the lack of competition and agency during the activity, as well as the balance between the natural and the virtual world. Susan provided fewer statements about her immersive experience, but also highlighted user-friendliness, and commented that the topic of the investigation and the location-aware aspect of the activity captured her interest. She provided indications for her motivation since, as she reported that during the activity she felt anticipation to move forward and to identify new data. Susan did discuss her emotional connection with the activity, as she reported that in some cases she was carried away or she felt that she was experiencing the activity as something real. Based on these statements, both students could be characterized as of high immersion.

On the other hand, Pair 2 (David and Janet) remained at the lowest level of immersion (see Table 3 for a summary of the assessment of their immersion experience), since the learning activity did not manage to capture their interest. Even though David had positively evaluated the user-friendliness of the app, he negatively evaluated several aspects relating to the interface of the app, such as the text-based information presented and the fidelity of the graphics. He also negatively evaluated the narrative employed in terms of its topic, the narrative plot and the lack of competition, as well as the locality, in terms of the arrangement of the hotspots and the lack of balance between the natural and the virtual world. Since most of the activity did not capture his interest, he also reported a lack of emotional connection with the activity, stating that he could hardly identify with the main character of the narrative-driven investigation. Hence, even though he had indicated that on some occasions he felt motivated to reach a solution to the problem, he provided no indications about his engagement with the learning process. Similarly, Janet reported that her attention and interest were hardly captured by the interface, the narrative and the locality. Therefore, as she admitted, there were times that she felt bored to engage with the learning process (e.g. examine the data sources provided). Given that these students did not provide indications of reaching the immersive states of emotion, motivation and engagement, while at the same time they adopted a, mostly, negative stance towards the content of the activity, both students could be characterized as of low immersion. Table 3 shows the percentage of statements devoted to different aspects of immersion by the students in each pair. The sub-categories under each state of immersion (content, emotion, motivation, and engagement) were reached using a thematic analysis approach.

Table 3: Characterizing students' immersion based on the analysis of the post-session interviews

		High immersion pair				Low immersion pair			
		Jack		Susan		David		Janet	
		(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)
<b>Immersion State 1: Content</b>		<b>55.3</b>	<b>21.3</b>	<b>33.2</b>	<b>33.2</b>	<b>10</b>	<b>76.6</b>	<b>14.2</b>	<b>71.6</b>
<i>Interface</i>	User-friendliness	8.5	0	8.3	0	10	0	0	0
	Augmentation of reality	8.5	0	8.3	0	0	0	7.1	0
	Realism, animation and interactivity of graphics	0	4.3	0	0	0	0	0	14.3
	Realism and fidelity of virtual characters	0	2.1	0	8.3	0	10	0	28.8
	Text-based information	0	0	0	0	0	6.7	0	14.3

<i>Narrative</i>	Topic of investigation	14.8	0	8.3	0	0	6.7	0	0
	Level of challenge	0	0	0	0	0	23.2	0	0
	Diversity and usefulness of the data	12.8	0	0	8.3	0	0	0	7.1
	Competition	0	8.5	0	0	0	3.3	0	0
	Agency and first-person perspective	0	8.5	0	8.3	0	0	0	0
	Narrative plot	0	0	0	0	0	13.3	0	0
<i>Locality</i>	Nature-based location	4.3	0	0	0	0	0	7.1	0
	Mobility and location aware nature of the activity	6.4	0	8.3	0	0	6.7	0	7.1
	Balance between the physical and virtual world	0	2.1	0	0	0	0	0	0
	Hotspots' arrangement	0	0	0	8.3	0	6.7	0	0
	<b>Immersion State 2: Emotion</b>	<b>0</b>	<b>0</b>	<b>25.3</b>	<b>0</b>	<b>0</b>	<b>6.7</b>	<b>0</b>	<b>7.1</b>
<i>Authenticity</i>	Experience the activity as something real	0	0	8.3	0	0	0	0	0
<i>Excitement</i>	Carried out by the activity	0	0	17	0	0	6.7	0	7.1
<b>Immersion State 3: Motivation</b>		<b>4.3</b>	<b>0</b>	<b>8.3</b>	<b>0</b>	<b>6.7</b>	<b>0</b>	<b>0</b>	<b>0</b>
<i>Continuous</i>	To discover something from your data	4.3	0	0	0	6.7	0	0	0
<i>Challenge</i>	To discover new data	0	0	8.3	0	0	0	0	0
<i>Anticipation</i>									
<b>Immersion State 4: Engagement</b>		<b>19.1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>7.1</b>
<i>Data collection</i>	Collecting data and new information	4.3	0	0	0	0	0	0	0
<i>Reflecting</i>	Analyzing, interpreting and combining your data	14.8	0	0	0	0	0	0	7.1
<b>TOTAL</b>		<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

The above characterization of immersion was complimented through the analysis of the pairs' discourse during the learning process (see Figure 1). This analysis corroborated students' post-activity statements about their immersion.

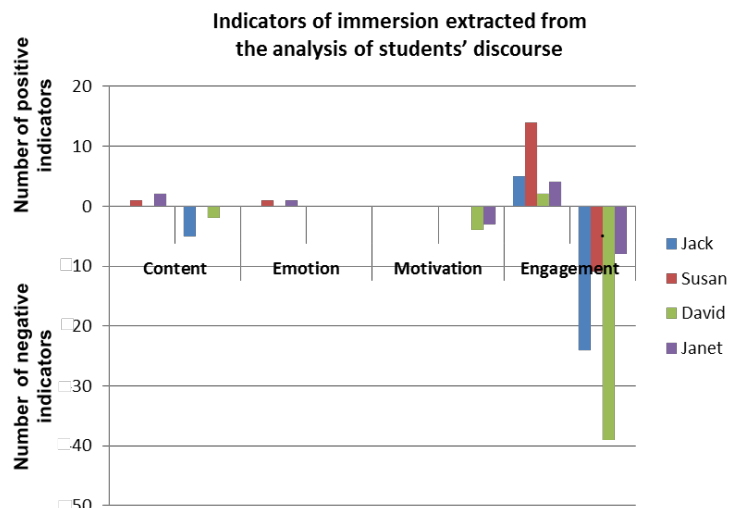


Figure 1. Indicators of immersion extracted from the analysis of students' discourse during the AR activity.

As shown in Figure 1, Pair 1 discourse (Jack and Susan) provided no indications of low motivation while Pair 2 discourse (David and Janet) offered several indications of low motivation during the activity. In addition, Pair 1 seem to be distracted and disengaged much less during the activity than Pair 2.

### Does immersion relate to students' learning?

A descriptive analysis of the students' actions, as recorded in the log file of each pair, indicated that both pairs visited all the hotspots. However, the high immersion pair (Pair 1, Susan and Jack) differed from the low immersion pair (Pair 2, David and Janet). Pair 1 allocated almost double the time at hotspots in examining all the data sources provided, and triple the time in examining specifically the data sources with inscriptions, such as tables, graphs and diagrams, which needed to be analyzed and interpreted (see Figure 2).

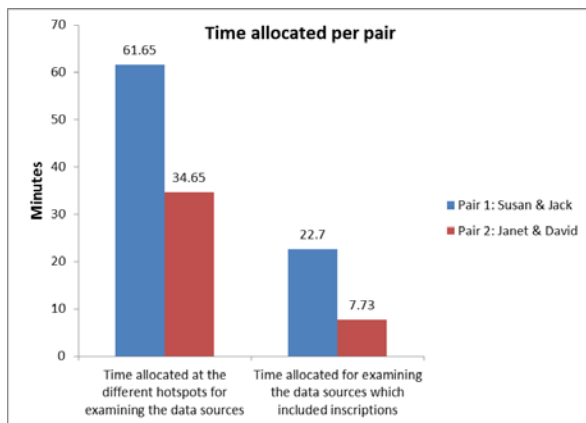


Figure 2. Time allocation per pair at hotspots.

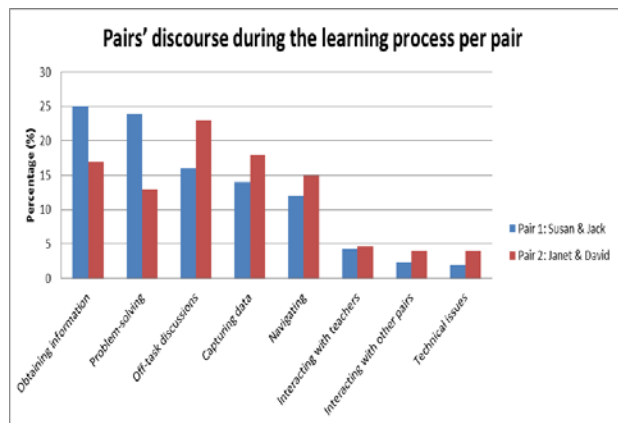


Figure 3. Time allocation per pair at hotspots.

The analysis of each pair's discourse during the learning process indicated that the learning activity of the two pairs also differed: Susan and Jack (high immersion pair) seemed to be more engaged with the activity than Janet and David (low immersion pair). As shown in Figure 3, while for Jack and Susan (high immersion pair) the coded episodes relating to the categories of obtaining information and problem-solving covered 25% and 24% of the total discourse coded for the group respectively, in the case of David and Janet (low immersion pair) these percentages were much lower, covering 17% and 13% of the total number of coded episodes. In addition, the percentages of the coded episodes relating to off-task discussions for Janet and David were much higher (23%) as compared to the percentages of the high immersion pair (16%). According to students' discourse, the actions of obtaining background information and problem-solving were also much different between the two pairs. Susan and Jack, who were highly immersed, paid more attention to the data; as shown in Excerpt 1, these students invested much effort in making sense of the information collected, by reading, for instance, the text more than once.

Excerpt 1

*Virtual character:* I have an analysis regarding the water quality for you. The analysis focuses on the detection of aquatic invertebrates of the lake...

*Jack:* Did you understand what he just said?

*Susan:* What did he say about the aquatic invertebrates? Rewind the video for a moment...

*Jack:* Ok... Let's hear it once again from the previous point.

In contrast, Janet and David seemed to pay less attention on making sense of the data sources when dealing with new information, as shown in Excerpt 2.

Excerpt 2

*Janet:* Several chemical substances...

*David:* There is no need to give much emphasis here. Please read it more quickly.

*Janet:* Ok! Several chemical substances like DDT or lindane, blah, blah, blah... This phenomenon is called bioaccumulation... blah, blah, blah. DDT is transferred to zooplankton... blah, blah, blah...

Another difference was the extent to which students' discussions focused on the problem-solving action, as the two pairs approached the activity very differently. Janet and David, who were not highly immersed, not only allocated less time on reasoning about the subject but, as presented in Excerpt 3, in most of the cases they did not make an effort to interpret the information and relate it to how it could be employed as evidence to confirm or reject a hypothesis, or connect new information with data they had already seen.

Excerpt 3

*Janet:* So, now we have the lindane pesticide which is still employed during some occasions for the



*agricultural crops.*

*David: Yes, ok...*

*Janet: Lindane...*

In contrast, Susan and Jack, who were highly immersed, were in continuous discussion about how the new information obtained could confirm a plausible explanation or not. As shown in Excerpt 4, students would often discuss the different emerging hypotheses regarding the cause of the decline at the duck population, such as the use of pesticides or the use of fertilizers resulting to eutrophication, trying to reach an evidence-based decision.

#### Excerpt 4

*Susan: Yeah... But keep in mind that the cause for the problem is probably one... Now we are divided between the nitrates and the phosphates and the eggshell thinning.*

*Jack: Ok... Let me think... nitrates and the phosphates...*

*Susan: To what reason did we attribute the eggshell thinning?*

*Jack: To the lindane...*

*Susan: To the lindane... You see? But lindane is a pesticide...*

*Jack: Yes. They use it as a pesticide.*

*Susan: So the problem could be attributed either to spraying or to fertilizers.*

To sum up, Susan and Jack, who were characterized as a pair of high immersion, were deeply engaged in the process of interpreting and combining the collected data. On the other hand, the analysis of the low immersion pair's discourse and actions indicated that Janet and David defined the whole investigation process more as a scavenger hunt, and collected the same data as quickly as possible, without focusing on analyzing or interpreting the collected data. Hence, while by the end of the investigation, Susan and Jack correctly concluded that the decline of the duck population could be attributed to bioaccumulation, Janet and David did not manage to reach an evidence-based conclusion.

## **Discussion and implications**

The present study sought to investigate immersion in relation to science learning in a location-based AR activity. In this context, we purposefully focused on two pairs of high school students, who expressed diametrically opposite views regarding their immersion, attempting to: (a) characterize students' immersive experiences, and (b) investigate the learning process of each pair, to examine the relation of immersion to students' learning. The analysis of the two selected pairs' learning process indicated several differences. While by the end of the investigation, the first pair correctly concluded that the decline of the duck population could be attributed to bioaccumulation, the second pair did not manage to reach an evidence-based conclusion. While the outperforming pair was immersed in the process of analyzing and interpreting the collected data, the second pair defined the whole investigation process as a scavenger hunt, by simply gathering data as quickly as possible, but without reflecting on the collected data. These extremes observed in the learning behaviors of the two pairs are aligned with reports of previous studies, which concluded that while in some cases some students employing location-aware AR apps could be deeply engaged with the true meaning of scientific inquiry, others could present indications of disengagement by transforming the learning process into a meaningless "treasure hunt" activity (e.g. Dunleavy, Dede, & Mitchell, 2009; Squire & Jan, 2007; Squire & Klopfer, 2007).

The observed differentiation between students' performance could be attributed, in our case, to students' immersion, as this was reflected in the immersion indicators emerging for each of the pairs. According to the immersion indicator of the outperforming pair, students were positively engaged in the immersive levels that Scoresby and Shelton (2011) suggested: content, motivation, emotion and engagement. In contrast, the students in the second pair did not find the activity content interesting and remained at the lowest level of immersion. These findings provide empirical support for Cheng and Tsai's (2013) assumption that immersion is expected to relate to students' behaviors in AR learning, while also extending previous research efforts, from the field of game-based virtual worlds, supporting that immersion may influence students' performance, such as problem-based behaviors (Cheng et al., 2015; Hsu & Cheng, 2014). However, considering that findings from this case study are based on only two pairs of students, our future work will analyze the data derived from the remaining student pairs who also engaged with the location-based AR activity. Future work will also look at low and high immersion students, as characterized using the immersion indicators described in this study, to examine the role of scaffolding in fostering students' higher levels of immersion. The present

study contributes to the literature by providing empirical evidence about the relation between immersion and learning in location-based augmented reality settings, which is an area that has received little attention in the literature. A better understanding of how learning occurs in informal learning contexts, such as outdoors, location-based augmented reality settings, can support the creation of hybrid spaces for learning in and out of school contexts, and the development of augmented reality learning environments.

## References

- Attride-Stirling, J. (2001). Thematic networks: An analytic tool for qualitative research. *Qualitative Research*, 1(3), 385-405.
- Cabiria, J. (2012). Augmenting engagement: Augmented reality in education. In C. Wankel, & P. Blessinger (Eds.), *Increasing student engagement and retention using immersive interfaces: Virtual worlds, gaming, and simulation* (pp. 225-251). New York: Emerald Group Publishing Limited.
- Cheng, K.-H., & Tsai, C.-C. (2013). Affordances of augmented reality in science learning: Suggestions for future research. *Journal of Science Education and Technology*, 22(4), 449-462.
- Cheng, M.-T., She, H.-C., & Annetta, L. A. (2015). Game immersion experience: Its hierarchical structure and impact on game-based science learning. *Journal of Computer Assisted Learning*, 31(3), 232-253.
- Dede, C. (2009). Immersive interfaces for engagement and learning. *Science* 2, 323(5910), 66-69.
- Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology*, 18(1), 7-22.
- Georgiou, Y., & Kyza, E. A. (2013). *The TraceReaders*. [Augmented reality application]. Limassol: Cyprus University of Technology.
- Hsu, M.-E., & Cheng, M.-T. (2014). Bio detective: Student science learning, immersion experience, and problem-solving patterns. *22<sup>nd</sup> International Conference on Computers in Education*, 171-178.
- Hickey, D. T., Ingram-Goble, A. A., & Jameson, E. M. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education and Technology*, 18(2), 187-208.
- Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human Computer Studies*, 66(9), 641-661.
- Ketelhut, D. J., B. C. Nelson, J. Clarke, and C. Dede. (2010). A multi-user virtual environment for building and assessing higher order inquiry skills in science. *British Journal of Educational Technology* 41 (1): 56-68.
- Klopfer, E. (2008). *Augmented learning: Research and design of mobile educational games*. Cambridge, MA: MIT Press.
- Macphail, A. (2001). Nominal group technique: A useful method for working with young people. *British Educational Research Journal*, 27(2), 161-170.
- Nilsson, E. M., & Svingby, G. (2009). Gaming as actions: Students playing a mobile educational computer game. *Human IT*, 10(1), 26-59.
- O'Shea, P., Mitchell, R., Johnston, C., & Dede, C. (2009). Lessons learned about designing augmented realities. *International Journal of Gaming and Computer-Mediated Simulations*, 1(1), 1-15.
- Scoresby, J., & Shelton, B. E. (2011). Visual perspectives within educational computer games: Effects on presence and flow within virtual immersive learning environments. *Instructional Science*, 39(3), 227-254.
- Squire, K., & Klopfer, E. (2007). Augmented reality simulations on handheld computers. *Journal of the Learning Sciences*, 16(3), 371-413.
- Squire, K. D., & Jan, M. (2007). Mad city mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education and Technology*, 16(1), 5-29.

## Acknowledgments

We would like to thank all the students who participated in this implementation. We would also like to acknowledge the support of Dr. Andreas Ch. Hadjichambis, Dr. Demetra Hadjichambi, and of the personnel of the Cyprus Center for Environmental Research and Education (CY.C.E.R.E.) during the implementation of the "Mystery at the lake" investigation at the facilities of the center. Special thanks are also due to Andria Agesilaou for her assistance in parts of the data coding process.

# A Self-Organizing Network of Schools That Transform Teacher and Student Learning Through Socio-Technical Co-Evolution

Nancy Law, University of Hong Kong, nlaw@hku.hk

Leming Liang, University of Hong Kong, lmliang@hku.hk

Kenneth Cheng, Buddhist To Chi Fat She Yeung Yat Lam Memorial School, ywcheng@seltas.edu.hk

**Abstract:** Scaling up educational innovations through networks has attracted much interest in diverse research and education policy communities. Literature on scaling are often associated with top-down or partnership models of change, and the goals, resources and technology tools used are generally defined and developed by stakeholders outside of schools. This paper reports on the sustained efforts of a self-organizing network of special needs schools in Hong Kong that has worked together for more than a decade to realize the vision of providing the same educational opportunities to children with various degrees of learning disability. The analysis focuses on how their engagement in the development of a collaborative platform for teacher learning started a journey of socio-technical co-evolution that resulted in exponential scaling of the innovation both qualitatively and quantitatively. The evolution trajectory of this network shows characteristics and susceptibilities similar to those in the socio-technical innovation literature.

## Introduction

Scaling up educational innovations (Coburn, 2003; Clarke & Dede, 2009) has been a central theme for education research for several decades, starting from the educational leadership and reform literature (e.g. Fullan & Hargreaves, 2009), and attracting increasing interests from the learning sciences community (Vuorikari, Kampylis, Scimeca & Punie, 2015; Fishman, Penuel, Allen, Cheng, & Sabelli, 2013). Scaling up of innovations are particularly important as countries around the world are launching various curriculum and pedagogical reform efforts since the turn of the millennium, to bring about changes in curriculum standards, and/or to bring a stronger focus on some higher level generic capacities such as collaboration, communication, creativity and critical thinking, often referred to as 21<sup>st</sup> century outcomes (Partnership for 21st Century Skills, 2009). Scalability has also been identified as a key issue in e-Learning implementation (Law, Yuen, & Fox, 2011) as use of digital technology per se would not bring about enhanced learning outcomes, and much depends on the pedagogy adopted (Watson, 2001; Fisher, 2006). In order for e-Learning to bring about transformative learning outcomes requires deep pedagogical transformation (Somekh & Davis, 1997). Hence the challenge to scaling up ICT-enabled learning innovations is primarily one of learning at multiple levels, as sustained changes in classroom practice requires aligned changes to take place within the education ecosystem from classroom to school to district and system levels (Davis, 2008; Law, Niederhauser, Christensen, & Shear, 2016).

Innovation networks have been found to be a productive model of supporting teacher learning for innovation (Hamel, Turcotte, & Laferrière, 2013; Vuorikari et al., 2015) as it provides room for teachers to engage in peer learning and engage in productive knowledge building in the innovation process. However, as educational change is located within a complex ecosystem with many established rules, regulations, practices and expectations as well as organizational, physical and technological infrastructures, the architecture for learning (Wenger, 1998) plays an important role in such situated learning contexts. Architecture for learning can be broadly defined as the “organizational structure, mechanisms and artefacts that are available to facilitate interactions and to consolidate change at different levels of the education system” (Law, Yuen, & Lee, 2015, p. 3). Comparative studies of innovation development under different architectures for learning have revealed that coupling mechanisms (Spillane, Parise, & Sherer, 2011) and the kind of organizational structures and interaction mechanisms (Stein & Coburn, 2008) impact strongly on the effectiveness and sustainability of reform efforts.

Design-based implementation research (DBIR) (Fishman et al., 2013) is an approach to scaling educational innovations that builds on the design-based research approach developed in the learning sciences community (Brown, 1992; Collins, Joseph, & Bielaczyc, 2004) to connect research and practice in an organic nexus to address the challenge of applying research-based learning principles to guide learning innovations in authentic classroom contexts. At the core of the DBIR approach is the concept of infrastructuring (Penuel, 2015), which recognizes the crucial role of the architecture for learning and interactions in the success of scaling efforts and argues for the need to engage in organizational sensemaking

through partnership between researchers and educators in order to be able to make dynamic changes to infrastructures and designs at different levels of the organization to achieve the innovation goals.

In this paper, we report on a study of a self-organizing network of schools that serve children with different levels of learning disabilities. The researchers first met this network in the context of a government commissioned evaluation of a three-year e-Learning Pilot Scheme in which this network was a grant-holder for one of the 21 funded projects.

## Research context and methods

In 2011, the Education Bureau (EDB) of Hong Kong launched a three-year e-learning Pilot Scheme (2011-2014) in order to develop, try out and evaluate when and how e-Learning works best to bring about effective interactive learning, self-directed learning, and to cater for learner diversity in different curriculum and school contexts in Hong Kong (EDB, 2011). Two studies were carried out in relation to this e-Learning pilot scheme. Study 1 was a longitudinal evaluation of the e-learning pilot scheme (2011-14, referred to here as Years 1-3) commissioned by the EDB. Study 2 was a follow-up study (2014-16, referred to as Years 4-5) to investigate the sustainability and scalability of the pilot projects after the end of the funding period. There were wide diversities in the 21 funded projects, and the projects selected for Study 2 were already the most successful studies that showed potentials for sustainability.

A common feature of the 21 pilot projects was the inclusion of an e-Learning resources/tools development component as the technology base for the e-Learning implementation in these projects. The EDB required all the projects to include partnership involvement with the business sector, which generally served the role of technology development to serve the aspirations of the schools' pilot projects. Altogether 21 pilot projects (61 schools in total) were selected for funding; 9 were individual school projects and 12 joint-school projects. Study 1 found that the e-Learning technology developed in some of the projects was marginally used during the 3-year pilot period, and were rarely used again after the end of the 3-year pilot period. Preliminary analysis of Study 2 data found that while the e-Learning technology tool may still be used by one or two schools in some of the projects, all of the project networks stopped functioning after the project funding ended, except for one of the projects. This project (to be referred to as Project S) was an "outlier" in that after the funding ended, the schools still continued to fund further technology development from their regular school budget, and both the scale of adoption and level of pedagogical transformation taking place in the schools actually progressed exponentially. The present study is a case study of Project S, with the purpose of investigating (1) the trajectory of development of the project both in terms of technology development and e-Learning practice implementation over the five years, (2) how this school network evolved in terms of the architecture of learning that supported the innovation, and (3) if there is a connection between the innovation evolution and the network infrastructure that supported it.

In Studies 1 and 2, the research team interviewed the principal, the project core team members and teachers participating in the e-learning pilot at the beginning and at the end of each school year from Years 1 to 5. From Year 2 onwards, the participants were asked about: (1) the general status and any change of features and goals of e-learning in the school; (2) the organizational structures, organizational routines and interaction mechanisms within and outside the school related to the school's e-Learning initiatives; (3) teachers' learning opportunities and outcomes relating to e-learning, and whether these were related to their responses to (2). To evaluate teachers' learning outcomes in terms of changes in e-learning pedagogical design, teaching and assessment practices, we collected in each year the teaching plan and students' work for one curriculum unit selected by the teacher that used e-Learning. In Years 4 and 5, we conducted in addition classroom observations of one e-learning lesson within the teacher nominated curriculum unit after discussion and negotiation with the project team and teachers. The focus of the observations was on the pedagogical approaches adopted by the teacher when using ICT, and the extent to which students were given opportunities to use ICT in their learning that were oriented towards building 21<sup>st</sup> century competencies.

During the interviews in Years 4 and 5 with the principals and teachers in Project S, the research team was told that the project success can be largely attributed to the efforts of two important teams: the Network of principals from a number of special needs schools in Hong Kong (referred to as Network S) that was actually established in 2006, and the Project Accelerator Team (AC Team) that was first established in summer 2013, as well as the unflinching support from a retired University academic from UK who served as a consultant for Network S since its inception and for Project S. At the request of the research team, the project leaders were very generous in making available the entire set of minutes and related documents of these two teams, which became a primary source of data for the research team to understand what kind of architecture for learning was established for the Network, how it evolved over time, and how the architecture impacted on the innovation development and teachers' learning. Additional interviews were conducted with

the key members of Network S and the AC Team, as well as the UK consultant, to understand their role in these two organizational structures and their views on how the project evolved over time.

For the purpose of this study, we operationalize the concept of architecture for learning in our analysis as comprising four important elements: (1) Organizational structures that direct and guide interactions; (2) Mechanisms for sharing, interactions and decision-making; (3) Artefacts that serve as reifications of outcomes of interactions to propagate decisions and advances in understanding; and (4) Technology infrastructure that supports communications, interactions and knowledge management of individuals and communities (Law et al., 2015). Educational institutions and innovation networks are complex systems, and a characteristic of such systems is that history matters. This is no different for Project S. We will first present our descriptive analysis of the S Network, which predated Project S by 7 years, and was the initiator and change agent for the Project, in order to provide the necessary contextual background for understanding how the project and its architecture for learning evolved.

### **Network S: A self-organizing school network committed to providing equitable learning opportunities for children with learning disabilities**

The “incubation” for Network S started in 2003, when the Hong Kong Government launched its comprehensive curriculum reform, which emphasized the goal of nurturing students’ lifelong learning abilities. Some of the special needs school heads were disappointed that no guidelines or support were given on how special needs schools should incorporate the reformed curriculum into the teaching and learning of students with different levels of disability. These principals share the vision that irrespective of the nature and profoundness of a child’s learning disability, s/he should be entitled to a pathway of learning that would give them access to the same curriculum outcome goals as all other children in Hong Kong. This shared curriculum philosophy was referred to by the principals as SAME, to stand for Systematic Approach to Mainstream Education.

In 2006, the Hong Kong Government changed the secondary school structure from seven years to six and launched a new school curriculum in conjunction with that change as an integral part of the overall curriculum reform. Network S was formally established in 2006 when nine of the special needs school principals joined together to bid for government funding under a University-School Support Program (USSP) to develop a curriculum framework so that students with Special Education Needs (SEN) can still be able to access the mainstream curriculum. To achieve the goal of building a direct “bridge” for intellectual disabilities to access the general curriculum, learning activities and resources need to be adapted and customized according to the developmental status and special needs of each child. The goal of this USSP project was to develop the following deliverables:

- A learning progression framework (comprising fine attainment levels) for the general curriculum;
- An assessment system called SCALE (Same Curriculum Assessment for Learning Effectiveness);
- A common curriculum framework and Scheme of Work (SoW) for teachers across SEN schools to share ideas and resources, collaborate with and learn from each other.

To carry out this USSP project, Network S established two inter-connected teams: one for principals and the other for teachers. The Principal Team took leadership in steering the direction of the project, making important decisions and providing professional advice to the teachers, seeking advice from an overseas consultant from the UK and consultants from a local university. This team conducted meetings on a monthly basis. The Teacher Team was also referred to as the Writing Team, which comprised specialist subject teachers from the nine project schools and took responsibility for writing the SoWs, lesson plans and accompanying learning resources. In fact, there were a number of Writing Teams, one for each subject area. Generally, each team would spend about two weeks on its writing tasks. Each Writing Team was coordinated by an Organizing Manager (OM, usually a specialist teacher from one of the schools) and a Strategic Manager (SM, usually one of the principals whose expertise was in that particular subject).

The deliverables from these projects provided a solid and common artefact base for teachers to use in planning their teaching activities to match the learning needs, attainment levels and learning patterns of their students. In 2010, Network S was successful in being funded for a subsequent USSP project that focused on developing resources for lesson planning: co-constructing teaching resources around the teaching units set out in the strands and key stages of each of the key learning areas.

During the process of collaboration, core members of Network S found that they were unable to maximize the uptake and utilities of the resources created without these being integrated into an advanced information technology system. The Network leaders also believed that e-Learning could enrich and enliven the learning of SEN students. Hence, Network S decided to submit an e-Learning pilot project proposal to

further enhance and build on what the Network has already developed to achieve its vision and goals.

### **Project S: History of its 5-year development**

Project S was a joint bid from 10 SEN schools in Network S. Their project goal as indicated in the proposal was to develop an online platform, for sharing and collaboration in adapting and customizing the learning activities and resources to cater for the special needs of each child, building on the resources that have already been developed. During Year 1, the project focus was on collating resources according to the attainment levels framework established. In Year 2, the project team found this to be inadequate and re-focused the project on designing learning activities for effective use of the collated resources. In Year 3, the project team considered it necessary to facilitate a paradigm shift in teachers: focusing on changing teachers' pedagogical practices in order to achieve the goal of catering for learner diversity through e-learning. So the school-based and joint-school activities were changing from a resource and activity focus to learning design, peer observation of lessons and documentation of evidence of students' learning outcomes. Learning design with a four-level structure (Curriculum, Scheme of Work, school, and class levels) was explicitly identified as a framework for teachers to follow. The primary focus for Project S in Year 3 was to scaffold teachers' lesson design and classroom practice. The technology Platform was modified so that teachers were no longer able to access the curriculum and assessment resources available from the Project unless they go through the SoW, school level plan and class lesson plan (i.e. the four level design structure). This change was specifically introduced to force teachers to think carefully about learning design considerations through changing their lesson design practice. In Year 4, the focus moved to student-centered learning and the collection and use of analytics on students' performance data. The Project also started to transform many of the functions on the web-based platform to support mobile-based applications, making it more convenient for teachers to collect evidence of students' formative performance. More collaborative functions were also added, such as providing parental access to students' learning records, allowing students to upload their own assignments, etc. In Year 5, as the demands for e-Learning resources and collaborative activities increased, Project S established a formal collaboration arrangement with Google and another digital education portal in Hong Kong so that the resources and services from these partners can be integrated into the Project platform. Besides, the platform was gradually revised to cater for diverse e-learning development progress and context among the project schools.

Table 1 summarizes the key project changes over the five years. There are several noteworthy features in the project development trajectory. First of all, the deliverables (artefacts such as SoW, lesson plans and learning resources) developed in the years prior to the project start served as an important part of the architecture for learning as reifications (Stein & Coburn, 2008) to scaffold teachers' learning, design and classroom practices. Secondly, these artefacts were developed by the Network and so have authenticity and ownership for the schools and teachers in the Network. Thirdly, there was a learning process for the Project team, moving from a resource model of change to a strong focus on expertise development (design expertise) and on changing pedagogical paradigm and practice from Year 1 to Year 3. In Year 4, the focus was on building support for feedback on students' learning through the portfolio type assessment support platform linked to the attainment levels framework (SCALE). In Year 5, the change was further consolidated through adding platform functions that support collaboration with parents as partners in facilitating children's learning. Fourthly, by making the students learning outcomes evident and linking them to the SCALE framework, the achievability of the Network vision was made tangible and convincing. This led to an exponential increase in the uptake of the platform use as well as in the increasingly student-centred pedagogical practices adopted by the Network teachers. Fifthly, the evolution of the technology platform reflects a deepening understanding of the "nature of the beast" in terms of e-Learning adoption as a pedagogical innovation requiring a clear focus on changing teachers' practices. Further, the project leaders have cleverly changed the interaction design of the platform so as to enforce changes in teachers' work practices in lesson design: demanding that teachers pay explicit attention to pedagogical design considerations before the selection of activities and resources. In fact, the platform also required teachers to consider students' specific learning needs in the design process by requiring teachers to specify the specific students targeted when assigning learning activities, and the learning levels of the students also have to be made explicit to match the level of the learning activities. Hence this project is not simply one of developing an e-Learning support platform, but one involving socio-technical co-evolution.

### **Project S: Architecture for learning**

To lead the e-Learning Pilot Project, the joint-school Principal Team in Network S served as the driving change agent, and a teachers' network was also set up for implementing the project on a day-to-day basis.

The latter had specialized sub-networks under it for IT coordination and for the different subjects in Years 1 and 2. These two teams were interconnected through a formal coordination mechanism of regular meetings to exchange ideas, concerns and explore solutions. During these meetings, the teachers would report on problems the Teacher Teams identified to the Principal Team. The Principal Team then held meetings to discuss the issues raised, which could concern administration, resources, technology, or pedagogy, and come up with solutions to feedback to the Teacher Team. In addition, both teams would conduct lesson observations across schools. In Year 2, the project leaders found that the communication between the technical group and the subject groups was not effective as these two groupings of teachers did not have sufficient expertise on both technology and subject teaching to understand each other's concerns. Further,

Table 1: Project S developmental trajectory, Years 1 to 5.

Year	Innovation development focus	Platform features developed	Innovative practices implemented	Students' learning outcomes observed	Scale
1	Transforming the SoW, attainment level, and related learning resources to e-copies for more convenient sharing among network members.	Teachers creating, uploading and sharing teaching resources	No pedagogical practices at this stage. Teachers were trained to prepare lesson plans & other learning resources on the platform	Students work appeared to be paper-based drill and practice. No 21 <sup>st</sup> century competency was observed.	Teachers involved in SoW writing team (Chinese and PSHE)
2	Starting the introduction of e-learning resources into classroom teaching	<ul style="list-style-type: none"> <li>- The major progress of this period was to try-out &amp; revise lesson plans through classroom teaching.</li> <li>- Teachers' use of this platform increased &amp; more resources were created and shared.</li> <li>- An e-forum was established for teachers to share their good practices.</li> </ul>	Introduced collaboration activities among students, such as online discussions, peer assessment.	<ul style="list-style-type: none"> <li>- Some high-ability students started to use tablets.</li> <li>- Differentiated outcomes goals: high-ability students prepared a short presentation &amp; weak ones did a paper-based assignment.</li> <li>- Some students were able to engage in collaboration online.</li> </ul>	Try-out teachers from writing team (Chinese and PSHE)
3	Facilitating paradigm shift in teachers: encourage the modification of pedagogical practices to cater for learner diversity	A clear structuring of a four-level learning design pathway through the online platform (Curriculum, Scheme of Work, School, and Class levels) for the teachers.	Teachers were required to use an online teaching plan to organize the teaching materials, and students' learning materials and learning artefacts.	Collaboration; inquiry within groups; information literacy such as searching, organizing and ethical use of information (for high ability students).	Use of the e-Learning platform in teaching practices in all Network schools.
4	Student-centred learning and analytics on students' performance data (how the data collected on the platform could better serve student-centered learning)	Capture students' learning process in the form of qualitative evidence of students' performance; development of mobile apps to support the web applications.	Teachers continued to explore implementation of collaborative learning among students, use of mobile devices in outdoor activities & differentiated designs for learner diversity.	Collaborative inquiry in groups; information literacy such as searching, organizing and ethical use of information (for high ability students).	The scale increased in each network school
5	Sustainability of the project; e-learning design to cater for diverse learner developmental context in the project schools; learning analytics and big data	<ul style="list-style-type: none"> <li>- Develop media capture apps (iOS, student version)</li> <li>- Student login data added for analysis</li> <li>- Single sign-on between project platform and Google/HKEdCity for easy access to external</li> </ul>	Teachers continued to explore implementation of collaborative learning among students, use of mobile devices in outdoor activities & differentiated designs for learner	<ul style="list-style-type: none"> <li>- Collaborative inquiry in groups; information literacy such as searching, organizing and ethical use of information (for high ability students).</li> <li>- Using IT tools as productivity tools in</li> </ul>	The scale increased in each network school

		resources.	diversity.	everyday life (For low ability students.)	
--	--	------------	------------	---	--

more and more problems were emerging from the teachers in the participating schools in the process of implementation, but the technical team did not have adequate knowledge and expertise to address the diverse problems arising in the different school contexts. To address these challenges, a new organizational structure called “Accelerator Team” (AC) was created in Year 3, comprising of senior teachers who have knowledge both in technology and pedagogy from each of the participating schools. The role of the AC Team was, on the one hand, to refine and improve the function of the technology platform based on teachers’ feedback, critiques and suggestions the team members solicited in their own schools; and on the other hand, to provide in-situ support to the school teachers in the form of peer coaching and school-based training workshops. From Year 3 onwards, the AC Team had been playing a crucial role in connecting the teachers and the project leaders, providing instant support to and receiving feedback from the classroom teachers on the ground, and proposing revisions to the platform and implementation strategies to the project leader network on the top.

In reviewing the Network documents, it was clear that the Principal Team played a leadership role in driving the project in two important aspects: (1) steering the strategic direction of the project in alignment with the Network’s educational vision, (2) assignment of roles and allocation of human and other resources throughout Years 1 to 5. While the principals’ leadership was important, the compositions and modes of work of the specific working teams (hereafter referred to as innovation teams) were also critical. The innovation teams played the quintessential role of leading from the middle: these teams comprise the teachers who designed, constructed, and piloted the e-Learning platform, and who interacted with classroom teachers to develop pedagogical implementations using the technology platform and tools on a day-to-day basis. They made and implemented school-based plans to realize the project vision and goals, and mediated between the project leadership, school leadership and classroom teachers in the participating schools.

In Years 1 and 2, the innovation teams were structured similarly to those set up in the USSP projects conducted by Network S in the years before Project S came into existence. Essentially, the project was carried out by a small number of task groups, each led by a principal with teacher members drawn from the various Network schools. This was adequate when each component in the project was relatively well-defined, without strong interdependence, and no technology development was involved. In Year 2, the overseas consultant pointed out that there was a need for a central team with strong technological and pedagogical expertise to work together, and that the key challenge was a pedagogical one. There was a pressing need for the core working team to put a strong focus on teacher professional development. With professional and networking assistance from the overseas consultant, the Network S principals and some core innovation team members went on a study visit to the UK to learn about how student-centered learning can be implemented for children with special needs, and how SoW and attainment levels can be used to scaffold teachers’ lesson design and assessment work. These events triggered a significant structural change in the innovation team structure, and the network schools agreed to contribute an experienced teacher with e-Learning and/or curriculum innovation expertise to set up an Accelerator Team (AC Team) in Year 3.

The AC team members were handpicked for their pedagogical experience and relatively sophisticated understanding of the role of technology in supporting student learning. Once constituted, the AC Team began by formulating a total reconceptualization of the platform functions to focus on supporting teachers’ lesson design practices. Further, none of the AC team members had full-time commitments to Project S. Hence they all had teaching duties and roles within their own schools, and were able to sense quickly how the platform features were received by their colleagues in the context of their day-to-day practices. The establishment of the AC Team was instrumental to the successful refocusing of the project directions. In Year 4, the AC Team structure changed again. The reason for the change was two-fold. Firstly, without government funding for the project, some of the schools found it difficult to contribute their prized staff to work on the project. Secondly, the Network decided that the focus from Year 4 should be on developing student-centered practices, and that further technology development would be scaled down. Hence, every school was encouraged to assign one teacher to the AC team, whose role was to help and introduce the platform functions to teachers, and to scaffold innovation-focused professional development activities within their own schools. This AC Team 2 met regularly to review implementation progress in each of the schools, and to report problems and suggestions from grassroots teachers within the Network schools. The new AC Team structure was very successful in stimulating adoption by teachers in the Network. Many of the improvements such as the desirability of mobile applications as tools for collecting evidence of students’ learning was gathered as teachers’ voice in the innovation process.

## Summary and discussion



A Technology-Enhanced Pedagogical Innovation (TEPI) is a journey and a process. An innovation journey is by definition one in which both the destination and the pathway are not clear. It is also a collaborative problem-solving process involving agents at multiple levels: classroom teachers, teacher leaders, school leaders and external partners such as consultants and technology developers to address inter-related problems that emerge on the way. Network S is a self-organizing network that came together voluntarily to achieve an educational vision through embarking jointly on Project S as a TEPI. To achieve success, the Network has to be able to undertake successful *self-organized learning* in navigating through the many challenges encountered during the journey. In this final section, we will summarize our key learning from this case study of Project S.

First, the TEPI is a process of sociotechnical co-evolution. Over the five years, as illustrated in Table 1, the functions and roles of the e-Learning platform changed alongside the changes in teachers' experimentation with the technology and also with the changes in the architecture for learning set up to implement the project. In Year 1, the focus of the innovation was to develop a platform for resource sharing (e.g., teaching plan, teaching materials, useful e-learning tools, etc.), and the teachers spent a lot of time learning to transform paper-based resources into electronic format, with little impact on classroom practices. In Year 2, the innovation focus shifted to the pedagogical use of ICT in classrooms, and some teachers experimented with using the e-Learning platform in this project (referred to as Platform S) for planning their lessons, including efforts to cater for learner diversity. The availability of discussion forum and peer assessment resources on the platform stimulated the adoption of these activities by some teachers. There were also scattered efforts to design and implement differentiated learning activities based on students' based on the fine-grained SCALE specifications of learning outcomes. In Year 3, the newly constituted AC team designed a well-structured teaching plan e-form that capitalized on the teachers' interests in using the teaching and learning resources to enforce a four-stage model of lesson design. Some teachers who initially doubted the value of e-Learning for their students changed their views after interacting with other teachers and experimenting with new practices in their own classrooms. In Year 4, the focus shifted to using student data to improve their learning, and new mobile applications and interfaces were developed in Platform S to support user-friendly ways of documenting students' performance. These new tools serve as resources for students to reflect on their own learning, and to inform parents about their children's development. In Year 5, sustainability and compatibility of Project S was discussed in the Network schools and affirmed as a priority by most of the member schools. Good practice cases were shared among the network teachers to attract more teacher adoption.

A second observation is the need for a strong, cohesive and respected core at the network leadership level that constantly monitors the innovation direction and re-focuses its efforts once it deviated from the primary vision and goal. As Network S moved from the development of curriculum resources such as SoW and SCALE to the implementation of e-Learning, there was an initial shift towards a techno-centric focus on the development of digital resources for teaching and learning. The international consultant's visit in the second year of the project was instrumental in re-focusing the project team on pedagogical concerns, and the affirmation that professional development of teachers should be the primary strategic implementation goal. The seven years of prior collaboration among the Network leadership and the consultant provided a trusting relationship that underpinned the re-focusing effort and the establishment of the AC Team as a consequence of the consultant's intervention.

A third observation is that the capacity of a network to undertake TEPI has to be built up over time. The two prior USSP projects conducted by Network S fostered crucial innovation capacity for Project S as a pedagogically focused innovation. First, the SoW and SCALE developed served as the curriculum and pedagogical bases for developing the TEPI. Secondly, the Principal Team and Writing Teams that conducted those two projects served as a foundational social infrastructure to lead the implementation of Project S. Thirdly, the working relationships among the Principal Team and Writing Teams, and the successful implementation of these two prior projects helped to lay a good social milieu for Project S.

A fourth observation is the need for a "middle-layer" organizational structure and interaction mechanisms to materialize the innovation vision of the Network leadership through concrete implementation plans that engage classroom teachers "on the ground". In fact, variations in project implementation efficacy across schools in the Network often reflect differences in the strength and suitability of the middle management at the school level.

Fifth, the composition of the Network middle management teams (e.g. the Writing Teams, AC Team 1 and AC Team 2) need to have the necessary expertise and be tasked with the appropriate mission for the specific phase of the innovation, and hence will need to be changed/adjusted as necessary. Such change is also part of the sociotechnical co-evolution.

Network S is an autonomous self-organizing network connecting a group of schools serving SEN students. Project S, led by the Network, went through stages of development reflecting a co-evolution of the curriculum artefacts, organizational infrastructure and interaction mechanisms in bringing about innovations in pedagogical practices and the technology platform.

## References

- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141-178.
- Clarke, J., & Dede, C. (2009). Design for Scalability: A Case Study of the River City Curriculum. *Journal of Science Education and Technology*, 18, 353-365. doi:10.1007/s10956-009-9156-4
- Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, 32(6), 3-12.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design Research: Theoretical and Methodological Issues. *Journal of the Learning Sciences*, 13(1), 15-42.
- Davis, N. (2008). How May Teacher Learning be Promoted for Educational Renewal with IT? In J. Voogt & G. Knezek (Eds.), *International Handbook of Information Technology in Primary and Secondary Education Technology in Primary and Secondary Education* (pp. 507-519): Springer
- EDB (2011). *Selection results of Pilot Scheme on e-Learning in Schools announced*. <http://www.info.gov.hk/gia/general/201101/03/P201012310128.htm>
- Fisher, T. (2006). Educational transformation: Is it, like beauty? in the eye of the beholder, or will we know it when we see it? *Education and Information Technologies*, 11(3), 293-303.
- Fishman, B. J., Penuel, W. R., Allen, A. R., Cheng, B. H., & Sabelli, N. H. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. *National Society for the Study of Education*, 112(2), 136-156.
- Fullan, M., & Hargreaves, A. (Eds.). (2009). *Change Wars*. Bloomington, IN: Solution Tree Press.
- Hamel, C., Turcotte, S., & Laferrière, T. (2013). Evolution of the conditions for successful innovation in remote networked schools. , 6(3), 1. *International Education Studies*, 6(3), 1-14.
- Law, N., Niederhauser, D. S., Christensen, R., & Shear, L. (2016). A Multilevel System of Quality Technology-Enhanced Learning & Teaching Indicators. *Journal of Educational Technology & Society*, 19, 72-83.
- Law, N., Yuen, A., & Fox, B. (2011). *Educational Innovations Beyond Technology: nurturing leadership and establishing learning organizations* New York: Springer.
- Law, N., Yuen, J., & Lee, Y. (2015). *Precarious School Level Scalability Amid Network Level Resilience: Insights from a multilevel multiscale model of scalability*. Paper presented at the American Education Research Association Conference 2015, Chicago.
- Partnership for 21st Century Skills. (2009). *P21 Framework Definitions*. Retrieved from [http://www.p21.org/storage/documents/P21\\_Framework\\_Definitions.pdf](http://www.p21.org/storage/documents/P21_Framework_Definitions.pdf)
- Penuel, W. R. (2015). *Infrastructuring As a Practice for Promoting Transformation and Equity in Design-Based Implementation Research*. Paper presented at the Keynote presented at the International Society for Design and Development in Education (ISDDE) 2015 Conference, Boulder, CO.
- Somekh, B., & Davis, N. (1997). *Using information technology effectively in teaching and learning : studies in pre-service and in-service teacher education*. London ; New York: Routledge.
- Spillane, J. P., Parise, L. M., & Sherer, J. Z. (2011). Organizational Routines as Coupling Mechanisms Policy, School Administration, and the Technical Core. *AERJ*, 48, 586-619.
- Stein, M. K., & Coburn, C. E. (2008). Architectures for learning: A comparative analysis of two urban school districts. *American Journal of Education*, 114(4), 583-626.
- Vuorikari, R., Kämpylis, P., Scimeca, S., & Punie, Y. (2015). Scaling Up Teacher Networks Across and Within European Schools: The Case of eTwinning. In C. K. Looi & L. W. Teh (Eds.), *Scaling Educational Innovations* (pp. 227-254). Singapore: Springer.
- Watson, D. M. (2001). Pedagogy before Technology: Rethinking the Relationship between ICT and Teaching. *Education and Information Technologies*, 6(4), 251-266.
- Wenger, E. (1998). Communities of practice: Learning as a social system. *Systems thinker*, 9(5), 1-5.

## Acknowledgment

We thank the principals, teachers, students and the international consultant Mr. Keith Humphreys in Project S, and the contributions of Dr Yeung Lee (CITE, HKU) during the various stages of this study. This work is funded by General Research Fund grant #17404314 and Humanities and Social Sciences Prestigious Fellowship Scheme (HSSPFS) grant #HKU 37600414 to the first author, both from HK RGC.

# A Meta-Synthesis of CSCL Literature in STEM Education

Jessica McKeown, Indiana University, jeschamb@indiana.edu  
Cindy E. Hmelo-Silver, Indiana University, chemelosi@indiana.edu  
Heisawn Jeong, Hallym University, heis@hallym.ac.kr  
Kylie Hartley, Indiana University, hartley2@indiana.edu  
Roosevelt Faulkner, Indiana University, rtfaulkn@indiana.edu  
Navo Emmanuel, Indiana University, nemmanue@indiana.edu

**Abstract:** This research aims to synthesize the extensive literature on Computer Supported Collaborative Learning (CSCL) in STEM education published between 2005-2014. Our synthesis focuses on the interactions of collaboration, technology, and pedagogies to see how different combinations may contribute to learning. A Latent Class Analysis was used to categorize existing research and points to a six-cluster solution. We have synthesized across and within the three largest clusters to 1) help us identify robust themes in this field and 2) help us better understand the positive outcomes associated with these aspects of CSCL in STEM education. The results suggest that different combinations of technology, pedagogy, and collaboration types require different strategies to scaffold students' learning. This research provides a frame for synthesizing the effects of CSCL in synchronous and asynchronous STEM education and with various technologies and pedagogical designs.

Visions of technology as a social entity are now ubiquitous and tout collaborative learning as a key benefit (Roschelle, 2013). There is extensive research regarding the use of CSCL in science, technology, engineering, and mathematics (STEM) education (Stahl et al., 2014; Jeong, Hmelo-Silver, & Yu, 2014). Many individual studies have reported encouraging results with different types of technologies used in a range of pedagogical approaches, with different forms of collaboration. Roschelle, Bakia, Toyama, and Patton (2011) have argued that we need to understand the “compound resources” at play in complex learning environments. However, there has been little systematic review on the impact of CSCL, especially ones focusing on the interactions of collaboration, technology, and pedagogies used (Jeong & Hmelo-Silver, 2016; Kirschner & Erkens, 2013). To understand the impact of CSCL research, it is important to examine the evidence based on the effectiveness of CSCL with both CSCL and measures of effectiveness broadly defined. Research in CSCL and the learning sciences should be especially well positioned to examine the complexity of these learning environments given the emphasis on mixed methods and design-based research (Jeong, Hmelo-Silver, & Yu, 2014; Roschelle et al., 2011), but as Kirschner and Erkens (2013) point out, we are not there yet. The editors of *ijCSCL* have recently noted that CSCL is becoming a mature field and that we need to understand the landscape of the field (Ludvigsen, Cress, Law, Rosé, & Stahl, 2016).

## Theoretical framework

Our work uses meta-synthesis as a frame for reviewing research, allowing for the integration of research across qualitative and quantitative studies, and combinations of evidence across multiple studies (Suri & Clark, 2009; Bair, 1999). One of the strengths and challenges of this type of analysis is that it compares research that may not have common metrics. Research in CSCL focuses on learning as a cognitive and/or social process and studies learning designs, learning processes, and pedagogic practices that support technology-mediated collaborative processes in communities of practice. CSCL research is guided by a variety of theoretical frameworks that include information processing, social constructivist, sociocultural, social psychology, and communication theories (e.g., Jeong, Hmelo-Silver, & Yu, 2014; O'Donnell & O'Kelly, 1994). This synthesis cuts across theoretical frameworks to focus on how different combinations of technology, approaches to collaboration, and pedagogy contribute to learning, as these are the pillars of CSCL. Technologies that promote learning through collaboration mirror major shifts in education, which characterizes learning as being social and collective rather than individual. The changing pedagogies that support these perspectives and evolving technologies have merged to create many new CSCL opportunities in classrooms (Jeong & Hmelo-Silver, 2016; Miyake, 2007). Jeong and Hmelo-Silver (2016) have recently theorized that technology needs to have particular affordances to support CSCL. These include 1) establishing a joint task, 2) communication, 3) sharing resources, 4) engaging in productive processes, 5) engaging in co-construction, 6) monitoring and regulation, and 7) finding and building groups and communities. These affordances can be realized when the interactions of

technology, pedagogy, and modes of collaboration are considered. Our work considers this complex nature of CSCL and seeks synthesis within these interactions.

## Methods

### Data sources

To guide our systematic review of CSCL literature in STEM domains, we defined CSCL as two or more people using technology to work together toward a shared learning goal, and used this definition while searching and screening papers. We searched through two databases, ERIC and Web of Science, in addition to seven key journals regarded by experts to be major outlets for publishing CSCL research (Jeong, Hmelo-Silver, & Yu, 2014). Over 1,500 qualitative and quantitative papers focusing on various education levels published between 2005-2014 were screened to ensure each paper met the following criteria: (a) STEM education, (b) empirical research. Of the screened papers, 708 papers met our criteria and were then coded for a range of study characteristics; educational level, collaboration type, pedagogy, and technology (e.g., Jeong et al., 2014). Interrater reliability was checked by having two members of the research team independently code 20% of the sample. Kappas were satisfactory ranging from .76-.82. This data was then submitted to an LCA analysis, described next.

### Sampling approach using LCA analysis

To identify and characterize groups of similar cases we used LCA, Latent Class Analysis (Hagenaars & McCutcheon, 2002; Linzer & Lewis, 2011). LCA is a technique to find latent subgroups in the data based on the co-occurrence of variables. LCA takes the coded data and assigns each row a cluster membership by selecting the highest posterior probability of that row across all clusters. To classify the papers into one and only one cluster, the original posterior probabilities are used with the highest probability selected. As the number of clusters is unknown, several models are fit and one is selected per fit indices. Our results show that the Akaike Information Criterion (AIC) as well as the Negative Log Likelihood Ratio (Deviance) pointed at a six-cluster solution, which is shown in Table 1 below. Space precludes additional details regarding the LCA analysis. Clusters were named based on dominant themes in the cluster.

Table 1: Description of the six clusters with overall frequencies, percentage of total articles and sample size.

Clusters	Cluster Description	N (% of total)	# of Articles Sampled
1. Mediated Inquiry with Dynamic Feedback	Collaboration: Mediated Pedagogy: Inquiry and exploration learning Technology: Dynamic tools or miscellaneous tools	246 (34%)	22
2. Teacher-directed Synchronous Collaboration	Collaboration: Synchronous Pedagogy: Teacher-directed Technology: Synchronous communication	74 (10%)	10
3. Synchronous Exploration and Construction	Collaboration: Synchronous Pedagogy: Discussion or inquiry and exploration learning Technology: Synchronous communication	51 (7%)	7
4. Mixed Tools and Pedagogies	Collaboration: Asynchronous or synchronous Pedagogy: Other Technology: Groups and communities	38 (5%)	4
5. Asynchronous Teacher-directed Discussion	Collaboration: Asynchronous Pedagogy: Discussion or teacher-directed Technology: Asynchronous communication	154 (21%)	18
6. Generative Asynchronous Inquiry	Collaboration: Asynchronous or mediated Pedagogy: Inquiry and exploration learning or teacher-directed Technology: Sharing and co-construction	145 (20%)	11

To sample a subset of the paper for more in depth analysis, a stratified random sampling technique was used (Sampath, 2001). A stratified random sampling technique takes the size and variability of the subgroups in the data, in our case, clusters. Because all variables in this analysis are categorical, the Average Deviation Analog (ADA) was employed to compute the variability of the clusters (Wilcox, 1973). The sample size and the ADA's were used to feed into an Optimal Allocation algorithm, which is a class of stratified random sampling techniques that accounts for differences in size and variability among clusters (Sampath, 2001). As Table 1 indicates, clusters 1, 5, and 6 are the most robust in terms of size and multiple iterations of the analysis with different subsets of articles. In this paper, we present the synthesis for these three clusters given (a) they are the most common and (b) space limitations that preclude including the smaller clusters here.

## **Synthesis findings**

We identified the dominant themes that emerged in each cluster. In this process, we took note of the educational levels of the studies in each cluster as they can be an important moderator of CSCL.

### **Cluster 1: Mediated inquiry with dynamic feedback**

This cluster (sample  $n = 22$ ) emphasizes face-to-face mediated collaboration with inquiry and exploration pedagogies using dynamic technological tools (e.g. tools that provide a response to user actions) such as immersive technology, games, and simulations. This cluster also represents papers using a variety of miscellaneous technological tools such as mobile devices, open-source software to conduct remote controlled experiments, and digital game design software. More than half of papers in this cluster were coded as K-12 (59%), which was highest percentage for K-12 education among the six clusters, suggesting this combination of pedagogy and technology is more common in K-12 than combinations in other clusters. Sampled studies reported that learning under these conditions led to significant learning gains, positive student engagement, meaningful interactions between students, and improved group collaboration and communication skills.

What may have led to such significant learning gains and positive engagement? One overarching theme in this cluster is guided instruction, with varying levels of guidance ranging from open-ended to highly structured suggestions (Avraamidou, 2013; Chiang, Yang, Hwang, 2014; Jaakkola & Nurmi, 2008; Kong et al., 2009; Kuo et al., 2012; Loke et al., 2012; Santos-Martin et al., 2012; Lai & White, 2012; Tsa et al., 2012; Yang, & Chang, 2013). Closely related to this theme of guided instruction is immediate feedback and discussion; participants at a variety of educational levels received immediate feedback from facilitators (Avraamidou, 2013; Kong et al., 2009; Lantz, & Stawiski, 2014; Santos-Martin et al., 2012), peers (Chen, et al., 2012; Chiang et al., 2014; Gallardo-Virgen, & DeVillar, 2011; Kuo, Hwang, & Lee, 2012; Lai & White, 2012), and/or software (Chen, et al., 2012; Dzikovska, et al., 2014; Echeverría et al., 2012; Holmes, 2007; Roschelle et al., 2010; Loke, et al., 2012; Nelson & Ketelhut, 2008). For example, as students engaged in problem-based learning (PBL), facilitators provided assistance and feedback throughout the learning processes (Avraamidou, 2013; Santos-Martin, et al., 2012).

Working with authentic problems and materials was also common in this cluster (Avraamidou, 2013; Chiang et al., 2014; Loke, et al., 2012; Tsai, et al., 2012). Using real materials along with a simulation tool also proved to be an effective way to use CSCL for improved learning and positive student engagement (Jaakkola & Nurmi, 2008; Kong, Yeung, & Wu, 2009; Santos-Martin et al., 2012). In one instance, students using virtual simulation tools were guided to explore their own ideas and test them against models (Li et al., 2006), thus encouraging learning through mistakes while minimizing risks. Simulation tools also afforded students more opportunities to conduct experiments compared with using real apparatus (Kong et al., 2009). Similarly, augmented reality games allowed students to play with concepts and ideas (Echeverría et al., 2012). Together, simulated tools or augmented reality games allowed students opportunities to practice and redesign without great demand on time, money, or physical tools. In sum, learning with authentic problems was supported by guided instruction and immediate feedback from the tools and discussion in this cluster. The role of technology was in helping students to work in more authentic settings and have opportunities to directly test their ideas and solutions, with the tools providing dynamic feedback.

### **Cluster 5: Asynchronous teacher-directed discussion**

The Asynchronous Teacher-Directed Discussion cluster (sample  $n = 18$ ) represents papers emphasizing asynchronous collaboration, discussion, or teacher-directed pedagogies and asynchronous communication technology such as discussion boards, a knowledge forum, or email. Many of these sampled papers included a variety of types of collaboration, pedagogies, and/or technologies in addition to those that are emphasized in the cluster description shown in Table 1. Most papers in this cluster were coded as higher education (70%),

suggesting that this combination of CSCL is quite common at that level. Most papers within this cluster focused on trying out new learning management systems and collecting student feedback, not learning outcomes.

Articles within this cluster primarily focus on online learning environments and the characteristics of these environments that promote learning, thus justifying the sense that researchers and educators alike seek to better understand the technology that best supports this type of learning environment. For example, papers in this cluster studied new learning management systems (Lopez-Morteo & Lopez, 2007; Yang & Liu, 2007), assessment for distance learning users (van Aalst & Chan 2007), student satisfaction with a virtual environment (So & Brush 2008), or with specific communication tools (Overbaugh & Casiello, 2008). Within this cluster, researchers studied how learners participated and interacted within an online discussion forum (Swigger, Hoyt, Serçe, Lopez, & Alpaslan, 2012; Vercellone-Smith, Jablow, & Friedel, 2012; Manca, Delfino, & Mazzoni, 2009; van Aalst, 2009), how assigning group roles or tutors affected student learning within an online platform (De Wever et al., 2010), and how group size impacted participation and learning a new programming language (Shaw, 2013). Other papers focused on learning gains of users in an asynchronous learning environment (Lakkala, Lallimo, & Hakkarainen, 2005), and building skills necessary for collaborative engagement in an asynchronous course (Yukawa, 2006; Žavbi, & Tavčar, 2005). Discussion forums have an important role in these kinds of online environments; they provide a space for reflection in which students interpret and talk about the problems and tasks at hand, but also space for some personal contacts. One of issues that emerged from our synthesis of this cluster is the “single pass” strategy that students may use in discussion forum assignments. Students often made contributions to a discussion topic, but moved on from the topic before a conclusion was made (Hewitt, 2005). Such strategies can cause premature closure or death of threads before the issues have been fully explored and efforts are needed to address students’ use of such strategy.

Our synthesis also indicated that asynchronous collaboration seems to be most effective when collaboration and communication are flexible (e.g. students are able to use both synchronous and asynchronous communication tools, or students can choose which modes of communication and/or collaboration to use; Yang & Liu, 2007; Lopez-Morteo & Lopez, 2007; Yukawa, 2006), when there is embedded support and scaffolds (Lakkala, Lallimo, & Hakkarainen, 2005; Martínez, Del Bosch, Herrero, & Nuño, 2007), when roles are assigned or chosen (De Wever et al., 2010; Lin, Lin, & Huang, 2008), and when good communication practices are established within a group (So & Brush, 2008; Swigger et al., 2012). In sum, synthesis of research in this cluster suggests that learning in online environments requires attention to several issues such as group size, available communication channels, and communication practices and strategies with the platforms. As in cluster 1, teacher guidance is important, particularly in providing feedback and in supporting desired communication practices.

### **Cluster 6: Generative asynchronous inquiry**

This cluster represents papers emphasizing asynchronous and mediated collaboration with inquiry and exploration or teacher-directed pedagogies that use sharing and co-construction or integrated environmental technologies (sample  $n=11$ ). Papers in this cluster were coded as 26% K-12 and 69% higher education. Such inquiry and exploration pedagogies used within this sample include project-based learning (Pifarre & Cobos, 2010) and case-based learning (Hämäläinen & Arvaja, 2009; Vicari, Flores, Seixas, Gluz, & Coelho, 2008). Meanwhile, teacher-directed pedagogies include problem solving (Krause, Stark, & Mandl, 2009), blended learning (Goktas, & Demirel, 2012; Hämäläinen, & Arvaja, 2009; Neumann, & Hood, 2009), and distance learning (Inayat, Amin, Inayat, & Salim, 2013). Knowledge building is a common pedagogy guiding instruction among these sampled papers, whether in addition to using an inquiry and exploration pedagogy like PBL (Pifarre, & Cobos, 2010). or simply using knowledge building pedagogy on its own (Huang, & Nakazawa, 2010; Marée, van Bruggen, & Jochems, 2013). The technology used in knowledge building is very much co-constructive in this sample. Such sharing and co-construction technologies used among these sampled papers include participatory technology (Goktas, & Demirel, 2012; Huang, & Nakazawa, 2010; Neumann & Hood, 2009) and representational tools (Marée, van Bruggen, & Jochems, 2013). By their nature, integrated environments offer instructors and students a variety of tools that can be used asynchronously or in mediated face-to-face environments. Therefore, it is no surprise to see this technology used in collaboration spaces. Among these sampled papers using integrated environments, we see one emphasized asynchronous collaboration (Pifarre & Cobos, 2010), one emphasized mediated collaboration (Krause, Stark, & Mandl, 2009), and one that incorporated both types of collaboration (Inayat, Amin, Inayat, & Salim, 2013). Other sampled papers include those that used case-based pedagogies, but with other technology such as intelligent tutoring (Vicari, Flores, Seixas, Gluz, & Coelho, 2008) or other software (Hämäläinen & Arvaja, 2009).

In another example from this cluster using a wiki co-construction environment, students reported more interaction with peers than their instructor, indicating that this technology can help instructors move into a

facilitator role (Huang & Nakazawa, 2010). In doing so though, students noted how it is important that when the instructor/facilitator does want to communicate with students about revising their work, they do so in the co-constructive environment. Meanwhile, a major conclusion of Marée, van Bruggen, and Jochems (2013) is the possibility for undergraduate science students to learn more with less guidance from teachers by using multimedia enriched concept maps with built-in instructions for collaboration. Second, this sample of CSCL literature also offers some promising implications about specific technologies and pedagogical practices. When students participate in asynchronous discussion, particularly when they engage as student-facilitators, it is important that they understand the different types of thread patterns and how questioning, summarizing, pointing and resolving may affect discussion thread development and closure (Chan et al., 2009). Pedagogically, in ICT courses, it is important to integrate the technology being discussed so participants better understand not only its purpose, but also how to use it from a first-hand perspective (Goktas & Demirel, 2012). Finally, Krause et al., (2009) support the notion that feedback may promote more reflection, especially when it offers explanations that encourages deeper understanding, therefore, feedback whether that be from software or the instructor should be thoughtful and thorough and encourage students to think beyond remembering information.

Like cluster 5, many of these sampled papers investigated how students used and perceived specific technology. The conclusions that emerged from these sampled papers tended to emphasize perceptions of co-construction or integrated environmental technologies in the classroom rather than learning outcomes. These conclusions suggest that the use of collaborative group activities, instructors' timely feedback, and support materials embedded within an integrated system all related to student satisfaction with a variety of STEM related vocational e-learning courses, (Inayat, et al., 2013). The articles that did focus on student outcomes suggest that we can draw conclusions about technology supports for learning that encourages specific positive outcomes such as improved metacognitive skills (Pifarre & Cobos, 2010) and positive student engagement and classroom attendance (Neumann & Hood, 2009). Similar to the finding in cluster 1, when guided instruction and immediate feedback are integrated within these specific pedagogies and technologies, it can lead to improved student learning (Krause, Stark, & Mandl 2009; Marée, van Bruggen, & Jochems, 2013) and task completion (Hämäläinen, & Arvaja, 2009).

Results, however, indicated that there is a delicate balance between too much and not enough feedback or guidance. Consistent with the findings in other clusters, a lack of feedback can negatively affect students' learning outcomes (Krause, Stark, & Mandl 2009). Meanwhile, too much feedback can lead to discussion thread premature closure (Chan, Hew, & Cheung, 2009). Without enough guidance or clarity regarding the importance of positive collaboration, students may have high task activity, but not necessarily good collaboration (Hämäläinen, & Arvaja, 2009). In sum, student perception of asynchronous inquiry CSCL is overall positive. Timely guidance from teachers also play an important role in increasingly student learning outcomes as well as favorable perception of the environment, but the results also highlight the importance of keeping the guidance at an optimal level.

## Discussion

Through a synthesis effort, this research has attempted to fill a gap in the CSCL literature. By focusing on the relationships among technologies, pedagogies, and types of collaboration in these complex learning environments, we address the concerns raised by Roschelle et al. (2011) and Kirschner and Erkens (2013). The results show how these facets of CSCL interact as well as themes that cut across educational levels.

Using the combination of face-to-face mediated collaboration, inquiry and exploration pedagogies with dynamic or other technologies allow teachers to move lessons beyond the traditional classroom into other learning spaces that may provide more situated learning experiences where students are still engaged in face-to-face communication. This type of instruction may lead to learning gains, positive engagement, and promote meaningful interactions between students across STEM domains and education levels; however, it does appear to be more common in K-12 educational research. While these technologies have the potential to make learning meaningful, relevant, and interesting, there are some things to consider before employing this type of CSCL in the STEM classroom. CSCL researchers and teachers need to design guided instruction that allows the student to explore the content and receive positive and regular feedback, whether through the technology, in person directly, or through student peers. Another consideration is that CSCL environments such as these can be very resource intensive; learning spaces need to have access to all necessary technologies as well as space for face-to-face mediated collaboration that encourage inquiry and exploration pedagogies.

The use of asynchronous technologies and collaboration methods along with the teacher-directed and discussion pedagogies allows teachers and students to learn and interact with more flexibility surrounding when, where, and how they elect to participate; something that seems to be more common in higher-education perhaps simply because this flexibility is not afforded to many K-12 learners. This type of learning may lead to learning

gains and help students build skills necessary to successfully engage cooperatively in an asynchronous course. However, to be an effective learner in distance education using asynchronous collaboration, learners need to be capable of self-regulating, something that may be more challenging for younger learners. Another issue, particularly in graded discussion forums, is students moving on from a conversation before a conclusion is made. This ultimately leads to difficult questions being passed over and might contribute to poor integration across course topics. Using CSCL effectively requires considering many different combinations of technology, pedagogy, and ways of collaborating. As such, professional development is critical for supporting teachers on strategies to integrate these components of CSCL for effective classrooms that promote positive student outcomes.

Finally, the literature reviewed offers some promising implications for a broad range of combinations of technologies, pedagogies, and types of collaboration. First, it suggests how the interaction of these technologies with inquiry and exploration pedagogies can help support instructors in their move toward a facilitator role and thus encourage students to be active and constructive among themselves. Second, this literature suggests that asynchronous pedagogy needs to be scripted or have some type of guidance embedded for students to stay engaged and informed. Third, instructors take on a facilitator role in most of these learning environments, which may require some training regarding knowing when to guide students toward resources or lets students grapple with ideas they are struggling with. More generally, across these clusters, opportunities for learner agency are important, whether working on authentic tasks or in having opportunities to choose communication/collaboration modes. Similarly, support and guidance for collaboration appear to be a necessary condition for CSCL effectiveness across all six clusters, although the exact forms of guidance are likely to differ by the age of the learners and with online compared with face-to-face settings.

To further explicate under what conditions technology may support learning in STEM education CSCL, we are conducting additional analyses to identify relationships between collaboration, technology, and recently proposed seven affordances of CSCL technology for learning (Jeong & Hmelo-Silver, 2016). We believe this will provide a more nuanced look at the mechanics behind each identified cluster. The next step to come in this research in terms of research and practice would be to carefully detail how this meta-synthesis can directly impact classroom practices.

## References

- Avraamidou, L. (2013). The use of mobile technologies in project-based science: A case study. *Journal of Computers and Mathematics and Science Teaching, 32*(4), 361-379.
- Bair, C. R. (1999, November). *Meta-synthesis*. Paper presented at Annual Meeting of the Association for the Study of Higher Education. San Antonio, TX.
- Chen, Y. H., Looi, C. K., Lin, C. P., Shao, Y. J., & Chan, T. W. (2012). Utilizing a collaborative cross number puzzle game to develop the computing ability of addition and subtraction. *Educational Technology & Society, 15*(1), 354-366.
- Chan, J. C. C., Hew, K. F., & Cheung, W. S. (2009). Asynchronous online discussion thread development: examining growth patterns and peer-facilitation techniques. *Journal of Computer Assisted Learning, 25*(5), 438-452.
- Chiang, T. H.C., Yang, S. J.H., Hwang, G. (2014). Students' online interactive patterns in augmented reality-based inquiry activities. *Computers & Education, 78*, 97-108.
- De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2010). Structuring asynchronous discussion groups: Comparing scripting by assigning roles with regulation by cross-age peer tutors. *Learning and Instruction, 20*(5), 349-360.
- Dzikovska, M., Steinhauer, N., Farrow, E., Moore, J., & Campbell, G. (2014). BEETLE II: Deep natural language understanding and automatic feedback generation for intelligent tutoring in basic electricity and electronics. *International Journal of Artificial Intelligence in Education, 24*(3), 284-332. DOI 10.1007/s40593-014-0017-9
- Echeverría, A., Améstica, M., Gil, F., Nussbaum, M., Barrios, E., & Leclerc, S. (2012). Exploring different technological platforms for supporting co-located collaborative games in the classroom. *Computers in Human Behavior, 28*(4), 1170-1177.
- Gallardo-Virgen, J. A., & DeVillar, R. A. (2011). Sharing, talking, and learning in the elementary school science classroom: Benefits of innovative design and collaborative learning in computer-integrated settings. *Computers in the Schools, 28*(4), 278-290.
- Goktas, Y., & Demirel, T. (2012). Blog-enhanced ICT courses: Examining their effects on prospective teachers' ICT competencies and perceptions. *Computers & Education, 58*(3), 908-917.



- Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis*. New York: Cambridge University Press.
- Hämäläinen, R., & Arvaja, M. (2009). Scripted collaboration and group-based variations in a higher education CSCL context. *Scandinavian Journal of Educational Research*, 53(1), 1-16.
- Holmes, J. (2007). Designing agents to support learning by explaining. *Computers & Education*, 48(4), 523-547.
- Huang, W. H. D., & Nakazawa, K. (2010). An empirical analysis on how learners interact in wiki in a graduate level online course. *Interactive Learning Environments*, 18(3), 233-244.
- Inayat, I., ul Amin, R., Inayat, Z., & Salim, S. S. (2013). Effects of collaborative web based vocational education and training (VET) on learning outcomes. *Computers & Education*, 68, 153-166.
- Jaakkola, T., & Nurmi, S. (2008). Fostering elementary school students' understanding of simple electricity by combining simulation and laboratory activities. *Journal of Computer Assisted Learning*, 24(4), 271-283.
- Jeong, H., & Hmelo-Silver, C. E., & Yu, Y. (2014). An examination of CSCL methodological practices and the influence of theoretical frameworks 2005-2009. *International Journal of Computer-Supported Collaborative Learning*, 9(3), 305-334.
- Jeong, H., & Hmelo-Silver, C. E. (2016) Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help?, *Educational Psychologist*, 51:2, 247-265, DOI: 10.1080/00461520.2016.1158654
- Kirschner, P. A., & Erkens, G. (2013). Toward a framework for CSCL research. *Educational Psychologist*, 48(1), 1-8.
- Kong, S. C., Yeung, Y. Y., & Wu, X. Q. (2009). An experience of teaching for learning by observation: Remote-controlled experiments on electrical circuits. *Computers & Education*, 52(3), 702-717.
- Krause, U. M., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction*, 19(2), 158-170.
- Kuo, F. R., Hwang, G. J., & Lee, C. C. (2012). A hybrid approach to promoting students' web-based problem-solving competence and learning attitude. *Computers & Education*, 58(1), 351-364.
- Lai, K., & White, T. (2012). Exploring quadrilaterals in a small group computing environment. *Computers & Education*, 59(3), 963-973.
- Lakkala, M., Lallimo, J., & Hakkarainen, K. (2005). Teachers' pedagogical designs for technology-supported collective inquiry: A national case study. *Computers & Education*, 45(3), 337-356.
- Lantz, M. E., & Stawiski, A. (2014). Effectiveness of clickers: Effect of feedback and the timing of questions on learning. *Computers in Human Behavior*, 31, 280-286.
- Li, S. C., Law, N., & Lui, K. F. A. (2006). Cognitive perturbation through dynamic modelling: A pedagogical approach to conceptual change in science. *Journal of Computer Assisted Learning*, 22(6), 405-422.
- Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, 42(10), 1-29.
- Loke, S. K., Al-Sallami, H. S., Wright, D. F., McDonald, J., Jadhav, S., & Duffull, S. B. (2012). Challenges in integrating a complex systems computer simulation in class: An educational design research. *Australasian Journal of Educational Technology*, 28(4).
- Lopez-Morteo, G., & López, G. (2007). Computer support for learning mathematics: A learning environment based on recreational learning objects. *Computers & Education*, 48(4), 618-641.
- Ludvigsen, S., Cress, U., Law, N., Rosé, C. P., & Stahl, G. (2016). Future-looking conversations in CSCL. [journal article]. *International Journal of Computer-Supported Collaborative Learning*, 11(3), 255-262. doi: 10.1007/s11412-016-9242-6
- Manca, S., Delfino, M., & Mazzoni, E. (2009). Coding procedures to analyse interaction patterns in educational web forums. *Journal of Computer Assisted Learning*, 25(2), 189-200.
- Marée, T. J., van Bruggen, J. M., & Jochems, W. M. (2013). Effective self-regulated science learning through multimedia-enriched skeleton concept maps. *Research in Science & Technological Education*, 31(1), 16-30.
- Martínez, R. A., Del Bosch, M. M., Herrero, M. H. P., & Nuño, A. S. (2007). Psychopedagogical components and processes in e-learning. Lessons from an unsuccessful on-line course. *Computers in Human Behavior*, 23(1), 146-161.
- Miyake, N. (2007). Computer supported collaborative learning. In R. Andrews & C. Haythornwaite (Eds.), *Sage Handbook of E-learning Research* (pp. 248-266). London UK: Sage.
- Nelson, B. C., & Ketelhut, D. J. (2008). Exploring embedded guidance and self-efficacy in educational multi-user virtual environments. *International Journal of Computer-Supported Collaborative Learning*, 3(4), 413-427.

- Neumann, D. L., & Hood, M. (2009). The effects of using a wiki on student engagement and learning of report writing skills in a university statistics course. *Australasian Journal of Educational Technology*, 25(3), 382-398.
- Overbaugh, R. C., & Casiello, A. R. (2008). Distributed collaborative problem-based graduate-level learning: Students' perspectives on communication tool selection and efficacy. *Computers in Human Behavior*, 24(2), 497-515.
- Pifarre, M., & Cobos, R. (2010). Promoting metacognitive skills through peer scaffolding in a CSCL environment. *International Journal of Computer-Supported Collaborative Learning*, 5(2), 237-253.
- Roschelle, J. (2013). Special Issue on CSCL: Discussion. *Educational Psychologist*, 48(1), 67-70.
- Roschelle, J., Bakia, M., Toyama, Y., & Patton, C. (2011). Eight issues for learning scientists about education and the economy. *Journal of the Learning Sciences*, 20, 3-49.
- Roschelle, J., Rafanan, K., Estrella, G., Nussbaum, M., & Claro, S. (2010). From handheld collaborative tool to effective classroom module: Embedding CSCL in a broader design framework. *Computers & Education*, 55(3), 1018-1026. doi:10.1016/j.compedu.2010.04.012
- Sampath, S. (2001). Sampling theory and methods. New Delhi, India: CRC press.
- Santos-Martin, D., Alonso-Martínez, J., Carrasco, J. E. G., & Arnaltes, S. (2012). Problem-based learning in wind energy using virtual and real setups. *Education, IEEE Transactions on*, 55(1), 126-134.
- Shaw, R. S. (2013). The relationships among group size, participation, and performance of programming language learning supported with online forums. *Computers and Education*, 62, 196-207.
- So, H. J., & Brush, T. A. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education*, 51(1), 318-336.
- Stahl, G., Koschmann, T., & Suthers, D. D. (2014). Computer-supported collaborative learning. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (2nd ed., pp. pp. 479-500). NY: Cambridge University Press.
- Strijbos, J.W., Martens, R.L., & Jochems, W.M.G. (2004). Designing for interaction: Six steps to designing computer-supported group based learning. *Computers & Education*, 42, 403-424.
- Suri, H., & Clarke, D. (2009). Advancements in research synthesis methods: From a methodologically inclusive perspective. *Review of Educational Research*, 79, 395-430.
- Swigger, K., Hoyt, M., Serçe, F. C., Lopez, V., & Alpaslan, F. N. (2012). The temporal communication behaviors of global software development student teams. *Computers in Human Behavior*, 28(2), 384-392.
- Tsai, P. S., Tsai, C. C., & Hwang, G. J. (2012). Developing a survey for assessing preferences in constructivist context aware ubiquitous learning environments. *Journal of computer assisted learning*, 28(3), 250-264.
- van Aalst, J., & Chan, C. K. (2007). Student-directed assessment of knowledge building using electronic portfolios. *The Journal of the Learning Sciences*, 16(2), 175-220.
- van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning*, 4(3), 259-287.
- Vercellone-Smith, P., Jablowski, K., & Friedel, C. (2012). Characterizing communication networks in a web-based classroom: Cognitive styles and linguistic behavior of self-organizing groups in online discussions. *Computers and Education*, 59(2), 222-235. <http://doi.org/10.1016/j.compedu.2012.01.006>
- Wilcox, A. R. (1973). Indices of qualitative variation and political measurement. *The Western Political Quarterly*, 26(2), 325-343.
- Yang, Z., & Liu, Q. (2007). Research and development of web-based virtual online classroom. *Computers & Education*, 48(2), 171-184.
- Yang, Y. T. C., & Chang, C. H. (2013). Empowering students through digital game authorship: Enhancing concentration, critical thinking, and academic achievement. *Computers & Education*, 68, 334-344.
- Yukawa, J. (2006). Co-reflection in online learning: Collaborative critical thinking as narrative. *International Journal of Computer-Supported Collaborative Learning*, 1(2), 203-228.
- Žavbi, R., & Tavčar, J. (2005). Preparing undergraduate students for work in virtual product development teams. *Computers & Education*, 44(4), 357-376.

## Acknowledgements

This work was funded by National Science Foundation Grant #DRL-1439227. Conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank Alejandro Andrade for his able assistance with Latent Class Analysis.

# Exploring Computational Modeling Environments as Tools to Structure Classroom-Level Knowledge Building

Michelle Wilkerson, University of California, Berkeley; mwilkers@berkeley.edu

Becca Shareff, University of California, Berkeley; rlshareff@berkeley.edu

Brian Gravel, Tufts University, brian.gravel@tufts.edu

Yara Shaban, Tufts University, yara.shaban@tufts.edu

Vasiliki Laina, University of California, Berkeley; vlaina@berkeley.edu

**Abstract.** Although computational modeling is noted as a powerful way to engage students in scientific knowledge construction, many studies focus on individuals or small groups. Here, we explore computational modeling as an infrastructure to support classroom level knowledge building. We present data from a two-week study where two fifth grade classrooms modeled evaporation and condensation. We focus our analysis on one group that experienced success with the activity, and another that struggled; these groups' intended models emphasized random motion and aggregation respectively, two important but complementary molecular behaviors. Both groups' ideas were incorporated into a collective model designed in consultation with the entire class. We show that computational modeling (1) often required explicit support, but when leveraged productively (2) served a *representational* role by supporting the elaboration of student ideas about physical mechanism, and (3) served an *epistemic* role by allowing students to compare, synthesize, and build on other's contributions.

**Keywords:** scientific modeling; classroom discourse; science education; computational modeling; knowledge building; epistemic games

In this paper, we are concerned with *computational modeling*. We define this to be the practice of iteratively constructing, refining, and thinking with representations of scientific systems that are encoded as computer-executable code. By engaging in these practices, learners are expected to discover the inner workings of scientific and mathematical systems: First elaborating their understandings of a given system through constructing a computer model, then “debugging” that knowledge by testing and refining the model (Papert, 1980; Penner, 2000). With proper facilitation and support, computational modeling is generally understood to be an effective way to engage learners in model-based inquiry and knowledge construction (van Joolingen, de Jong, & Dimitrakopoulou, 2007; Wilensky & Reisman, 2006). Here, we argue that it can also be transformative at the *classroom* level by providing a framework for students to collectively contribute, evaluate, and synthesize scientific ideas through the production and refinement of shared playable artifacts. In this way, computational model construction environments can serve as analogs to the types of argument and explanation focused knowledge-organizing tools often used to support classroom learning communities (Scardamalia & Bereiter, 1994), though considerably different in structure and focus.

To illustrate, we present a multiple case study from a computational modeling unit about condensation and evaporation enacted in two fifth grade science classrooms at an urban-rim public school in the Northeastern United States. We focus on two groups within one classroom that are representative of student experiences during the unit. As we will show, these two groups developed models drawing from different experiences and explanations of condensation and evaporation, and their models foregrounded complementary scientific mechanisms. One group was quickly successful in building a working model, while the other struggled to do so. The scientific mechanisms illustrated by these two groups' models, along with contributions from other groups in the classroom, were consolidated into a collective model (constructed by facilitators in consultation with students) that offered more explanatory power than any one group's model alone. This collective model was then taken up by the classroom to reason about other, related science phenomena. We argue that for this classroom, computational modeling (1) often required explicit support, but when leveraged productively (2) served a *representational* role by supporting the elaboration of student ideas about physical mechanism, and (3) served an *epistemic* role by allowing the class to compare, synthesize, and build on each other's contributions.

## Background

There is a growing body of work focused on computational model construction as a learning activity (for a recent review, see VanLehn, 2013). This work has shown that with proper support and facilitation, engaging in model construction can help students make sense of complex scientific phenomena, and productively engage in model-based inquiry. However, the focus of this research has typically been on the conceptual learning of

individual students, or on practices and discussions that emerge among pairs or small groups of students as they model. For example in a recent study, Xiang and Passmore explored students' engagement in cycles of model-based inquiry supported by ABM programming while working in pairs (Xiang & Passmore, 2015). This focus on individual or pair model construction persists across research group and modeling environment (e.g. Löhner, van Joolingen, Savelsbergh, & Hout-Wolters, 2005; VanLehn, Wetzel, Grover, & van de Sande, 2016 to mention a few). Similarly, studies that explore *scaffolding* to support model construction activities often focus on students working "separately and individually" (Basu, Sengupta, & Biswas, 2015, p. 307) or in pairs (Fretz et al., 2002). Our own explorations of model construction activities thus far have similarly focused on discussion of specific scientific content among small groups (Wilkerson-Jerde, Gravel, & Macrander, 2015).

Work that has investigated the role of computational models and simulation at the *classroom* level, on the other hand, typically does not focus on the computational model itself as a site for knowledge construction in the same way as studies focused on content learning or modeling practice. Instead, they focus on providing classroom communities with *access* to computational models, to serve as fodder for discussion or other collective activities. One line of work has explored the development of public galleries or immersive environments where students can interact with, organize, and create of computational data and artifacts (Van Joolingen, De Jong, Lazonder, Savelsbergh, & Manlove, 2005). Another line of work has investigated how teachers can orchestrate productive whole-group discussion, argumentation, and sense-making using pre-constructed simulations as representational tools to inform and test theories about scientific phenomena (Berland & Reiser, 2011; Hmelo-Silver, Liu, Gray, & Jordan, 2015). Yet others investigate the relationship between individual or small group work with computational tools, and collaborative engagement in discussion, juxtaposition, and sharing of resulting artifacts (Hegeudus & Moreno-Armella, 2009).

The work reported in this paper lies at the intersection of these two literatures. It explores whether or how computational model construction activities themselves can serve as a site for community knowledge building. Looking across several projects focused on knowledge representations to support learners' engagement in the epistemic practices of science, Sandoval and colleagues (2000) separated projects into two classes. One class of knowledge representations, which they called "epistemic representations", focused on the construction and organization of student arguments and explanations. The other, which they called "discipline-specific models", focused on engaging learners with conceptual aspects of a discipline (for example, by interacting with or constructing their own representations of content in probability or chemistry). They argued that these two classes were distinct, but suggested that "[a]n interesting line of research... could be to consider how [discipline-specific models] might communicate epistemological ideas more explicitly" (p. 39). We conjecture that discipline-specific representations and other representational infrastructures not specially designed to make argumentation explicit *can* nevertheless serve as a site for students to co-construct scientific knowledge and learn productive epistemic moves in the process, with proper scaffolding and facilitation.

## Theoretical framework

To investigate our conjecture, we leverage Collins & Ferguson's (1993) theory of *epistemic forms* and *epistemic games*. Epistemic forms are representational structures that can be populated by practitioners to organize, reflect upon, and expand their knowledge—such as lists, tables, or graphs. Epistemic games are the ways of thinking that allow them to effectively populate and make use of those forms—for example, reasoning about what might be reasonable axes on a graph, or recognizing and developing methods to fill in missing data in a table. Epistemic forms and games are cultural conventions that are shared by communities of practice, and particular forms are well-suited to answer particular questions. A time series graph, for instance, can make evident temporal and covariational relationships that might otherwise be difficult to identify.

Collins and Ferguson illustrate the notion of epistemic forms and games by invoking the periodic table. The table was an effort to organize an as-of-yet unstructured collection of elements whose physical and chemical properties were difficult to understand. It revealed regularities within spatially proximate groups of elements, and empty spaces in the table predicted the existence of other undiscovered elements. As additional discoveries were made, scientists modified the table, eventually recognizing its relationship to the electron shell structure of atoms. This example highlights four key aspects of epistemic forms: They can be populated with what 'players' know now; they make evident what players may need to know and investigate; they can be modified and contributed to by other players; and they reveal links to other representations and domains.

Bielaczyc and colleagues (Bielaczyc & Ow, 2014) further describe *multiplayer epistemic games*. These are games in which epistemic moves are distributed among a community, such the epistemic form serves as an infrastructure to bring differential expertise together. They argue that a major part of building communities of learners is to explicitly scaffold their shared participation in epistemic games, and that knowledge emerges at the level of the whole from interactions among players. Classrooms and infrastructures designed to support multi-

player epistemic games emphasize student work as a driver for curricular content, enable epistemic games to be played as a collective, and support epistemic moves and games through explicit pedagogical moves.

Here, we posit that computational modeling environments can serve as epistemic forms to support multi-player epistemic games in the classroom. Specifically, we argue that they can support the comparing, testing, and synthesis of one another's proposals regarding what physical mechanisms underlie certain scientific events (such as evaporation or condensation, in our case). The elaboration and comparison of proposed mechanisms using computational models is commonplace in scientific practice (Chandrasekharan & Nersessian, 2015; Grimm et al., 2005). And, agent-based models have been identified as a "model type" or scientific tool alongside concept maps, tables, lists, and mathematical formulae (NGSS, 2013; White, Collins, & Frederiksen, 2011). Therefore, the question that motivates this paper is: **What is the potential for computational modeling environments to support knowledge building at the classroom level?**

## Study context

The data we present are drawn from a design-based research study conducted with two classrooms at a public, urban-rim K-5 school in the Northeastern United States. The school serves a population of students with a variety of identified racial/ethnic, economic, and special needs backgrounds, reflected in the classrooms we worked with. However, the gender balance of our focal classroom in this paper deserves comment. Only 5 of the 15 students who participated in the study identified as girls (one student did not consent to data collection); students we focus on in this paper all identified as boys. Given that the research was conducted in a public classroom during the school day, this imbalance is reflective of naturally occurring differences in student populations and does not reflect a self-selection or participation bias in the research or activities. We focus our analysis on two groups from one class (Figure 1). We video recorded all whole-class and student group interactions, screen captured students' interactions with computers, and collected all written work for analysis.



**Figure 1.** Disperse Group (left), Water Cycle Group (center). A projector was positioned in the front of the classroom (right), and was used to show animations and simulations for whole-group discussion.

The two-week enactment involved two modeling activities that are loosely adapted from activities developed for the IQWST and MoDeLS projects (Schwarz, Reiser, Acher, Kenyon, & Fortus, 2012), and connected to important science standards (Building and Using Models; Matter and Its Interactions; NGSS, 2013). In the first week, we introduced students to the launching question "Why does a cold bottle of soda become wet on the outside?". On the first day, students discussed the question as a class, and then created drawings that illustrated their ideas using templates (Figure 2). On the second, they created animations using craft materials and critiqued others' productions. On the third day they created simulations, and on the fourth they viewed and discussed simulations as a group. The second week followed a similar sequence of activities around the question "What happens to puddles on a sunny day?".

As part of the SiMSAM project, we have developed a web-based animation and simulation application that is used as the computational modeling environment during activities. Students can create stop-action movies using craft materials or drawings. Once a movie is created, they can crop objects from frames of the movie to become programmable "sprites". They then drag the "sprites" onto a simulation interface and use programming-by-demonstration and menu options to define behaviors like interaction, duplication, or random motion.

The two groups we choose to focus on for the purposes of this paper represent cases drawn from a larger multiple case study (Stake, 2006). The quintain, or central shared phenomena, that the cases are chosen to shed light on concerns the ways in which students' group artifacts contribute, conceptually or materially, to an eventual collective product that is endorsed by the class as a community. We choose these two groups because they offer a contrastive illustration of this central question. The groups both constructed simulations that reflected different experiential knowledge resources – related to steam and weather – and that eventually contributed very different but equally essential elements to the shared classroom model. Second, they experienced varying levels of success working with the modeling environment. Together, the focal groups reflect two extremes of a diversity of approaches across the class as a whole, in terms of knowledge leveraged, focal mechanisms expressed, and levels of success.

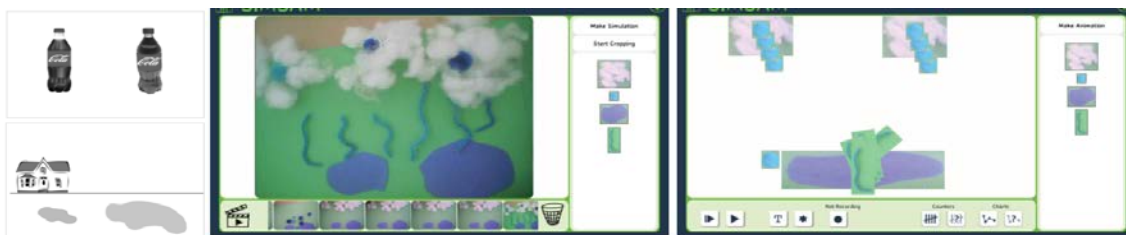


Figure 2. Templates for drawing tasks (left), and SiMSAM animation (center) and simulation (left) interfaces.

## Methods

Our analyses focus on (1) elaboration of student ideas, (2) the degree to which the tool supported students' articulation and revision of ideas, and (3) collaborative knowledge building through use of the tool. To identify student ideas, we used verbal analysis to identify what knowledge learners mobilized when making sense of prompts and activities. To investigate the nature of student engagement in relevant epistemic games, we attend to the degree to which their discussions and actions provide evidence that they are reasoning mechanistically (Russ, Scherr, Hammer, & Mikeska, 2008) and engaging in modeling practices (Schwarz et al., 2009). Finally, to explore collective knowledge construction, we identified when artifacts (such as a simulation or representational convention), theories (such as the particulate nature of water), and vocabulary (such as use of the term "vapor") became taken up as shared by the classroom, and how much those artifacts, theories, or vocabulary were made possible or emerged through collective use of the modeling environment.

## Results

In this section, we report on students' participation in constructing simulations. We will show that in both cases, the modeling environment offered students a way to externalize and elaborate their ideas. We then report the development and adoption of a collective model that featured contributions from several classroom groups. Throughout the results, we present transcript of group and classroom discussion. To make clear the role of facilitators in these discussions, when the classroom teacher or research staff (Michelle and Brian who are also authors on this paper) appear in transcript records, they are marked with an asterisk.

### Dispersion group

During the first activity, students were asked to explain why the exterior of a cold bottle gets wet on a hot day. While the three students' initial drawings had been quite different, Kenny had identified "fog" as a point of agreement across the drawings and this became a focus for the Disperse group's animation.

- 1 Kenny: So we agreed on like, there's coming fog onto it, right? That's what we all agreed on. We
- 2 thought, that we could, that this [*holds cotton tufts*] could be fog and then it could slowly
- 3 be coming down. And then like only little bits of it, but then after a while more and more.
- 4 Miles: We should put like every picture a little bit farther [*acts out steps with cotton tufts*]

Kenny described fog "coming down...like only little bits" and "after a while, more and more." Miles enacts how this description can be expressed using the stop-action animation format with craft materials, suggesting that in "every picture"—each frame—the bits of fog could move a little bit farther. From the drawings and Kenny's comment, we see emerging descriptions of "fog" as a discrete, scattered entity that moves gradually. We also note Miles' early adoption of gradual, discrete movement as the "epistemic game" he uses to populate animation as a form. The group's final animation showed tufts of cotton gradually moving toward and sticking to the Coke bottle, creating droplets of water on the bottle, and then scattering away from the bottle.

The next day, the group began to build their simulation using cropped images of a soda bottle and "fog" (cotton tufts). Kenny initially tried to reproduce the animation by programming cotton tufts to move toward the bottle. Miles, however, noticed that there was an option in the simulation environment to make objects *interact* with one another, rather than merely move. This option works by triggering only when two objects come in physical contact with one another within the simulation environment.

- 5 Miles: Wait no let's do it to interact actually, cuz the fog goes on the Coke bottle and then goes
- 6 out after.
- 7 Edgar: Well the fog is coming and then it's gonna, yeah, go away.
- [1.5 minutes later]
- 8 Teacher\*: What are you guys discussing?
- 9 Kenny: We're trying to like move it [*toward the bottle*] and then hit the coke bottle and then

- 10           come back.  
 11   Kenny:   We want this and then when it hits it [*coke bottle*] it just stops and then we're gonna  
 12           keep on adding more and then were gonna put that [*puddle*] down there [*bottom of coke*  
 13           *bottle*] and then after that we're gonna make the white things [*"fog" objects*] disperse.

Here, Miles suggests the “interact” feature is a more sensible representation for their purposes, since the group agrees that the point of contact between “fog” and the bottle is critical. Miles’ attention to the features he has available in the simulation, and their potential connections to the phenomenon the group is modeling, suggests continued alignment between the epistemic game he is playing and the form he is populating. Meanwhile, Kenny notes a behavior of fog that he wishes to illustrate in the simulation—*dispersion*, or a random scattering of particles. This scattering is programmed into the computational model as a slight random wiggle, so that “fog” objects meander toward the soda bottle, and then bounce against it to represent “sticking”.

The representation of “fog”, or water particles, moving randomly and “dispersing” persists into the second activity concerning evaporation. The group constructed a simulation in which “steam”, represented as blue dots, moves randomly between puddles at the bottom of the screen and clouds at the top. When asked to describe their construction, all three invoked notions of cooking, steam, and combustion:

- 14   Edgar:    So the steam is pretty much the evaporation. The steam is like the evaporation. The steam  
 15           is the blue stuff.  
 16   Miles:    The steam is like, fake. The steam is like, invisible  
 17   Brian\*:    Say I'm so small I can see what's going on inside the puddle, right? What's going on inside  
 18           the puddle, when you said it boils up, what's happening?  
 19   Kenny:    Inside, inside the puddle what's happening is the heat, it's hitting the water and then you  
 20           know when you put um, water in a pot and it starts to boil. The steam is actually the  
 21           evaporation going into the air. But then eventually you can't see the steam when it's  
 22           boiling because eventually, um it, um, it combusts and goes away into the air so it can  
 23           stay there, because, yeah.

Throughout the Disperse group’s participation, there is evidence that they are leveraging knowledge of the behavior of fog and steam as substances that scatter, “disperse”, and “combust” to inform their model. These notions are easily translated over the course of a few days into random motion of particles in the simulation. This random motion, as we will see in the next section, becomes a major contribution to the class’ shared model.

### Aggregation group

When discussing what they wanted their group animation explaining condensation on a soda bottle to look like, Luis offered an early proposal. He suggested that on a hot day a cloud forms; then, that cloud “turns cold” so that it “could drop ... ice” which would melt causing water to appear on the bottle.

- 1    Luis:      So I thought, like, if it was a really hot day. So I thought this cloud would like form  
 2           because like it might like like if if, cuz you know evaporation like it has like clouds. So  
 3           then I thought if, the cloud turns cold it could drop like ice, if you put it back in the  
 4           refrigerator, and then ice would be on it like you could see the ice, and then it would all  
 5           melt for like the water to appear.

After several minutes of discussion, Ryan proposes this plan to the classroom teacher, and the group settles in to creating an animation featuring a sort of mini water cycle that carries water to the outside of the bottle:

- 6    Ryan:      So we're gonna try to like, you know, just pictures of the ice and water just sitting there,  
 7           and then the ice starts coming, and we're gonna try and say in like one sentence about  
 8           you know, you know, it gets so cold it builds up ice.  
 9    Teacher:    It builds up ice where?  
 10   Ryan:      Builds up ice outside the glass.  
 11   Teacher\*:    Outside the glass  
 12   Ryan:      Mhm, and then and then we, these little things are gonna be water droplets that's gonna  
 13           start dropping and then after a while we're gonna make ice like on the table.

While it is unclear from the transcript alone that Ryan is echoing Luis’ idea of clouds, the resulting animation produced by the group aligns especially well with traditional descriptions of the water cycle. It featured an open glass that was filled with water, then filled with a cloudlike substance (represented with cotton), a movement of the cotton to the outside of the glass, and a releasing of blue drops from the cotton onto the outside of the glass.

Unlike the Disperse group, who generated their animation using small, discrete movements of objects, the Water Cycle group generated their animation using “scenes” that roughly corresponded to evaporation, condensation, and precipitation. While creating an animation from “scenes” is commonplace and makes reasonable use of the stop-action moviemaking tool, it did not set the group up well for agent-based modeling as an epistemic form. This first became evidence as started to decide what objects to crop in order to assign those objects rules to define their movements and interactions.

- 14 Ryan: What moves?  
15 Ryan: I don't really think anything moves.  
16 Sergio: Ya I don't think anything moves.  
17 Luis: This is gonna be hard.

We interpret this to mean that no isolated objects move in the group’s animation, and therefore it did not make sense to crop anything. We are careful to note that describing scientific processes in terms of phases or scenes is not necessarily a faulty epistemic move. However, it is not well aligned with agent-based modeling as a form. Because of this, the group struggled to generate a working model to illustrate their explanation. Remembering this, the classroom teacher offered explicit guidance on how to construct animations that could more easily be turned into simulations during the second activity in which students modeled evaporation.

- 18 Teacher: So remember, you guys have got to use materials to show this stuff happening. So if you  
19 just draw a cloud on there, it's never going to move. So if you want to show a cloud  
20 moving, you need an object. Like the, like a puff ball or something.  
21 Luis: We can use the eraser [*to erase the pencil drawings*].  
22 Teacher: But, you can move the puff ball.

After these explicit instructions and continued support from the classroom teacher, the Water Cycle group was able to create a working simulation to describe evaporation:

- 23 Teacher: Ok, so you want to make some droplets coming out of the puddle?  
24 Ryan: And once it like it hits it (cloud) then it disappears.  
25 Teacher: Once it hits what?  
26 Sergio: The cloud.  
27 Teacher: Oh ok. Do we want to make the clouds bigger so that you can—[Ryan: Yea] Ok.  
28 Luis: Yea, that's pretty good.  
29 Ryan: Then when it hits it, the clouds are gonna like get bigger.

This time, the students readily describe behaviors and interactions among objects in their simulation, and relate those behaviors and interactions to the phenomenon they wish to describe. For example, Ryan proposes to model clouds as “droplet accumulators” that grow larger when they are hit by a droplet of water that disappears—ostensibly because it has joined, or gone inside of, the cloud. Like the Disperse group, the Water Cycle group’s focus on clouds as collectors of water was sustained across both activities, despite their struggles with simulation as a representational form. After explicit instruction on how to play the appropriate epistemic games, they were able to operationalize this notion of cloud as a water collector by exploring interactions between droplet and cloud objects. By the end of the day, the group had decided to duplicate clouds, rather than making them grow, to show that droplets are added to rather than taken inside of clouds.

### Development of the collective model

Toward the end of the evaporation activity, we projected the Water Cycle group’s simulation to be discussed by the class. Students immediately began to interpret the simulation and make suggestions:

- Sheree: I think it represents when the sun evaporates the water, um the clouds they start to make new ones because of the water vapor.  
Sarah: But one thing I don't understand is when, when the droplets go up, the puddles don't disappear. So are you saying that the puddles are still there?  
Miles: I think it's like the water droplets are just going straight up, and then it's [clouds] just gonna get bigger and bigger

We took note of students’ suggestions, and brought a model that incorporated many of them into the Water Cycle group’s simulation the next day (Figure 3). The new model illustrated different group’s ideas across both modeling activities: The notion of “droplets” or small, discrete representations of water had been adopted



unanimously by the class; random motion was contributed by the Dispersion group; the notion of droplets adding to clouds was contributed by the Aggregation group; and other contributed behaviors were also featured.



**Figure 3.** Simulation presented to students that combines random motion of particles (from Disperse Group), accumulation of water particles in clouds (from Water Cycle Group), and shrinking puddles that emit water particles (volunteered during discussion by another group in the class).

The class became excited when the model was presented, and their subsequent descriptions and analyses included combinations of mechanism that had been distributed across various student groups' simulations:

Michelle\*: So I heard that a lot of people were sort of excited about this one. Does anyone want to talk about why? Maybe voices we haven't heard?

Luis: Because the water droplets are like going up and the puddles are shrinking.

Michelle\*: And what do you think that represents that you agree with?

Luis: That, like, all the water droplets are like going away because the puddles are drying up.

James: Yesterday we couldn't get the puddles to go down, get smaller and the raindrops to go up. But in this one the puddles got smaller and she didn't have to place the things, it just you know.

Kenny: The sun rays go into the water, the water starts to boil, it goes into the air, it makes clouds, and it's evaporation. And it happens over and over again.

Sheree: Evaporation is when the sun takes water, and then it turns it into water vapor and then it turns it into a cloud, or adds on to a cloud.

There are two things to notice here. First, the students engaging with the shared simulation and making sense of it in the context of their own ideas (note Miles and Kenny's descriptions of boiling to describe the random motion assigned to a different situation, and Luis' interpretation of the edits to his own). Second, they attend to some of the unique affordances of computational models as a representational form, noting that interactions such as the puddles growing smaller without needing to do anything manually is preferable.

There is evidence that students took up this shared model and its main components: particles, random motion, and cloud-as-collector. In a follow up activity students explained why the water level of a heated covered beaker does not fall while the level of a heated uncovered one does. All groups in this class created models with particles, random motion, and descriptions of water "going to clouds" or entering the water cycle.

## Conclusions and implications

In this paper, we argued that computational modeling environments can serve as sites for collaborative knowledge construction, much like the collaborative concept mapping and diagramming tools that are common in communities-of-learners classrooms. We illustrated how two groups used animation and simulation infrastructures to externalize, elaborate, compare, test, and synthesize ideas about the mechanisms that underlie evaporation and condensation. These mechanisms were rooted in the students' existing knowledge, and contributed elements of a model that became taken up at the classroom level. This work suggests specific supports for enacting computational modeling activities. The classroom teacher, when working with the Water Cycle group, explicitly and intentionally focused on *epistemic moves*—how the group was populating their animation in anticipation of converting it to a simulation—and not other aspects of their work such as content. The use of templates to structure student models, and the inherent modularity of computational models, made it especially easy for us to share, contrast, and combine students' ideas. This allowed students to recognize their contributions within a new or different model. Research should extend beyond attention to individual and small group learning to see how computation can become part of the epistemic fabric of science classrooms.

## References

Basu, S., Sengupta, P., & Biswas, G. (2015). A scaffolding framework to support learning of emergent phenomena using multi-agent-based simulation environments. *Rsch in Sci Education*, 45(2), 293-324.

- Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. *Science Education*, 95(2), 191–216. <https://doi.org/10.1002/sce.20420>
- Bielaczyc, K., & Ow, J. (2014). Multi-player epistemic games: Guiding the enactment of classroom knowledge-building communities. *International Jnl of Computer Supported Collaborative Learning*, 9(1), 33–62.
- Chandrasekharan, S., & Nersessian, N. J. (2015). Building cognition: The construction of computational representations for scientific discovery. *Cognitive Science*, 39(8), 1727–1763.
- Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational Psychologist*, 28(1), 25–42.
- Fretz, E. B., Wu, H., Zhang, B., Davis, E. A., Krajcik, J. S., & Soloway, E. (2002). An investigation of software scaffolds supporting modeling practices. *Research in Science Education*, 32(4), 567–589.
- Grimm, V. et al., (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, 310(5750), 987–991. <https://doi.org/10.1126/science.1116681>
- Hegedus, S. J., & Moreno-Armella, L. (2009). Intersecting representation and communication infrastructures. *ZDM*, 41(4), 399–412. <https://doi.org/10.1007/s11858-009-0191-7>
- Hmelo-Silver, C. E., Liu, L., Gray, S., & Jordan, R. (2015). Using representational tools to learn about complex systems: A tale of two classrooms. *Journal of Research in Science Teaching*, 52(1), 6–35.
- Löhner, S., van Joolingen, W. R., Savelsbergh, E. R., & Hout-Wolters, B. Van. (2005). Students' reasoning during modeling in an inquiry learning environment. *Computers in Human Behavior*, 21, 441–461.
- Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: Basic Books, Inc.
- Penner, D. E. (2000). Cognition, computers, and synthetic science: Building knowledge and meaning through modeling. *Review of Research in Education*, 25(1), 1–35. <https://doi.org/10.3102/0091732X025001001>
- Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, 92(3), 499–525. <https://doi.org/10.1002/sce.20264>
- Sandoval, W. A., Angeles, L., Bell, P., Coleman, E., & International, S. R. I. (2000, April 25). Designing knowledge representations for learning epistemic practices of science. New Orleans, LA.
- Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences*, 3(3), 265–283. <https://doi.org/10.1207/s15327809jls0303>
- Schwarz, C., Reiser, B. J., Acher, A., Kenyon, L., & Fortus, D. (2012). MoDeLS: Challenges in defining a learning progression for scientific modeling. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning Progressions in Science: Current Challenges and Future Directions* (pp. 101–137). Rotterdam: Sense.
- Schwarz, C., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Acher, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. S. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654. <https://doi.org/10.1002/tea.20311>
- Stake, R. E. (2006). *Multiple case study analysis*. New York: The Guilford Press.
- NGSS Lead States (2013). *Next generation science standards: For states, by states*. Washington, DC: NAP.
- van Joolingen, W. R., de Jong, T., & Dimitrakopoulou, A. (2007). Issues in computer supported inquiry learning in science. *Journal of Computer Assisted Learning*, 23(2), 111–119.
- Van Joolingen, W. R., De Jong, T., Lazonder, A. W., Savelsbergh, E. R., & Manlove, S. (2005). Co-Lab: Research and development of an online learning environment for collaborative scientific discovery learning. *Computers in Human Behavior*, 21(4), 671–688. <https://doi.org/10.1016/j.chb.2004.10.039>
- VanLehn, K. (2013). Model construction as a learning activity: A design space and review. *Interactive Learning Environments*, 21(4), 371–413. <https://doi.org/10.1080/10494820.2013.803125>
- VanLehn, K., Wetzel, J., Grover, S., & van de Sande, B. (2016). Learning how to construct models of dynamic systems: An initial evaluation of the Dragoon intelligent tutoring system. *IEEE Transactions on Learning Technologies*, 1382(c), 1–1. <https://doi.org/10.1109/TLT.2016.2514422>
- White, B. Y., Collins, A., & Frederiksen, J. R. (2011). The nature of scientific meta-knowledge. In M. S. Khine & I. M. Saleh (Eds.), *Models and Modeling* (pp. 41–76). Dordrecht: Springer Netherlands.
- Wilensky, U. J., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories — an embodied modeling approach. *Journal of the Learning Sciences*, 24(2), 171–209. <https://doi.org/10.1207/s1532690xci2402>
- Wilkerson-Jerde, M., Gravel, B. E., & Macrander, C. A. (2015). Exploring shifts in middle school learners' modeling activity while generating drawings, animations, and computational simulations of molecular diffusion. *Journal of Science Education and Technology*, 24(2–3), 396–415.
- Xiang, L., & Passmore, C. (2015). A framework for model-based inquiry through agent-based programming. *Journal of Science Education and Technology*, 24(2–3), 311–329.

# Using Rotating Leadership to Visualize Students' Epistemic Agency and Collective Responsibility for Knowledge Advancement

Leanne Ma, OISE/University of Toronto, leanne.ma@mail.utoronto.ca

Samuel Tan, Ministry of Education, Singapore, Samuel\_TAN@moe.gov.sg

Chew Lee Teo, Ministry of Education, Singapore, TEO\_Cheew\_Lee@moe.gov.sg

Muhamad Ansar B. Kamsan, Ministry of Education, Singapore, Muhamad\_ansar\_kamsan@moe.edu.sg

**Abstract:** As in knowledge-creating organizations and Collaborative Innovation Networks (COINs), students in Knowledge Building classrooms work creatively with ideas in a self-organized fashion, with all members engaged in advancing emergent community goals. In this study, we examined the online knowledge work of 9-year-olds studying light and shadows. Data triangulation at multiple levels of analysis (e.g., community, students, teacher) was used to validate the COIN concept of rotating leadership to assess students' collective responsibility for knowledge advancement. Overall, we found many students leading the group at different points in time, facilitating the spread of diverse ideas that enhanced the breadth of community knowledge and the depth of individual learning. Teacher perceptions of classroom dynamics uncovered additional details of student leaders, such as their level of engagement and their learning outcomes. The practical implications of rotating leadership for assessing Knowledge Building community dynamics – such as epistemic agency and collective responsibility – are discussed.

## Introduction

Over the last decade, educational reform initiatives, such as Partnership for 21<sup>st</sup> Century Skills (2005), Design Thinking for Educators (2013), and Building Cultural Capacity for Innovation (2016), have been adopted in schools around the world to address various needs of innovation-driven societies (OECD, 2010). Of particular interest are schools in Ontario and Singapore, which have been successful at sustaining innovative practices in classrooms while continuously improving student achievement (Mourshed, Chinezi, & Barber, 2010). This study builds on existing research on educational innovations in Ontario schools (e.g., Zhang, et. al., 2011) and Singapore schools (e.g., Tan, So, & Yeo, 2014), with a focus on knowledge creation in primary education.

Knowledge Building (Scardamalia & Bereiter, 2014) represents a longstanding international initiative aimed at transforming schools into knowledge-creating organizations. Simply defined, Knowledge Building “is giving students *collective responsibility* for idea improvement” (Bereiter & Scardamalia, 2014). Knowledge Building takes an idea-centered, principle-based approach to classroom teaching (Hong & Sullivan, 2009) and is further distinguished from more generic forms of collaborative learning (O'Donnell & Hmelo-Silver, 2013) by transfer of responsibility for knowledge advancement to students so that responsibilities traditionally reserved for teachers (e.g., setting knowledge goals, monitoring knowledge growth, and assessing individual- and group-learning) are shared and negotiated between all members. Teachers support students' *epistemic agency* by encouraging students to take ownership of their learning at the highest levels (Scardamalia, 2002). Students generate *authentic problems* of understanding and collaborate in flexible groups that emerge and disband as needed in order to sustain collective efforts at *improving ideas* and creating coherence among *diverse ideas* in the community knowledge (*rise above*) (Zhang, Scardamalia, Reeve, & Messina, 2009).

The flexible, self-organizing dynamics in Knowledge Building classrooms mirror those of knowledge-creating organizations (Nonaka & Takeuchi, 1995) and Collaborative Innovation Networks (Gloor, 2006), where innovative cultures are fostered bottom-up through the valuing of members' intentionality, autonomy (and *epistemic agency*) during creative group processes. Despite their varied contexts (i.e., schools, organizations, virtual networks), a comparison of their underlying similarities allows further elaboration of our understanding of knowledge-creating dynamics (Ma, Matsuzawa, & Scardamalia, 2016). For example, Collaborative Innovation Networks (COINs; Gloor, 2006), as found in open-source communities, health care institutions, and knowledge-creating organizations, work collaboratively in powerful online environments to drive innovations around the world. Members share collective responsibility for their knowledge work and monitor group progress with a set of online metrics that measure their degree of connectivity, interactivity, and sharing (Gloor, 2006). Gloor and colleagues (2003; 2007) found that whereas highly efficient teams operate in a centralized fashion with a stable set of leaders, highly innovative teams, such as COINs, operate in a decentralized fashion with various emergent leaders rotating leadership over the course of a project. Moreover, rotating leadership is considered a key indicator of group creativity of COINs (Gloor, 2006). Recent work in

education (Ma, Matsuzawa, Chen, & Scardamalia, 2016; Ma, 2016) suggests that Knowledge Building classrooms operate in a COINs fashion, with rotating leadership as an emergent phenomenon of collective responsibility. Findings from four cases in Ontario consistently revealed that more than half the students emerged as leaders over the course of their Knowledge Building online. When leading, students were connecting “big” ideas in the community knowledge through various types of contributions, such as questioning, theorizing, introducing ideas, and synthesizing ideas – discursive moves consistent with the ways of contributing to Knowledge Building dialogue framework developed by Chuy and colleagues (2011). However, one limitation is that these studies did not include the teacher’s intended nor perceived pedagogical outcomes. The current study has two aims: to examine Knowledge Building community dynamics in a cross-cultural context and to integrate various data sources to inform our interpretation of rotating leadership in Knowledge Building. Thus, the study focuses on a Knowledge Building classroom in a Singaporean context and examines *collective responsibility for idea improvement* at three levels of analysis: the community, the students, and the teacher. The research questions are as follows:

1. At the community-level, is there rotating leadership? How does the emergence of student leaders correspond to pivotal points of knowledge advancement in the community knowledge?
2. At the student-level, which “big” ideas are student leaders discussing? How do those ideas correspond to the “big” ideas reported in their portfolios at the end of the intervention?
3. At the teacher-level, how does rotating leadership correspond to their perceptions of student engagement and community knowledge advancement during Knowledge Building?

## Methods

In acknowledging the Knowledge Building classroom as a complex, dynamic system, we framed the research questions at three levels of analysis and adopted a multi-level mixed methods design (Creswell & Plano Clark, 2011) in order to triangulate across the levels and gain a more holistic perspective of community dynamics.

### Classroom context

As mentioned above, Singapore represents a Knowledge Building hub of innovation. The current study took place in an all-girls primary school located in a middle income area. The school has been participating in the Knowledge Building initiative in Singapore over the last two years. Teachers at the school work in a professional learning team with the shared goal of fostering a culture of sustained work with ideas in their classrooms while addressing the rigid demands of the nationally mandated curriculum. In their professional learning team, teachers build knowledge about the science curriculum; create, share, and refine Knowledge Building practices; as well, they integrate Knowledge Forum technology in their classrooms in order to support students’ development of scientific literacy, collaboration skills, and self-directed learning. The Primary 3 science teacher has been engaged in design-based research (Barab, 2014) for 3 design cycles over 18 months in order to systematically improve their Knowledge Building practices by focusing on a set of targeted Knowledge Building principles each cycle. During the first cycle, the teacher focused on creating an idea-centered classroom culture (e.g., *authentic problems, improvable ideas, idea diversity*). During the second cycle, the teacher focused on constructing explanations of scientific phenomena (e.g., *improvable ideas, idea diversity, Knowledge Building discourse*). During the third cycle, the teacher focused on fostering collective responsibility for knowledge advancement (e.g., *collective responsibility, community knowledge, constructive use of authoritative sources*). This study represents the fourth design cycle, where the teacher supported students in sustaining their Knowledge Building discourse as they worked toward increasing the explanatory power and coherence of their ideas (e.g., *epistemic agency, improvable ideas*). During this pedagogical intervention, students discussed features of powerful scientific explanations (e.g., claim, evidence, reasoning, logic) and worked collaboratively to improve the conceptual coherence of their community knowledge. Students generated ideas, co-constructed explanations, critiqued each other’s theories, considered alternative perspectives, and actively revised and synthesize ideas in the community knowledge. As much as possible, the teacher provided minimal direction in order to allow their students to take ownership of their Knowledge Building.

The Primary 3 science class consisted of 35 students (9 year-olds) coming from a variety of ethnic and cultural backgrounds (e.g., Asia, Europe, Africa). 4 students were excluded from this study due to moving overseas and other logistical reasons; therefore 31 students were included for analysis. Because the Singapore Primary Science Curriculum begins in Primary 3, students had not been exposed to a formal science curriculum prior to the study, but they were familiar with Knowledge Building and Knowledge Forum, which had become integrated into daily science classroom practices since the first design cycle. Knowledge Forum (Scardamalia, 2004) is an online networked environment designed for collaborative knowledge creation, with scaffolds in

place to support continual idea improvement. Ideas are represented as multimedia objects in conceptual spaces called views. Ideas can be searched, annotated, referenced; improved with “build-on” notes; and synthesized into more comprehensive theories with “rise-above” notes. Knowledge Forum served as the central community space for students to visualize, actively monitor, and advance the ideas in their community knowledge.

Over the span of one month, the students engaged in Knowledge Building about light and shadows, developing theories about how light travels, how shadows are formed, how the eye sees, and how light energy works. Students wrote 170 notes across 3 views in Knowledge Forum: Light and Shadows, How can we see?, and Manipulating Light. Using embedded, formative assessment tools, the teacher facilitated reflective discussions with the class as a whole in order to assess and sustain group progress. For example, the scaffold growth tool was used to visualize the distribution of students’ online contributions and determine next steps to advance the community knowledge. The promising ideas tool was used to identify ideas that had the greatest potential to enhance existing scientific explanations. Promising ideas and theories were then tested and refined through various inquiry activities supported by the teacher. Students compared shadows at different times of day in order to understand how the size of shadows vary as a function of the angle of the light source in relation to the object. Their understanding was further elaborated after they watched a video of a shadow show, which prompted their thinking about the relationship between the distance of the screen, the light source, and the object. Students then designed and conducted experiments to understand how light reflects off different surfaces and how light enables the eye to see. Working in small groups, they investigated the relation between light and shadows by manipulating a set of variables of their choice, such as light source (distance, intensity, angles), objects (shapes, texture, colour, size), and screen (distance). At the end of their Knowledge Building, students created portfolios to summarize the “big” ideas they had learned.

## Data sources and analyses

Multiple data sources and analytic methods were used in order to address the different questions at each level of analysis. At the community-level, the student discourse on Knowledge Forum was used to conduct temporal network analysis and explore rotating leadership, then discourse analysis was performed on the first 100 notes to identify pivotal points of knowledge advancement. All 170 notes were spellchecked and exported into KBDeX (Knowledge Building Discourse Explorer; Oshima, Oshima, & Matsuzawa, 2012), which performs social-semantic network analysis based on a list of content-related words and produces network visualizations of the student discourse. The learners network (see Figure 2a) and the words network (see Figure 2b) are created via the co-occurrence of key words in each network, with the strength of connections represented by the thickness of edges between nodes. KBDeX also creates temporal visualizations of network metrics, such as betweenness centrality, which indicates a member’s influence relative to other members in the group on a scale from 0 (low) to 1 (high) (Gloor et. al., 2003). This KBDeX feature was used to conduct temporal network analysis.

The temporal visualization of betweenness centrality (see Figure 1) was then used to identify the first four leaders that emerged during Knowledge Building. At the student-level, analysis involved using KBDeX to explore the position of each student leader within the learners network and the keywords used in the word network, then content analysis was conducted on the notes and the portfolio of each student leader. The student notes helped shed light on which “big” ideas student leaders discussed and how those students contributed to the community knowledge. In turn, the student portfolios helped clarify the relation between student leadership and learning outcomes, such as idea improvement, by making the individual learning process more transparent. The teacher-level analysis involved content analysis of the teacher journal then a follow-up interview. The journal entries helped provide insight on the Knowledge Building process and whether or not the teacher perceived students were assuming *collective responsibility* for *idea improvement*. The follow-up interview involved the teacher discussing community dynamics and student outcomes relative to the rotating leadership visualizations created in KBDeX (Figures 1 to 5). It concluded with the teacher reflecting on the potential for these visualizations to support them and their students in Knowledge Building.

## Findings

In this section, we describe the findings at the community-level (e.g., temporal network analysis, discourse analysis), student-level (e.g., social network analysis, content analysis), and teacher level (e.g., content analysis, interview).

### Community-level analyses

Temporal network analysis suggests that many students were simultaneously influential at different times. Figure 1 shows the betweenness centrality of all students over the course of their Knowledge Building, where the Y axis represents the betweenness centrality value, and the X axis represents the turn in discussion. A

different coloured line is used to represent each student, and the oscillation of coloured lines illustrates the phenomenon of rotating leadership. Of the 31 students, 19 students took a leading position. The legend in Figure 1 indicates that Students 1, 11, 22, and 19 emerged as the most influential leaders during the first 100 turns in discussion. These student leaders and their contributions are examined in further detail in the subsection below.

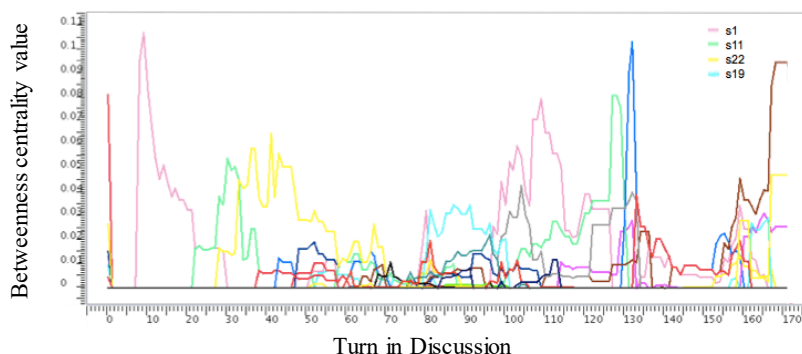


Figure 1. Temporal visualization of betweenness centrality of all students from turn 1 to turn 170.

Discourse analysis on the first 100 notes across the three views revealed that student ideas and theories surrounding various problems of understanding became increasingly more scientific and sophisticated. In the discussion about light and shadows, students generated theories about the size of shadows based on the distance between the object and the light source. Their theories were then improved by examining how properties of the object, such as opaqueness and texture, affect the way light travels and in turn, how shadows are formed. The discussion about how the eye sees grew out of a student’s observation that we cannot see without light. Students then consulted authoritative sources to understand how lenses refract light to help the eye see objects and how raindrops refract light to help the eye see colours in the rainbow. In the discussion about the properties of light, students originally hypothesized that light travels in a straight line. After reading about electromagnetic radiation and photons, students began considering an alternative theory that light travels in a wave. Later on, the discussion was reframed from the perspective of natural and artificial sources of light. Students then began theorizing that light is energy and further improved their theory by comparing various sources of energy, such as chemical and electrical energy, at the end of the first 100 notes. The ideas that emerged in the pivotal points of knowledge advancement overlap with the key concepts taught in the “Gadgets Work Wonders” module in the Singapore Upper Secondary Science Curriculum (Ministry of Education Singapore, 2013, p. 32-37), where secondary students (i.e., 15 to 17 year-olds) explore the relations between energy, waves, forces, and electricity.

## Individual-level analyses

### Student 1

Student 1 was leading between turns turn 9 to turn 30 and peaked at turn 10 (betweenness centrality=0.107).

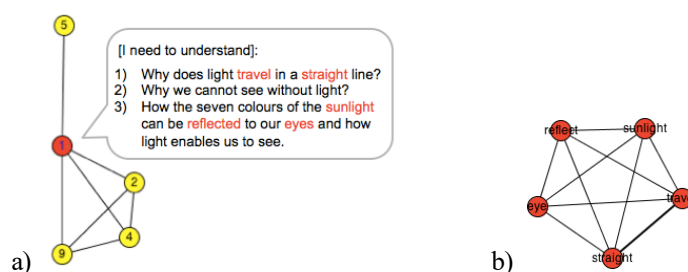


Figure 2. a) Student network and b) word network visualizations at turn 10 when Student 1 was leading.

Figure 2a shows the note Student 1 contributed that resulted in their occupation of an influential position in the student network. Student 1 contributed the first note in Knowledge Forum by defining their problems of understanding and asking deep questions that sparked the interest of their peers so much that this note was highlighted as a promising idea in the Lights and Shadows view. Figure 2b shows that Student 1 introduced

important ideas in the community discourse, such as “sunlight”, “reflection”, and “eye”, which would later appear in discussions about sources of light and energy, experiments manipulating light, and how the eye sees. It is interesting to note that while Student 1 had the greatest influence, their peers also influenced their learning. For example, Student 11 introduced the idea of “transparency”, which helped Student 1 develop a more nuanced understanding of how light travels and creates shadows. Below is an excerpt from Student 1’s portfolio:

Light moves in a **straight** line, creating shadows when the **path** of light is blocked. More solid things will have a darker shadow. Things that are more clear have a lighter shadow, and **transparent** things will have none or very little shadow. Light can pass through **transparent** things the most easily. Our **eyes** react to light when we see something we see the light it **reflects**, or the light it emits.

### Student 11

Student 11 was leading between turn 24 to turn 39 and peaked at turn 32 (betweenness centrality = 0.054).

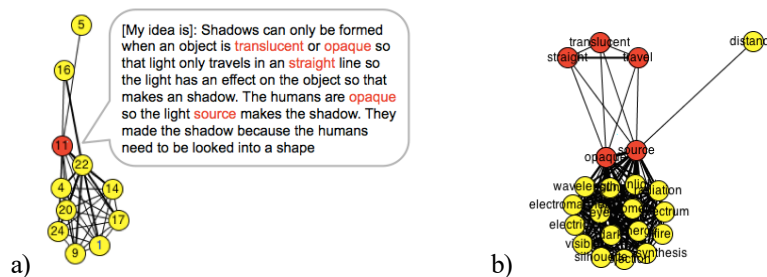


Figure 3. a) Student network and b) word network visualizations at turn 32 when Student 11 was leading.

Figure 3a shows the note Student 11 contributed that resulted in their occupation of an influential position in the student network. The problem of understanding at this time was about light and shadows, and students were wondering about how different sizes of shadows are made. Students investigated the size of shadows from the perspective of the light source, theorizing that the distance of an object from the source of light will affect the size of the shadow. Student 11 developed an alternative theory about shadows, by reframing the discussion in relation to the properties of the object, such as “translucent” and “opaque” objects. Figure 3b shows how Student 11’s ideas are connected to the ideas in the larger discussion of light and shadows. While Student 11 introduced new ideas to the community knowledge, they also learned about ideas introduced by their peers. For example, Student 19 and Student 22 discussed extensively the idea of light as “energy”, which helped Student 11 develop a nuanced understanding of the properties of light. Below is an excerpt from Student 11’s portfolio:

[Putting our knowledge together]: Light is a **source** of **energy** and is not **matter** as it does not occupy space or have **mass**. Light helps us see things as it bounces off of other objects and **shines** into our **eyes** and when it is **reflected** at an **opaque** object, it will make a shadow... Light travels in a **straight** line but it can be directed to different directions in a zig zag manner with **mirrors**.

### Student 22

Student 22 was leading between turn 30 to turn 71 and peaked at turn 43 (betweenness centrality = 0.065).

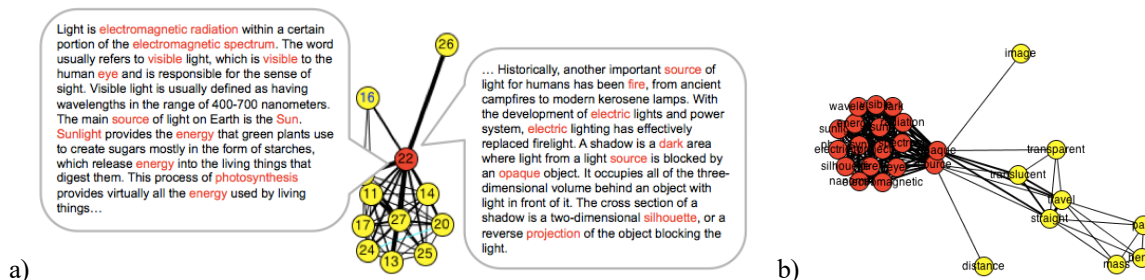


Figure 4. a) Student network and b) word network visualizations at turn 43 when Student 22 was leading.

Figure 4a shows the two notes Student 22 contributed that resulted in their occupation of an influential position in the student network. In the first note, they introduced new ideas about properties of light, explaining that light is “electromagnetic radiation” and how the eye is able to see portions of the “electromagnetic spectrum”. Student 22 also highlighted the importance of light in nature by connecting the class discussion of light (in physics) to “photosynthesis” (in biology). In the second note, they synthesized their ideas about sources and light and their peer’s ideas about “silhouettes” and “opaque” objects. Figure 4b shows how the many ideas contained in Student 22’s notes are connected to the class discussion about how light travels and how the eye sees. Although Student 22 had the longest influence, it is unclear who influenced their learning because Student 22’s portfolio was not available for analysis.

### Student 19

Student 19 was leading between turn 82 to turn 102 and peaked at turn 89 (betweenness centrality=0.035).

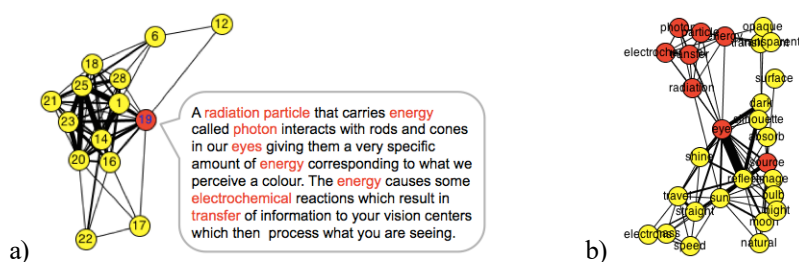


Figure 5. a) Student network and b) word network visualization at turn 89 when Student 19 was leading.

Figure 5a shows the note Student 19 contributed that resulted in their occupation of an influential position in the student network. The problem of understanding at this time was about light and shadows. Student 19 introduced the idea of light as “energy” and vision as an “electrochemical” reaction. Figure 5b shows how Student 19’s ideas are connected to the class discussion about how the eye sees and sources of light. It is interesting to note that Student 11 grappled with the most sophisticated idea about light as “energy” and “photons”, and their understanding of the properties of light was slightly more scientific and complex than that of their peers. Below is an excerpt from Student 19’s portfolio:

[Putting our knowledge together]: Light travels in a straight line. There is not only a straight line, but there are many straight lines. Light bounces off the object into our eyes. Shadows can only be formed with translucent and opaque objects. Transparent object allows all light to pass through. Shadows are not only black or grey but they can be multi-coloured.

### Teacher-level analyses

In their journal, the teacher reflected on the classroom practices they used to encourage students to take initiative to create coherence of ideas. Some effective practices included: explicitly teaching how to construct scientific explanations using the “claim, evidence, reasoning” and “premise, reasoning, outcome” models; applying heuristics to write clearly on KF (such as crafting explanations for younger audiences so that students would unpack sophisticated ideas in their notes); and developing criteria with students for judging promising ideas. The teacher supported students in evaluating the state of their community knowledge by querying whether their problems of understanding had been addressed, whether new questions had emerged, and asking students to compare their ideas with other “big” ideas in the curriculum. Moreover, the teacher was intentional about supporting students’ ideas without overshadowing them. For example, instead of giving students a set of questions to answer for their light experiments, the teacher encouraged students to pursue their own questions in order to gain a deeper understanding of the causal mechanisms of the phenomena under study. Over the course of the intervention, the teacher observed that students became more deliberate in their KF posts, posing questions constructively, and taking on higher levels of *epistemic agency* and *collective responsibility*.

During the follow-up interview, the teacher shared their classroom experiences and reflected on the rotating leadership visualization. The teacher was surprised that the most active students during class were not picked up in KBDeX. However, upon further consideration, they suggested that the rotating leadership may tap



into the KB principle of *epistemic agency*, picking up “value-adding” students without their knowledge. The teacher also noted that some students picked up as leaders were quieter or academically weaker students. Those students asked deep questions on KF, worked diligently throughout, and went beyond their expectations. What’s more, the weaker students did well on their summative assessment, suggesting that their active participation in KF supported individual learning. Below is an excerpt of teacher comments during the interview:

The rotating leadership visualizations provided another layer of analysis for myself as the identities of the ‘leaders’ of discussion were different to that of what I had in mind. As the visualizations were able to make connections on how the lesson progressed and how members of the class contributed to the progress, it would have been a helpful tool when implemented perhaps half-way through the unit. I could have used the analysis to direct the discussions towards more instrumental contributors [who enhanced the] connectedness of ideas... Moving forward, I would want to allow students to identify their own leaders for discussion.

## Conclusions and future directions

This study represents the fifth case of rotating leadership in Knowledge Building. Our findings demonstrate that collective responsibility for knowledge advancement, as depicted by oscillating patterns of betweenness centrality in the student network, is consistently found in Knowledge Building classrooms, even across cultural contexts with diverse populations. In this class, 19 out of 31 students were leaders. When leading, students asked deep questions, generated theories, shared novel ideas, and/or synthesized existing ideas in order to increase the conceptual coherence in the community knowledge. Furthermore, these ideas were uptaken by other students in their portfolios, suggesting that rotating leadership also facilitates the spread of ideas between students. As students pursue ideas in complex, unpredictable directions, a major challenge is charting their trajectory of idea improvement in order to assess individual progress and group progress. Here, we highlight a point of convergence between the current case and the Ontario case of 9 year-old students studying light (Ma, Matsuzawa, & Scardamalia, 2016), despite the different directions they took for their Knowledge Building. Students in both classes shifted from understanding light as travelling in a straight line to light travelling in a zigzag. While the Ontario class further shifted to light travelling in a wave, it is important to note that the Ontario students spent three months studying light, whereas the Singapore students spent one month. Thus, a comparison of their learning outcomes would not be entirely fair. Given an additional two months, however, we would not be surprised that these students would reach similar levels of understanding, if not beyond.

In addition to replicating the rotating leadership phenomenon in Singapore, we extended past research by adding the teacher perspective to understand the community dynamics of Knowledge Building classrooms. In particular, the teacher noticed a shift in students’ intentionality toward their learning during this intervention when they were given more autonomy, suggesting that rotating leadership taps into students’ *epistemic agency* as well as *collective responsibility*. It is interesting to note that the teacher noticed different students leading during face-to-face versus online discussions. However, the discrepancy between the teacher perceptions and rotating leadership visualization suggests that student leaders facilitated the self-organization of idea improvement on Knowledge Forum, without any direct intervention by the teacher. At the same time, it is also important for the teacher to be aware of these students and their contributions. One possible explanation is that rotating leadership also taps into the KB principle of *democratizing knowledge*, making important notes and ideas transparent to all. This is especially important for bridging the gap between stronger and weaker students in the classroom. The fact that weaker students who were picked up as leaders experienced greater learning gains in the summative assessment seems to directly support the notion that learning is a byproduct of Knowledge Building (Bereiter & Scardamalia, 2010). Knowledge Building represents one way to engage all students deeply and meaningfully in the classroom through the valuing of student voice. We intend to invite student voice to further illuminate the relation between rotating leadership and collective responsibility.

## References

- Barab, S. (2014). Design-based research: A methodological toolkit for engineering change. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 151-17). New York: Cambridge University Press.
- Bereiter, C., & Scardamalia, M. (2010). A brief history of knowledge building. *Canadian Journal of Learning and Technology*, 36(1). Retrieved from <http://cjlts.csj.ualberta.ca/index.php/cjlt/article/view/574/276>
- Bereiter, C. & Scardamalia, M. (2014). Knowledge building and knowledge creation: One concept, two hills to climb. In Tan, S. C. & Tan, Y. H. (2014). *Knowledge creation in education* (pp. 35-51). Singapore: Springer.

- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research*. Thousand Oaks: Sage Publication.
- Chuy, M., Resendes, M., Tarchi, C., Chen, B., Scardamalia, M., & Bereiter, C. (2011). Ways of contributing to an explanation-seeking dialogue in science and history. *QWERTY - Interdisciplinary Journal of Technology, Culture and Education*, 6(2), 242–260.
- Gloor, P. A. (2006). *Swarm Creativity: Competitive Advantage through Collaborative Innovation Networks*. New York: Oxford University Press.
- Gloor, P. A., Laubacher, R., Dynes, S. B., & Zhao, Y. (2003, November). Visualization of communication patterns in collaborative innovation networks-analysis of some w3c working groups. In *Proceedings of the twelfth international conference on Information and knowledge management* (pp. 56-60). ACM.
- Hong, H. Y., & Sullivan, F. R. (2009). Towards an idea-centered, principle-based design approach to support learning as knowledge creation. *Educational Technology Research and Development*, 57(5), 613-627.
- Kidane, Y. H., & Gloor, P. A. (2007). Correlating temporal communication patterns of the Eclipse open source community with performance and creativity. *Computational and mathematical organization theory*, 13(1), 17-27.
- Ma, L., Matsuzawa, Y., Chen, B., & Scardamalia, M. (2016). Community knowledge, collective responsibility: The emergence of rotating leadership in three knowledge building communities. In C. K. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016, Volume 1* (pp. 615-622). Singapore: The International Society of the Learning Sciences.
- Ma, L., Matsuzawa, Y., & Scardamalia, M. (2016). Rotating leadership and collective responsibility in a grade 4 knowledge building classroom. *International Journal of Organizational Design and Engineering*, 4(1/2), pp. 54-84.
- Ma, L. (2016). The emergence of rotating leadership for idea improvement in a grade 1 knowledge building community. In M. P. Zylka, H. Fuehres, A. Fronzetti Colladon, & P. Gloor (Eds.), *Designing Networks for Innovation and Improvisation: Proceedings of the 6th International Collaborative Innovation Networks (COINs) Conference* (pp. 13-20). Switzerland: Springer International.
- Mourshed, M., Chinezi, C., & Barber, M. (2010, November). *How the world's most improved school systems keep getting better*. London: McKinsey and Company.
- Nonaka, I. and Takeuchi, H. (1995) *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University Press.
- O'Donnell, A., & Hmelo-Silver, C., (2013). What is collaborative learning?. In C. Hmelo-Silver, C. Chinn, C.K.K. Chan, & A. O'Donnell (Eds.), *The International Handbook of Collaborative Learning* (pp. 1-16). New York: Routledge/Taylor & Francis.
- OECD (2010). *The OECD innovation strategy*. Paris: OECD Publishing.
- Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith, & C. Bereiter (Eds.), *Liberal Education in a Knowledge Society* (pp. 67-98). Berkeley: Publishers Group West.
- Scardamalia, M. (2004). CSILE/Knowledge Forum®. In *Education and Technology: An Encyclopedia* (pp. 183-192). Santa Barbara: ABC-CLIO.
- Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 397-417). Cambridge University Press, New York. Tan, S. C., So, H. J., & Yeo, J. (Eds.). (2014). *Knowledge Creation in Education*. Singapore: Springer.
- Singapore Ministry of Education (2013). *Science Syllabus: Lower and Upper Secondary Normal (Technical) Course*. Retrieved from <https://www.moe.gov.sg/docs/default-source/document/education/syllabuses/>
- Zhang, J., Hong, H. Y., Scardamalia, M., Teo, C. L., & Morley, E. A. (2011). Sustaining knowledge building as a principle-based innovation at an elementary school. *The Journal of the Learning Sciences*, 20(2), 262-307.
- Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge-building communities. *The Journal of the Learning Sciences*, 18(1), 7-44.

## Acknowledgments

We are grateful to the Knowledge Building teacher and students for their time and generosity. Their creative ideas and insights have been invaluable to our Knowledge Building.

# GroupWork: Learning During Collaborative Assessment Activities

William T. Tarimo, Brandeis University, wtarimo@brandeis.edu  
Timothy J. Hickey, Brandeis University, tjhickey@brandeis.edu

**Abstract:** The positive benefits of collaborative and cooperative learning are well documented in educational literature and research. In this paper, we present GroupWork - a group-based assessment tool within TeachBack, and an evaluation of our experience in using it as part of an interactive flipped classroom. GroupWork was used to facilitate real-time group-based formative assessment and other problem-based activities in the classroom. Specifically, we examine the short-term and long-term impact of this implementation on students' learning outcomes and experiences compared to performing formative assessments where students work individually. Results show significant performance improvement for each student and for each assessment question when students are allowed to discuss their solutions with peers. The formative assessment grades after collaborative problem solving are shown to have a strong correlation with final course grades, suggesting that collaborative problem solving may have a lasting benefit.

## Introduction

This paper presents a case study and the results of an experiment where students collaborated in small groups during in-class formative assessment activities. The experiment studies whether collaboration during formative assessment activities has better learning outcomes for the participants compared to when they work individually. In assessing the immediate effects on the learning outcomes, we compared students' overall performance when responding individually to formative assessment problems and their corresponding performance when working in collaborative groups. In assessing the impact of collaboration on overall learning outcomes on the course, we looked for evidence of correlations between performance in these assessment activities and final performance in the course. We also surveyed students at the end of the course to get their perspectives on their learning and experiences during group-based formative assessments. All the assessment activities were done using the GroupWork tool within TeachBack (Hickey & Tarimo, 2014; Tarimo, Deeb, & Hickey, 2015; Tarimo, Deeb, & Hickey, 2016; Tarimo & Hickey, 2016; Tarimo, 2016) and the majority of the activities were real-time formative assessments during class time. TeachBack is an in-class web application that offers various interactive tools for active and interactive learning and teaching in the classroom.

Over the last three years of empirical studies using TeachBack, the Questions feature (a lightweight *clicker* functionality) has been used to facilitate assessment activities in the classroom where students respond individually followed by class-wide discussions. For example, a previous study (Tarimo, Deeb, & Hickey, 2016) has shown statistically significant correlation between levels of participation and performance in in-class formative assessments and overall mastery of the material and concepts in a course. In this experiment, we wanted to see whether adding collaboration to these activities would lead to better learning outcomes and experiences for the students.

Conducting this experiment would also inform us on the feasibility and practicality of using web based interactive tools like GroupWork in facilitating collaborative and cooperative learning and assessment activities in classroom settings. The findings would also provide the evidence necessary to propose whether pedagogy design should include collaborative and cooperative activities.

## Literature review

The proposed implementation of collaborative formative assessment using GroupWork is based on various instructional activities and learning theories that empirical research and educational literature have shown to support improved learning and teaching outcomes.

First, is the opportunity for students to work collaboratively and cooperatively during learning. Presently, competitive and individualistic (Roger & Johnson, 1988) are the more dominant modes of student-student interactions during learning. However, compared to competition and individualism, cooperation and collaboration have been shown to result in students achieving better learning outcomes, more effectiveness interpersonally and students developing positive attitudes towards learning, each other, instructors and subject areas. This practice is also referred to as Peer Instruction (PI), (Crouch & Mazur, 2001; Crouch, Watkins, Fagen, & Mazur, 2007), where class time includes assessment and learning activities where students work individually and in groups.

Second, having activities that allow a group of students to work together stresses the idea of co-construction of knowledge and mutual engagement of participants. A constructivist classroom focuses on student-centered discourse, over the typical teacher centered classroom, where students drive discussions and the teacher serves as a guide on the side (Palincsar, 2005).

Third, our extended use of formative assessment activities is a form of problem-based-learning based on cognitive apprenticeship (Collins, Brown & Newman, 1987) where students learn in the context of solving complex and meaningful problems. The role of instructor is to create appropriate questions, and then guide students on the learning process by encouraging them to think deeply through discussions in evaluating responses to the problems.

And finally, GroupWork is closely modeled on the Think-Pair-Share methodology of classroom-based collaborative active learning (Lyman, 1981; Lyman, 1987; McTighe & Lyman, 1988; Kagan, 1989). In a typical Think-Pair-Share activity, students work on a problem posed by the instructor, first individually, then in pairs or small groups, and finally all participate in a class-wide discussion. Some of the benefits of this strategy include promoting engagement, allowing students to express their reasoning, reflect on thinking, and obtain immediate feedback on their understanding and misconceptions (Kothiyal, Majumdar, Murthy & Iyer, 2013). Moreover, the Think-Pair-Share technique is recommended as an instructional activity that engages learners in higher-order thinking, and as a feedback mechanism, both for students and teachers (Cooper & Robinson, 2000).

## Experimental design

### The class

The experiment was carried out during the fall semester of 2015 in a flipped advanced level Computer Graphics course with enrollment open to undergraduate and graduate students. The course exposes students through a hands-on introduction to the science and practice of rendering three dimensional (3D) images using both resource intensive ray-tracing methods and real-time techniques using the GPU. The course was also intensive in mathematics and programming content and exercises in Java and JavaScript. The course was a typical *flipped* classroom where class meetings started with pre-class reading assignments and reflections, and class times were devoted to interactive teaching and learning activities.

During that semester, 41 students completed the course which consisted of 26 interactive class meetings, with each class lasting for 80 minutes and meeting twice in a week. TeachBack was used in the class to facilitate some of the instructional activities and interactions in the classroom, and this required all students to bring laptops or tablets to class. This included formative assessments, students giving feedback to the instructor, the back-channel assistance forum for asking and answering questions as well as discussions on class materials, and taking notes by students.

### Collaborative assessment using GroupWork

The majority of the exercises were given one-by-one in real-time during classes as a way to assess students' understanding and to inform the instructor on whether to reiterate the just covered concepts or move on to new material. The exercises also counted as an active activity during classes in getting students to actively engage with the material, gauge their own learning, reveal misconceptions, and more importantly engage in discussions and share the learning with others in the class.

Answering questions in GroupWork follows three steps. During the first step, each student takes the time to think and attempt the problem individually and then submit a personal answer which cannot be changed. During the second step, group members share their personal answers and have a collaborative discussion where the team attempts to improve on their initial answers and perhaps reach an agreement on better responses. Groups are made of three students, and they are created using a random assignment of members for each class meeting based on attendance for the day. During the group discussions, the team members can each create one or more new answers based on the level of agreement in the group. Afterwards, each group member is required to submit a final answer by voting on any of the answers from the team. Students are not forced to agree on a single group answer, and that's why a voting system is used, requiring each team member to vote/indicate a final answer. This strategy is intended to encourage a sense of individual accountability and autonomy as shown in cooperative learning theories.

Figure 1 shows a students' view of a question in GroupWork with access to answers panel and the group mini-forum. The interface is designed to include a group discussion area for each question so that group members can still have a discussion in the cases where members are not physically seating next to each other or some members are attending class remotely. In a related study reported in (Tarimo & Hickey, 2016), we studied the effects of offering optional remote attendance to class meetings where students are required to follow along

with a live-streaming of the class and participate in all class activities (including the group activities reported in this paper) using TeachBack. This hybrid style was practical and feasible; however, we did not implement an entirely online mode.

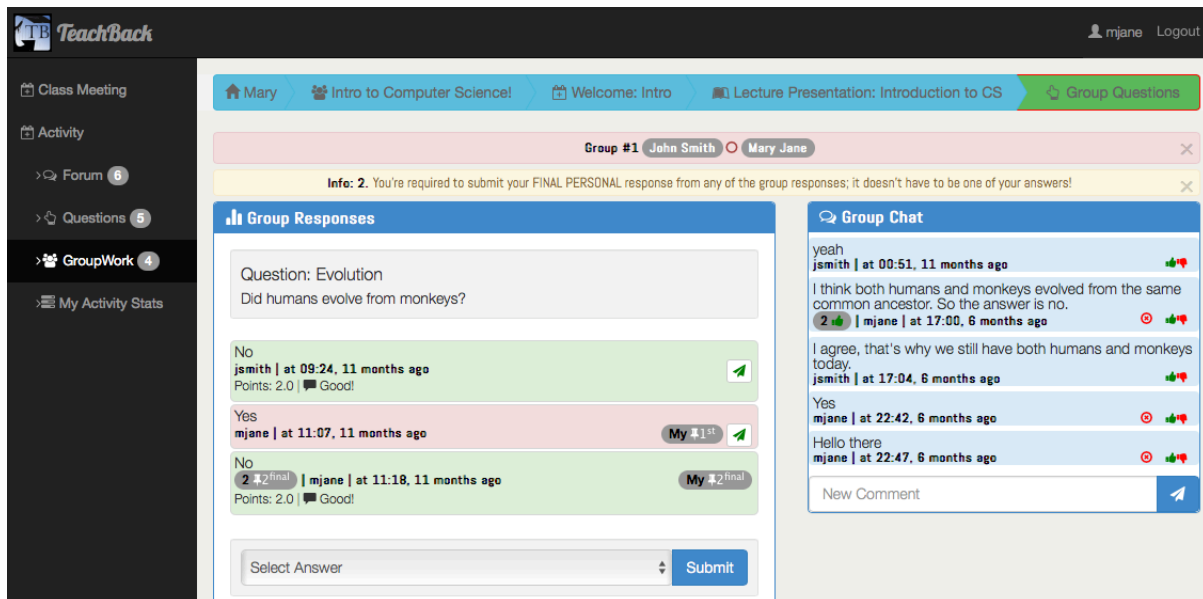


Figure 1. TeachBack: Students' question view in GroupWork.

An important design feature in GroupWork is making sure that group members attempt to answer questions individually before engaging with other group members, or see each other's attempts. In order to achieve this, the interface is designed such that a group member cannot see answers or messages from the rest of the group until that student has submitted his/her individual answer. This design closely models the Think-Pair-Share methodology.

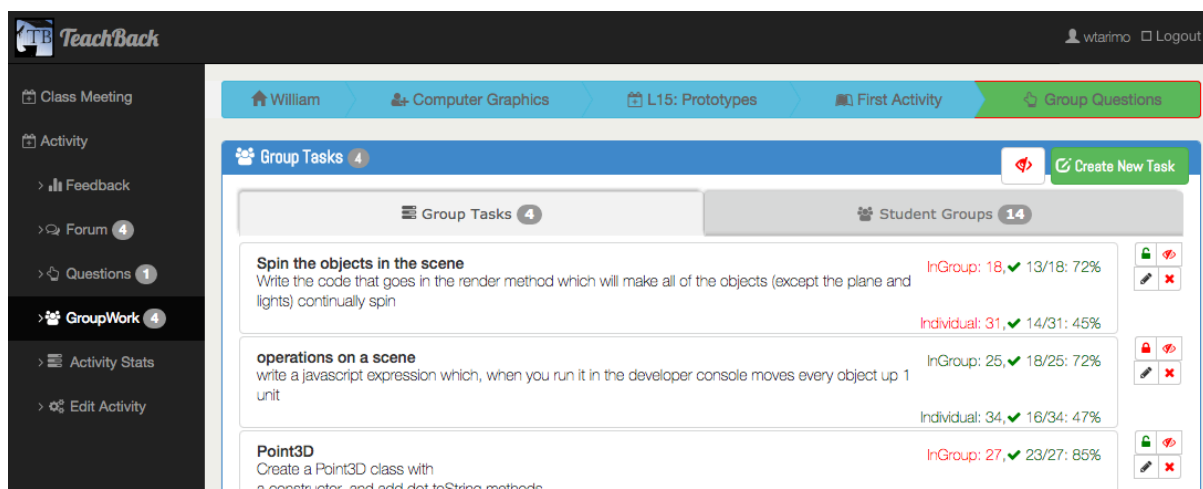


Figure 2. The main GroupWork page listing questions with participation and grading summaries.

In the main GroupWork page, shown in Figure 2, each listed question displays the response counts and grading summary for individual and in-group answers. During class, the counts update in real-time as students respond to questions, this allows instructors to more easily gauge and manage the progress of the assessment. During the course reported in this experiment, an instructor would review the answers and a teaching assistant (TA) would do the actual grading at the same time.

The final stage in a GroupWork exercise is a class-wide discussion that is led by the instructor to go over responses to questions. The interface is designed such that instructors can view and grade both types of answers, see Figure 3. After a question is graded, the percentage of correct answers is also displayed for the

individual and group answers. As seen in Figure 2, this detail provides a convenient way to see and compare the students' performance when responding individually and in groups.

## Data collection

The first analysis involved looking for the evidence to support whether students perform better in formative assessment activities when they work collaboratively in small groups compared to when they work individually. For this, we used the grading data from GroupWork for all graded problems that were given throughout the semester. Each individual and in-group problem response was graded using the same rubric. The correctness of each answer was assessed on the scale of either being fully correct (2.0 points), partially correct (1.0 points) or fully incorrect (0 points), this can be seen in Figure 3.

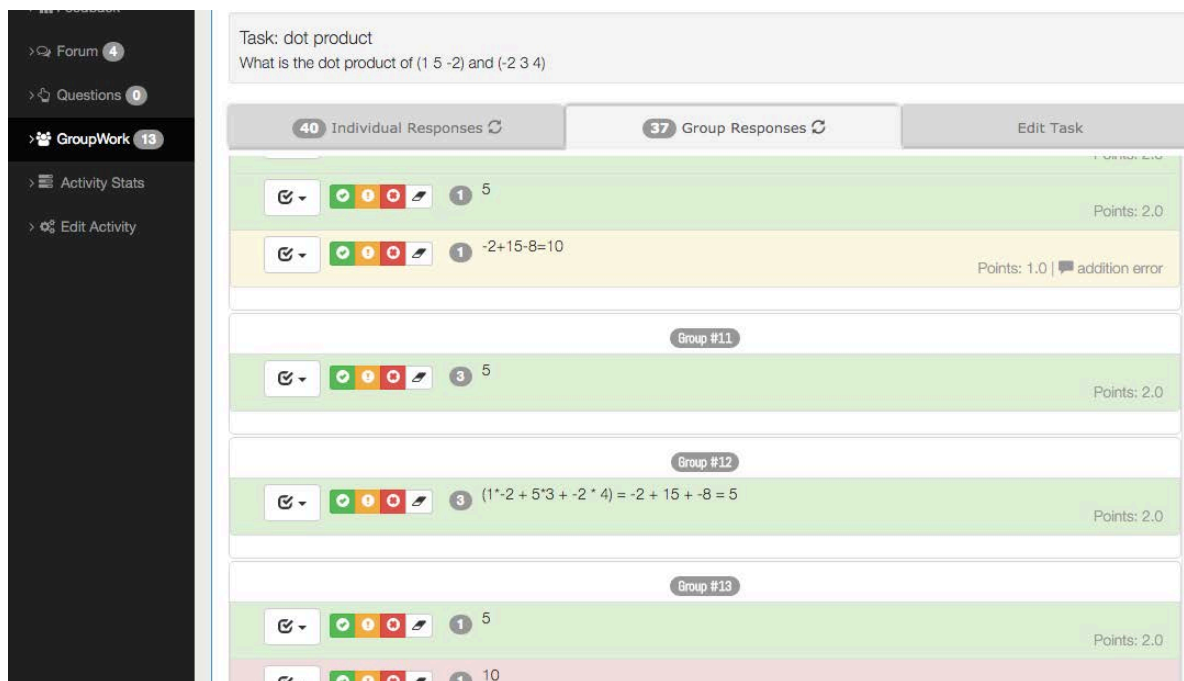


Figure 3. A GroupWork page showing group responses to a question.

The second analysis looked for evidence of correlation between performance in collaborative formative assessments and overall mastery of concepts covered in the course. Overall learning was measured as the course grade earned at the end of the semester which included cumulative grades from weekly or bi-weekly small programming assignments, three major programming projects, and a final written exam. The final exam was a cumulative exam which was administered at the end of the course and covered most of the materials taught in the course. 20% of the official course grade accounted for class participation (based on participation in TeachBack activities and pre-class readings), however, we omitted this participation component in the course grade that was used in our analysis as it didn't exactly reflect mastery of concepts.

At the end of the course a survey was given to students in order to gather their perspectives and evaluation of their learning experiences from the use of collaborative formative assessment and how it was implemented using the GroupWork feature of TeachBack.

## Results

### Learning during collaborative formative assessments

For each student, we looked at the average performance when the student responded individually and when the student responded after the group collaboration stage. Average performance was calculated as the average number of points earned; that is, the sum of points earned divided by the number of questions attempted.

Figure 4 shows a scatter plot of average points earned by each student when responding individually versus when working in groups. The plot demonstrates that, on average, most students tend to earn more points as a result of group collaboration, and this is true even for students who score highly during the individual part.

We used a two-tailed *t* test to compare the mean of the average number of points from the individual and in-group populations. The class average in points when working individually is 0.995 points per question, and the average while working in collaboration is 1.223. The improvement in the mean of 0.228 is statistically significant with a *P* value of less than 0.0001, with a 95% confidence interval of 0.170 to 0.286. This improvement represents an effect size of 0.618, which is calculated as a Cohen's *d* Effect Size (Cohen, 1988). According to Cohen, an effect size of 0.618 is just above the 'medium' range. And according to a survey of academic research on instructional interventions, an effect size of 0.5 or higher would represent significant gains as it is higher than the results found in most instructional interventions (Albanese, 2000; Dubin & Taveggia, 1968).

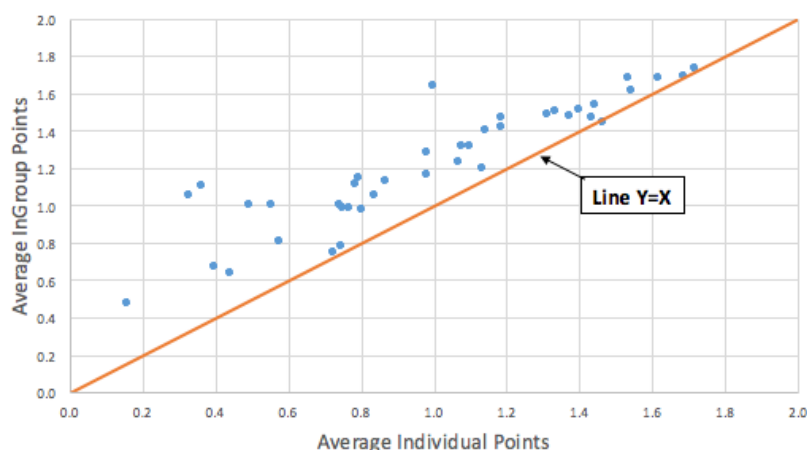


Figure 4. A scatter plot of average individual points vs average in-group points for all students.

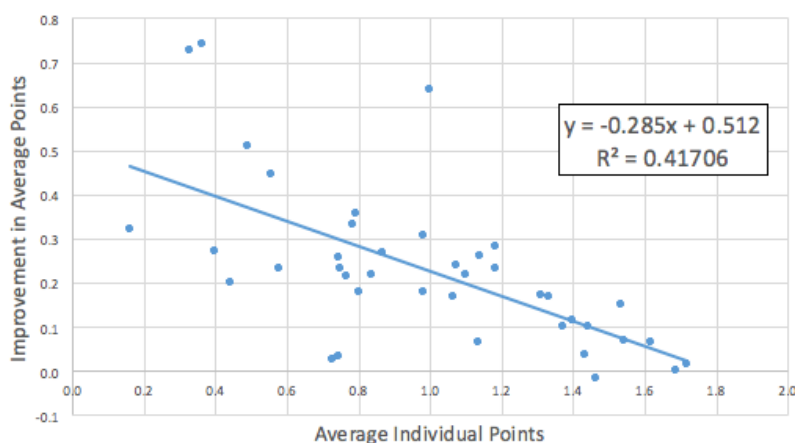


Figure 5. Average improvement in points from individual to in-group responses.

Looking at the improvement in performance resulting from collaboration, Figure 5 shows that students who tend to do poorly when working individually are the ones who take the most advantage of group collaboration. For example, a student with an average individual performance of 0.85 (out of 2.0) points improved performance by 0.28 points, whereas a student with individual performance of 1.7 points only improved by 0.02 points. This is contrary to, for instance, the hypothesis that collaboration would affect all students in the same way.

Secondly, for each GroupWork problem, we compared the average class performance from individual responses and in-group responses. Figure 6 shows a scatter plot of average points earned in each GroupWork question when students responded individually versus when they responded within groups. In only one of the 60 questions asked we see no average improvement in students' performance when they worked in groups.

Therefore, regardless of the question type or difficulty, students' performance for each question improved as a result of collaboration compared to working individually. Again, a two-tailed *t* test was used to compare the means of the average number of points for the two populations representing individual and in-group performances. The mean of average number of points earned increased by 0.233 from 1.003 points per

question to 1.236, and the *t-test* test showed this difference to be statistically significant with a *P value* of less than 0.0001. The 95% confidence interval is 0.190 to 0.276. The corresponding *effect size* of the improvement is calculated as 0.609, which is again considered an impressive gain from an instructional intervention.

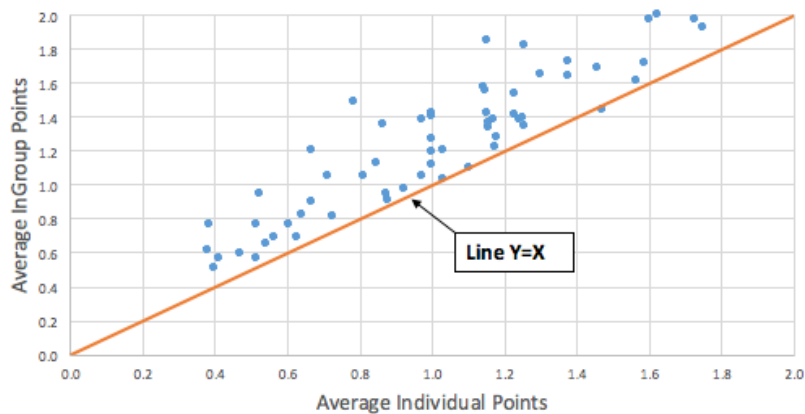


Figure 6. A scatter plot of average individual points vs average in-group points for all GroupWork questions.

### Correlation with overall learning in the course

In this analysis, we looked for evidence of correlation between performance in collaborative formative assessments and overall mastery of concepts covered in the course. Mastery of the material was measured as the course grade earned at the end of the semester which included cumulative grades from weekly or bi-weekly small programming assignments, three major design and programming projects, and the final written exam. Linear regression was used to test the hypothesis that performance in collaborative formative assessment could predict the learning outcomes in the course. Separate regression analyses were used to study the correlation between the final course grade and each of the individual and in-group average points in GroupWork problems. The results of the analyses are shown in Table 1. Both regression models indicate statistically significant, ( $P < 0.0001$ ), positive correlation between course grade and each of the individual and collaborative performances in formative assessment. Indeed, the correlation coefficients are 0.647 and 0.568 between course grade & average individual points, and course grade & average in-group points, respectively.

Table 1: Linear Regression Analysis: Course Grade vs. Average Individual Points and Course Grade vs. Average In-Group Points

Average Points	R <sup>2</sup>	P-Value	Coefficient
Individually	0.4188	< 0.0001	13.074
In Group	0.3232	0.0001	14.429

Figure 7 shows a visual representation of the regression analysis using scatter plots. On the left is a scatter plot of course grade versus average individual points, and on the right, is a scatter plot of the course grade versus average in-group points. And even though both models are statistically significant, performance while working individually is a slightly stronger predictor of course performance compared to performance when working in groups. This observation is still consistent with an earlier study which found significant positive correlation between performance in a course and performance in formative assessment activities where students worked individually. For this class, final course grade is comprised of summative assessments in which students worked individually, this explains why there is a stronger correlation between course grade and individual performance in the GroupWork exercises compared to in-group performance. Moreover, despite the fact that each group member is required to submit a personal answer after group discussions, the group discussions and collaboration has a significant influence on subsequent group answers which in turn averages down the differences in mastery levels among group members.



## Students' opinions

Students were given a survey at the end of the course in order to get their learning experiences and evaluation of GroupWork and its use in facilitating formative assessment exercises. 37 of the 41 students participated in the survey. Using a range of 1 (not useful) to 5 (very useful), students were asked to assess the extent to which the use of GroupWork based collaborative assessment activities was useful towards their learning. 80.5% of the students responded with a rating of 3 or above, with 55.5% believing that it was useful with a rating of 4 - 5. Moreover, using a scale of 1 (not satisfied) to 5 (very satisfied), 75% of the respondents reported being satisfied with how GroupWork tool was used in the course with a rating of 3 - 5. 44.4% of the respondents reported a satisfaction rating of 4 - 5.

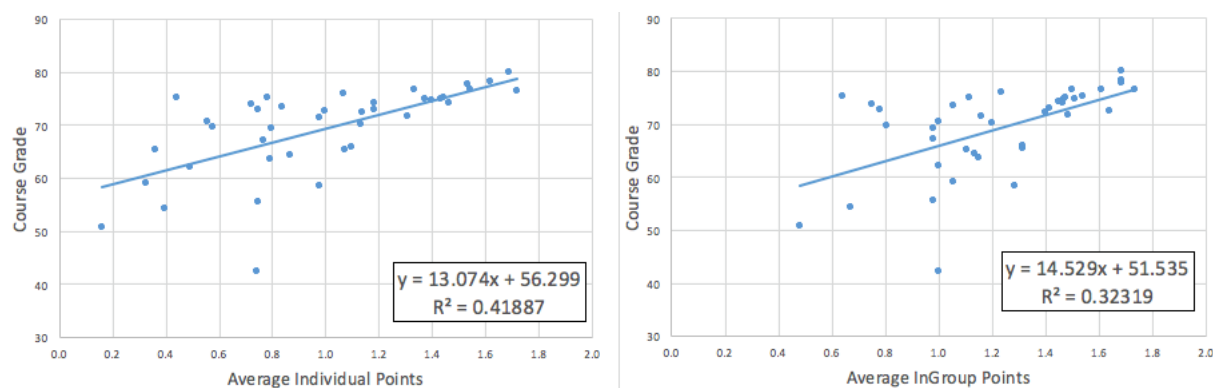


Figure 7. A scatter plots of course grade vs. average individual points, and course grade vs. average in-group points for all students.

Among the aspects of the implementation that students liked were how it forces them to start by attempting problems individually before engaging with a group in discussion and collaboration. The requirement to attempt problems individually was commented to be especially useful in tackling complex concepts as it revealed various personal misconceptions and confusions that would then be corrected during group and class discussions. Moreover, enforcing individual attempts results in many alternative answers to problems which would then be refined to fewer in-group answers. Students pointed out that having access to various alternative answers to a problem opens a new opportunity to learn from seeing how others think and approach answering questions.

Students also appreciated the introduction of group-based formative assessment using GroupWork as it “provided the opportunity for collaborative activities in class”. GroupWork activities were part of most of the class meetings during the semester, and as a result, student collaboration and formal discussions were part of the class routine. This introduced the component of collaboration to the pedagogy that would otherwise not be there as in traditional classrooms. Working with ever changing groups of classmates was pointed out by students in the survey as enabling them to “get to know other(s)” in the class.

## Conclusions

The findings of this experiment demonstrate the benefits of cooperative, collaborative and problem-based learning in group-based formative assessment activities. The data analysis compared students' performance in formative assessment activities when they worked entirely individually and when they worked in small groups using the Think-Pair-Share methodology. Results showed that students reached a significantly higher level of achievement when formative assessment activities included collaboration. In general, all students performed better when they worked in groups than when they worked individually alone. Similarly, students' performance was higher for each assessment problem when students worked in collaboration.

Furthermore, the results show that students benefit in various amounts from collaboration based on individual student's mastery levels. Even though all students gather benefits from collaboration, it is the students who would usually do poorly when working individually who benefit the most. This sounds like a trivial outcome but in this case the benefit comes from having heterogeneous groups where groups are randomly generated for each class meeting. When it comes to choosing students assignment into groups based on individual capabilities, heterogeneous groups tend to be more powerful than homogeneous groups (Roger & Johnson, 1988). This is because in collaboration learning comes from discussion (presentation and arguing of

different perspectives and alternatives), explanations (to other group members, especially those in need of help), justification of answers, and shared resolutions.

The results demonstrated positive correlation between overall performance in the course and students' average performance in formative assessments when working individually and in groups. Performance working individually has a slightly stronger correlation compared to in-group performance. And since we were interested in studying the impact of collaboration on overall learning, a better study approach would have been comparing course performances between experimental and control populations of students, one with collaborative formative assessment and the other with only the traditional individualistic formative assessment.

In this work, we have shown an implementation of collaborative formative assessment that is practical and feasible. Students thought the implementation of was beneficial and useful towards their learning. Students were generally satisfied with the technical implementation of GroupWork even though they had recommendations for improvement. The demonstrated benefits to the teaching and learning outcomes imply that this instructional approach is worthy of further adoption, investigation and testing in various subject areas and curriculum designs.

## References

- Albanese, M. (2000). Problem-based learning: why curricula are likely to show little effect on knowledge and clinical skills. *Medical education*, 34(9), 729-738.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* Lawrence Erlbaum Associates. Hillsdale, NJ, 20-26.
- Collins, A., Brown, J. S., & Newman, S. E. (1988). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. *Thinking: The Journal of Philosophy for Children*, 8(1), 2-10.
- Cooper, J. L., & Robinson, P. (2000). Getting started: Informal small-group strategies in large classes. *New directions for teaching and learning*, 2000(81), 17-24.
- Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American journal of physics*, 69(9), 970-977.
- Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. *Research-based reform of university physics*, 1(1), 40-95.
- Dubin, R., & Taveggia, T. C. (1969). *The teaching-learning paradox: A comparative analysis of college teaching methods*. Eugene; Center for the Advanced Study of Educational Administration, University of Oregon, 1968. 78 pp.
- Hickey, T. J., & Tarimo, W. T. (2014). The affective tutor. *Journal of Computing Sciences in Colleges*, 29(6), 50-56.
- Kagan, S. (1989). The structural approach to cooperative learning. *Educational leadership*, 47(4), 12-15.
- Kothiyal, A., Majumdar, R., Murthy, S., & Iyer, S. (2013, August). Effect of think-pair-share in a large CS1 class: 83% sustained engagement. In *Proceedings of the ninth annual international ACM conference on International computing education research* (pp. 137-144). ACM.
- Lyman, F. (1981). The responsive classroom discussion: The inclusion of all students. *Mainstreaming digest*, 109-113.
- Lyman, F. (1987). Think-pair-share: An expanding teaching technique. *MAA-CIE Cooperative News*, 1(1), 1-2.
- McTighe, J., & Lyman, F. T. (1988). Cueing thinking in the classroom: The promise of theory-embedded tools. *Educational Leadership*, 45(7), 18-24.
- Palincsar, A. S. (2005). 12 Social constructivist perspectives on teaching and learning. *An introduction to Vygotsky*, 285.
- Roger, T., & Johnson, D. W. (1988). Cooperative learning: Two heads learn better than one. *Transforming Education: Awakening the Full Human Potential... in Everyone*, Context Institute, Langley, WA, 34-36.
- Tarimo, W. T., Deeb, F. A., & Hickey, T. J. (2015, May). A flipped classroom with and without computers. In *International Conference on Computer Supported Education* (pp. 333-347). Springer International Publishing.
- Tarimo, W. T., Deeb, F. A., & Hickey, T. J. (2015). Computers in the CS1 Classroom. In *Proceedings of the 7th International Conference on Computer Supported Education (CSEDEU) (2)*, 67-74.
- Tarimo, W. T., Deeb, F. A., & Hickey, T. J. (2016). Early detection of at-risk students in CS1 using TeachBack/Spinoza. *Journal of Computing Sciences in Colleges*, 31(6), 105-111.
- Tarimo, W. T., & Hickey, T. J. (2016, October). Fully integrating remote students into a traditional classroom using live-streaming and TeachBack. In *Frontiers in Education Conference (FIE), 2016 IEEE* (pp. 1-8). IEEE.
- Tarimo, W. T. (2016). *Computer-Supported Agile Teaching* (Doctoral dissertation, Brandeis University).

# Explanation-Giving in a Collaborative Tangible Tabletop Game: Initiation, Positionality, Valence, and Action-Orientation

Alyssa Friend Wise, New York University, alyssa.wise@nyu.edu

Alissa Nicole Antle and Jillian Warren  
aantle@sfu.ca, jlw29@sfu.ca  
Simon Fraser University

**Abstract:** Explanations given to each other by 20 pairs of 5<sup>th</sup> grade children while playing a tangible tabletop sustainability game were analyzed inductively for key themes relating to their use of language, gesture and system tools. Half the pairs had been assigned roles (human development or natural resources manager) with associated system controls. Findings showed that explanations by pairs in both conditions often employed collectivist language (“we”) in conjunction with positive reflections on the game-world state using the provided Impact Tool which gave feedback while system was paused. Pairs in the roles condition also gave explanations in response to partner actions and more frequently included negative and action-oriented prospective language about what should be changed moving forward. Roles pairs additionally used questions to seek confirmation or action from their partner and made comments from the perspective of the inhabitants of the fictional world. Implications for the research and design of collaborative tabletop learning systems are discussed.

**Keywords:** Tangible systems, discourse, student roles, positive interdependence

## Collaborative learning with interactive tabletops and tangibles

Interactive tabletops and tangibles are technologies that allow for physical interaction by users directly with the digital surface (Higgins et al., 2011) and/or via digitally augmented objects that are recognized by the system (Ullmer & Ishii, 2000). They are particularly exciting for supporting face-to-face collaborative learning for multiple reasons (Fernaes & Tholander, 2006; Dillenbourg & Evans, 2011; Speelpenning et al, 2011) such as the ability to have multiple simultaneous users and create a shared transaction space for reference, negotiation, and action. Of particular interest in this work is the ability to facilitate joint attention through the visibility of action, the possibility to engage multiple modes of communication (e.g. speech, gesture, system action), and the opportunity to generate positive interdependence through the physical embodiment of distributed control. Despite these affordances, opportunities for collaboration aren’t always taken up by children. In addition to such obvious problems as domination by one child, independent parallel play, and competition (Fleck et al., 2009; Marshall et al., 2009), children can also work together to complete tasks successfully but without deep engagement around the issues involved. For example, in our prior work studying children’s collaboration using a tangible tabletop sustainability game, we found that pairs worked together for the entire duration of system use, but only talked in depth about their sustainability choices 5% of the time (Wise et al., 2015). For the remaining 95%, the children engaged in other forms of interaction in which they made game-based decisions, but without substantive explanation of their thinking. This mode of operation is parallel to quick consensus building in the context of argumentation, in which learners are simply “accepting the contributions of the learning partners in order to move on with the task” (Weinberger & Fischer, 2007, p. 77). This stands in contrast to richer forms of consensus building which are integration-oriented or conflict-oriented in nature. Fundamental to both of these activities is the opportunity for students to experience a change in perspective based on hearing the ideas and reasoning of their peers (Teasley, 1997). Thus getting children to share their thinking may have benefits for reflective processes, collaboration, and learning (Price et al., 2003), and a greater understanding of what leads children to explain their ideas in a tangible tabletop learning environment and the characteristics of this language that might lead to it being accepted by their partner is of central concern to the CSCL community.

## Giving explanations, positive interdependence and the Youtopia system

One way to encourage children to share their thinking is by setting up a situation which requires their complementary participation for success. Tangible systems can be used to create such situations through the technological interdependence of different objects’ use (i.e. more than one input action must be taken sequentially in order to create a successful system response) (Ullmer & Ishii, 2000; Antle, 2015). This can be

further layered with social interdependence in which the objects are assigned to children along with a set of duties, rights and responsibilities (Wise et al., 2015). When multiple actions are needed to produce a certain outcome and children require the assistance of their partner to enact these actions, it should encourage them to articulate and explain their thinking to each other. Previously we have described a system designed to enact these ideas: Youtopia is a tangible and multi-touch tabletop activity about sustainable land-use planning that includes co-dependent access points and the option to add scripted roles (with associated tools) to the existing contingencies of tangible use (Antle et al., 2013; Fan et al., 2014). In a prior study we measured the degree of explanation-giving by 20 pairs of 5<sup>th</sup> grade children using Youtopia in an authentic school environment with and without roles (Wise et al., 2015). Results showed pairs in the assigned roles/controls condition gave a greater number of explanations to their partners about what they wanted to do in the game. In this follow-up study we use inductive qualitative methods to unpack this finding and probe the reasons why this occurred.

## Methods

### Research question

What kinds of talk, gesture and tool use characterize explanation-giving in pairs of children playing a tangible tabletop sustainability game with and without assigned roles/controls?

### Youtopia design

Youtopia is a hybrid tangible and multi-touch tabletop application about sustainable land-use planning in which children have the opportunity to design their own world, exploring how different land-use decisions affect the amount of food, housing and energy provided to the population, and the impact these decisions have on the level of pollution in the environment. Following the principles of Emergent Design (Antle et al., 2014), our interaction goals were for children to explore the relationships between different land-use decisions, see their effects on the world, discuss with their partner the inherent tradeoffs between meeting human needs and causing pollution, and through this make informed decisions to create a world they would want to live in. Children begin with an undeveloped map of mountains, grasslands, forest and a river. They interact with the tabletop through two kinds of physical stamps that designate different land-use types (see Figure 1a): natural resource stamps (tree icon); and human development stamps (wrench icon) [see Table 1]. Each stamp also has a picture and a label describing the land-use type, and color is used to indicate land-uses that relate to the same human need.

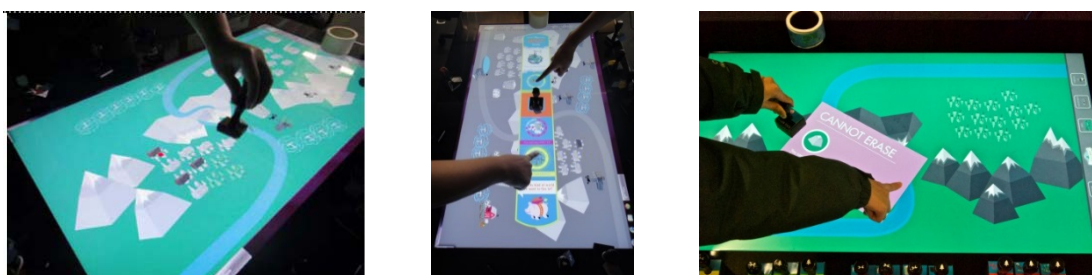


Figure 1. (a) Stamping a land-use (b) Assessing world-state from Impact Tool (c) Attempting to erase a land-use.

Land-use types have predefined relationships to each other and to the terrain designed to reflect real world relations (see Table 1). For example, a farm can only be built on grasslands and requires irrigation connecting it to the river. Thus, while each stamp is used individually, system interdependencies require multiple stamps to be used together to build the Youtopia world. Children can learn about these relationships and their effects on the world through trial and error with educative error explanation tabs, using the open circular tangible “info ring,” and by assessing the state of the world with the Impact Tool that provides visual information about pollution levels and the proportion of the population that has shelter, food and energy (see Figure 1b). The Impact Tool freezes the map (so no other interactions with the system can take place while it is in use) and includes a diagnostic touch functionality: children can touch any of the four rings indicating levels of food, shelter, housing and pollution to highlight the elements in the world contributing to it. The Impact Tool also contains an image of a pig that asks children “Is this a world you want to live in?” with the goal of eliciting a discussion of tradeoffs. An Eraser Tool (see Figure 1c) is available to remove land uses children want to eliminate or replace without penalty, allowing for experimentation and incremental building. A full description of the system design is given in Antle et al. (2013) and a short video of functionality is available at [tinyurl.com/youtopiagameplay](http://tinyurl.com/youtopiagameplay).

Table 1: Types of Youtopia Land-use Stamps

Area of Human Need	Natural Resource Stamps (tree)	Human Development Stamps (wrench)
Food (green labels)	Garden, Farm ←	Irrigation**
Shelter (pink labels)	Harvest Lumber →	Houses, Town Houses, Apartments
Energy (yellow labels)*	Coal Mine →	Coal Plant
Environment (orange labels)**	Forest, River & Mountain Reserves	Hydro Plant***

Arrows indicate land-uses required for other land-uses. \* Energy land-uses *increase* pollution. \*\* Environment land-uses *reduce* pollution \*\*\* Irrigation and hydro plants use-up river water; this is reflected in a visual display of the water level

## Youtopia use

Forty 5<sup>th</sup> grade children (ages 10-11, 18 boys / 22 girls) from two classrooms used Youtopia in pairs (N=20). All children had extensive experience collaborating and participated in a class unit on sustainability issues four months earlier. Use of Youtopia was introduced as a review of the sustainability unit in which children would get to spend up to 25 minutes building a world they wanted to live in to share with the class. Half of the pairs were randomly chosen to be given assigned roles/controls with role types balanced across gender in the overall sample. For these pairs, one child was randomly assigned to be the “natural resource manager” with all the “tree” stamps (lumber; garden; farm; coal mine; nature, river, mountain reserves) placed in front of them on one side of the table; the other child was randomly assigned to be the “human development manager” with all the “wrench” stamps (irrigation; (town)house; apartment; coal plant; hydro dam) placed in front of them on the other side of the table. Use of assigned tools was not explicitly enforced; the design relied on social norms around ownership of items on tables (e.g. you use the cutlery placed in front of you) and the explicit labeling of the tools with the wrench/tree icons that identified them as associated with a particular role. Tools not associated with a role (Impact Tool; info ring; eraser) were placed at the end of table between the children. For pairs who were not given assigned roles/controls, all of the stamps were placed at the end of table between the children.

## Data collection and explanation identification

Videos of all twenty pairs’ sessions were collected. Youtopia use (using one of two identical installations) was conducted in separate rooms apart from the classroom to minimize distraction and extraneous noise. A high-definition digital video camera captured a landscape view of the children and oblique view of the tabletop.

One (roles) pair’s data was removed from analysis as an outlier due to a lack of substantial talk or engagement throughout the session. Time-periods in which children explained their thinking or reasoning related to decisions about what resources and developments to use were identified in the remaining video data. For example “Let’s build houses, not apartments-they use less lumber so we can make more nature reserves” would be coded as an explanation but “I think we should have houses not trees” would not. Three researchers achieved acceptable inter-rater reliability ( $\kappa > .63$ ) with all differences reconciled. A total of 58 explanations were identified across the 10 non-roles pairs while 80 explanations were found for the 9 roles pairs.

## Analysis

Data analysis was conducted inductively following the constant comparative method (Gibson & Brown, 2009) with particular attention to the concerns of video-based data (Derry et al., 2010).

### Phase 1: Transcription

All 138 explanations were transcribed into a text file. With the goal of understanding collaborative processes, the time-span of an explanation included events that led up to the explanation as well as any response or reaction expressed by the partner. Transcripts thus captured (a) all verbalizations during the coded time period and two to four related turns of talk before and after; (b) indication of the physical actions taking place during / between turns of talk including (i) body position, (ii) tool use, (iii) gestures, (iv) facial expressions, and (v) child location around the table. In addition, the transcriber provided a global overview of each event that indicated what aspect of the Youtopia world the explanation was about, what occurred leading up to the comment, and any response or reaction expressed by the partner. An example for one explanation is shown in Figure 2.

### Phase 2: Open and confirmatory coding of non-roles explanation transcripts

Three researchers individually read all the transcripts for the non-roles pairs and using constant comparison generated a list of potential themes seen in the data. The goal at this stage was to be as inclusive as possible in

retaining themes. The researchers then shared and discussed proto-themes, condensing and combining similar ideas to create a master list of 11 possible themes. Each potential theme was then subject to individual scrutiny with a search through the transcripts for confirming or disconfirming evidence. Each of the three researchers took primary responsibility for a theme, but work was conducted collaboratively with extensive consultation. Importantly, attention was paid not only to how frequently evidence of a theme was found across the whole corpus, but also its existence across multiple episodes and groups. In this round of coding, two similar themes relating to use of the Impact Tool and the initiation of explanations were combined and two related to the expression of emotion were dropped due to lack of evidence. This left a total of 8 themes.

C1 (f) Human Devel.	C2 (m) Natural Resour.	Verbalizations and Physical Actions Surrounding C2 Explanation in R3@10:15
	X	And then what should we do? [ <i>puts Impact Tool on screen</i> ]
X		[ <i>reading from Impact Tool</i> ] Is this the kind of world...? [ <i>pointing to the text</i> ] No. We need more power.
	X	I don't know about using coal because we don't want to make it too polluted.
X		A dam.
	X	Sure, you have it.
X		Hydro dam [ <i>picks up stamp and places it on river</i> ] there.

Figure 2. Sample transcript of an explanation by child C2 in Roles Pair R3 at 10 min 15 sec into the session.

### Phase 3: Open and confirmatory coding of roles explanation transcripts

The above process was repeated with the roles transcripts while carrying forward the 8 already identified themes. In the open coding, 4 additional potential themes were identified. Thus, a total of 12 possible themes were examined in the confirmatory coding. As the data was examined, two of the newly identified potential themes were incorporated into an existing theme due to similarity in content, and one theme related to the use of the Eraser Tool was dropped due a lack of evidence. In addition, the final new theme and three existing themes which all related to aspects of articulation of arguments and tradeoffs were combined into a single theme. The existing themes whose scope was altered or expanded during analysis of the roles condition were taken back for follow-up examination with the non-roles pairs' transcripts. This resulted in a total of 6 themes relating to explanation-giving whose presence or absence and characterization are reported across roles and non-roles pairs.

## Findings

The six themes described below speak to how explanation-giving was initiated as well as the perspectives the explanations represented and the temporality, valence and positionality expressed within them.

### Theme 1: World-state-versus partner-initiated comments

Explanation-giving often occurred as a response to the world-state; in the roles condition explanations also commonly occurred in response to a partner action or comment

In *Non-Roles* pairs, explanations were most often made in response to the state of the world. This was commonly prompted by use of the Impact Tool which provided feedback on pollution levels and the meeting of human needs (e.g. "So how are we turning out? [*puts Impact Tool on screen*] Food is pretty good. I think it is really good because there is no pollution." C1 in N8@15:58). Children also gave explanations of their thinking about the world state as depicted on the map (e.g. "I don't like how we took up all the trees but [*rubs hands on face*]...like...I think it's good." C1 in N1@18:03). In a more limited number of cases, non-roles children gave explanations in response to something said or done by their partner (e.g. "But we don't want to put it near a mountain, it will kill the goats and stuff. That's sad." C1 in N9@11:42 responding to their partner's suggestion about where to put a coal plant). In *Roles* pairs the same kinds of situations initiated explanations with one notable difference: there were a substantial number of both world-state *and* partner-initiated explanations. In many cases these explanations appeared to be related to stamp ownership by the two roles (e.g. "Yea, because then most people will have food." C2 in R6@18:54 responding to their partner, who had the garden stamp but had not yet used it, making a hesitating statement that only "maybe" stamping it was a good idea). Similar initiation of explanations occurred even when the action in question would be taken with the Eraser Tool which both children had access to. This indicates a tacit acknowledgement of the ownership of decisions related to particular land use types in roles pairs (e.g. "These ones? [*motioning to some apartments on the map*] maybe we \*should\* lose some apartments. They're too cramped up in that area." C2 in R5@16:19, who was responsible for human development, responding to their partner's idea that they should remove some of the apartments).

## Theme 2: Retrospective versus prospective temporality

Most explanations were retrospective reflections on the state of the world, but in the roles condition there were also action-oriented prospective explanations about what should be done to change it

Explanations in *Non-Roles* pairs were mostly retrospective reflections on the existing state of the world (e.g. “Wait, if you look at our pollution. See now our pollution is fine.” C1 in N3@18:55). At times these also included a call to action about what should be done (e.g. “There’s little pollution! We’re in the good zone! Let’s just put one more.” C2 in N6@16:28). In addition, there were a small number of statements of a timeless nature describing what children valued in the world (e.g. “There should be at least many [people who have food].” C2 in N2@8:57) and ones which described an anticipated future (e.g. “Yeah, it will be nice to have some animals.” C2 in N6@3:21). In *Roles* pairs the temporality of explanations distributed into the same four categories; however, there was a much greater presence of prospective calls-to-action about what should be done. These occurred both in combination with reflections on the existing state (as was seen for non-roles pairs, e.g. “Oh man, a lot of the water’s gone [*looking at the brown river*], do you think we need to kill the hydro dam? C1 in R8@12:56”) but also on their own as stand-alone comments (which was seen less frequently non-roles pairs, e.g. “I don’t know about using coal because we don’t want to make it too polluted.” C2 in R3@10:20).

## Theme 3: Collectivist versus partner-directed language

There was a strong use of collectivist language (“we”) in both conditions; in the roles condition language referring to the learning partner (“you”) was also commonly seen

Within *Non-Roles* pairs there was a much greater use of first person plural language (e.g. “Okay so that means we need [to use] more trees which is kind of sad.” C1 in N1@3:30) than first person singular (e.g. “I think it is really good because there is no pollution.” C1 in N8@13:05). When employed, use of the first person singular was almost always followed by an opinion verb such as think or feel (as in the example above) and often also connected with a reference to the collective (e.g. “I feel like we’re using up too much of the trees.” C2 in N4@9:55). A similar pattern was seen for *Roles* pairs with the addition of a greater use of the second person singular (“you”) in combination with a reference to the collective (e.g. “Do you want to try another irrigation to get more food ‘cause that’s the only way big problem we have?” C2 in R1@25:00).

## Theme 4: Positive versus negative language

Non-roles pairs primarily used positive language in giving explanations while roles-pairs also used negative language to oppose things their partner had said or done, and used questions to seek confirmation, agreement or action from their partner

Children in *Non-Roles* pairs gave explanations largely using positive terms. Typically a child made a statement about the state of the world, followed by a positive comment (e.g. “Only a little [pollution]. That’s good.” C2 in N4@4:54). Other positive statements referred to the state of the world directly (e.g. “We have a pretty darn good world here.” C2 in N6@17:10). Non-roles children less frequently used negative terms in their explanations or opposed something their partner had said. When they did, these tended to be tied to action-oriented phrases about what should be changed (e.g. “There’s not enough water though, maybe we need to erase the fish.” C1 in N5@11:53). In contrast, children in *Roles* pairs expressed their explanations using a variety of neutral, positive and negative statements as well as questions. Notably, there were many more explanations in that had a negative valence; children often offered an explanation as they opposed something their partner had just said or done (e.g. “But this like pollutes though, remember?” C2 in R2@12:43 responding to their partner stamping a coal plant on the map). Often these opposing statements involved negative terms and/or suggested undoing what the other learner had just done (e.g. “Maybe, ugh ... Water brown?” [*erases the irrigation they have just stamped in response to their partner’s suggestion that they build more gardens and a farm*] “I don’t want to do that ‘cause then the fish don’t have much water.” C2 in R5 15:02). Opposition was also enacted at times in the form of a question which opened up a space for the partner to share their thoughts in response to the difference in opinion (e.g. “Why is the garden so far from the houses?” C2 in R2@3:14 in response to C1 placing an apartment stamp). Directing questions to the learner partner was also a frequently occurrence in roles pairs’ explanations, even when opposition was not present. For example children often sought confirmation, agreement or action from their partner either directly (e.g. [*placing Impact Tool*] “Not everyone has shelter... Do you want to take out some parks?” C1 in R8@8:48 [C2 who is responsible for natural resources responds by erasing three nature reserves]) or indirectly (e.g. “Energy would be coal plant and hydro dam [*points at both stamps which belong to their human developer role*]. Both cause pollution. This is less [*waves hand above hydro stamp*]. Maybe hydro dam? Where can it fit?” C1 in R9@18:59).

## Theme 5: Holistic versus positioned perspective

Children in non-roles pairs gave explanations when the environment was healthy or human needs were met, often shifting to focus on the balance between these towards the end of a session; in contrast children in roles pairs gave explanations when pollution was high or human needs were not met and many of the pairs aligned their comments with their roles for the majority of the session.

Children in *Non-Roles* pairs gave explanations about three main aspects of the Youtopia world. First, when the environment was relatively healthy (e.g. “Okay there’s little pollution. That’s way better.” C2 in N1@7:25). Second, when human needs were met (e.g. “Oh wow... [*pointing to full shelter indicator on impact display*] ...that’s good!” C1 in N7@8:14). Third, about the balance or trade-offs balance between a healthy state of the environment and human needs being met (e.g. “Well I want everyone to have energy but then we might have more pollution.” C2 in N2@16:42). More often than not, these comments were positive in nature as noted in the previous theme (e.g. [*looking at impact display rings*] “There’s lots of pollution but everyone has energy now.” C1 in N3 @23:06) though at times children also recognized the difficulty of the choices that needed to be made (e.g. “This is really hard trying to keep the water full and having stuff as well.” C1 in N10@13:06). In some non-roles pairs children talked about trade-offs and advocated for a balanced world from the beginning; however in others they initially advocated for either people’s needs or the environment, only shifting to consider balance later in the session (e.g. “I want everyone to have energy.” C2 in N2@16:45 versus “All I wanna do is add more energy but then that’s gonna add more pollution.” C2 in N2@20:08). In contrast, children in five of the Roles pairs advocated according to their assigned role for most of the session. As described in the two prior themes, these explanations were often phrased in the negative and oriented towards taking action. For example, the natural resources manager gave explanations when pollution was high or water levels were low (e.g. “Oh no no no, we don’t need that happening [*erases irrigation*] we need to preserve some water [*tries a river reserve, water turns brown again, erases it the reserve*] that’s not good.” C2 in R5@15:31) while the human development manager gave explanations when human needs were not met (e.g. [*looking at Impact Tool*] “Not everyone has shelter...do you want to take out some parks?” C1 in R8@18:54). In the other four pairs, children gave explanations that were both aligned and in contrast with their assigned role. Near the end of their session, they shifted away from the roles to address questions of balance, making comments that showed a recognition of the trade-offs involved (e.g. “Now there’s a little pollution but I think we’re almost at full energy.” C2 [natural resource manager] in R7@16:14).

## Theme 6: Connected versus detached language

Children in roles pairs gave explanations that included the perspectives of the world’s inhabitants

An additional theme found in the examination of explanations in *Roles* pairs was comments about the experiences or feelings of the inhabitants of the world. There were a substantial number of these statements relating to a variety of things such as: living conditions (e.g. [*smiles*] “Oh, they’re neighbors.” C1 in R5@5:17); the availability and proximity to food (e.g. “But should we do like houses around the garden? So they stay alive with food and stuff...” C2 In R2@4:28); the impact of pollution (e.g. “Do you want to put this over there so these people don’t have like the pollution from that?” C1 in R1@20:51); and lifestyle concerns (e.g. “I think we should make a few because people like nature reserves don’t they...” C1 in R7@5:08). What the comments all had in common was thinking about or from the perspective of the people living inside of the system. Such comments were seen much less frequently in the *Non-Roles* condition where the vast majority of explanations took a detached “gods-eye” view (e.g. “Okay there’s little pollution. That’s way better.” C2 in N1@7:23).

## Discussion

### Explanations and interdependent roles/controls

The impact of positive interdependence induced via assigned roles with associated controls on explanation-giving was the core focus of this research. Prior work had found that the overall *amount* of explanation-giving was increased by this strategy (Wise et al., 2015). The present finding that, for roles pairs only, a substantial portion of explanations were given in response to a comment or action made by their partner, begins to offer some insight into why. Roles pairs asked questions of their partner to seek confirmation, opposed things their partner had said or done, and used “you” language to request (or direct) their partner to take a particular action. These behaviors appear to stem directly from the distributed ownership and interdependence of the system controls. For example, opposition was stimulated by the fact that actions taken by one child always had implications for the other (Antle, 2015). Similarly, questions or requests for action were necessitated by the fact that providing for human needs (food, shelter, energy) required using at least one stamp assigned to each



partner. Notably, while use of assigned tools was not enforced, no violations of the assignment occurred. Drawing on Rick et al.'s (2009) finding that children took more responsibility for the parts of the tabletop surface closer to their relative position, this may be in part due to the initial presentation of role stamps on opposite sides of the table. These norms of social ownership (Speelpenning et al, 2011) related to land-use were strong as they even extended to use of the shared Eraser Tool to remove elements associated with each role.

The extent to which the different character (and additional quantity) of roles pairs explanations has implications for the children being open to, or actually changing, their ideas is an important areas for future research. Children's efforts to seek confirmation or agreement from their partner about what to do in the game suggest some attempt towards the desired end of thinking beyond their own personal views (Teasley, 1997). This is also supported by the finding of comments considering the welfare of the world inhabitants by roles pairs. In addition, oppositional explanations are potentially valuable for stimulating learning since they push the children to reconsider their ideas about what an ideal world looks like in ways that positive evaluations do not (Weinberger & Fischer, 2007). Conversely, many children in the roles pairs stayed on "their" side of the issues (human needs or the environment) for the majority of the session, rather than progressing to a more balanced position acknowledging the tradeoffs between them. This suggests a design strategy first promoting an oppositional stance and then later switching to a togetherness stance. One approach to this could be via game phases where roles/controls are scripted/assigned at the beginning but deliberately released partway through.

### Explanations and the impact tool

Across all pairs, the Impact Tool was commonly used in conjunction with explanation-giving. There are several design features that may have contributed to this. First, use of the Impact Tool paused interaction in the game. The ensuing talk during these pauses suggests the effectiveness of Antle's (2015) recommendations to help children "step-out" of the tangibles action and create space for them to explain their thinking. Second, the tool brought up a status screen that provided feedback on the world-state (levels food, energy, shelter and pollution) and was jointly available to both children with a prompting question ("Is this a world you want to live in?"). Previously we found that to be shared by group members, tabletop elements must be readable and reachable from multiple sides of the table (Antle et al., 2011). This appeared to be effective here in eliciting children to share their evaluations of the world and the reasons for them with each other. The visualization providing this feedback was intentionally value-neutral (see Emergent Design Principles, Antle et al., 2014): the pig's speech bubble asked "Is this a world you want to live in?" and the circular scales for pollution, shelter, food, energy could be viewed from either a "half-full" or "half-empty" perspective. While non-roles roles pairs tended to use this referential anchor to comment on what was (already) good in the world, roles pairs also used it to point out could be better. The root cause of this difference is not clear, but it is potentially related to a greater feeling of responsibility for the world imparted with the assignment of roles (Wise et al., 2012). There is evidence for this in the connected language used by roles pairs to talk about the experiences of people in the world (rather than the detached language used by non-roles pairs). Whether a heightened sense of responsibility and forward-oriented talk is advantageous for learning remains to be examined; we suspect there may be benefits for cognitive engagement from children feeling accountable for their activity and actively comparing the world-state with the one they would like to build, rather than simply admiring the current one (Chi & Menekse, 2015). This can be tested empirically with designs that intentionally lead children to take one perspective or the other (for example the pig could ask different questions: "What is good about this world?" versus "What can make this world better?") and evaluate the resultant talk and learning. The optimal situation may involve fluid flow between the two perspectives; this could be encouraged in various ways (e.g. the system rotates between different prompts or prompts whichever perspective is less represented). A third feature of the Impact Tool that may have been important specifically for roles pairs was the diagnostic touch functionality that showed the elements contributing to world's food, shelter, housing and pollution. This functionality was often used to identify and explain the causes of a dissatisfying situation leading to a suggestion for change. The presence of an Eraser Tool that allowed children to endlessly undo and redo actions also may have played an enabling role in supporting cycles of experimentation, evaluation, and change (Fleck et al., 2009). This follows Antle's et al. (2014) recommendation of multiple, bidirectional pathways through activities as a way to encourage emergent dialogue and the exploration of values. The proposition that diagnostic and undo functionalities play a role in supporting explanation-giving (when a prospective view is taken) can be tested via comparisons of children's system use with and without the relevant features available.

### Conclusions and future work

This study investigated the character and differences in explanation-giving by pairs of children playing a tangible tabletop sustainability game with and without assigned roles/controls. Findings showed an important

role of the Impact Tool in provoking explanations related to its pausing, feedback and diagnostic features. The interdependency created in roles pairs led to greater opposition and question-asking as part of explanations. Improvement-oriented explanations that included references to Youtopia inhabitants' experience may be due to an increased sense of responsibility to the game-world. Enduring positionality among roles pairs needs to be addressed with future design iterations. An additional area for future inquiry is the use of the specific linguistic features identified as proxies to semi-automate the characterization of explanation-giving.

## Acknowledgements

We gratefully acknowledge funding from NSERC, SSHRC, and the GRAND NCE and wish to thank the grade 5 students and teachers we worked with at Stratford Hall Independent School.

## References

- Antle, A.N. (2015) Scratching the surface: Opportunities and challenges from designing interactive tabletops for learning. In *Learning Technologies and the Body: Integration and Implementation in Formal and Informal Learning Environments*. NY: Routledge.
- Antle, A.N., Tanenbaum, J., Macaranas, A. and Robinson, J. (2014) Games for change: Looking at models of persuasion through the lens of design. In *Playful user interfaces: Interfaces that invite social and physical interaction* (pp. 163-184). NY: Springer.
- Antle, A.N., Bevans, A., Tanenbaum, J., Seaborn K. and Wang, S. (2011). Futura: Design for collaborative learning and game play on a multi-touch digital tabletop. In *Proceedings of TEI'11* (pp93-100). ACM.
- Antle, A.N. & Wise, A.F. (2013) Getting down to details: Using theories of cognition and learning to inform tangible user interface design, *Interacting with Computers*, 25(1), 1-20.
- Antle, A.N. et al. (2013). Youtopia: A collaborative, tangible, multi-touch, sustainability learning activity. In *Proceedings of IDC 2013* (pp. 565-568). ACM.
- Chi, M. T. H., & Menekse, M. (2015). Dialogue patterns that promote learning. In *Socializing intelligence through academic talk and dialogue* (pp. 263-274). Washington, DC: AERA
- Derry, S. et al., (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *JLS*, 19(1), 3-53.
- Dillenbourg, P., & Evans, M. (2011). Interactive tabletops in education. *ijCSCL*, 6(4), 491-514.
- Fan, M., Antle, A. N., Neustaedter, C. & Wise, A. F. (2014). Exploring how a co-dependent tangible tool design supports collaboration in a tabletop activity. In *Proceedings of GROUP 2014*. (pp 81-90). ACM.
- Fernaues, Y. & Tholander, J. (2006). Finding design qualities in a tangible programming space. In *Proceedings of the Conference on Human Factors in Computing Systems* (pp. 447-456). ACM.
- Fleck, R. et al. (2009). Actions speak loudly with words: Unpacking collaboration around the table. In *Proceedings of ITS '09* (pp. 189-196). ACM.
- Gibson, W. J., & Brown, A. (2009). *Working with qualitative data*. London: Sage Publications.
- Higgins, S. E., Mercier, E., Burd, E., & Hatch, A. (2011). Multi-touch tables and the relationship with collaborative classroom pedagogies: A synthetic review. *ijCSCL*, 6(4), 515-538.
- Marshall, P., Fleck, R., Harris, A., Rick, J., Hornecker, E., Rogers, Y., & Dalton, N. S. (2009). Fighting for control: Children's embodied interactions when using physical and digital representations. In *Proceedings of the Conference on Human Factors in Computing Systems* (pp. 2149-2152). ACM.
- Price, S., Rogers, Y., Stanton, D., & Smith, H. (2003). A new conceptual framework for CSCL. In *Designing for change in networked learning environments* (pp. 513-522). NY: Springer.
- Rick, J. et al. (2009). Children designing together on a multi-touch tabletop: An analysis of spatial orientation and user interactions. In *Proceedings of IDC '09* (pp106-114). ACM.
- Speelpenning, T., Antle, A. N., Doering, T., & Van Den Hoven, E. (2011). Exploring how tangible tools enable collaboration in a multi-touch tabletop game. In *HCI-INTERACT 2011* (pp. 605-621). Springer.
- Teasley, S. (1997). Talking about reasoning: How important is the peer in peer collaboration? In *Discourse, tools and reasoning: Essays on situated cognition* (pp. 361-384). Berlin: Springer.
- Ullmer, B. & Ishii, H. (2000). Emerging frameworks for tangible user interfaces. *IBM Sys. Journal* 39, 915-931.
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46(1), 71-95.
- Wise, A. F., Antle, A. N., Warren, J., May, A., Fan, M. & Macaranas, A. (2015). What kind of world do you want to live in?: Positive interdependence and collaborative processes in the land-use planning game "Youtopia." In *Proceedings of CSCL 2015* (pp. 236-243). ISLS
- Wise, A. F., Saghafian, M. & Padmanabhan, P. (2012). Towards more precise design guidance: Specifying and testing the functions of assigned student roles in online discussions. *ETR&D*, 60(1), 55-82.

# Mediating Access: How Visually-Impaired Users Leverage Collaborative Learning to Keep Up With Mobile Phone Innovations

Joyojeet Pal, University of Michigan, joyojeet@umich.edu  
Anandhi Viswanathan, Independent Researcher, anandhi.viswanathan@gmail.com  
Aditya Johri, George Mason University, johri@gmu.edu

**Abstract:** We illustrate learning strategies used by visually impaired mobile phone users to support device information gathering and troubleshooting. Observing experiences of adult mobile users in Bangalore, India, we find that functional access and device familiarity is acquired through a range of digital and interpersonal consultations. We find that generic smartphones are increasingly becoming primary digital assistive devices due to their in-built accessibility features, yet the available information and user guides on these devices is biased towards sighted users. Consequently, device users rely on collaboratively learning from peers for general device familiarity and troubleshooting. We propose that collaborative learning will likely be a central to technology adoption and maintenance as artifacts and interfaces evolve constantly, forcing people for whose needs the devices may not have originally been designed, to learn and adapt constantly.

**Keywords:** accessibility, smartphones, India, collaboration, tech support, Android, iOS

## Introduction

Much CSCL research has focused primarily on educational and learning issues of importance to advanced industrialized societies, often necessarily so because advanced technology has been available primarily in these contexts. Salient issues addressed by researchers have focused on higher-level skills in science and mathematics and on formal learning environments. This overall research agenda has been determined by institutional constraints, funding imperatives, and areas of national needs within the context of developed countries.

The “typical” context of consumer technology use, particularly mobile phones, is a moving target as smartphones fall in cost and come to be the expected norm for basic communications. A result of this has been widespread uptake of smartphones by populations for whom these devices were not originally designed – including people using the devices in atypical ways, such as through non-visual interfaces in the case of people with visual impairments. In recent years, smartphones have also seen rapid adoption in parts of the Global South — a term that recasts formerly used terms such as Third World and Developing Nations to refer to people living in poor nations as well as the poorer parts of wealthy, industrialized nations (Pagel, Ranke, Hempel & Köhler, 2014). Smartphone use in the Global South has however been different in a few ways from the typical “western” setting – first the devices tend to be lower-end, thus less processing power and memory, second, they run on narrower network bandwidth, impacting what is seen as the typical operating environment for these devices.

The differences in device environments have consequences for mobile users with visual impairments living in the Global South. Within the context of assistive technologies (AT), mobile phones have come to play a critical role even though these devices were not necessarily designed for this user population. Mobiles have allowed individuals to better control their communication environments, find and maintain networks, apply for jobs, and access news and entertainment. With ever evolving mobile phone ecologies, though, AT users have to constantly adapt to newer devices and interfaces. This can be problematic because people with visual impairments lack access to training and tech support for AT through much of the Global South for two reasons. First, the institutions through which people in the West get access to technological training, such as accessible public libraries or educational institutions, are not equipped to offer access to and training for AT. Second, the overall lack of awareness of AT among the general public through much of the Global South means that sighted friends, family and colleagues of people with disabilities have limited knowledge of AT and do not know how to help others learn. Consequently, people with visual impairments have to collaborate with each other to learn to use and manage their technology either through direct interactions and online resources.

Mainstream market-oriented consumer technology production has meant that accessibility has usually been an after-thought in the design process. Consequently, people with disabilities have traditionally relied on collaborating with each other to select, learn to use, and adapt new technologies, and circulate this knowledge within their own community. Smartphone use by people with visual impairments presents scenarios in which interface, hardware, and network elements present challenges outside the realm of sighted smartphone use on device use and management, necessitating consultative learning within the accessibility community.

## Prior work and research objective

While seemingly novel within the context of current research efforts, analysis of non-industrialized settings has a significant lineage within research on learning. Starting with studies by Cole and Scribner (1981), Lave (1987), Saxe (1990), empirical findings from non-industrialized settings have significantly shaped learning sciences research. Ideas such as situated learning, legitimate peripheral participation, cognitive apprenticeship, and the integration of learning and everyday practices, as discussed for instance by Lave and Wenger (1991), have emerged primarily from ethnographic studies of non-industrialized communities (Lave, 1987). Yet, if we examine the impact of learning research or even attempts to design learning in non-industrialized settings, the outcome has been minimal (for exceptions see Evans et al., 2008; Kral & Heath, 2013). There seems to be either the implicit assumption that what is good for a the 'mainstream' early adopter is transferable to other settings, or that learning is so contextually bounded that it is not productive to even try to transfer lessons to settings different from those commonly found. This is further complicated when the learner population is at an intersection of marginalities, where informal settings may not only be the primary accessible mechanism for acquiring applicable skills but may also present a more potent mechanism, as illustrated by findings from the family math project (Martin, Goldman, & Jimenez, 2009).

People with disabilities often lack formal mechanisms of support, especially in our field site — India, which in turn means they need to rely on their individual and community resources. The motivation to learn how to use new technologies is high as personal technologies, such as mobile phones, offer the theoretical possibility of bypassing structural inequities faced by AT users and provide social and economic access. The challenges in adapting these technologies for non-sighted use has been well documented (Kane et al, 2011), and that despite people turning to their community for tech support, there is a dearth of learning support tools (Rodrigues et al, 2015). While there is much research on mobile accessibility, most work focuses on innovation and usability (Olivera 2011) little has focused on how visually-impaired users learn and continue to support their accessibility needs.

To this end, we present an exploratory empirical case study that pinpoints some of the specific areas of recurrent device use challenges on product information, hardware management, and interface usability for which people use various in-person and virtual consultative learning techniques. Most mainstream sources that have reference or learning material about mobile devices such as sellers or discussion sites have training guides, reviews, and specifications that primarily speak to the typical use cases. Moreover the flood of information through crowdsourced contributions make such sites difficult to scan quickly for performance information beyond basic accessibility features.

The range of devices available in the market, and their short shelf life on online sales platforms, and the constantly updating versions of mobile OS (especially Android) make evaluation of models and their performance hard to gauge. Whereas the old model of mobile AT, in which people with visual impairments used devices with separately installed software exclusively for sight-free interaction meant that the knowledge of the fairly stable software, rather than the hardware environment was key to managing devices, and the community had unique awareness and closed groups with expert users to help troubleshoot. The smartphone paradigm shifts the focus of accessibility management to the overall device and the app environment, increasing the need for ensuring appropriate knowledge filtering mechanisms to serve device learning and maintenance needs. In poorer parts of the world, the lack of public accessibility centers or educational institutions, coupled with the tendency of consumers to purchase inexpensive devices, further exacerbates technical challenges related to accessibility.

## Field setting and methodology

The work presented here is a subset of a larger scale study (Pal, Viswanathan & Song 2016, Pal et al 2017) research conducted with 81 participants with visual impairments in Bangalore, India that included a survey administered in 2015. The survey, described in Table 1, was followed by in-depth semi-structured interviews with a subset of 26 respondents; interviews took place over two rounds till April 2016. A requirement for participation was that respondents had to own and use daily a mobile device with assistive technology, typically either a functioning screen-reader or magnification software.

We required that all respondents be working-age adults. Our sampled respondents ranged in age from 21 years to 61 years. We recruited survey respondents through snowballing, with original contact through disabled people's organizations (DPOs), and online accessibility groups. As we see in the sample description in Table 1, 55 individuals in the sample were male and 26 were female. This is partly due to the difficulty in accessing who own their AT, which can be expensive, and women with disabilities have lower access to employment or control over household finances. Our sample also over-represents college graduates compared to what is typical for persons with visual impairments in India

Roughly a third of respondents had some functional vision, but only 8 of the 81 sampled persons used magnification as their primary interface; the vast majority used some form of speech output. We collected data on whether respondents had lost sight early or late in life, since late vision loss typically means an individual did not go to educational institutions equipped to train individuals in the use of assistive technologies, and also is a factor in one's social contacts with other visually-impaired people. Past sighted experience using technology like computers can also indicate comfort with interfaces and a conceptual understanding of technical metaphor.

Table 1: Sample description of survey participants

Measures	Male (55)	Female (26)	Total (81)
Median Age	29 years	32 years	31 years
Low Vision (Moderate – Severe Vision Loss)	34.5%	30.8%	33.3%
Profound Vision Loss / No Sight Perception	65.5%	69.2%	66.7%
Late Vision Loss (after age 12)	25.5%	30.8%	27.2%
College Graduates	78.2%	88.5%	81.5%
Median Cost of Mobile Device	US\$ 192	US\$ 223	US\$ 203
Median Years Using a Mobile Device	9.0	8.5	9.0
Median Years using a Smartphone	1.0	0.5	0.6
Smartphone Users	67.3%	80.8%	71.6%
Mobile Browser Users	63.6%	46.2%	58.0%
Mobile Social Media Users	69.1%	53.8%	64.2%

71.6% of respondents in our sample were smartphone users. The remainder used feature phones, which are not basic mobiles with simple call and text but rather have limited processing and browsing capability. These phones typically run on the Symbian platform. The median years of using a smartphone is less than 1 year for the sample — thus our data were collected at a point when many people were newly transitioning away from feature phones and starting up on the app space and a range of internet-based services.

Outcomes of the survey data relate to a series of issues around the use of various functionalities of mobile devices by people with visual impairments – which we have discussed extensively, particularly around issues with managing interactions on touchscreen devices (Pal, Viswanathan & Song 2016). Following the surveys, our qualitative research focused on themes that emerged out of the survey analysis. Interviews probed issues around technical barriers and support, technology adoption and transition, technology-related adjustments at workplaces, and the role of social networks in technology use.

All interviews were conducted in English, Tamil or Kannada and were transcribed verbatim. Following the transcriptions, we coded and analyzed the interviews for themes, and in this research we focus primarily on technical barriers and support. The purpose of this paper is not to use qualitative methods to highlight the frequency or types of technical barriers or support but rather to delve into a single pre-identified theme – technical challenges -- how these challenges manifest and how mobile users go about managing them.

## Findings

Respondents reported different challenges depending on the devices they used. For feature phones, we found that problems tended to be related to management and stability of third-party software and TTS (text-to-speech) packages such as system crashes, and a smaller number of interface related challenges around keeping track of key commands, shortcuts and compatibility with online material. Challenges around the use of smartphones were relatively more device-centric including issues with system settings, hardware configurations, and cache management, but also interface concerns such as input, and app usability. We primarily focus here on the smartphone use challenges and strategies used to manage them.

## Concern areas for smartphones

### Product reviews

Respondents talked about needing reviews from within the VI community because the majority of product reviews online (when available in an accessible format) were written focusing on the sighted user experience, except in the cases of apps that were specifically designed for accessibility. Respondents posted questions, often through mailing lists or social media, about specific apps to learn from someone else within the community with first-hand experience before making a decision to purchase. Even within the VI community, the range of skill levels and device environments of users meant that product reviews needed nuanced explanation on initial set up, performance etc. We found that the notion of product reviews itself is relatively new since the past feature phone

environment that people upgraded from had fewer products to choose from, and the “review” was a consultation with an expert, usually a trainer at a non-profit working with people with disabilities, who provided input on device and software choices. The breadth of the smartphone device market meant that even experts knew little about individual device models and their performance

### Hardware specifications

Related to product reviews, respondents noted challenges with assessing the appropriateness of hardware specifications for devices they planned to purchase. This was reported to be particularly true for Android-based smartphones, of which there is a wide range with distinct capabilities, as opposed to iOS devices, which have less variation between devices. In absence of the appropriate hardware requirements for certain functions, especially voice-based processing and output, a device may be accessible only in theory. The cyclical popularity of certain brands makes them attractive from a cost point-of-view, especially as specific telephony carriers offer deals on them. We found in our sample both globally significant brands such as the Samsung Galaxy and Moto G series phones, as well as less-expensive brands including Xiaomi, Micromax, and Asus phones. Respondents noted that the risk of picking up one of these brands was that they were newer, and it was harder to learn consultatively from people who had owned one for a long time. This was important to estimate performance on Bluetooth pairing, battery life over time using assistive technologies (which may take more power than typical sighted navigation), and built-in software such as file and music management, which also differ across devices.

### Input management

Input management was a significant concern for smartphone users, especially those who were relatively new converts. There are multiple options — using external hardware, voice input with speech recognition, or virtual keyboard settings like drag-and-release input (in which a finger must roll over a virtual keyboard that reads out alphabet until the user comes to the right key, upon which the specific letter is recorded), double-tap (in which the finger taps twice on a virtual keypad when at the right letter). Unlike with sighted smartphone users, where switching from one mode of input to another is unlikely or unnecessary, non-sighted use involves trying out and sometimes switching even after one is a fairly adept user of a specific type of input. Other add-ons that work well for sighted users such as auto-complete typically get in the way during speech-based navigation.

### Output management

Users reported issues with adding separate TTS packages (both for feature phones and smartphones) because these packages can be acquired separately from the in-built TTS to work with the existing screen-reading software. TTS management came up frequently in our interviews in part due to language issues particular to India. Users reported having Hindi, Kannada, or Tamil TTS (of which there were multiple options) which ran into problems with compatibility and quality of output, for which some technical consultation was necessary.

### Bundled services

Devices with custom builds from a telephony service provider required users to acquaint themselves with the specific functional environment of devices that may differ from the standard flavor of Android. Such bundled software are typically part of productivity suites or push ads, these can lead to accessibility-related annoyances where users need to customize their AT or TTS speed, as the bundled settings came into effect with each reboot.

### Strategies

Other than figuring out a device on one’s own, there are two most used strategies that respondents report having used. The first is referring to non-profits that work with people with disabilities, and the second is talking to others in the community for technical help either directly or online. We found that respondents do not turn to the typical sources sighted people reach out to. Only 5 out of 81 respondents mentioned going to the showroom or to the device company’s representative for assistance. Sighted users were sometimes contacted for smartphone problems, though this posed privacy problems as the device owner could not tell what the sighted intermediary was doing. Only one feature phone user (of 21) sought help from a sighted person for a major failure. In general, some form of discursive troubleshooting was the norm for both minor and major breakdowns and help scenarios.

### In-person collaborative learning

For most respondents, the smartphone environment was seen as a game-changer in terms of the ability to do things, even if they faced challenges with adoption and use. Some respondents switched to smartphones willingly on counsel of friends and others did so as there was no longer a choice with feature phones being phased out of the market. Interviews frequently highlighted respondents’ excitement about using technology.

*In the beginning I did ignore my family for at least a couple of weeks when my attention was completely absorbed by the phone. I have always been excited by new technology, now that I am confident of using touch – I keep experimenting with new apps. And I am waiting for the upgrade to Android Marshmallow....it has made it possible for me to have continuous conversations with my friends. (Female, 39 years)*

Accessible mobile devices enable collaboration by making the participation in social networks easier. Social networks can be critical in learning about devices and making decisions on platforms and apps, but also in just casual browsing or conversations that are not specific to purchase moments, but add to one's familiarity with technology. Social networks can be critical for people who acquire a visual impairment late in life as there is generally little awareness of accessible technology in the mainstream media. Respondents discussed challenges getting good information on accessible technology through sighted friends, especially in the pre-smartphone era.

*I don't have many friends in the VI community as I am a late blind person. My friends are mostly sighted... At the time I got this mobile, my major concern was accessibility. I'd heard from people that Android phones had a lot of constraints when it came to accessibility. (Male, 39 years)*

People who acquire visual impairments in later life may have familiarity with visual interfaces and better access to sighted intermediaries is helpful in mediated learning, but their lack of contact with other blind people reduces access to detailed first-person commentary or the option of deciding collaboratively on appropriate purchases. In the above case, purchasing the wrong Android, meant getting stuck with an unusable smartphone.

Even on platforms like iOS and Android where the accessibility features are easily available on anyone's smartphone, troubleshooting with sighted intermediaries is not always helpful, since people do not typically look up their own accessibility features when they don't use it themselves. The average workplace in India does not have a lot of people with vision impairments, further distancing the sighted mainstream from accessibility concerns. With this lack of public awareness and exclusion of accessibility from public institutions, the non-profit sector has played a central role in Bangalore's accessibility sphere. Respondents spoke of turning for accessibility consultation to non-profits that offered training for screen reader use on desktop devices. These organizations served as repositories of information and expertise on AT, both through institutional knowledge of training, and through pools of graduates who used AT. Non-profits were an important means of acquiring AT in the past since they had agreements with distributors to bundle AT software into a certain feature phone models and sell those at bulk prices (a result of which has been the popularity of the Nokia N series phones among respondents, which were offered in such bundles). However, the onset of smartphones has also meant that the AT acquisition has moved from institutional purchases to something negotiated directly by consumers.

Users now often turn to friends who they meet informally and tap into advice when encountering technical barriers. We found that social connections influenced decisions on devices as well as carrier networks, since in India there are no calling fees for in-network calling, so people keep multiple SIM cards which they top-up, but switch networks on data use as and when deals come by, since AT use can be data heavy.

*Usability was a primary concern while transitioning. I still continued to use the regular keypad phone as my primary phone for some time even after I got the touch phone. I got another SIM card for my new phone and I was using this touch phone to browse the internet and practice navigation, while I learnt to use the touch phone ... I checked with my other friends to see how they were able to use their phones. ... I thought to myself, if they can do it so can I. (Male, 31 years)*

Smartphones can also play a role in making the assistive device more of an exchangeable, relatable device in respondents' domestic settings among sighted family members, because these were often platforms that others in the household knew (Android, iOS), as opposed to the older feature phones that worked with screen readers.

*I learnt to use the (keypad) phone mostly all by myself. It was not so difficult to learn after understanding the basic key strokes. If at all there was any difficulty, I used to ask my friends who were also using Talks and got solutions to the problem [With the smart phone] I also had help from my mother...to understand different icons. Once I showed her all the different smiley emoticons and got her to describe them to me so I knew what they all looked like. (Female, 29 years)*

In this instance of collaborative learning, the mother helped with visual items and in the process got familiar with her daughter's phone. The initial experience of exploring the interface trained both newbies to smartphones use. Consequently, the daughter felt a sense of security that a trusted person was familiar with her technology environment, and knew what kinds of help her mother could offer, versus when she needed to consult a friend.

### Online collaborative learning

The mobile purchase is invariably prefaced by some consultation. With feature phones these consultations were with friends or DPOs. Online question-asking and research was common for smartphone purchases, especially when these were made through Indian e-commerce sites, which make it harder to “try out” and return. This put the onus on buyers to research thoroughly before purchasing. But as we find with the quote below, price was still a major driving factor. Buying phones with poor processing power meant needing to make do with a cheap mobile as a secondary device to a desktop device rather than a viable replacement for workplace activities.

*I was looking for a phone in the Rs. 7000 (US\$110) category only. At that time, this was one of the better options available... I did not get to check out the phone before buying it because it bought it from Flipkart... there is a lot of pressure about taking calls and assisting customers. I use two different ear phones, one for taking customer calls and the other for listening to JAWS on the computer. This is very challenging. (Female, 35 years)*

The two primary means of online and collaborative learning that emerged in discussions were threaded emails and social media. Threaded email lists emerged out of the DPOs which initially set these up for their members or extended community. Over time, these grew to be major “go to” places for consultation where people both posted specific questions, but also followed threads to listen in on what mobiles were recommended by others.

*[When I have a problem] I would try my hand. I would use all my tools, when nothing happens only then will I go to Enable India. If Enable India is not able to help me or something then I will try asking Access India. It is a mailing list. Sometimes people out here, the visually challenged people over here in IBM, they themselves could be of help. (Male, 30 years)*

As we found with the respondent above, posting to mailing lists was done after exhausting first-person contacts. In part, this is driven by the social norms of an open online forum: Respondents were cagey about asking questions deemed too basic because a number of the mailing lists have a number of advanced users of mobile technology, and many conversations can be fairly technical.

The advent of Facebook and WhatsApp groups came up repeatedly in interviews as having impacted people’s ability to get quick tech support. Mailing lists were inherited from the desktop era, and isolating a single email ID and sending questions to individual respondents rather than the entire group required some work. On social media, the easy addition of contacts (and their integration into platforms like iOS and Android) made asking questions to both the entire group and to individual members easier. Social media groups offered the option of collaborative learning where new technologies were reviewed and posted with replies or comments that allowed for natural threads of conversations and clarifications, which tended to be cumbersome on email.

*Finding out about the gestures and learning to use them took some time. I also did not know the terminologies that were being used for using the touch phone. Learning the location of the icons and remembering their positions was a challenge ... I struggled very badly for the first 2 weeks at least. I used to keep checking online for tips to use Android phone. I came across a Facebook group for visually impaired Android users. Over there I saw a post for joining a WhatsApp group for visually impaired Android users. I applied to join the group and was added. This WhatsApp group changed my Android experience. There are some very advanced users who give tips on how to use Android phone with TalkBack. (Male, 31)*

The growth of WhatsApp in India within the mainstream population pushed the demand for Android devices among people who wanted to be included while circumventing paying for SMS services, which are typically charged per message. WhatsApp was reported as accessible, including advanced features like creating or exiting groups. It also allowed users to easily put together their own sub-groups at their own levels of expertise for help.

We also found that collaborative learning was not necessarily shifting from email to social media, rather people used both simultaneously. Respondents participated in smaller groups with more intimate questions and posted general device queries on larger, open mailing lists such as the “eyes-free” group on Google (a resource for Android users) and the Daisy Forum, a group that started with discussions on document formats but moved to general technology assistance. These groups have evolved to a role of answering high-level generic questions of relevance to the entire VI community such as commentary on new OS developments or device in the market, policy issues, and reviews of major products or apps.

## **Discussion**



No learning strategy operates exclusively online or face-to-face; the two have morphed in form over time and operate in tandem. Gaining technological fluency in particular has now become a continuous process in which both online and offline networks and information access play a crucial role (Barron & Kafai, 2006). Our study, first and foremost, foregrounds the complexity of learning about new devices and applications for users of AT and illustrates how these users are able to leverage their “learning ecology” to gain fluency (Barron, 2004). Their ecology consists of other users who face similar challenge, non-profits that work with users of AT, and a larger online community to which they have access, and depending on prior fluency as well as the change in technology, different resources are leveraged for learning.

For initial interface learning with both feature phones and smartphones, we find a consistent pattern of consultation with others in the community, but with the key difference that non-profits have traditionally been more involved in training with feature phones. The feature phone environment bears an important similarity to the desktop screen reader learning environment in which understanding the basic operational framework requires structured learning of the basics, thus many respondents reported taking orientation classes. Beyond basic operations, shortcuts and workarounds are critical to effective use, which makes consultations with human contacts an important source of tips and tricks. The “closed” nature of feature phones with AT, that only people with visual impairments use them, makes the community the sole source of almost all learning resources.

Unlike with feature phones, which for at least a few users were the first interaction with screen reading, all respondents who had moved to smartphones had some prior experience using feature phones with screen-reading software. While there were a few training classes for Android run by the same non-profits that offered screen-reader training, most respondents reported no formal training, despite significant misgivings about the basic interface. Here, the one-on-one consultations with friends and colleagues were reported as helpful because the challenge with touchscreen-based smartphone interfaces is getting over the initial shock of having no tactile feedback. We find a mix of excitement alongside apprehension for people deciding to make the switch. Such an environment is arguably well suited for collaborative learning because users are intrinsically motivated and their concerns are best managed through the homophily effect of helping their colleagues learn to use the devices. We find thus a combination of collaborative information-gathering online during the information gathering phase (i.e. experiences of other visually impaired users posting publicly about their transitions to smartphone use) and collaborative informal learning with friends and colleagues after one acquires the device (i.e. first hand tutorial sessions or consultations with other smartphone users).

We also find that learning issues vary by stage. Early learning issues differ from some of the longer-term challenges. On hardware management in particular, we found that users received significant amounts of conflicting information about challenges with certain devices because of battery and storage issues that often did not emerge until several months after starting to use the devices. Mailing lists are a source of high-level information on devices, but mainly cover the major, widely used devices. Lists can be inadequate for the more relevant specifications information that is needed at the time of purchasing a device. Thus knowing one should buy a Moto smartphone is potentially harmful information because a low-configuration (E series, for instance) would appear to work well initially but soon run out of memory and battery with the relatively power-hungry accessibility apps. Information on such devices would also be directly misleading when looking at reviews on online shopping sites such as Amazon or Flipkart, because these would be high ratings assuming sighted use.

Respondents reported developing preferences on input techniques such as drag-and-release in conversations with friends rather than through online interactions. An online interaction was to post a high-level query on a mailing list or WhatsApp about a device or interface followed by a request for anyone willing to answer questions to contact off-list. We found more willingness among people to ask direct questions about phone models, and smartphone apps, bundled services, or questions about TTS packages on lists, but less willingness to discuss troubleshooting issues like input or output management. The reasons for these distinctions in question-posing behavior are likely related to community norms – mailing lists are widely used for skill-agnostic questions such as reviews or generic questions around interface use that have significance for the user experience of the community at large, rather than very specific individual troubleshooting issues.

## **Conclusions and implications**

Accessibility work will likely constantly have to play catch-up with the emerging landscape of interactive experiences. Some of the problems faced by the participants in this research are unique to Global South settings due to income and institutional barriers. However, issues around hardware specifications for new devices and product reviews in the app space are problems that are true for people who need to use accessibility features on a regular basis across geographies. This study also has implications for our understanding of collaborative learning, much of which has focused on children or on workplace learning, where drivers of learning typically include concurrent groups that work toward a goal, at least a loosely understood one. The changing nature of assistive

technologies from an older model where generic devices were given additional adaptive capabilities to make them accessible, to a new environment where mainstream smartphones have inbuilt accessibility capabilities brings people with visual impairments closer to the sighted mainstream on one hand, but also creates new challenges for their learning environments. In this study we have shown how access is mediated through human and digital resources for access, and that for the users of accessible technology, the real challenges lie in building learning strategies to confront an ever-evolving technology environment. Collaboration and communication, it appears thus far, will be a central need for success.

## References

- Abeebe, M. V., De Cock, R., & Roe, K. (2012). Blind faith in the web? Internet use and empowerment among visually and hearing impaired adults: A qualitative study of benefits and barriers. *Communications* 37(2): 129–151.
- Barron, B. (2004). Learning ecologies for technological fluency: Gender and experience differences. *Journal of Educational Computing Research*, 31(1), 1-36.
- Barron, B. & Kafai, Y. (2006). Clubs, homes, and online communities as contexts for engaging youth in technology fluency building activities. In Proceedings of ICLS '06, pp. 1022-1028.
- Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of Learning Sciences*, 2(2), 141-178.
- Cole, M. & Scribner, S. (1981). *The Psychology of Literacy*. Cambridge, MA: Harvard University Press.
- Evans, M., Johri, A., Gasson, G., Cagiltay, K., Pal, J., & Sarkar, P. (2008). ICT4D and the Learning Sciences. In the Proceedings of International Conference of Learning Sciences 2008. Vol. 3, pp. 229-236.
- Gerber, E. (2003). The benefits of and barriers to computer use for individuals who are visually impaired. *Journal of Visual Impairment & Blindness*, 97(0).
- Hawkins, J., & Pea, R. D. (1987). Tools for bridging the cultures of everyday and scientific thinking. *Journal for Research in Science Teaching*, 24, 291-307.
- Kane, S. K., Jayant, C., Wobbrock, J. O., & Ladner, R. E. (2009). Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. In Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility (pp. 115-122). ACM.
- Köhlmann, W. (2012, July). Identifying barriers to collaborative learning for the blind. *International Conference on Computers for Handicapped Persons* (pp. 84-91). Springer Berlin Heidelberg.
- Kral, I., & Heath, S. B. (2013). The world with us: Sight and sound in the “cultural flows” of informal learning. An Indigenous Australian case. *Learning, Culture and Social Interaction*, 2(4), 227-237.
- Lave, J. (1987). *Cognition in practice*. New York, NY: Cambridge University Press.
- Lave, J. & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. New York, NY: Cambridge University Press.
- Martin, L., Goldman, S., & Jimenez, O. (2009). The Tanda: A Practice at the Intersection of Mathematics, Culture, and Financial Goals. *Mind, Culture & Activity*, 16(4).
- Oliveira, J., Guerreiro, T., Nicolau, H., Jorge, J., & Gonçalves, D. (2011). Blind people and mobile touch-based text-entry: acknowledging the need for different flavors. In The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility (pp. 179-186). ACM.
- Page, H., Ranke, K., Hempel, F., & Köhler, J. (2014). *The use of the concept 'Global South' in Social Science & Humanities*. Humboldt University of Berlin.
- Pal, J., Viswanathan, A., & Song, J. H. (2016). *Smartphone adoption drivers and challenges in urban living: Cases from Seoul and Bangalore*. In Proceedings of the 8th Indian Conference on Human Computer Interaction (pp. 24-34). ACM.
- Pal, J., Viswanathan, A., Chandra, P., Nazareth, A., Kameshwaran, V., Subramonyam, H., Johri, A., Ackerman, M., O'Modhrain, S. (2017) *Agency in assistive technology adoption: Visual impairment and smartphone use in Bangalore*. In Proceedings of the SIGCHI 2017 (Forthcoming). ACM.
- Rodrigues, A., Montague, K., Nicolau, H., & Guerreiro, T. (2015). Getting smartphones to TalkBack: understanding the smartphone adoption process of blind users. *Proceedings of ACM SIGACCESS*, pp.23-32.

## Acknowledgments

We thank our study participants for their time and input. Dr. Johri's contribution is based, in part, upon work supported by the U.S. NSF Award EEC#1424444. We thank Dr. Mamoru Iwabuchi and the team at RCAST University of Tokyo for valuable discussions at early stages of this work.

# Scoring Qualitative Informal Learning Dialogue: The SQuILD Method for Measuring Museum Learning Talk

Jessica Roberts, University of Illinois at Chicago, [jessicaannroberts@gmail.com](mailto:jessicaannroberts@gmail.com)  
Leilah Lyons, University of Illinois at Chicago and New York Hall of Science, [llyons@uic.edu](mailto:llyons@uic.edu)

**Abstract:** Museums are increasingly developing computer-supported collaborative learning experiences and are in need of methods for evaluating the educational value of such exhibits. Exhibit designers, like web interaction designers, have long been employing A/B testing of exhibit elements in order to understand the affordances of competing designs *in situ*. When the exhibit elements being tested are intended to support open-ended group exploration and dialogue, existing knowledge-based metrics like measuring the amount of content recalled don't quite apply. Visitor groups can explore the educational content in idiosyncratic ways, meaning that not all groups have the same exposure to content, and the learning arises from the visitors' conversations. In order to evaluate learning outcomes for CSCL exhibits, we present a method for quantifying idiosyncratic social learning, Scoring Qualitative Informal Learning Dialogue (SQuILD), and demonstrate how it was applied to A/B testing of a collaborative data visualization exhibit in a metropolitan museum.

## Introduction

Interaction designers have long relied on A/B testing to compare competing designs in real settings of use on measurable outcomes like duration of interaction and number of events logged during a session. These measures give valuable information about the use and benefits of interactive systems, but they are typically most informative for single user applications like websites. In multi-user systems such as collaborative interactive museum exhibits, interpersonal interactions are typically key aspects of the experience, yet have been traditionally difficult to quantify (Block et al., 2015). Museum designers have used A/B tests to evaluate exhibit elements like alternate versions of label designs (Serrel, 1996), using metrics like dwell time or information recall (Bitgood, 2000), but an increasing body of literature studying learning in museums suggests that such assessments miss the full breadth of learning occurring in these settings.

Instead, recent work has attended more closely to the role dialogue plays in informal learning. Researchers "view one of the richest forms of learning in a museum to be evident in the patterns of discourse and activities that groups engage in - such as labeling, theorizing, predicting, recognizing patterns, testing ideas, and explaining observations" (Atkins et al., 2009). Visitors' dialogue during their interactions with an exhibit is thus presumed to be the primary vehicle for learning, yet we have limited means for quantitatively analyzing this dialogue to support *in situ* A/B testing, which is a critical method for establishing the ecological validity of an exhibit design. This paper describes a methodology that solves three key problems in quantifying data talk: identifying socially productive learning talk in open-ended dialogic activities, segmenting spontaneous dialogue to permit cross-group comparisons of talk, and developing a scoring approach that allows alternate styles of learning conversations to be valued.

## Background

### Challenge 1: Identifying open-ended learning with interactive exhibits

Museums are increasingly using novel interactive exhibits to create engaging spaces where visitors can shape their own narratives of the experience (Roberts, 1997). Unfortunately for exhibit designers and learning researchers, it is impossible to design a knowledge-based post-test to demonstrate content mastery for an open-ended exhibit, in which we have no expectations that all visitors will explore the same content. Other measures like visitor interviews are also problematic, as "there is little correspondence between people's post hoc characterizations of their experience and the activities in which they engage when visiting exhibitions" (Heath & vom Lehn, 2008). Studying the dialogue of visitors while they are engaged with the exhibit may be more reliable, and researchers have made gains toward understanding what is productive talk given the free-choice environment, for example reading text aloud, asking and answering questions, connecting new information to prior knowledge, and giving explanations to companions (Allen, 2002; Ash, 2003; Leinhardt & Knutson, 2004; Kisiel, Rowe, Vartabedian, & Kopcek, 2012). The foundation laid by this prior work, largely developed before exhibits were highly-interactive, needs to be built upon to address challenges specific to highly interactive computer-based exhibits, where visitors have increasing agency in shaping their own interaction experiences.

## Challenge 2: Segmenting dialogue to permit cross-group comparisons

The grain size for segmentation is a key decision in any analytical process. Due to the spontaneous nature of joint exploration of a museum exhibit, many ideas are split among two or more visitors as they work together to make sense of the data. Visitors interrupt each other and in some cases interrupt themselves mid-idea as they notice new information. The fragmented nature of museum dialogue is a known challenge for assessing learning in this context (Allen, 2002), particularly when an analysis aims to quantify talk by counting instances of a particular kind of speech act. Some analyses of visitor dialogue address this challenge by coding simply for the presence or absence of a particular kind of talk (e.g. making a prediction) at all during a session (Allen, 2002; Atkins et al., 2009), but such an analysis runs the risk skewing the quantification toward under-representation: a visitor group that had an in-depth conversation with many predictions would receive the same score as a group that made only a single prediction. Some meaningful segmentation is necessary to measure and compare session dialogue. A common segmentation strategy is to divide speech into conversational turns and code and count those turns to quantify them (Chi, 1997). Because of the frequent interruptions and repetitions common in informal learning talk, this delimitation technique would skew the quantification toward over-representation of certain kinds of talk. Larger delimitations, meanwhile, such as segmenting by theme or referenced data, would obscure the intricacies of the productive dialogue. A new segmentation method is necessary.

## Challenge 3: Respecting socially-constructed learning when quantifying dialogue

The ultimate goal of the analysis described here is to create a valid quantitative measure of spontaneous dialogue during exhibit interactions for the purpose of conducting A/B testing of multiple exhibit designs, e.g. competing form factors for control of an interactive exhibit (Roberts & Lyons, in progress). As noted above, challenges in segmenting dialogue already make quantification of talk difficult, and complicating the process even further is the irreducible tension that while not all dialogue may be equally well aligned with exhibit learning goals, all visitor groups are guaranteed to engage with exhibits in manners that suit their current interests and level of understanding. What constitutes meaningful learning talk for one group might fall short of designers' goals for the exhibit. For example, it is widely accepted that reading a label aloud is a productive form of talk in museums (Borun, Chambers, & Cleghorn, 1996; Kisiel et al., 2012; Allen, 2002; Atkins et al., 2009), but should such reading merit the same quantitative score as, for example, a comparison of two datasets, given that the aim of the exhibit is to foster such comparisons? But conversely, should a "conversation" where the only talk is one visitor making a single comparison be valued more highly than the extended conversation of a group that involves the reading aloud of a lot of low-level exhibit content and linking of that content to personal experiences? A productive analysis should ideally respect the socially constructed nature of museum learning and "give credit" to both quality and quantity of talk, acknowledging all productive talk while retaining qualitative distinctions among different kinds of talk.

## SQuILD: A method for quantitatively comparing informal learning talk

Here we present the SQuILD (Scoring Qualitative Informal Learning Dialogue) method, and we provide examples of applying it to an interactive data map museum exhibit.

### Addressing challenge 1: Identifying learning in open-ended informal dialogue

We began by identifying common threads in the literature on visitor talk to develop five categories of substantive talk—management, instantiations, evaluations, integrations, and generations. We then used an open coding process, informed by the literature of our exhibit's content domain (graph interpretation), to determine specific sub-codes within each of those categories (see Table 1). While our sub-codes may not be directly applicable to an exhibit with different content, the five categories could easily be adapted to a variety of interactive exhibits. Future work embracing this methodology would likely find that developing unique sub-codes within these presented categories would retain the structure of the method while adapting to the specific content focus of the research.

### Manage codes

It is to be expected that when multiple people are interacting with an exhibit, particularly when that exhibit is based on a novel technology, some amount of talk will directly address the interaction with the exhibit. Talk that related to the establishment of joint attention, negotiation of action, or scaffolding exhibit use was coded as *management*. These kinds of behaviors are of interest to researchers of museum learning because they speak to how visitors are working together and mediating each others' experiences. For example, Allen (2002) categorized these kinds of actions as "strategic" with only two sub-codes: "use" and "metaperformance." Borun et al. (1996) attended to observable coordination behaviors like "call over." Multiple studies have attended to

facilitative behaviors such as explaining, asking and answering questions, and suggesting actions (Ash, 2003; Eberbach & Crowley, 2005, Diamond et al., 1986; Atkins et al., 2009). Researchers of technology-based multi-user interactives are similarly concerned with interpersonal interactions like interference (Falcão & Price, 2009), negotiation of exploration (Davis et al., 2013), and collaboration (Williams, Kabisch, & Dourish, 2005).

### Instantiate codes

Here the term “instantiation” indicates when a user makes information part of the conversation by saying it aloud. The instantiation of information provides opportunities for the individual visitors to internalize that information (i.e., learn from the exhibit) and can lay the foundation for further reasoning among learners on a museum visit (Kisiel, et al., 2012). Saying something aloud is an important part of the social learning process: putting ideas into the shared social space and helping establish joint attention (also referred to as “grounding”). Per sociocultural learning theory, learners must articulate ideas via communication before learning can take place (Vygotsky, 1978). Processes of noticing and establishing joint attention among visitor group members have been found to be productive in facilitating learning talk in museums (Povis & Crowley, 2015; Leinhardt & Crowley, 1998), and reading labels aloud was identified as a “significant behavior” linked to increased group learning by Borun et al. (1996).

### Evaluate codes

Evaluation statements go beyond merely instantiating content to make some kind of judgment or assessment about a piece of information by assigning some kind of value, whether qualitative or quantitative. Such personal qualitative evaluations are arguably very important in informal learning settings, where developing one’s identity is seen as just as much of a goal of the meaning making process as absorbing content (Rounds, 2006). In this context, evaluations can be simple standalone comments or part of a more complex statement. The most common sub-code of evaluative statement in our exhibit was *characterize*. Examples of *characterize* evaluation statements are those remarking that there are “a lot” or “not very many” of something, or describing a population as being “everywhere.” The characterizations could be spatial or quantitative in nature.

### Integrate codes

While evaluation statements refer to a single idea, the final two categories connect multiple pieces of information in some way. Friel, Curcio, and Bright (2001) refer to the act of looking for relationships in data as “interpretation.” The SQULLD coding framework adopts the more precise term *integration* from Murray, Kirsch, & Jenkins (1997) to describe the act of pulling together multiple pieces of information presented in an exhibit. Statements that integrate are those that make explicit connections or comparisons between multiple pieces of information: for example, in our exhibit, between two different datasets, between a dataset and the geography, between a dataset and itself over time, etc. Connections and comparisons are integrative talk widely acknowledged to be valuable in museum settings (e.g. Allen, 2002; Atkins et al., 2009; Falk & Dierking, 2000.)

### Generate codes

*Generate* statements “[go] beyond the data” (Curcio, 1987) to combine information from the exhibit with visitors’ own prior knowledge and experiences. Falk & Dierking’s (2000) Contextual Model of Learning posits that what learners gain during a learning experience is inextricably tied to what the personal context they brought into the experience—prior knowledge, experiences, motivations, identities, etc. Allen (2002) incorporates what she calls “connecting talk” into her framework for analyzing visitor conversations at an exhibit, but unlike the *connections* described above as an integrate code, the type of connections she is referencing are making use of outside information, by connecting an exhibit to life, prior knowledge, or other exhibits. She describes this stitching-together of information from different sources as “powerful and ubiquitous means of learning in informal settings.”

## Addressing challenge 2: Segmenting dialogue through idea units

Dialogue is a group activity. Some ideas are spoken by only one visitor and are contiguous and completed in a single conversational turn. Others are co-constructed by multiple visitors as they collaboratively investigate the exhibit’s content. To reach the appropriate level of granularity, this method adopts the *idea unit* as its unit of analysis, introduced by Jacobs et al. (1997) as “marked by a distinct shift in focus or change in topic.” We amend this to more closely capture dialogue emerging in the midst of a group activity by defining an idea unit as marked by a distinct shift in focus or change in topic *or purpose*. This adjustment segments visitor conversation into chunks according to what that speech is doing in the group interaction. Idea units can range in length from a single word, e.g., reading aloud a category name, to a multi-sentence utterance. To illustrate the concept, below

are two excerpts of dialogue from two visitor sessions. The first shows somewhat straightforward linear idea units, as annotated below:

- |     |   |  |
|-----|---|--|
| [1] | A: I want to see how it changes.  | [states intention]                                     |
| [2] | A: Like that area over there changed a lot in regards to... demographics, you see it? | [draw joint attention to areas that changed over time] |
| [3] | B: Yeah.  |  |
| [4] | A: And up there.  |  |
| [5] | B: More spread out.   | [characterize data]                                    |
| [6] | A: But you see the greatest change here on this side.                                 | [identify area of particular interest]                 |

This excerpt was divided into four idea units. These idea units vary in length and in one case span multiple turns and speakers, but they are fairly straightforward. Some idea units are less obvious, because they are detached and inter-spliced. Take this segment from another pair:

- A: So whatever's, I'm assuming there must be railway or, oh wait, isn't that a road? That goes across, across the water. So there's-
- B: It's a bridge.
- A: So my guess is, oh it's a waterway or a roadway or whatever. Waterway maybe. But that area's most likely industry.

Visitor A's main goal is to pose his theory about the area being industrial but he keeps interrupting himself trying to correctly describe the roadway. This segment is counted as two overlapping idea units, as the participants are doing two meaning making moves in these three turns: decoding the map representation, represented with a dashed underline below, and posing an inference about the area based on the data ("So whatever's...So there's..So my guess is...But that area's most likely industry," double-underlined below).

- A: So whatever's, I'm assuming there must be railway or, oh wait, isn't that a road? That goes across, across the water. So there's-
- B: It's a bridge.
- A: So my guess is, oh it's a waterway or a roadway or whatever. Waterway maybe. But that area's most likely industry.

This segmentation into idea units prevents stutters and echoing (e.g., the repeated starts to the inference "So whatever's", "So my guess is...") from unfairly weighting a statement beyond its contribution to the dialogue, which can occur in a speaking-turn-based quantification of talk (Chi, 1997). Idea unit coding is particularly useful when characterizing the overall educational quality of a group's conversation, rather than trying to draw attention to the individual contributions or cognitive acts of each speaker. Given the sociocultural perspective much work in museum learning is taking (namely, learning is evidenced in the group's talk, and benefits the group as a whole), idea units are more appropriate than a turn-based approach.

Idea unit coding is best done directly from video to retain the context of visitors' comments (see Figure 1). Separating dialogue from the visitors' experience by making—and later coding from—a transcript removes the context in a way that obfuscates or even completely alters the meaning of the statement. Context is particularly important for learning talk occurring at dynamic interactive exhibits, where visitors can alter the exhibit state with their actions, and their dialogue responds to the changing state of the display.

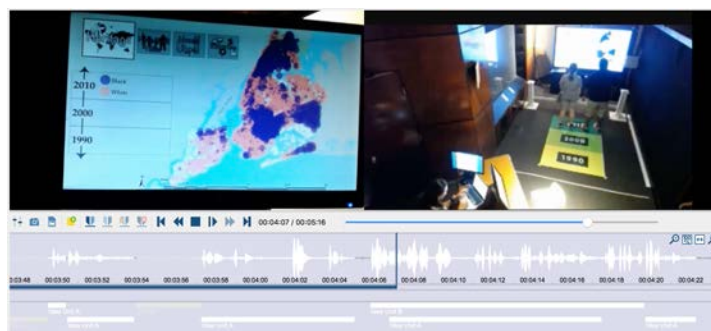


Figure 1. Idea units that overlap each other in time can be segmented in video, shown here as white bars in the coding software MaxQDA.

### Addressing challenge 3: Quantifying depth and nuance in visitor talk

Quantifying visitor talk begins by assigning the five categories of codes identified above to the identified idea units. In our work, any idea units that did not match any of the above categories and sub-codes were marked *non-substantive* and were disregarded in the analysis. Some idea units were coded with a single code. Many idea units, however, were coded with multiple codes: though the statements were one logical idea, they were deep and complex enough to warrant multiple codes. This process of simultaneous coding (Saldaña, 2009) maintains the richness of the talk, rather than reducing a statement to a single code. For example, consider the statement:

“Okay, there’s a lot of White in 2000, rather than Mexicans.”

This statement as an overall idea unit compares the two heritage groups. Within that broad goal, it does multiple things. It *INSTANTIATES* the *datasets* (“White” and “Mexicans”) and *decade* (“2000”), it *INTEGRATE-connects* the dataset to the decade “White in 2000”, it *INTEGRATE-compares* the datasets (“White” and “Mexican”) using “rather than,” and it *EVALUATE-characterizes* “White” as being “a lot.”. Coding the statement only as a single code—in this case *INSTANTIATE-compare*—would give it the same value as a much less rich statement like, “It looks like there are more of them.” Only by simultaneous coding can we give credit to the multiple “hooks” this complex statement provides for further discussion.

### Assigning values to codes by tying them to exhibit learning goals

The coding framework described above stays close to the data in identifying how people are talking by flagging conversational acts that are likely to contribute to shared meaning-making at an interactive data exhibit. In any open-ended exhibit, multiple kinds of talk are considered to be highly relevant to the intended learning. Other kinds of talk are important but less directly aligned with the learning goals. Therefore, this methodology employs a form of magnitude coding (Saldaña, 2009; Miles & Huberman, 1994) by sorting the sub-codes into high, medium, or low categories according to their relation to the goals of the exhibit, as determined by the research team (see Table 1; note that the weighting is particular to this exhibit.). These relevance categories are assigned numerical weights in order to quantify and compare the substance of visitor talk across conditions. Using magnitude coding as a way of “quantizing” a phenomenon (Tashakkori & Teddie, 2010) permits the use of inferential statistics (Bernard, 2006; Saldaña, 2009) in order to compare the experimental conditions.

Table 1: Sub-codes were sorted according to their relevance to learning objectives

Low Relevance (1)	INSTANTIATE category INSTANTIATE dataset INSTANTIATE decade INSTANTIATE geography INSTANTIATE representation INSTANTIATE self	INTEGRATE connect multiple INTEGRATE connect simple MANAGE ask interpretive question MANAGE direct co-visitor's movements	MANAGE narrate intentionality MANAGE negotiation of control MANAGE purpose of exhibit MANAGE suggest action
Mid Relevance (2)	EVALUATE characterize EVALUATE win GENERATE contextualize GENERATE identify knowledge gap	INTEGRATE challenge interpretation INSTANTIATE outside knowledge	MANAGE ask guiding question MANAGE clarify MANAGE direct co-visitor's attention
High Relevance (3)	EVALUATE question census categories GENERATE confirm GENERATE make prediction	GENERATE negotiate meaning GENERATE notice surprising pattern	GENERATE pose inference INTEGRATE compare

The codes assigned to the “low relevance” category – mostly *INSTANTIATE* and *MANAGE* subcodes – are all activities that are useful for grounding and coordinating the group learning experience and may serve as springboards for future dialogue, but are in and of themselves not strongly related to the learning goals of the exhibit. These statements were assigned a weight value of one. “Mid relevance” codes took steps to more directly make sense of the presented data by characterizing and contextualizing it (including instantiating outside knowledge to help with sense-making), clarifying the representational forms, and directing co-visitor’s attention to an interesting element of the exhibit (which rises above a simple *instantiate* code because it conveys to the listener that the targeted element is worthy of joint discussion). These codes were given a weight score of two. “High relevance” talk included statements that related presented data to prior knowledge or expectations, predicted or inferred information, compared datasets with each other or over time, and questioned the source of the data (such as how the census counts a particular category). This kind of talk is exactly the kind of exploration and meaning making the exhibit is intended to support, and thus were assigned a weight of three. Quantitative “content scores” were calculated by summing the weighted values of all codes applied to a session.

### Validity and limitations of magnitude coding

The assumption being made by this approach is that the overall richness of codes corresponds to the overall richness of the shared learning experience throughout the session, and that idea units are used to avoid over- or under-representing that richness. There is no assumption that a session's value should be determined by the number of "high value" (i.e. multi-coded) idea units like the example above, or that calculating the average value of idea units over a session is a meaningful measure. Other methodologies exist to closely scrutinize individual discourse statements. Instead, we recommend summing all codings applied over a session to assign a quantified value to what learners were able to do in the session. Looking at another example:

"Oh yes, lots of West Indians in Brooklyn, that is true."

This statement *INSTANTIATES-dataset* ("West Indians", 1) + *INSTANTIATES-geography* ("Brooklyn", 1) + *EVALUATE-characterizes* ("lots of", 2) + *INTEGRATE-connect:simple (dataset to geography)* ("West Indians in Brooklyn", 1) + *GENERATE-confirms* ("oh yes ... that is true", 3) = content score of 8. Whether this was delimited as one 8-point idea unit versus a 3-point idea unit ("Oh yes... that is true") plus a 5-point idea unit ("lots of West Indians in Brooklyn"), the impact on the session score is the same. Because of this flexibility in segmenting idea units, the methodology does not recommend analyzing scores of individual idea units (e.g., to compute metrics like "average value per idea unit"), but only the total dialogue in a session.

Session scores are intended to illuminate differences among conditions in their ability to support visitors in productive exploratory talk. Even the codes identified as "low relevance" are still productive learning talk. In this context, high-value codes often (but not always) build on low and mid-value talk, and a good statement often contains all three. For example, consider again the 8-point statement above. An alternative method for evaluating visitors' dialogue might be to focus on the proportion of codes at each relevance level (high, mid, and low) applied in each session. If you look at proportions, that example becomes 3/5 (or 60%) low-relevance, and 20% mid-relevance and 20% high-relevance. So by that system, a much simpler statement like "It looks like there are more of them" is a 100% high-level statement, because it is an *INTEGRATE-compare* without any elaboration. But in terms of what it's adding to the conversation—keeping in mind sociocultural perspective on learning most studies of museum learning adopt—it is nowhere near the same level of substance. The simpler high-level statement gives the speaker's companions fewer "hooks" to build on: they can only respond to the comparison, whereas the more complex statement gives companions a number of different directions to take the conversation. Given the open-ended nature of the interactions and the underlying assumption that each group will be approaching the exhibit from a unique background and with unique goals, it is to be expected that productive interactions will not be the same for each group. We do not assert there is an ideal ratio of low to mid to high statements. Five example sessions graphed in Figure 2 below demonstrate different profiles of productive conversation.

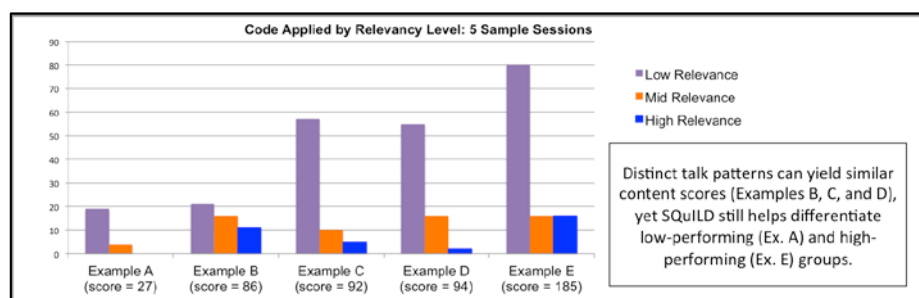


Figure 2. The SQulLD method is not intended to value a particular ratio as "best." Instead, dialogue is evaluated based on how well it aligns to learning goals and how many "hooks" visitors have to engage with each other and the presented content.

The average content score for all 119 coded sessions of groups using our interactive data visualization exhibit was 69.4 ( $SD = 42.6$ ). Example A in Figure 2 was a low-performing group, with a content score of 27. This interaction involved a largely one-sided dialogue, with one active participant narrating her activities and making some interpretive statements but receiving very little input from her companion. Examples B, C, and D, by contrast, all scored roughly half a standard deviation above the mean, but they achieved those scores in different ways. The pair in Example B had the highest proportion of high-relevance talk of any of our examples but overall fewer codes applied, resulting in a session score of 86. Examples C and D both had fewer high-relevance codings but made up for them with more low- and mid-relevance codes, resulting in scores of 92 and



94. By comparison, both members of the high-performing group in Example E were actively engaged in data interpretation, building off each others' comments and their own observations. The richness of their discussion is evidenced by the high numbers of codes applied in all three categories and their high overall score of 185. The utility of this weighted coding system is that it allows different kinds of engagement (like Examples B, C, and D above) to be acknowledged as productive while still distinguishing low (e.g. Example A) and high (Example E) performing groups.

The final point to consider in applying a magnitude coding scheme is the numerical values assigned to each code level. The research team felt values of 1-2-3 for low-mid-high codes best reflected the relationship among code levels, but a full analysis conducted by Roberts & Lyons (in progress) vetted this assumption by testing two alternative scoring proportions. In the 119 visitor sessions analyzed for that study, the results of the A/B testing were consistent regardless of the scoring proportion, i.e. the same design "won" in all scoring versions (Roberts & Lyons, in progress). The SQuILD methodology, by meaningfully segmenting dialogue and applying codes relevant to open-ended discussion and weighted according to their alignment with the exhibit's learning goals, provide a valid quantitative measure for conducting A/B testing and informing exhibit design decisions.

## Discussion

Understanding visitors' dialogue as they interact naturalistically with museum exhibits is of great interest to museum researchers but has traditionally been difficult to quantify for A/B testing. SQuILD accomplishes this through the combination several techniques. First, the context of the dialogue is maintained by coding all talk directly from the video recordings, which both illuminates the referents in visitors' conversations and allows segmentation of dialogue into a meaningful unit for the spontaneous, flowing discourse occurring in museums: the "idea unit." Because idea units as defined here, modified from Jacobs et al. (1997), can flow across multiple users as visitors co-construct the dialogue, they most accurately reflect the content of the shared meaning-making occurring in these interactions. Segmenting dialogue this way addresses a problem addressed by Allen (2002) in dealing with visitor discourse that tends to be "fragmented, ambiguous, or lacking clear referents" and that frequently involves repetition of words and phrases as members of a group echo each other. Allen dealt with this issue by coding only for the presence or absence of a type of talk during the entire interaction. Breaking the discourse into idea units that are then coded individually for presence or absence of a type of talk allows a clearer picture of the content of the dialogue to emerge.

By using simultaneous coding (Saldaña, 2009), a single idea unit can be coded with any number of codes, capturing not only the content of each statement but also the depth and complexity of the talk. Because each code applied, though productive, is not equally relevant to the learning goals of the exhibit, SQuILD utilizes magnitude coding (Saldaña, 2009; Miles & Huberman, 1994; Tashakkori & Teddie, 2010) to differentiate codes by their relevance to the exhibit's learning goals, much the way a teacher uses a rubric to quantify a student's piece of creative writing. By assigning weighted values to the codes applied, differences in the dialogue generated in each session are quantifiable and available for statistical comparisons.

We successfully employed this methodology in an *in situ* A/B comparison of different interaction designs for an collaborative data visualization exhibit (Roberts & Lyons, in progress), which allowed us to perform a number of statistical analyses exploring the relationship between exhibit use and the learning talk, and helped persuade stakeholders that a favorite design, despite being innovative and vetted in lab studies, was actually less successful for supporting learning talk. We expect that the SQuILD methodology will be useful to other researchers and designers developing open-ended computer-supported collaborative exhibits, in making similar assessments of the ecological validity of their design decisions on visitor learning. As museums embrace the dialogic model of education, they must concurrently embrace research methods suited to that model. The SQuILD methodology presented here takes steps toward that goal.

## References

- Allen, S. (2002). Looking for learning in visitor talk: A methodological exploration. *Learning conversations in museums*, 259-303.
- Atkins, L. J., Velez, L., Goudy, D., & Dunbar, K. N. (2009). The unintended effects of interactive objects and labels in the science museum. *Science Education*, 93(1), 161-184.
- Ash, D. (2003). Dialogic inquiry in life science conversations of family groups in a museum. *Journal of Research in Science Teaching*, 40(2), 138-162.
- Bernard, H.R.(2006). *Research methods in anthropology:Qualitative & quantitative approaches*. AltaMira Press.
- Bitgood, S. (2000). The Role of Attention in Designing Effective Interpretive labels. *Journal of Interpretation Research*, 5(2), 31-45.

- Block, F., Hammerman, J., Horn, M., Spiegel, A., Christiansen, J., Phillips, B., ... & Shen, C. (2015, April). Fluid grouping: Quantifying group engagement around interactive tabletop exhibits in the wild. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 867-876). ACM.
- Borun, M., Chambers, M., & Cleghorn, A. (1996). Families are learning in science museums. *Curator*, 39(2), 262-270.
- Chi, M. T. H. (1997). Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. *Journal of the Learning Sciences*, 6(3), 271-315.
- Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18, 382-393.
- Davis, P., Horn, M. S., Schrementi, L., Block, F., Phillips, B., Evans, E. M., ... & Shen, C. (2013). Going deep: Supporting collaborative exploration of evolution in natural history museums. In Proceedings of 10th International Conference on Computer Supported Collaborative Learning.
- Diamond, J., Luke, J. J., & Uttal, D. H. (2009). *Practical Evaluation Guide: Tools for Museums & Other Informal Educational Settings* (2 ed.). Lanham, MD: AltaMira Press.
- Eberbach, C., & Crowley, K. (2005). From living to virtual: Learning from museum objects. *Curator: The museum journal*, 48(3), 317-338.
- Falcão, T. P., & Price, S. (2009). What have you done! The role of 'interference' in tangible environments for supporting collaborative learning. In Proceedings of the Conference on Computer Supported Collaborative Learning, CSCL'09. ISLS, 324-334.
- Falk, J. H., & Dierking, L. D. (2000). *Learning from museums: Visitor experiences and the making of meaning*. Altamira Press.
- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Math Education*, 32(2), 124-158.
- Heath, C., & vom Lehn, D. (2008). Configuring 'Interactivity' Enhancing Engagement in Science Centres and Museums. *Social Studies of Science*, 38(1), 63-91.
- Jacobs, J. K., Yoshida, M., Stigler, J. W., & Fernandez, C. (1997). Japanese and American teachers' evaluations of mathematics lessons: A new technique for exploring beliefs. *The Journal of Mathematical Behavior*, 16(1), 7-24
- Kisiel, J., Rowe, S., Vartabedian, M. A., & Kopczak, C. (2012). Evidence for family engagement in scientific reasoning at interactive animal exhibits. *Science Education*, 96(6), 1047-1070.
- Leinhardt, G., & Crowley, K. (1998). Conversational elaboration as a process and an outcome of museum learning. *Museum Learning Collaborative Technical Report (MLC-01)*. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.
- Leinhardt, G., & Knutson, K. (2004). *Listening in on museum conversations*. AltaMira Press.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Murray, T. S., Kirsch, I. S., & Jenkins, L. B. (1998). *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey*. US Government Printing Office, Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328. Web site: <http://nces.ed.gov>.
- Povis, K. T., & Crowley, K. (2015). Family Learning in Object-Based Museums: The Role of Joint Attention. *Visitor Studies*, 18(2), 168-182.
- Roberts, L. C. (1997). *From Knowledge to Narrative: Educators and the changing museum*. Washington: Smithsonian Institution Press.
- Roberts, J., and Lyons, L. (in progress). Bigger May Not Be Better: Impacts of Whole Body Interaction and Distributed Control on Visitors' Learning Talk at an Interactive Data Museum Exhibit.
- Rounds, J. (2006). Doing identity work in museums. *Curator: The Museum Journal*, 49(2), 133-150.
- Saldaña, J. (2009). *The coding manual for qualitative researchers*. Sage
- Serrell, B. (1996). *Exhibit Labels: An Interpretive Approach*. Walnut Creek, CA: Altamira Press.
- Tashakkori, A., & Teddlie, C. (2010). *Sage handbook of mixed methods in social & behavioral research* (2nd ed.). Los Angeles: SAGE Publications.
- Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.
- Williams, A., Kabisch, E., & Dourish, P. (2005). From interaction to participation: Configuring space through embodied interaction. Proceedings of the Ubicomp 2005, LNCS 3660 (pp. 287-304).

## Acknowledgments

This material is based upon work supported by the National Science Foundation under NSF INSPIRE 1248052.

# Learning to Model Ecosystems With Interaction Food Webs in Middle School Classrooms

Michelle Lui and Tom Moher  
mmylui@uic.edu, moher@uic.edu  
University of Illinois at Chicago

**Abstract:** Modeling is an important part of scientific practice that represents phenomenon but also allows for prediction and reasoning about systems. While students typically do not have issues with understanding simple linear relationships, difficulties arise when more complex relationships are introduced. Richer representational diagrams make visible the complexities within ecosystems and can help students better understand interacting forces. We co-designed an instructional sequence with software applications to support middle school students building interaction food webs as models of ecosystems. 48 students participated in the lessons and used the applications in groups of three to four students. 87% of groups were able to successfully produce an interaction food web. Posttests found that over 80% of students gave a completely or partially correct explanation, demonstrating reasonable proficiency in understanding interaction food webs. The results from each step of the instructional sequence are discussed with respect to the curriculum and technology design.

## Introduction

Modeling is increasingly being recognized as an important aspect of scientific literacy and practice. According to the Next Generation Science Standards (NRC, 2012), students can begin developing and using models as early as Kindergarten (at age 4-5). They recommend that by middle school (age 11-12), students progress to constructing, using, and revising conceptual models to communicate their understandings and making predictions about more abstract phenomena. Used to represent real-world phenomena (yet not corresponding completely), models are abstractions that allow scientists and students to test theories and develop a better understanding of complex systems (Schwarz et al., 2009). Expressed models (as opposed to mental models) can take on several different forms, including schematic diagrams, physical replicas, mathematical representations, analogies, and computer simulations (Baek et al., 2011; Gobert & Buckley, 2000). Examples of scientific models include molecular models, weather systems for predicting weather patterns, the water cycle model, and food web models that highlight interactions between organisms. All of these not only represents phenomena but also allows for prediction and reasoning about real-world systems (Schwarz et al., 2009; Soloway et al., 1995). For example, building a food web that models an ecosystem not only demonstrates student understanding about an ecosystem, it can also foster deeper understanding of how different organisms relate to one another during the construction process. Once created, students can use the food web to make predictions about how one species' population can affect others within the model. However, studies have demonstrated difficulties learners encounter with understanding relationships within systems (Gallegos, Jerezano & Flores, 1994; Mandinach & Cline, 1994). While students typically have no trouble understanding single causal and linear connections, relationships within systems pose a bigger challenge (Ben-Zvi Assaraf & Orion, 2005; Grotzer & Basca, 2003; Hmelo-Silver, 2007). In the case of food webs and ecosystems, students generally understand predator-prey food chains quite well but have difficulty interpreting food web dynamics if two populations are indirectly related (Gallegos, Jerezano & Flores, 1994). When examining food webs some students think about species relationships on an individual level rather than on a population level (Griffiths & Grant, 1985).

Richer representational diagrams can make visible the complexities within ecosystems and have potential to help students attend to multiple interacting forces within the system. Some ecologists use interaction food webs to represent relationships amongst species (Hui, 2002). In traditional food webs, arrows between species indicate the direction of energy transfer, whereas in interaction food webs, each set of relationship is depicted by a pair of arrows, with each arrow accompanied by + or - symbol (Figure 1). The symbol denotes the resulting population effect of the species the arrow is pointing *at*, should the population of the species the arrow is pointing *from* increases (e.g., Rabbit —(+)→ Wolf may be translated as, if the rabbit population increases, it will have a *positive* effect on the population of wolves). Not only can an interaction food web represent feeding relationships, it can also express competition and mutualistic relationships within ecosystems. Since the semantic meaning of the arrows are tied directly to population effects, the building and using of interaction food webs are coupled tightly to the practice of making predictions. Once students learn how to read and build interaction food webs, the representation has potential for supporting students' reasoning around complex

ecosystems. Along with science teacher partners, we co-designed an instructional sequence and accompanying software applications for middle school students in order to support students' reasoning around interaction food webs. This paper describes the curriculum design, technology design, and the outcomes of student learning.

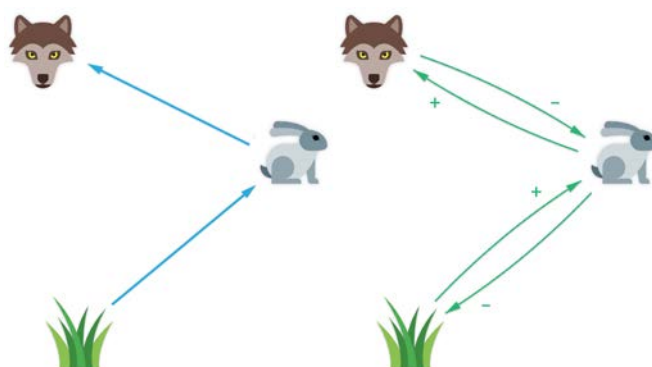


Figure 1. Example of traditional food web (left) and interaction food web (right).

## Related work

A number of prior studies have addressed modeling for students to understand ecosystems. The ScienceWare Model-It is a learning environment that allows high school students to study natural phenomena (such as a stream ecosystem) for building dynamic qualitative and quantitative models (e.g., visualizing relationships between phosphate on stream quality) in planning, building, and testing modes (Jackson et al., 1996). Qualitative results showed that scaffolding within software design supported model construction and that students created models of reasonable complexity and sophistication expected of their grade level. In another study, authors found that most students engaged in cognitive strategies such as analyzing, relational reasoning, synthesizing, testing and debugging, and explaining while using Model-It, however additional support is needed for all students to progress beyond superficial relationships (Stratford, Krajcik, & Soloway, 1998). In Aquatic Ecosystems, a two-week intervention with 311 middle school students, students participated in a technology-supported inquiry unit organized around structure, behavior, and function (SBF; Hmelo-Silver et al., 2014). Modeling technology supports included NetLogo simulations, and an Aquarium Construction Toolkit modeling application. Students significantly improved their understanding of the aquarium ecosystem in terms of structure, behavior, and function, as well as micro and macro relationships. Authors attribute the success of their unit in part to the use of a distinct conceptual representation that allows students to adopt the SBF conceptual framework into their language for expressing complex ideas about ecosystems. While the Model-It studies highlight the need for scaffolding and curricular support around modeling, research on Aquatic Ecosystems demonstrate success with representations that reinforce language for expressing complex ideas. Our instructional sequence and technology design takes the above findings into consideration, and is an example of an additional curricular support designed around providing students with representations and language for model building and reasoning.

## Wallcology

The designs described in this paper are created as part of the Wallcology unit, in response to challenges students encountered in a previous iteration (Slotta, Lui, Cober & Moher, 2017). The larger project centers on a complex phenomenon embedded in the physical classroom environment (e.g., walls), providing an evidentiary base for student inquiry. In the Wallcology phenomenon, simulations of fictitious organisms are "embedded" in the classroom walls. "Wallsopes" (i.e., computer monitors) provide internal views of the walls that reveal different ecosystems of varying abiotic (wood, plaster, and brick) and biotic components (vegetation, herbivores and predators). An underlying computational model developed in consultation with an expert population biologist governs these species and their environmental conditions. Student groups observe and perform investigations on specific ecosystems, with each ecosystem inhabiting a subset of species (e.g., 4-5) that make up a cohesive community of species (e.g., 11 species). Students are guided by collective inquiry activities (Cober et al., 2012) in order to model all of the species' relationships as a class – with the complete set of relationships across all of the species in the community discoverable only by aggregating everyone's observations and investigation finding. In the previous iteration, students had difficulties creating food webs and models, which took curricular time intended for students to actually work with their models.

Building upon prior research on modeling ecosystems and our own past experiences implementing the Wallcology unit, we designed explicit supports for modeling, including tablet applications where students explore feeding relationships, competition and indirect relationships (e.g., effects along a trophic cascade) before they begin engaging with the Wallcology simulation. Understanding that food chains must be taught not as a simple set of isolated organisms, but as an interactive population embedded in an ecological context (Gallegos, Jerezano & Flores, 1994), and that elementary-school students have capability to develop systems thinking skills (Evagorou et al. 2009), we developed an instructional sequence for modeling ecosystems based on an interaction food web representation. This paper describes this curriculum and technology. Our aim is to help students understand complex ecosystems, and to give students the tools to successfully make predictions and reason around food web models. The research questions that drive the current study include: (1) how well can middle school students understand and build interaction food webs? And (2) to what extent can they make predictions and reason around relationships depicted within interaction food webs?

## Methods

This research employs a design-based methodology, characterized by iterative cycles of design, evaluation and revision of an intervention for study in authentic classroom settings (Brown, 1992; Design-Based Research Collective, 2003). Researchers and two high school teachers developed curriculum activities, content materials and specialized software using a co-design method (Penuel, Roschelle, & Shechtman, 2007). The study described in the current paper took place in private urban middle school in Chicago across three Grade 6 Science classes, which were taught by one of the co-design teacher partners. A total of 48 students (aged 11-12, with 16 students per class) participated in the lessons in groups of three to four during their regular science class over a period of two days. Each class period ranged from 50 minutes to 1 hour and 50 minutes, depending on the class and day. A posttest was conducted in the class period following day 2. Several data sources were used in the analysis: the models the students constructed, software-generated log files of each group's activities with the application, and audio and videotapes of student discussions and class activity.

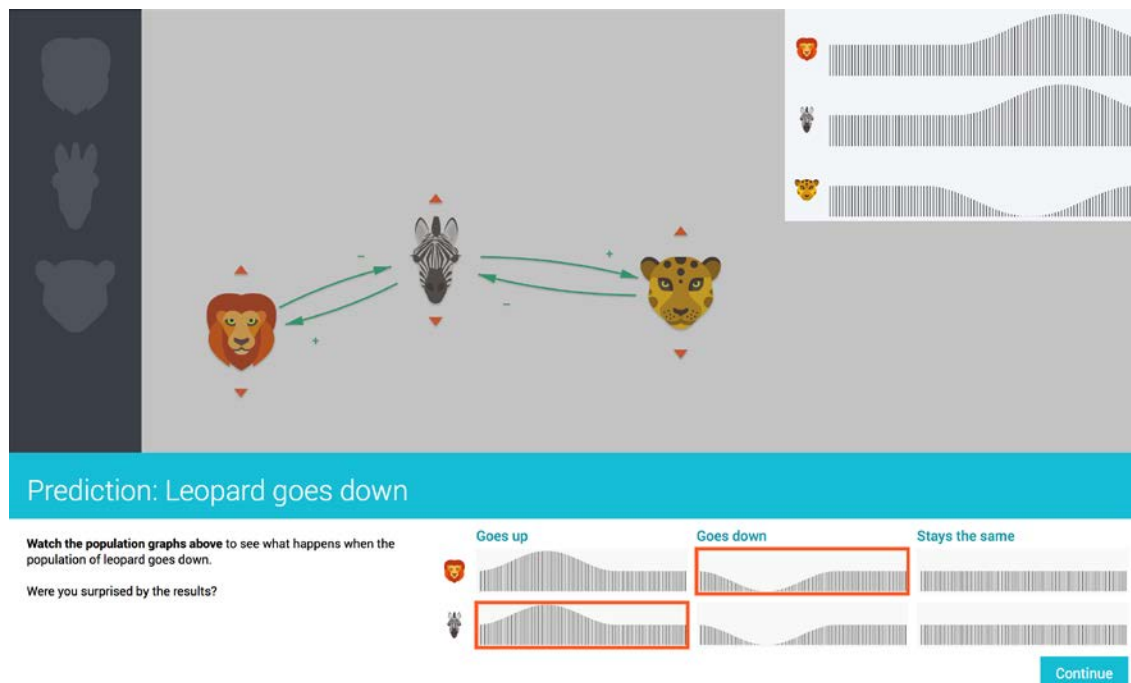
## Design

### Day 1: Introduction to interaction food webs

On day 1, students worked in groups of three to four (five groups per class) to learn about the interaction food web representation using a custom software application. Each group was provided with a laptop computer. There are four stages in the program, taking groups through successively complex relationships (Table 1). The first two stages guided students through with step-by-step instructions, while the latter two stages allowed students more latitude to explore various relationships and meaning of + and – symbols. At each stage, a palette on the left is populated with recognizable species (e.g., lion and zebra). As students dragged species into the work area, a graph depicting its population over time is revealed. When species that are related to one another are placed on the work area, arrows are automatically established. The work of the student then, is to make changes to species populations. Up and down arrows for each species are made available (i.e., increase or decrease populations respectively). When a population manipulation is attempted, students are tasked with making predictions about the population effects on the other species in the work area. Only then will the population graphs reveal the resultant effects (Figure 2).

Table 1: Stages in introduction to interaction food web application (day 1)

Level	Aim	Species Available
1	To explore population effects with food chains	Lion, zebra
2	To explore population effects with interaction food webs notation	Lion, zebra
3	To explore population effects with interaction food webs (feeding relationships only)	Lion, zebra, leopard
4	To explore population effects with interaction food webs (feeding and competition relationships)	Lion, zebra, grass, acacia tree



**Figure 2.** Introduction to interaction food webs software interface (level 3). The application shows a prediction panel (which pops up from bottom) when students select a manipulation of species population. In the screen above, the red ‘down’ arrow below the leopard has been selected, indicating that an attempt was made to decrease the leopard population. Once predictions about lion and zebra population have been made, students can compare their answers with the animated population-over-time graphs (top right).

## Day 2: Modeling with interaction food webs

On day 2, students worked in newly formed groups of three to four (five groups per class) to build their own interaction food webs using a modeling application on a tablet computer (Figure 3). The modeling application is based on the software design from day 1, with differences being that: (1) students could choose the relationships they wished to depict and (2) non-specific “species” (i.e., shapes to represent species) were used in place of recognizable species. Similar to the application on day 1, up and down arrows were made available for each species, and when a population manipulation is attempted, students were tasked with making predictions about the population effects of other species in the work area. Once completed, the population graphs reveal the ensuing effects. Students explored the meaning of + and – of relationship arrows through an iterative process of prediction and revision. Each group was assigned four different sets of species examined during their recent field trip to a local ecosystem (e.g., sanderling, Karner blue butterfly, butterfly milk weed) and was asked to create a traditional food web (with energy transfer arrows). Then, using the modeling application, they were to determine the correct relationship representations and expand the basic food web into an interaction food web. The instructional sequence concluded with group presentations of population interaction web and class discussions about population effects of the species.



**Figure 3.** Students working with the modeling application on day 2.

## Results and discussion

### Day 1: Introduction to interaction food webs

As groups (n=15) worked through each level, they made predictions about population effects when they manipulated another species' population. The accuracy of groups' predictions was used to evaluate their performance. The mean prediction accuracy was 81.27% (SD = 13.35%), broken down by level in Table 2.

Table 2: Prediction accuracy on day 1

Level	Performance Accuracy (%)	SD (%)
1	86.71	18.66
2	91.22	16.58
3	79.73	21.68
4	76.06	17.72
Average	81.27	13.35

Given the difficulty of the tasks, taking into consideration the relative young age of the learners, these results are promising. Prior studies noted that the concept of populations of organisms in the wild is established in children older than age 13, but remains challenging for them to reason about how populations depend upon each other and compete with one another (Driver et al., 1997).

Further examination of students' performance pattern over successive levels, reveals that groups with the highest accuracy scores in levels 3 and 4 also attained 100% in the first two levels (top four groups). However for groups with the lowest scores in levels 3 and 4, their averages in levels 1 and 2 ranged from 60% to 100% (bottom four groups). Students' group discussions during the activity suggested that they were using what they know about the species to discuss population changes (e.g., "lions eats zebras, so when there are more lions there are less zebras"). They had to make predictions, and were excited when their population relationships matched what the population graphs showed. There was little evidence of students attending to the arrows or the symbols on day 1, but additional video analysis is underway to understand groups' performance patterns over the course of the activity.

### Day 2: Modeling with interaction food webs

Students presented their final interaction food webs in groups, with each student taking a turn at explaining what would happen to the population of one species if the population of another species increased or decreased. Most students were able to correctly explain population effects. With respect to the interaction food webs that groups presented (Figure 3), 87% of the groups were able to produce a completely correct or partially correct interaction food web, 67% (10 of 15 groups) produced completely correct interaction food webs, 20% (3 groups) had minor errors, 2 groups neglected to include a symbol on one of the arrows but the representation was otherwise correct. 1 group reversed the symbols for one set of relationship, and 13% (2 groups) were unable to produce an interaction food web, instead presenting a traditional food web (Figure 4).

During class discussions, several students were active in debating the validity of their models. In one scenario, students discussed what would happen if one of two prey populations of a predator suddenly disappeared? One student suggested that the predator population would also decrease which would lead to the remaining prey population to thrive (since their predator population decreased). Another suggested that the predator population would simply make up more of its diet with the remaining prey population, which would result in a decreased population of the remaining prey population. This discussion led to the class discussing issues about the model being a simplified representation of the actual system, and that in reality, there would be other factors involved (such as how prevalent are each of the species' populations? How much of the predator's diet is made of one prey species vs. the other?). These discussions further demonstrated the students' sophistication in understanding interaction food webs as models of ecosystems and their ability to reason around them. Furthermore, the language that students used throughout their presentations and class discussions was indicative of their thinking about ecosystems on a population level rather than on an individual species level.

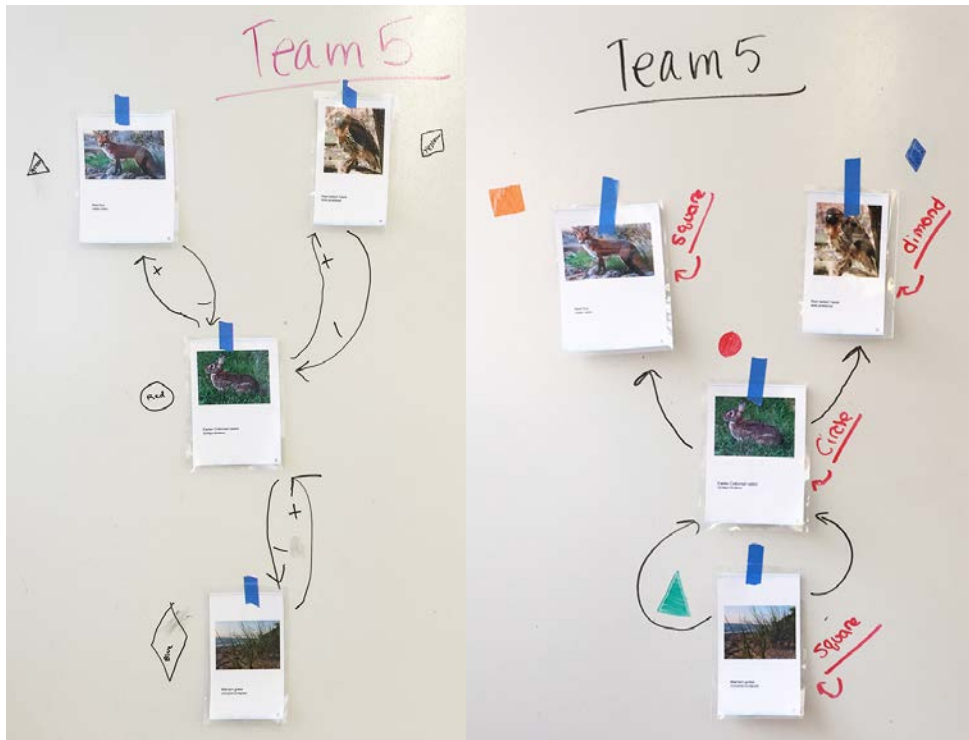


Figure 4. Examples of group interaction food webs. Correct version (left). Incorrect version (right).

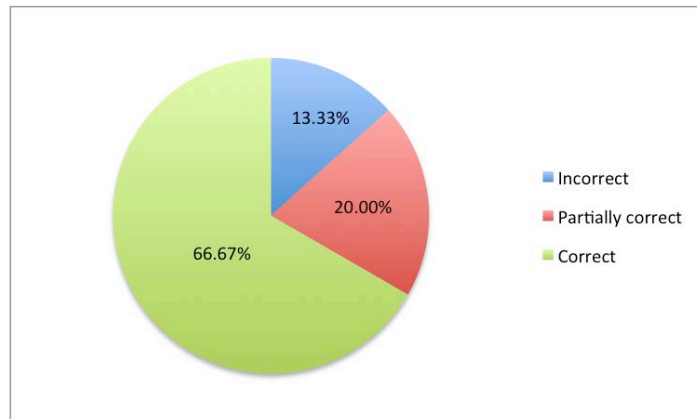


Figure 3. Accuracy performance of group interaction food webs.

## Posttest

41 students completed the posttest, in which students were asked to build a model of an interaction food web consisting of square, triangle, circle and diamond "species" and explain population effects from changes in two different populations. For example, if the population of triangles suddenly disappeared, what would happen to the square population?

A composite scoring was assigned based on whether both answers were correct, incorrect, or incomplete. The test was scored as correct if both explanations were correct, partially correct if one of the answers was correct, incomplete if explanations were not provided in both answers, incorrect if both explanations were incorrect. Results showed that over 80% of students gave a completely or partially correct explanation, demonstrating some level of understanding. Further breaking down results, 60.98% offered correct explanations, 19.51% offered partially correct explanations, 9.76% of the explanations were incomplete, and another 9.76% of explanations were incorrect (Figure 5).



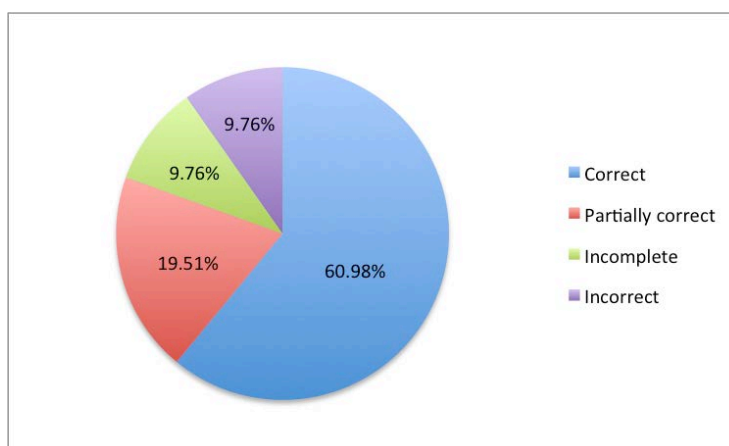


Figure 5. Frequency of explanation performance of posttest questions (n=41).

Of the students who produced incorrect or incomplete explanations (n=8), only one belonged in low performing groups on both day 1 and day 2 (defined as groups with lower than average performance scores on day 1, and groups who were unable to complete an interaction food web on day 2). Three students belonged in low performing groups on day 1, and the remaining students in this cohort belonged in high or average performing groups on both days. This suggests that these students may have missed core concepts learned on day 1.

Of the students who produced partially correct explanations (n=8), three students belonged in average performing groups on day 1, with the remaining students belonged in high performing groups on both days, echoing possibility that these students may have missed core concepts learned on day 1. Of the students who produced correct explanations (n=25), the majority 76% (n=19) belonged to high or average performing groups on both days. More than half (56%) belonged in high performing groups, however there were still some students who belonged to low performing groups, including: 3 (12%) who belonged in low performing groups on day 1 and 3 (12%) who belonged in low performing groups on day 2. It is possible that these students gleaned important information from the discussions that occurred during presentations, but further analysis will be conducted in order to understand these students' learning progressions.

## Conclusions

The majority of groups were able to successfully produce an interaction food web and analysis of posttest explanations found that over 80% of students were able to give completely or partially correct explanations about population effects along a trophic cascade. These results demonstrate that middle school students were able to attain reasonable proficiency in understanding and building interaction food webs. Students adopted appropriate language in discussing their interaction food webs, highlighting their predictions of population change throughout their presentations and class discussions. Given the complexity and abstract nature of the symbols at the heart of interaction food webs, these results suggest that, with proper scaffolding, middle school students may be capable of even more sophisticated systems thinking. A follow-up study of the Wallcology unit will reveal if students were successful in translating what they learned with the instructional sequence and technology applications described in the current study – to effectively build models and make predictions about much more complex ecosystems.

## References

- Baek, H., Schwarz, C., Chen, J., Hokayem, H., & Zhan, L. (2011). Engaging elementary students in scientific modeling: The MoDeLS fifth-grade approach and findings. In *Models and Modeling* (pp. 195–218). Dordrecht: Springer Netherlands. [http://doi.org/10.1007/978-94-007-0449-7\\_9](http://doi.org/10.1007/978-94-007-0449-7_9)
- Ben-Zvi Assaraf, O., & Orion, N. (2005). Development of system thinking skills in the context of earth system education. *Journal of Research in Science Teaching*, 42, 518- 560.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141-178.
- Cober, R., Fong, C., Gnoli, A., López Silva, B., Lui, M., Madeira, C., McCann C., Moher, T., Slotta, J., & Tissenbaum, T. (2012). Embedded Phenomena for Knowledge Communities: Supporting complex practices and interactions within a community of inquiry in the elementary science classroom.

- Proceedings of the 10th International Conference of the Learning Sciences (ICLS) - Volume 2, (pp. 64-71). International Society of the Learning Sciences (ISLS).
- Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5-8.
- Driver, R., Squires, A., Rushworth, P., Wood-Robinson, V. (1997) *Making Sense of Secondary Science: Research into Children's Ideas*. New York, Routledge.
- Evagorou, M., Korfiatis, K., Nicolaou, C., & Constantinou, C. (2009). An investigation of the potential of interactive simulations for developing system thinking skills in elementary school: A case study with fifth-graders and sixth-graders. *International Journal of Science Education*, 31(5), 655–674. <http://doi.org/10.1080/09500690701749313>
- Jackson, S., Stratford, S. J., Krajcik, J. S., & Soloway, E. (1996). Model-It: A case study of learner-centered software design for supporting model building. National Science Foundation, Arlington, VA. Retrieved from: [http://tccl.rit.albany.edu/knilt/images/6/66/case\\_study\\_student\\_centered\\_use\\_of\\_models.pdf](http://tccl.rit.albany.edu/knilt/images/6/66/case_study_student_centered_use_of_models.pdf)
- Gallegos, L., Jerezano, M. E., & Flores, F. (1994). Preconceptions and relations used by children in the construction of food chains. *Journal of Research in Science Teaching*, 31(3), 259–272. <http://doi.org/10.1002/tea.3660310306>
- Gobert, J. D., & Buckley, B. C. (2000). Introduction to model-based teaching and learning in science education. *International Journal of Science Education*, 22(9), 891–894. <http://doi.org/10.1080/095006900416839>
- Griffiths, A. K., & Grant, B. A. C. (1985). High school students' understanding of food webs: Identification of a learning hierarchy and related misconceptions. *Journal of Research in Science Teaching*, 22(5), 421–436. <http://doi.org/10.1002/tea.3660220505>
- Grotzer, T., & Basca, B. B. (2003). How does grasping the underlying causal structures of ecosystems impact students' understanding? *Journal of Biological Education*, 38, 16-35.
- Hmelo-Silver, C. E. (2014). Technology-supported inquiry for learning about aquatic ecosystems. *Eurasia Journal of Mathematics, Science & Technology Education*, 10(5), 405–413. <http://doi.org/10.12973/eurasia.2014.1170a>
- Hmelo-Silver, C., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: expert-novice understanding of complex systems. *Journal of the Learning Sciences*, 16, 307- 331.
- Hui, D. (2012) Food web: Concept and applications. *Nature Education Knowledge* 3(12):6.
- Hogan, K., & Thomas, D. (2001). Cognitive comparisons of students' systems modeling in ecology. *Journal of Science Education and Technology*, 10(4), 319–345. <http://doi.org/10.1023/A:1012243102249>
- Leach, J., Driver, R., Scott, P., & Wood-Robinson, C. (1996). Children's ideas about ecology 2: ideas found in children aged 5-16 about the cycling of matter. *International Journal of Science Education*, 18, 19-34.
- Mandinach, E. B., & Cline, H. F. (1994). *Classroom dynamics: Implementing a technology-based learning environment*. Hillsdale, NJ: Lawrence Erlbaum.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, core ideas*. Washington, DC: National Academies Press.
- Penuel, W. R., Roschelle, J. & Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. *Research and Practice in Technology Enhanced Learning*, 2(1), 51-74.
- Schwarz, C., Reiser, B., Fortus, D., Shwartz, Y., & Acher, A. (2009). Models: defining a learning progression for scientific modeling. Presented at the Learning Progressions in Science LeaPS Conference, Iowa City, IA. Retrieved from: <https://www.sensepublishers.com/media/593-learning-progressions-in-science.pdf>
- Slotta, J. D., Lui, M., Cober, R., Moher, T. (April, 2017). Modeling Wallcology: Technology scaffolds for building, predicting and reasoning around models of simulated ecosystems. In Karan, A. (Chair), *Supporting Science as a Modeling Practice in the Classroom through the Lens of NGSS*. Paper presented at the annual meeting of the American Educational Research Association (AERA), San Antonio, TX.
- Soloway, E., Krajcik, J., & Finkel, E. A. (April, 1995). The ScienceWare project: Supporting science modeling and inquiry via computational media & technology. Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA. Retrieved from: <http://www.uqac.ca/deptdse/3ped124/foundation/science.pdf>
- Stratford, S. J., Krajcik, J., & Soloway, E. (1998). Secondary students' dynamic modeling processes: Analyzing, reasoning about, synthesizing, and testing models of stream ecosystems. *Journal of Science Education and Technology*, 7(3), 215–234. <http://doi.org/10.1023/A:1021840407112>

# The Role of Visual Representations Within the Scientific Practice of Explanation

Rebecca Quintana, University of Michigan, rebeccaq@umich.edu

Tom Moher, University of Illinois at Chicago, moher@uic.edu

Jim Slotta, Boston College, slotta@bc.edu

**Abstract:** Two Grade 5/6 classes (n=47) participated in an 8-week biology unit where they investigated WallCology, a digital ecosystem. During a series of oral presentations, students presented findings from their digital experiments. Our goal is to understand the role of the visual representations (e.g., digital screens, graphs) that students referenced during these presentations. We use the Claim Evidence Reasoning framework to establish the suitability of the presentation task for scientific explanations and for evaluating students' explanations. We show that student groups were proficient in their scientific explanations and effectively used a variety of representation types across all three components of their explanations. We conclude that visual representations did not just influence scientific discourse, but were part of it.

## Introduction

A fundamental aim of science is to explain the world around us, in order to make sense of how and why various phenomena occur (Berland & Reiser, 2009). Scientists use a variety of representations (e.g., diagrams, equations, notations) to communicate what is currently known about phenomena, even whilst they are still working on a problem (Prain & Tytler, 2013). However, in typical science classrooms, learners are rarely afforded the opportunity to engage with visual representations in ways that align with scientific practice (Ainsworth, Prain, & Tytler, 2011). Instead, a common practice is to ask learners to construct representations *of something*, rather than to produce a representation *for something* (i.e., serving a concrete purpose, such as communicating to peers) (Greeno & Hall, 1997). If learners are to become more than passive consumers of science, educators need to envision scenarios where learners have the opportunity to invent, modify and discuss scientific representations (Danish & Enyedy, 2007). In this paper, we present a study where students are presented with the challenge of making sense of unfamiliar scientific phenomena and are offered the opportunity of “practicing representations” (Greeno & Hall, 1997) as they engage in the scientific practice of explanation with their peers.

## Theoretical foundations

A scientific explanation responds to a question, using data analysis and interpretation of scientific principles. The Claim Evidence Reasoning (CER) framework (McNeill & Krajcik, 2011) is designed to provide students with guidance on how to structure scientific explanations that are part of their scientific writing, discussions, and oral presentations. This instructional model includes three components (claim, evidence, and reasoning) and asks learners to create a coherent explanation by linking all three components together, with evidence providing support for a claim and reasoning linking evidence to the claim. To date, research that uses the CER framework has not focused on the role of visual representations within the scientific practice of explanation. To that end, this paper offers a new dimension that can contribute to this body of work.

### WallCology narrative

WallCology is a digital simulation that is part of the Embedded Phenomena framework (Moher, Uphoff, Bhatt, López Silva, & Malcolm, 2008). WallCology consists of fictional animated species that interact with each other and the environmental features of their ecosystem. Students investigate ecological concepts, such as how species interact with each other (e.g., producer consumer relationships) and the effect of environmental stressors on population levels.

In the first phase of the unit, students are introduced to the environment when their teacher asks them to imagine that there are four distinct ecosystems that exist behind each of their classroom walls. They are told the only way to view them is through specialized “wallsopes” (i.e., digital monitors). Students are tasked with observing the ecosystems and asked to pay close attention to the behaviors and interactions of digital species. Each ecosystem contains its own unique configuration of abiotic characteristics including temperature, available pipe, and available brick, as well as a unique subset of the 11 WallCology species. These factors combine to create a complex inquiry space for students to explore. The students work together as a scientific community to understand how each species interacts with its environment and to discover the food web relationships among

the species. In the second phase of the unit, a major perturbation (e.g., habitat destruction, climate change) impacts each ecosystem, causing dramatic changes to population levels within the ecosystems. Working in one of four table groups, students are then challenged to uncover the chain of events (starting from the initial perturbation) that led to the current state of the ecosystem. In the final phase of the unit (the investigations phase), student groups are given the opportunity to adjust the population levels of species within each ecosystem (by increasing or decreasing them) and to introduce new species into the ecosystem. After each investigation cycle (ecosystem manipulation), student groups present the results of each “investigation” to their peers in the form of an oral presentation. The role of visual representations within the scientific practice of explanation as demonstrated through these presentations is the focus of this study.

### Research questions

In this study, we explore how students engaged in scientific explanations as part of the presentations that they made within the investigation phase of the WallCology unit, with a focus on the types of visual representations that they elicited as evidence for their claims. The following research questions guided our analysis.

1. How does the WallCology curriculum and learning environment create opportunities for students to engage in the scientific practice of explanation?
2. What role do visual representations play within each component of students’ scientific explanations (e.g., claim, evidence, and reasoning)?

## **Methodology**

### **Participants**

This study was conducted at a university laboratory school in Ontario, Canada. The school has a long history of promoting inquiry-based approaches to learning and instruction, including knowledge building and learner-centered pedagogies. The student body consisted of learners from varied cultural and socio-economic backgrounds. Two Grade 5/6 classes participated (n=47), along with their teachers, Brad and Mark (pseudonyms). Both Brad and Mark had experience in facilitating collaborative forms of learning within their classrooms, with ten and five years of teaching experience respectively.

### **Learning environment and curriculum materials**

Two research groups from the University of Toronto and the University of Illinois at Chicago and classroom teachers from our partner schools participated in a co-design process (Roschelle, Penuel, & Sheckman, 2007) to design and develop the learning environment and technology-enhanced learning materials. Co-design processes can foster a sense of trust among design team members and can ensure that innovations are a good fit for a particular classroom context.

### Technology designs

The co-design team, including teachers, researchers, software developers and graphic artists, worked together to create a suite of web-based tools to support idea and knowledge exchange among students as they progressed through the WallCology narrative. Included within these tools was a brainstorm space, where students could share newly emerging ideas with peers, an observation space, where students could exchange structured notes (including visual media) and pairwise observations about producer-consumer relationships, a population graph tool, where students could specify species and environmental factors (e.g., temperature) to compare, and an investigation space, for student groups to propose a population change to one or more species in an ecosystem, predict how all other species will be affected by the change, justify predictions, and record the results of the investigation, including an explanation for the reported results.

In this study, we focus on students’ use of the *investigation space*, and consider how they used the presentation screens within that tool to spotlight the important elements of their investigations. The investigation space consisted of several screens that allowed students to proceed stepwise through each stage of their investigation to (1) describe the current state of the ecosystem (i.e., health of the ecosystem as evidenced by population levels), (2) write an investigation plan whose goal is to improve the overall health of the ecosystem, (3) make a prediction concerning the outcome of an investigation (i.e., manipulation to the population of one or more species), and (4) report the results of their investigation. After contributing descriptions, plans, predictions, and reports for all stages of their investigation, students could display a condensed view of their investigations using the tools *present screens*: plan, predict, and results. We designed these three views to allow students to easily show their peers an overview of their investigation by reducing the number of screens that they would

have to display during their presentations. Using these present screens, students could tap on attached media to enlarge them. Figure 1 shows a student group displaying an investigation present screen on the classroom SMART Board during one of their investigation presentations.

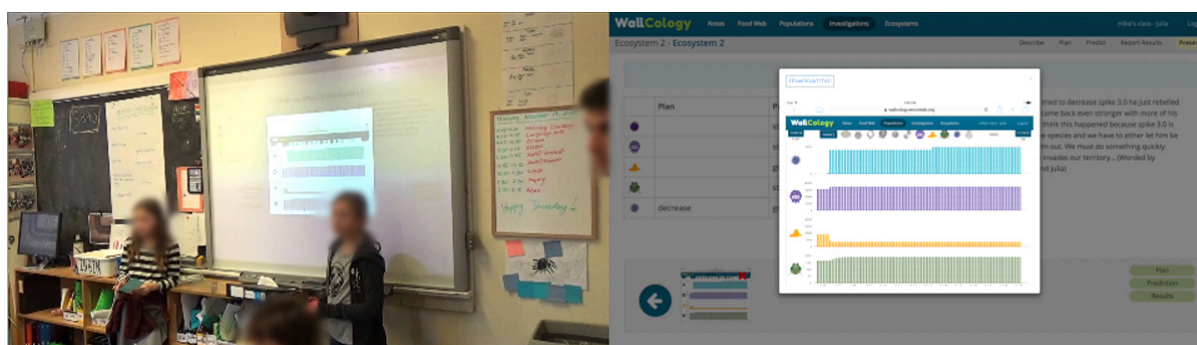


Figure 1. Student group enlarges media (a graph) during one of their investigation presentations.

### Activity designs

An important aspect of our activity designs was requiring students to engage in scientific explanations, particularly within the investigations phase of the unit. The classroom teachers introduced the investigations phase by asking students to consider the question: *What makes an ecosystem healthy?* Through whole class discussion, both classes elaborated on this concept and identified many features that would be indicators of the overall health of an ecosystem (e.g., balance of prey and predators, variety of species, able to support life). Teachers captured the important ideas from these discussions in concept maps that remained on display for the investigation phase.

Brad and Mark explained that table groups would have the opportunity to make adjustments to the digital ecosystem that they were responsible for using one of three mechanisms: introducing one or more new species, increasing a species that was already present in the ecosystem, or decreasing an existing species. The teachers reminded their students that in order to be successful in improving the health of their ecosystems, they would have to take into account the underlying cause of the changes in the ecosystems. For example, in Ecosystem 3, when an invasive predator entered the ecosystem, the indigenous predator was eliminated because the new predator had a significant advantage; it was a location generalist, whereas the indigenous predator was a location specialist (i.e., only existed in the brick habitat). The invasive predator had access to food sources that the indigenous predator did not. Since each ecosystem was impacted by a different perturbation, each table group needed to address a unique set of issues. However, because the interactions among species and the ways that species interacted with the abiotic features of their habitat were consistent across ecosystems, student groups needed a forum to share their scientific explanations with the entire knowledge community. The format for this knowledge exchange was the *investigation presentations*, where students used the present screens that are described above as a shared referent during their scientific explanations.

## **Analysis and findings**

### **Research question #1**

In the first phase of our analysis, we evaluated the WallCology learning task (i.e., investigation presentations) to establish that CER was “a good fit” to be used as a framework to evaluate students’ scientific explanations. We mapped the WallCology learning task to three categories outlined by McNeill and Krajcik (2011): (1) opportunities in the curriculum that are appropriate for CER, (2) the complexity of the learning task, and (3) classroom supports. To perform this mapping, we used a variety of data sources including provincial curriculum documents, field notes, photographs, and video recordings.

### Opportunities in the curriculum

To establish that WallCology presented a sufficient curricular *opportunity*, we looked at “learning performances” (i.e., a specification of what the students should be able to do, given both the content standard and scientific inquiry standard that is being targeted within a given activity) McNeill and Krajcik (2011). For this analysis, we were interested in examining the learning performances that related to the investigation presentation task. The co-design team consulted curriculum guidelines outlined by the Ontario Ministry of Education (2007) during the design phase and sought to map these curriculum goals to our activity designs. The

sixth grade “Understanding Life Systems: Biodiversity” unit outlines both content standards (i.e., expectations of student performance) and scientific inquiry standards (e.g., experimentation skills). According to McNeill and Krajcik (2011), a learning performance is defined as a product of content standards and scientific inquiry standards. Table 1 provides an overview of the learning performances that were part of the WallCology investigation presentation task.

Table 1: Overview of the learning performances that were part of the WallCology investigation task

Ontario curriculum expectations	Scientific inquiry/experimentation skills	WallCology learning performance
3.2. Demonstrate an understanding of biodiversity as the variety of life on earth, including variety within each species of plant and animal, among species of plants and animals in communities, and the physical landscapes that support them (p. 114).	<b>Initiating and planning</b> Students make predictions, based on prior knowledge from explorations and investigations about the results of an investigation (p. 13).	Students use their understanding of the WallCology ecosystems, including environmental characteristics and the species that inhabit each one, to <b>plan, make predictions</b> , and report on the results of each investigation.
3.7 <b>Explain</b> how invasive species reduce biodiversity in local environments (p. 114).	<b>Analyzing and interpreting data</b> Students identify patterns in the data, suggest explanations for discrepancies, and summarize the data. Students draw conclusions on the basis of data gathered (p. 14).	Students <b>explain</b> how an invasive species can reduce the biodiversity within their WallCology ecosystem, drawing on available data, such as their own observations and population graphs.
3.3 Describe ways in which biodiversity within species is important for maintaining the resilience of these species (p. 114).  3.4 Describe ways in which biodiversity within and among communities is important for maintaining the resilience of these communities (p. 114).	<b>Communicating</b> Student present steps in and results of an experimental procedure using numeric, symbolic, graphical, and/or linguistic methods (p. 14) to answer the question investigated (p. 16).	In the form of an oral presentation, students state a <b>claim</b> about the overall health of their ecosystem. They present their investigation plans/predictions/and reports to their peers, <b>answering the question</b> “How can we/did we improve the overall health of our ecosystem?”

As shown in Table 1, the WallCology learning performances associated with the investigation presentation task required that students engage in meaningful scientific explanation, combining curriculum expectations with scientific inquiry/experimentation skills. Students had to state a *claim* about how the health of their ecosystem will be improved through their manipulation, provide *evidence* in the form of known food web relationships and changes in population sizes of other species, and give *reasoning* about how changing the population of one species impacts multiple species in the food web, including both direct and indirect relationships.

### Complexity of the learning task

Complexity of the learning task can be described in terms of two aspects: (1) openness of question, (2) type of data, and (3) amount of data. The WallCology learning environment is on the more complex end of the spectrum in all three of these areas. Table 2 provides an overview of these three areas to demonstrate that students were working in a complex inquiry space. For example, students provided explanations in response to open-ended and semi-structured questions, rather than highly structured questions. Also, students used multiple forms of data, including quantitative, qualitative, and even contradictory data. The inquiry space was large and highly complex, with student groups having the ability to manipulate multiple species in one investigation and to perform multiple consecutive and accumulative investigations.

Table 2: Three aspects that describe the complexity of the learning task

Openness of question	Type of data	Amount of data
Students constructed explanations in response to <b>open-ended questions</b> : <ul style="list-style-type: none"> <li>How can we improve the health of our ecosystem?</li> </ul>	Students used <b>multiple forms of data</b> to construct their explanations. <b>Qualitative data:</b> <ul style="list-style-type: none"> <li>Students made observations about</li> </ul>	<ul style="list-style-type: none"> <li>Students could manipulate <b>up to 11 species</b> at one time.</li> <li>Students performed <b>multiple</b></li> </ul>

<ul style="list-style-type: none"> <li>• How can the factors under our control be manipulated to improve the health of our ecosystem?</li> </ul> <p>Students also made claims in response to <b>semi-structured questions</b>:</p> <ul style="list-style-type: none"> <li>• Is our ecosystem more or less healthy than it was prior to our last manipulation?</li> <li>• How has the introduction of an invasive species (the increase in temperature or habitat loss) contributed to a reduction in the health of our ecosystem?</li> </ul>	<p>the populations of species within their ecosystem and about the state of the habitats within their ecosystem.</p> <p><b>Quantitative data:</b></p> <ul style="list-style-type: none"> <li>• Students constructed population graphs using the WallCology graph tool.</li> <li>• Students constructed population graphs using the graph tool.</li> <li>• Students considered seemingly <b>contradictory data</b>.</li> </ul>	<p><b>consecutive and accumulative investigations.</b></p> <ul style="list-style-type: none"> <li>• Students attended to the investigations and ecosystems of other groups, not just their own.</li> </ul>
--	---	--

### Classroom supports

McNeill and Krajcik (2011) recommend that the components of the CER framework be displayed in the classroom as a memory aid for students as they provide written or oral explanations. Such charts could contain varying degrees of detail, from categorical titles (CER), to titles with descriptions and/or examples. In the classrooms we studied, there were no representations that referenced CER on display. However, the representations that captured the healthy ecosystem discussion (i.e., concept maps) remained on display throughout the investigation phase. These underlying concepts provided guidance for student groups as they constructed claims regarding the health or future health (i.e., after a manipulation) of their ecosystem.

According to McNeill and Krajcik (2011), curricular scaffolds are “sentence starters, sub-questions, graphic organizers, or prompts that help break down the task and provide students with hints on how to successfully complete the scientific explanation task” (p. 60). Our technology designs included CER specific language within the investigation phase prompts. McNeill and Krajcik (2011) also distinguish between general and content-specific scaffolds. General scaffolds provide support for scientific explanation but could be used with any content. Content-specific scaffolds provide support that is specific to the content area and to the task. The prompts that were provided to students in the investigation space were content-specific because they related directly to the tasks of planning, predicting, and reporting on a WallCology investigation. Table 3 provides a mapping of CER components to curricular scaffolds.

Table 3: CER components mapped to the WallCology curricular scaffolds

CER component	WallCology curricular scaffolds
Claim	What is your goal? What are you hoping to accomplish? What were the results of this investigation?
Evidence	You can add evidence to support your prediction (e.g., food web) For each species, choose whether the species went up or down.
Reasoning	Give your reasoning
	Justify your predictions. Why will some species go up or down? How do these changes affect other species in the ecosystem?

### Research question #2

The second phase of our analysis concerns the role of visual representations within the investigation presentation task. We analyzed all investigation presentations given by student groups in both Brad and Mark’s classes (n=32). We analyzed the content of student groups’ scientific explanations using a customized rubric, drawing on McNeill and Krajcik (2011). Table 4 presents a succinct version of this rubric. We cross-referenced each CER component with the type of visual representation(s) that students referenced during their presentations. We chose to include multiple forms of visual representations including system-generated resources (i.e., investigation present screens) and student-generated representations (e.g., population graphs, photographs). We also considered other knowledge resources that students referenced, including when they cited the “wallscope” themselves. We used StudioCode to analyze the investigation presentation videos and created segments according to the claim, evidence, and reasoning portions of student groups’ explanations. We coded segments with labels to describe representation types (learning environment, investigation screens, graphs, other representations, and none) and to describe the CER level that we assigned to each segment, using the customized rubric.

Table 4: Succinct view of customized CER Rubric

CER Component	Level		
	0	1	2
<b>Claim</b> – A conclusion that answers the original question: How can we make our ecosystem healthier?	Does not make a claim	Makes an accurate/reasonable but incomplete claim (i.e., does not relate to healthy ecosystem concepts and/or underlying cause of change to ecosystem).	Makes an accurate/reasonable and complete claim (e.g., relates to healthy ecosystem concepts and/or addresses underlying cause of change to ecosystem).
<b>Evidence</b> – Scientific data that supports the claim. The data needs to be appropriate and sufficient to support the claim.	Does not provide any evidence or provides only inappropriate evidence (i.e., inaccurate)	Provides appropriate evidence (e.g., based on observations, known facts about the ecosystem such as food web relationships, or analysis of graphs), but is insufficient to support claim.	Provides appropriate and sufficient evidence (e.g., based on observations, known facts about the ecosystem such as food web relationships, or analysis of graphs to support claim.
<b>Reasoning</b> – A justification that links the claim and evidence. It shows why the data counts as evidence by using appropriate and sufficient scientific principles.	Does not provide reasoning or provides reasoning that does not link evidence to the claim.	Provides reasoning that links the claim and evidence, but does not include appropriate and sufficient scientific principles (e.g., healthy ecosystem concepts or underlying cause of change to the ecosystem).	Provides reasoning that links evidence to claim. Includes appropriate and sufficient scientific principles (e.g., healthy ecosystem concepts, referencing underlying cause of change to the ecosystem).

Using the rubric shown in Table 4, we evaluated each component of student groups’ explanations, giving it a score of 0, 1, or 2. There were a total of 19 claims made in Brad’s class and 18 claims made in Mark’s class. In Figure 2, we present the results of our CER analysis. Student groups in both classes achieved more level 2s with their reasoning than with their evidence. In Brad’s class, they achieved more level 2s with their evidence than in making claims; in Mark’s class, this result was reversed.

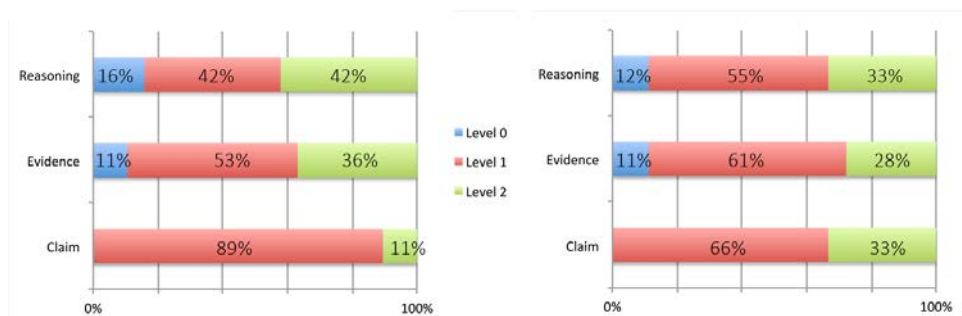


Figure 2. CER levels of student groups’ in Brad’s class (left) and Mark’s class (right).

Next we examined the percentage of instances when student groups made reference to one or more visual representations during each CER component. Student groups in both classes directly referenced visual representations most frequently during the *evidence* component of their explanation—approximately 80% of the time. However, student groups also referenced visual representations with high frequency during both the *claim* and *reasoning* components of their explanation, with some variation across classes (see Figure 3). To see if there was any possible connection between whether or not a student group utilized a visual representation and the level they scored in each CER component of their explanation, we calculated co-efficient values for these variables. In all instances, there were no positive correlations. Similarly, when we also looked for a connection between number of visual representations utilized and CER levels, we found no positive correlations.



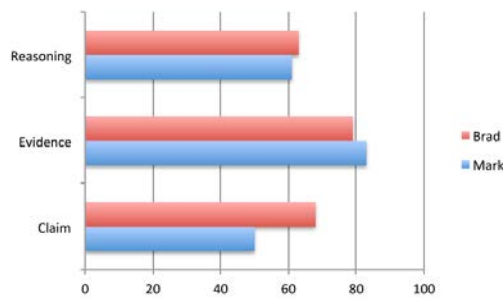


Figure 3. Percentage of time student groups made reference to a visual representation during CER components.

Finally, we examined the type of representation (e.g., investigation screens) that student groups referenced during each CER component (see Figure 4). While we were not able to show that there was any strong correlation between reference to a visual representation or a specific type of visual representation and higher CER level scores, our analysis shows that student groups successfully used the investigation screens to support all three CER components of their explanations. Students used a variety of other representations to support their explanations, sometimes using more than one type within a single CER component, and most often during the evidence component of their explanations.

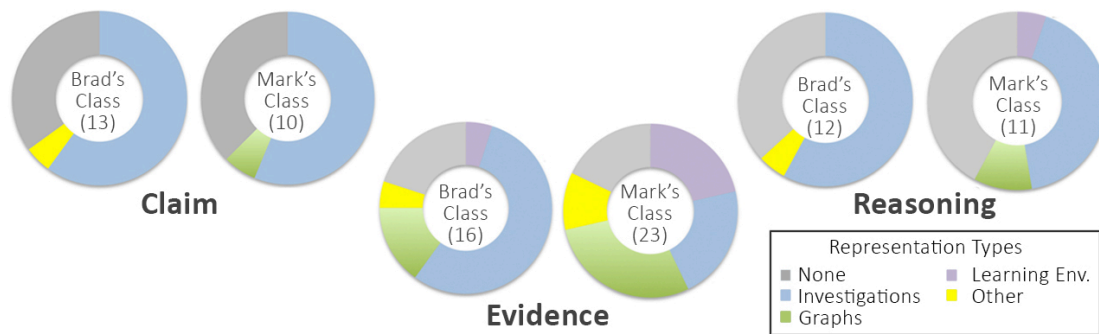


Figure 4. Breakdown of representation types referenced during each CER component for Brad and Mark's classes, with totals shown in brackets.

## Discussion

By mapping the WallCology curriculum designs to the three features outlined by McNeill and Krajcik (2011), we were able to portray a complex inquiry space that created space for students to engage in the scientific practice of explanation. An activity structure that allows students to persuade an audience that is unfamiliar with their data and work is not typical in science classrooms (Berland & Reiser, 2009). The WallCology curriculum design created opportunities for learners to construct complex scientific understandings through socially connected processes (Engle & Conant, 2002; Gray & Szalay, 2007). In the presentation task, student groups were only moderately familiar with the specifics of other ecosystem investigations, making them “unfamiliar readers” of each other’s work (Berland & Reiser, 2009). The implicit goal, therefore, of the student groups’ presentations was that of *persuasion*—to provide evidence (connected by reasoning) to support a claim that they had made about their ecosystem. By carefully evaluating another groups’ explanations, student groups could determine whether or not to incorporate new approaches and evidence into their own future investigations.

Greeno and Hall (1997) speak about “practicing representations” as part of social practice. It involves “learning to participate in the complex practices of communication and reasoning in which the representations are used” (p. 361). During the investigation presentations, student groups used various forms of visual representations as a communication tool. They were able to demonstrate their ability to understand and interpret traditional forms (e.g., graphs, food web relationship diagrams), but they also put these representations to use, and used them as a device to strengthen their explanations. Greeno and Hall (1997) distinguish between representations that are *of something* and representations that are *for something*. During the investigation presentations, student groups took representations that were *of something* (e.g., simulation screens that were

visible in the classroom) and retooled them into representations that were *for something*. In addition to “practicing representations” as a form of social practice, student groups used various representation types as a form of scientific practice. Although the investigation presentations were a form of culminating task, students groups presented to their peers while they were making new discoveries and refining their understanding. In this sense, students were engaging in work with representations in a manner that is similar to the way scientists use representations. Student groups coordinated a variety of representation types and used them with competence and fluency and to support their knowledge claims. Our study portrays a scenario in which visual representations played a vital role. They did not just influence or support scientific discourse; rather, they became woven into the fabric of the discourse.

### Limitations

During design discussions, the teachers on the co-design team indicated that they would like to include the language of “claim, evidence, and reasoning” in our prompts. The co-design team included scaffolds that asked students to stake a claim (by responding to a driving question) and to use the language of “evidence” and “claim.” However, the students themselves were not explicitly guided to use the framework during their presentations, nor did the teachers explicitly model the approach. Nevertheless, the classroom culture had long established knowledge building practices that provided learners with the necessary background to engage in the scientific discourse of explanation.

### **References**

- Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. *Science*, 333(6046), 1096-7.
- Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, 93(1), 26-55.
- Brown, A. L., & Campione, J. C. (1996). Psychological theory and design of innovative learning environments: On procedures, principles, and systems. In L. Schauble and R. Glaser (Eds.), *Innovations in learning: New environments for education*, (pp. 289-325). Mahwah, NJ: Lawrence Erlbaum Associates.
- Danish, J. A., & Enyedy, N. (2007). Negotiated representational mediators: How young children decide what to include in their science representations. *Science Education*, 91(1), 1-35.
- Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*, 20(4), 399-483.
- Gray, J., & Szalay, A. (2007). *eScience - A transformed scientific method* (PowerPoint slides). Retrieved from [http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB\\_eScience.ppt](http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt).
- McNeill, K. L., & Krajcik, J. S. (2011). *Supporting Grade 5-8 Students in Constructing Explanations in Science: The Claim, Evidence, and Reasoning Framework for Talk and Writing*. Boston, MA: Pearson.
- Moher, T., Uphoff, B., Bhatt, D., López Silva, B., and Malcolm, P. (2008). WallCology: designing interaction affordances for learner engagement in authentic science inquiry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. (pp. 163-172). ACM.
- Ontario Ministry of Education. (2007). *The Ontario Curriculum Grades 1-8: Science and Technology*. Retrieved from <http://www.edu.gov.on.ca/eng/curriculum/elementary/scientec18currb.pdf>
- Penuel, W.R., Roschelle, J., & Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. *Research and Practice in Technology Enhanced Learning*, 2(1), 51-74.
- Prain, V. & Tytler, R. (2013). Representing and learning in science. In R. Tytler, V. Prain, P. Hubber & B. Waldrup (Eds). *Constructing representations to learn in science*. Sense Publishers: Amsterdam, p 1-14.
- Studiocode Group. (n.d.). Retrieved October 29, 2016, from <http://www.studiocodegroup.com/>

### **Acknowledgments**

We heartily thank the teachers who partnered with us to design and implement the curriculum. We also thank their students, who enthusiastically engaged in this curriculum. The authors are grateful for the substantial support that was provided by many members of the Encore Lab, notably, Renato Carvalho, Armin Krauss, Colin McCann, and Meagan O’Hara. The material presented is based on work supported by the U.S. National Science Foundation under grants IIS-1065275 and DRL-1020027, as well as the Canadian Social Sciences and Humanities Research Council under grant 410-2011-0474.

# Through the (Thin-Slice) Looking Glass: An Initial Look at Rapport and Co-Construction Within Peer Collaboration

Jennifer K. Olsen, Human-Computer Interaction Institute, Carnegie Mellon University, jkolsen@cs.cmu.edu  
Samantha Finkelstein, Human-Computer Interaction Institute, Carnegie Mellon University, slfink@cs.cmu.edu

**Abstract:** Within peer collaboration, both cognitive and social phenomena have been identified as important components for success, though little is known about the relationship between these factors. In this work, we examined math collaboration discourse between 11 4<sup>th</sup> grade dyads in 30-second slices to investigate the relationship between *rapport state* and *reasoning state*. Prior to collaboration, students watched one of three instructional videos modeling either domain knowledge, collaborate reasoning, or both. There was no impact of video type on student talk behaviors, nor posttest scores. However, we found a correlation between high rapport states and strong reasoning states, as well as a marginal effect of more co-constructive reasoning leading to improved posttest scores. This work demonstrates that students' rapport states may play a role in students' reasoning states, and thus calls for a deeper investigation within the CSCL community about the role of rapport in peer learning.

## Introduction

Today's classrooms recognize the role of *talk* as not just a medium for conveying ideas, but also a process that creates *new* knowledge through the sharing of these ideas (Lee, Quinn, & Valdes, 2013). Peer collaboration can result in notable learning benefits for students (Lou, Abrami, & d'Apollonia, 2001), though prior work demonstrates certain conditions must be in place for these effects to be realized (Kollar, Fischer, & Hesse, 2006). Numerous conditions for success have been proposed, including both cognitive factors, such as transactively constructing new ideas with a partner (Chi & Wylie, 2014), and social factors, such as friendship status (Azmitia & Montgomery, 1993). Although both cognitive and social factors have been found to have an impact on learning, the problem of how these separate phenomena relate to each other has still been largely unaddressed. One notable exception is recent work demonstrating that friend-status impacts how peer tutors help their partners solve problems, with friend-tutors doing more question asking, while stranger-tutors do more knowledge-telling (Madaio, Ogan, & Cassell, 2016). By understanding how social and cognitive factors vary together, designers of learning experiences gain the ability to use one set of factors to leverage the other. Similarly, it is important to understand how each of these student factors is impacted by the design of the learning environment. In this paper, we investigate two primary questions: (1) what is the relationship between students' rapport states and reasoning states throughout a session, and (2) how does the way task activities are modeled to students impact either of these factors and their relationship? We analyzed the math collaboration discourse of 11 4<sup>th</sup> grade student dyads in 30-second slices over the course of one thirty-minute session. Dyads were presented with one of three instructional videos that either modeled domain knowledge, strong peer collaboration, or both so we could assess the impact on student behavior (Rummel, Spada, & Hauser, 2009).

From a cognitive perspective, collaborative learning is productive due to the reasoning that occurs through dialogue and provides students with deeper conceptual understandings of a domain (Webb, 2013). In the interactive-constructive-active-passive (ICAP) framework, Chi and Wylie (2014) argue that some of the success of peer collaboration is due to the type of cognitive engagement that the students voluntarily express together during the interaction. In ICAP, there are four levels of cognitive engagement, ranging from passive engagement (e.g., back-channel, or agreeing without providing new information) to interactive engagement (e.g., building off of the partner's ideas.) The more time a dyad spends engaging in talk that is closer to interactive co-construction rather than passive engagement, the more students will learn. We use this framework, discussed in more detail in the methods section below, as the inspiration for the *reasoning states* scheme we used to annotate students' on-task talk.

There are additionally a myriad of social factors that have been posited to impact student performance. For example, students' perception of how much they *belong* in a learning environment may impact the level to which they participate, and how *well* they participate. For example, students are far less likely to engage in school-ratified science reasoning when they feel that the expected behaviors within the learning environment are counter to their personal identities (Brown, 2006). On the other hand, there is evidence that by helping students feel like they are part of a learning community, we may be able to positively impact both their identity as a learner and their willingness to learn domain information (Gee, 2000). This connection between personal

identity and the learning environment may be one explanation as to why student learners perform better when working with friends rather than acquaintances (Azmitia & Montgomery, 1993). Similarly, the rapport between students as they collaborate can have an influence on student learning. As students build a relationship, the social behaviors that are engaged in during the learning process may need to gradually change (Ogan et al., 2012). Although both cognitive and social factors have been found to influence group learning, there is not much work into the relationship between these factors, nor how we may be able to leverage students' social states in the design of our technologies so that they could be used to improve student reasoning.

In collaboration settings, students do not spontaneously produce the sorts of talk associated with success without appropriate support (Fischer et al., 2006). Collaboration scripts, in which students are given specific instructions for what to say and what to work on, have been shown to successfully facilitate the collaborative process (Fischer et al., 2013). However, if the script is too restrictive for the student's current level of knowledge, it can lead to *over-scripting* (Dillenbourg, 2002). Additionally, scripting may be difficult to implement on a large-scale since it may need to be adapted to each new domain. To address these issues, researchers have examined the potential benefits of *modeling* collaboration (Rummel, Spada, & Hauser, 2009). Modeling allows students to observe the behavior of others successfully completing a task and then integrates these same behaviors into their own interactions through vicarious learning (Decker & Nathan, 1985). When compared against scripting, modeling has been shown to demonstrate higher positive learning gains for students (Rummel, Spada, & Hauser, 2009). Furthermore, vicarious learning through modeling has been associated with social effects – specifically with students feeling like they belong to a learning community (Stenning, 1999). Finally, while *epistemic* and *social scripts* have been compared for their impact on student learning behaviors (Weinberger et al., 2005), such studies have not been done with modeling.

This paper builds upon prior work to explore how *modeling type* impacts students' rapport and reasoning states, and how these behaviors impact subsequent student learning. We break down this goal into five hypotheses we aimed to address in our study presented below: (1) When students have more instances of high reasoning states co-occurring with high rapport states, they will demonstrate higher posttest learning gains, (2) Students will be less likely to demonstrate low reasoning states during high rapport instances, (3) Students will be more likely to demonstrate high reasoning states during higher rapport states, (4) Students who see one of the two models that include collaboration will more high rapport and high reasoning states than students who see a domain-only model, and (5) students who see models that include strong collaboration will demonstrate more learning gains than students who see a domain-only model.

## Methods

### Instructional videos and problem-solving environment

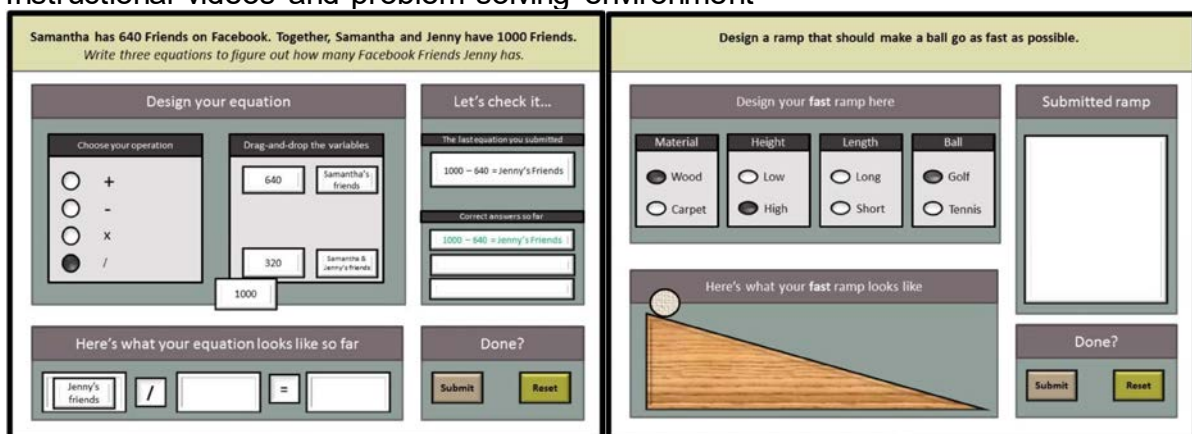


Figure 1. Before doing collaborative problem-solving, the students watched short videos of peers modeling the collaboration/problem-solving process in the domain (left) or out of the domain (right).

To model the collaboration for the students, we designed three different 4-minute vicarious learning videos for collaboration within the domain (Mixed), collaboration outside of the domain (Collaboration), and individual work within the domain (Domain). Each video was an animated version of a computer screen with students taking actions on the interface to solve the problem (see Figure 1) and audio of student(s) talking through the problem solving. For the videos that included modeling of the domain problem-solving skills (i.e., Domain and Mixed), students saw an interface that resembled the interface that was used in the intervention for the word

problems. For the video that did not model domain skills (i.e., Collaboration), the students saw an interface for solving a ramp physics problem. Within the videos, students demonstrated productive talk around the domain material by demonstrating good problem-solving steps either through a student thinking out loud (Domain) or dialogue with a partner (Collaboration and Mixed). For the collaboration, students demonstrated four types of talk adapted from the academically-productive talk framework (Michaels, O'Connor, & Resnick, 2008), including explicit reasoning (e.g., “we could try x, because y”), eliciting partner reasoning (e.g., “wait, why?”), transactivity (building off of partner ideas), and unity (referring to the collaboration with ‘we’ words). For the domain-related videos, we kept the scripts as identical as possible. For example, instead of cognitive conflict and disagreement in the Mixed video, the Domain video showed the student self-correcting a mistake.

Malik played 5 different games yesterday. Today, Malik played 9 different games. How many total games has Malik played?  
Please write three different equations that represent this problem.

Pick an operation for the equation. <input type="radio"/> + <input type="radio"/> - <input checked="" type="radio"/> x <input type="radio"/> /	Fill variables into the equation. Games Yesterday <input type="text" value="9"/>	Let's check that equation... $5 \times \text{Games Today} = \text{Total Games}$
Design your equation here. <input type="text" value="5"/> x <input type="text" value="Games Today"/> = <input type="text" value="Total Games"/>	Correct equations. $5 + 9 = \text{Total Games}$ $5 \times \text{Games Today} = \text{Total Games}$	

Reset Check

Done

Figure 2. An example of the translation task for the word problem. On the bottom left of the screen the students can construct equations based on the word problem at the top. On the bottom right of the screen, the students are provided with feedback on their solution and can see their correct solutions.

After watching the video of modeled collaboration/domain skills as a pair, the students worked on a word translation problem set with their partner. We focused on this domain as it would be age-appropriate for 4<sup>th</sup> graders and offers multiple solution paths. To identify problems that were appropriately difficult, we piloted our interface with three 4<sup>th</sup> grade student dyads and iterated wording to promote comprehension as necessary. The problem set contained a total of nine problems, all of which were focused on translating word problems into one-step equations. Five word problems contained addition/subtraction operations, and four problems contained multiplication/division operations. Each word problem contained one unknown value, half with the total being the unknown value. For each problem, the students were asked to write three different equations that represented the word problem out of 16 possibilities. The system would accept any combination of given variable names (three options) and values (two options) with the correct operation that represented the problem; the equation did not have to be in an order that would allow the student to solve the problem. To prevent typing errors in the system, the students were provided with a drag-and-drop interface (see Figure 2) for the values and radio buttons to select the operators. The students received correctness feedback on the entire equation after pressing the ‘Check’ button. The open-ended nature of this domain allowed the students to have discussions around the solutions rather than just looking for the one correct answer (Webb 2013).

### Experimental design and procedure

The study was conducted in a classroom setting with 22 4<sup>th</sup> grade students using a pull-out design. The students participated in the experiment during a free period in their regular school day for three days. Students completed pretest and posttest individually during 15-minute blocks on days 1 and 3, and on day 2, students worked with a partner on a math computer program that we designed for the experiment. We asked our partner teacher to pair students based on their perception of the students’ friendship (i.e., could work together) and ability levels (i.e., do not have drastically different ability levels).

Before beginning the problem solving, students watched one of the three vicarious learning videos described above. Students were told that this was an example video “completed by students just like you” to “give them an example of what using this program might be like.” Student dyads were randomly assigned to one

of the three conditions: Domain, Collaborative, or Mixed (see Table 1). The students watched the videos in pairs as collaboratively watching a worked example video may result in students learning as much as one-on-one tutoring (Chi, Roy, & Hausmann, 2008). During the study, partners shared a single laptop but both the trackpad and an external mouse were provided allowing students to negotiate control of the interface. During the intervention, the students were video and audio recorded.

Table 1. Descriptions of the three different vicarious learning models used in the study

Domain	Collaboration	Mixed
One student performing a think-aloud around the math domain content targeted within this task. This student clearly explains the domain content through worked examples.	Two students demonstrating strong collaboration behaviors on a science task that is unrelated to the math domain. The students demonstrate interactive engagement through co-construction of ideas.	Two students demonstrating strong collaboration behaviors on the math domain content targeted within the task. This model shows both relevant math worked examples, as well as co-construction.

### Hypotheses and dependent measures

We use this study design to address five primary hypotheses regarding the relationship between students' rapport states, reasoning states, and subsequent learning gains. We collected pretest and posttest measures on paper, and recorded students' dialogue as they worked through the intervention.

For analysis, files of each dyad dialogue were divided into 30-second segments for thin-sliced coding (Murphy, 2005). We chose thin-slice methodologies so we could rate students' talk holistically within a segment, rather than more traditional measures of annotation that mark individual utterances. Each segment was rated for two different types of behaviors that previous research has identified as being a potential mechanism explaining the relationship between collaboration and performance: reasoning state (Chi & Maneske, 2015) and rapport (Tickle-Degnen & Rosenthal, 1990). These annotations allowed us to address questions about how our vicarious learning conditions impacted dyad talk along these dimensions, as well as how students' talk behaviors impacted their learning results.

For reasoning state, a rating scale was developed based upon the ICAP framework (Chi & Whyllie, 2014) with categories from zero to four (see Table 2). The rating scale represents an ordinal categorization of five different reasoning states that captures the four different *levels* of the ICAP framework (i.e., passive, active, constructive, interactive) and follows the same hierarchy that interactive talk is more effective than constructive talk, which is more effective than active talk, which is more effective than passive (Chi & Maneske, 2015). For our rating scale, an inter-rater reliability analysis using the Kappa statistic was performed to determine consistency among raters within the 30-second thin-slice audio clips (Kappa = 0.89).

Table 2: Reasoning states aligning with overt cognitive actions seen in the student's dialogue and ICAP

Rating	Overt Actions	ICAP Framework
0	Neither student is talking about the problem or the solution (or no talk is present at all). Talk around work coordination, or is off-task	Passive-Passive
1	Only one student is talking and either repeats information already provided or suggests a new solution	Passive-Active, Passive-Constructive
2	Both students are talking but are only repeating information already provided	Active-Active
3	Both students are talking and at least one new solution is provided, however, the solutions are not related	Active-Constructive, Constructive-Constructive
4	Giving and receiving of questions/answers, referencing a partner's previous solution in a new suggestion, co-construction of ideas	Interactive

The rapport rating scale used was adapted from a rapport thin-slice coding scale developed by Sinha & Cassell (2015). It captures aspects of mutual attention, coordination, and positivity within the peer collaboration context (Tickle-Degnen & Rosenthal, 1990). The rapport rating scale consists of five rating categories from zero to four (see Table 3). Similar to the reasoning state rating scale, the categorizations are ordinal. An inter-rater reliability analysis was performed to determine consistency among raters (Kappa = 0.69).

Although both rating scales are measuring types of talk that may play a role in collaboration, they are analyzing different aspects of the student behavior and are not dependent on one another. The reasoning state rating scale takes into consideration how the students present solutions to the problems and build upon each

other's prior work while the rapport rating scale takes into consideration the social aspects of the student interactions and how well the students are engaging with their partner, regardless of the content. To ensure that these rating scales were independent in practice, we looked at cell distributions after annotation was complete to ensure that all cell combinations were present, which they were. In addition, to reduce the influence that one rating scale had on the category assignment from the other rating scale, after the inter-rater reliabilities were established, different researchers coded the reasoning state scale and the rapport rating scale.

Table 3: Rapport rating scale aligning with overt social actions seen in the student's dialogue

Rating	Overt Actions
0	Silence, or only talk to the experimenter or other students in the classroom, but not to each other. (These segments were considered categorically different than the rest of the ordinal scale, and not very low rapport. They were removed for some analyses presented below.)
1	Low rapport. This code is marked by lack of synchrony between students, low positivity, or lack of attention (e.g., "ugh, can you stop it? I'm working on this. I said stop.")
2	Neutral rapport. This code marked segments where students were demonstrating the 'bare minimum' for dyadic interaction to be successful (e.g., "Should we add here?" "yep." "okay.")
3	Positive rapport. Students' dialogue flowed smoothly, and students demonstrated some active interest in their partners' contributions (e.g., "Ooh, should we add here?" "okay yeah I thought so too!" "okay")
4	Very high rapport. These segments were marked by student social behaviors that appeared to proxy high levels of coordination. On the surface, these interactions may have been marked by strong levels of positivity (e.g., "we got it!" "yes!" "we're soooo good at this!") or, conversely, positively-received teasing (e.g., "ooooo you think you're going to get this one now? No way." [partner laughs]).

Finally, we collected students' pretest and posttest data with a paper assessment using two equivalent, counterbalanced test forms. The tests contained a total of eight problems. Four of the problems were isomorphic to the intervention and four of the problems were transfer problems. For two of the transfer problems, the students were asked to solve a word problem (not just write an equation) and received one point for the correct answer. For the other two transfer problems and the isomorphic problems, the students were asked to write three different equations that represented the given word problem. An equation was counted as correct if it used either the numbers or variables presented in the problem with the correct operators to make an accurate translation of the problem. The isomorphic problems were one-operation equations while the transfer problems were two-operation equations. For each correct equation (no credit for solving), the students got one point for a possible three points for each word problem. On the tests there were 20 possible points for the 8 questions.

## Results

Out of the 22 students in the study, 20 students were included in the analysis because of technical errors. There were four pairs assigned to the Mixed condition, three pairs assigned to the Collaborative condition, and three pairs assigned to the Domain condition. To check the distribution of knowledge across conditions, we compared student pretest scores and found no significant difference between conditions,  $F(17,2) = 1.42, p = .27$ .

### What is the impact of vicarious learning model on students' talk behaviors?

We hypothesized that students who heard a model of strong collaborative talk (Mixed or Collaboration) would subsequently demonstrate higher rapport and more instances of strong collaborative talk than those who were shown a video that exclusively modeled strong domain reasoning. To investigate the impact that condition had on student dialogue during the collaboration, we conducted a MANOVA analysis. Using the Pillai's trace, there was not a significant effect of condition on students' reasoning states nor rapport states,  $V = 0.23, F(4,14) = 0.46, p = .76$ . This analysis demonstrated that there was no impact of vicarious learning model on students' rapport or collaborative talk as measured by our thin-slice annotations.

### What is the impact of rapport-level on the likelihood of strong cognitive talk?

To test our hypotheses that (a) low rapport states would be *less likely* to have higher reasoning states, and (b) high levels of rapport will be *more likely* to have higher reasoning states, we conducted a Person's Chi-squared. Because there were not very many instances of category 2 for the cognitive rating scale (less than five), we combined categories 2 and 3 for a combined category that reflected when both students were talking but were not being interactive. In addition, we removed all segments with a '0' in the rapport scale as this marked utterances where students were not talking. This left us with four categories for both the cognitive rating scale (i.e., 0,1,3,4) and the rapport rating scale (i.e., 1,2,3,4) and a total of 540 segments.

We identified a significant association between the cognitive rating category of a talk segment and the rapport rating category of a talk segment,  $\chi^2(9) = 41.11, p < .05$  (see Figure 3). Based on the standardized residuals of the different cells, we found that there were significantly more talk segments that had a 0 category for cognitive talk and a 1 category for rapport talk than would be expected ( $p < .05$ ). In contrast, we found that there were significantly fewer talk segments that had a 4 category for cognitive talk and a 1 category for rapport talk than would be expected ( $p < .05$ ). These results indicate that when students are having interactions marked by notably low rapport (such as lack of coordination or notable negativity), they are less likely to be co-generating interactive math reasoning. In addition, there was a marginally significant effect for more talk segments that had interactive talk (reasoning state rating = 4) and moderately high rapport (rapport state = 3) than would be expected ( $p = .09$ ). These results indicate that while very high rapport doesn't co-occur with strong cognitive talk, better-than-average coordination indicated by a '3' on this scale may facilitate strong collaborative reasoning.

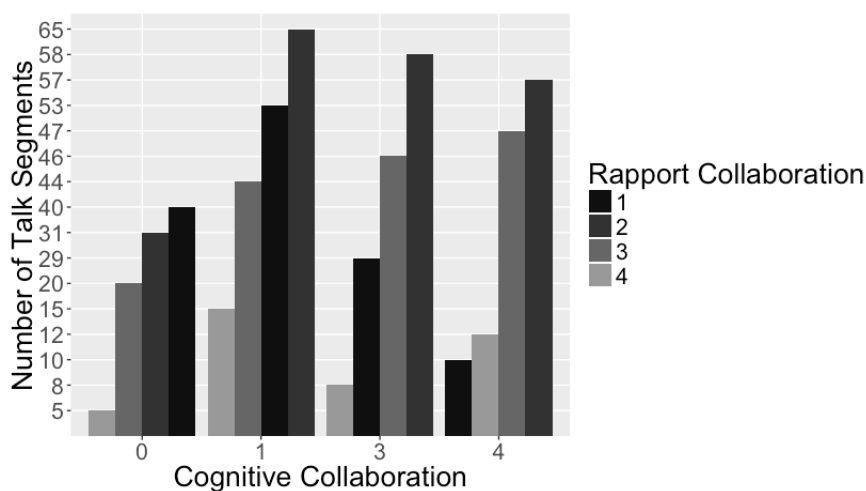


Figure 3. The distribution of reasoning state categories and rapport categories across talk segments.

### What is the relationship between student talk annotations and vicarious learning model on students' learning gains?

To investigate the impact that the modeling conditions and annotated student talk variables had on learning gains, we used a multilevel approach to take into account the repeated measures of the pretest and posttest and differences between dyads. For this analysis, we were interested in a particular combination of annotated talk moves – specifically the co-occurrence of interactive cognitive talk (rating 4) with high rapport (segments rated as a 3 or 4). We hypothesized that more instances of this combination would lead to greater learning gains. To test this, we included the percentage of talk instances for each dyad where the students had a cognitive score of 4 and a rapport score of at least 3 into the analysis. This percentage ranged from 0 to 0.41 across pairs. We used a hierarchical linear model (HLM) with student at the first level and dyad at the second level. At level 1, we modeled the pretest and posttest scores; at level 2, we accounted for individual differences, condition, and the pair's percentage of high rapport/collaborative talk. Within our model, we chose pretest and the Collaborative modeling condition as the baselines. For each variable, the model includes a term for each comparison between the baseline and other factors of the variable. We measured the effect size with Pearson's correlation coefficient ( $r$ ) where 0.1 is a small effect size, 0.3 is a medium effect size, and 0.5 is a large effect size.

Table 4: Means and (standard deviations) for the pretest and posttest scores across conditions

Condition	Pretest Score	Posttest Score
Collaborative Modeling Only	4.67 (3.78)	8.17 (4.11)
Domain Modeling Only	2.17 (2.14)	4.17 (2.48)
Collaborative and Domain Modeling	2.50 (2.45)	3.88 (2.95)

The results of the learning gains by condition are displayed in Table 4 and pretest and posttest analysis are shown in Figure 3. There was a significant difference between pretest and posttest scores,  $t(17) = 2.76, p < .05, r = 0.56$ , with the posttest scores being higher across all conditions. For the condition differences, there was



a marginally significant difference between Collaborative and Mixed models,  $t(6) = -2.12, p = .07, r = 0.65$ , with the Collaborative condition having the higher test scores and a non-significant difference between Domain and Collaborative models,  $t(6) = -1.73, p = .13$ . There was not a significant interaction between pretest/posttest and Collaborative/Mixed conditions,  $t(17) = -1.27, p = .22$ , and pretest/posttest and Domain/Collaborative conditions,  $t(17) = -0.84, p = .41$ . Regarding the impact of the percentage of high rapport/collaborative talk, there was a marginally significant main effect,  $t(6) = 2.27, p = .06, r = 0.68$ , with dyads with more high rapport/collaborative talk co-occurrence segments having higher posttest scores.

## Discussion and conclusion

In this work, we addressed two primary questions: (1) what is the relationship between students' rapport state and their corresponding reasoning state, and (2) how do vicarious learning models impact students' overall rapport and reasoning behaviors? By looking at how students' rapport states impact the likelihood of interactive reasoning, we gain a better understanding of how rapport plays a role in peer collaboration. Once this relationship is better understood, we may be able to design collaborative learning systems that productively leverage students' relationships during the learning process.

Our first result demonstrates that students' rapport state does in fact impact the likelihood of interactive reasoning. Low rapport (rating = 1) time slices are significantly less likely than expected to co-occur with interactive reasoning, and significantly more likely than expected to co-occur with talk the lowest reasoning states. In this work, time slices were deemed to have very low rapport (rating = 1) when an interaction seemed emotionally tense – for example, students did not seem aligned in their task goals, were attending in different directions, or were criticizing their partners' performance without positive response from the targeted student. We also found marginal significant results indicating that high (but not extremely high) rapport states (rating = 3) may be more likely than expected to co-occur with interactive reasoning, and that interactive reasoning was more likely to occur when rapport was *high* than when it was *neutral*. The results of this work demonstrate that students' relationship dynamic *at a given moment* is associated with the likelihood that high-quality co-constructive reasoning is also being produced. In other words, it may not be that *friends learn better than strangers*, which some prior work suggests (Azmitia & Montgomery, 1993), but that *friends may have fewer instances of low rapport than strangers*, and that this improved alignment results in more instances of idea co-construction when collaborating.

We additionally found that students demonstrated increased learning gains after receiving the vicarious learning model that focused on collaboration in a domain-separate task, rather than either a model that focused on domain-relevant information, or even collaboration within this domain-relevant task. This result is similar to the result by Weinberger et al. (2005), who found that students demonstrated greater learning gains after following a social script, rather than a domain script. Interestingly, the type of vicarious learning model did not impact students' average rapport-states over the course of the session, nor the overall number of time slices where interactive reasoning occurred. This indicates that while learning model type may have an impact on students' posttest scores, it did not have an impact on how the students spoke to each other while completing the task (as captured by the features that were annotated for in the data.) While we caveat these results due to both our small sample size and the marginal nature of the finding, we report it due to its large effect size ( $r = .6$ ), and encourage other members of the research community to investigate the impacts of collaborative over domain modeling within their own work.

These analyses serve as a first investigation into the relationship between students' rapport states and the corresponding reasoning states by looking at co-occurrence. It demonstrates that *social relationship in a given moment* plays a role in the likelihood of co-constructive idea generation, and that ignoring rapport between dyad members would be ignoring part of the equation. Moving forward, we will investigate more complex patterns in students' rapport states and how they may predict *future* reasoning states. We will also aim to investigate how different *patterns of reasoning* over the course of a session may predict learning gains in ways that may be more sophisticated than the "more is better" hypothesis that was tested in this work. Like learning itself, human relationships are complex, and uncovering the patterns of rapport that are predictive of success will give the CSCL and education communities a greater understanding of how to best leverage this often ignored phenomenon in students' learning experiences.

## References

- Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social development*, 2(3), 202-221.
- Brown, Bryan A. "It isn't no slang that can be said about this stuff": Language, identity, and appropriating science discourse." *Journal of Research in Science Teaching* 43.1 (2006): 96-126.

- Chi, M. T. H., & Menekse, M. (2015). Dialogue patterns that promote learning. In L. B. Resnick, C. Asterhan, & S. N. Clarke (Eds.), *Socializing Intelligence Through Academic Talk and Dialogue* (Chi. 21, pp. 263-274). Washington, DC: AERA.
- Chi, M. T., Roy, M., & Hausmann, R. G. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive science*, 32(2), 301-341.
- Chi, M. T. & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243.
- Decker, P. J., & Nathan, B. R. (1985). *Behaviour Modeling Training: Principles and Applications*. Praeger Publishers.
- Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. *Three worlds of CSCL. Can we support CSCL?* 61-91.
- Fischer, F., Kollar, I., Stegmann, K., Wecker, C., Zottman, J., & Weinberger, A. (2013). Collaboration Scripts in Computer Supported Collaborative Learning, in Hmelo-Silver, C., Chinn, C., Chan, C., & O'Donnell, A. (Eds) *The International Handbook of Collaborative Learning*, Routledge.
- Gee, J. P. (2000). Identity as an analytic lens for research in education. *Review of research in education*, 25, 99-125.
- Joshi, M., & Rosé, C. P. (2007). Using transactivity in conversation for summarization of educational dialogue. In *SLaTE* (pp. 53-56).
- Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts—a conceptual analysis. *Educational Psychology Review*, 18(2), 159-185.
- Lee, O., Quinn, H., & Valdés, G. (2013). Science and language for English language learners in relation to Next Generation Science Standards and with implications for Common Core State Standards for English language arts and mathematics. *Educational Researcher*, 0013189X13480524.
- Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, 71(3), 449-521.
- Madaio, M. A., Ogan, A., & Cassell, J. (2016, June). The Effect of Friendship and Tutoring Roles on Reciprocal Peer Tutoring Strategies. In *International Conference on Intelligent Tutoring Systems* (pp. 423-429). Springer International Publishing.
- Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in philosophy and education*, 27(4), 283-297.
- Murphy, N. A. (2005). Using thin slices for behavioral coding. *Journal of Nonverbal Behavior*, 29(4), 235-246.
- Ogan, A., Finkelstein, S., Walker, E., Carlson, R., & Cassell, J. (2012, June). Rudeness and rapport: Insults and learning gains in peer tutoring. In *International Conference on Intelligent Tutoring Systems* (pp. 11-21). Springer Berlin Heidelberg.
- Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning*, 4(1), 69-92.
- Sinha, T., & Cassell, J. (2015). We click, we align, we learn: Impact of influence and convergence processes on student learning and rapport building. In *Proceedings of the 1st Workshop on Modeling INTERPERSONAL Synchrony And Influence* (pp. 13-20). ACM.
- Stenning, K., McKendree, J., Lee, J., Cox, R., Dineen, F., & Mayes, T. (1999, December). Vicarious learning from educational dialogue. In *Proceedings of the 1999 conference on Computer support for collaborative learning* (p. 43). International Society of the Learning Sciences.
- Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological inquiry*, 1(4), 285-293.
- Webb, N. (2013). Information processing approaches to collaborative learning. In Hmelo-Silver, C., Chinn, C., Chan, C., & O'Donnell, A. (Eds) *The International Handbook of Collaborative Learning*, Routledge (pp 19-40).
- Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science*, 33(1), 1-30.

## Acknowledgments

We thank Nikol Rummel and Vincent Alevan for their guidance. This work was supported by Graduate Training Grant #R305B090023 from the US Department of Education (IES).

# Finding Collaboration Partners in a Scientific Community: The Role of Cognitive Group Awareness, Career Level, and Disciplinary Background

Julia Eberle, Ruhr-Universität Bochum, [julia.eberle@rub.de](mailto:julia.eberle@rub.de)

Karsten Stegmann, Ludwig-Maximilians University Munich, [karsten.stegmann@psy.lmu.de](mailto:karsten.stegmann@psy.lmu.de)

Frank Fischer, Ludwig-Maximilians-University Munich, [frank.fischer@psy.lmu.de](mailto:frank.fischer@psy.lmu.de)

Alain Barrat, Aix Marseille Univ, Université de Toulon, CNRS, CPT, [alain.barrat@cpt.univ-mrs.fr](mailto:alain.barrat@cpt.univ-mrs.fr)

Kristine Lund, ICAR, CNRS, Ecole Normale Supérieure Lyon, Université de Lyon, [kristine.lund@ens-lyon.fr](mailto:kristine.lund@ens-lyon.fr)

**Abstract:** Integrating newcomers and fostering collaboration between researchers with different disciplinary backgrounds is a challenge for scientific communities. Prior research suggests that both network-driven selection patterns (reciprocity and transitivity) and the active selection of specific others are important. Selecting appropriate collaboration partners may moreover require what we call cognitive group awareness, (i.e. knowledge about the knowledge of others). In a field study at two multi-disciplinary scientific events (Alpine Rendez-Vous 2011 and 2013) including N=287 researchers, we investigated selection patterns, looking specifically at career level and disciplinary background, and included a cognitive group awareness intervention. While we could not completely explain how researchers choose with whom they interact, we found that transitivity and interaction duration are relevant for later collaboration. Cognitive group awareness support was beneficial for fostering interdisciplinary collaboration. Career level was a less relevant factor. We discuss measures for supporting newcomer integration and community buildings based on our findings.

## Collaboration and integration of newcomers in scientific communities

Communities must constantly integrate new members to stay active and to further develop and cope with changing demands. Scientific communities can be considered to be a special form of communities of practice (Kienle & Wessner, 2006). According to Kienle and Wessner (2005), scientific communities consist of heterogeneous, often interdisciplinary groups of members who are usually geographically distributed. Members of scientific communities also sometimes have backgrounds in different disciplines and scientific cultures, resulting in the use of different methods and theories. What brings them together is a joint field of research interests. Therefore, scientific communities benefit from the integration of new members and their knowledge and ideas. In interdisciplinary scientific communities in particular, successful collaboration among the community members is another important factor for community cohesion and development. Successful interdisciplinary collaboration in a scientific community requires an integration of the contributing disciplines on some level, for example the mutual integration of concepts, theories, methodologies, and epistemological principles (van den Besselaar & Heimeriks, 2001). The development of mutual understanding and the building of shared representations are important for fruitful communication between experts of multiple domains (Fischer, 2000).

Yet many scientific communities struggle in integrating newcomers and developing interdisciplinary collaborations (see for example Kienle & Wessner, 2006, about challenges of the CSCL community) and it is unclear how to improve the situation. The huge body of research on scientific communities is so far mostly based on bibliometric analyses focusing on co-authorship or citation analysis of conference proceedings or journal papers (e.g. Lee, Ye, & Recker, 2012). The publication of articles is affected by many factors and bibliometric analyses so far were not able to identify factors that influence the integration of newcomers. Rather than looking at the results of successful collaborations (i.e. joint publications or references to each other's work), it seems necessary to study the onset of collaborations where newcomers have a chance to be included. Face-to-face meetings in conferences and workshops may provide regular opportunities for newcomers to start collaborations and thereby to become integrated into the scientific community (Kienle & Wessner, 2005).

## Potential selection patterns and influential factors on selection processes in scientific communities

To understand the behavior of researchers at scientific events, we should take into account that human interaction in face-to-face settings is based on several robust characteristics (e.g. physical and digital proximity,

social support, and community belonging) although there are differences based on the context in which interaction takes place (Isella et al., 2011). Research on interactional linguistics clearly illustrates how human interaction is co-constructed by phenomena such as gaze, prosody of language, gestures and general body movement (e.g. Goodwin, 2000) and this co-construction is viewed by many as a collaboration at the micro-level of communication; it is where any long-term collaboration begins.

Sociological and social network research suggests, furthermore, that the selection of partners for long-term relationships follows several patterns. Selection patterns have been studied in various contexts and the literature provides a variety of observed patterns. Some selection patterns are based on the opportunities provided by the given social network structure surrounding the selecting person, while others focus on personal preferences for specific persons. Reciprocity and transitivity as network-driven patterns and homophily as a person-driven pattern can be described as the three most important selection patterns (Baerveldt, van de Bunt, & de Federico de la Rúa, 2010). We will briefly describe them and discuss how they may be relevant for selection processes in scientific communities. Note that network-driven patterns and homophily effects are not mutually exclusive but can co-exist (Aiello, Barrat, Cattuto, Ruffo, & Schifanella).

### Reciprocity and transitivity

Network-driven selection patterns assume that the way in which persons choose relations is influenced by the (local) structure of their social network, and in particular that the alters chosen to establish a relation are preferentially chosen among those easily accessible in this network. Reciprocity is one of the simplest such selection patterns: If there is already a connection between two individuals, in which person A has chosen person B, for example as a friend or as someone to ask for advice, while person B does not yet perceive person A likewise, it is very likely that they will balance their relationship in the future. This can either mean that person B also chooses person A or that their relation will dissolve so that reciprocity is reached.

Transitivity is also a selection pattern driven by proximity in the network: Two individuals (A and B) who both are connected to a third person (C) are likely to build a relation as well. In social network terms, they are building a “transitive triple” or “closing a triangle”. Reasons for such transitivity patterns are manifold, e.g., person C can easily introduce A and B or A and B may share a common interest or activity which was the initial reason for their connection with C. C might also be interested in setting up a connection between A and B to stabilize the relation to both (Baerveldt et al., 2010).

### Selecting specific others

Social capital theory is one of the major approaches that explains person-driven selection patterns (Baerveldt et al., 2010). According to social capital theory (Coleman, 1988; Lin, 2001), people rationally select to engage with others either to maintain their own resources or to get access to resources of others. In the context of a scientific community, a vital resource accessible through social capital is, for instance, information about new developments in the field (Coleman, 1988).

The selection patterns for specific others can be categorized as homophilous or heterophilous. Homophily is a well observed, and probably the oldest and widest studied selection pattern (Baerveldt et al., 2010). Homophily is the tendency of people to be in contact with others who are similar to them (McPherson, Smith-Lovin, & Cook, 2001). Similarity/dissimilarity patterns influencing the selection of contact partners have been studied in regard to very different aspects, such as gender, religion, age, education, occupation and social class, behavior, attitudes, beliefs, abilities and aspirations (McPherson et al., 2001).

According to social capital theory, the selection of specific others is based on the maintenance or gain of resources. Maintaining resources is assumed to be the more dominant motive and leads to expressive actions towards others, meaning that people approach others to claim recognition for their resources or aim at receiving sentiments related to the maintenance of these resources (Lin, 2001). In a scientific community, such actions could include statements to gain recognition for one’s own expertise on a topic or sharing feelings about the complicated nature of a certain type of data collection. Such expressive actions require the least effort and bear the lowest risks among peers with similar resources and status, explaining why homophilous selections of interaction partners are most common (Lin, 2001). Researchers from other theoretical perspectives have also provided explanations for the generally observed tendency for homophilous selection patterns, e.g. from a reinforcement perspective, similar others may be more likely to reinforce behavior they show themselves and persons who reward us are preferred (Byrne & Clore, 1970).

### Information about others

Perceiving others as similar or dissimilar is likely to be influenced by which information about them is accessible. Some information is usually easily visible, such as gender or ethnicity, while other information is

invisible and harder to access (Baerveldt et al., 2010). Invisible information, such as expertise and knowledge of a person, are especially relevant for collaboration and collaborative learning (Cannon-Bowers & Salas, 2001; Wegner, 1987) as they can, for example, simplify grounding processes.

To select appropriate collaboration partners and interact meaningfully with them, knowledge about their knowledge seems to be necessary, i.e. it may require cognitive group awareness (Janssen & Bodemer, 2013). To select collaboration partners in a scientific community, relevant information about other researchers may include their area of expertise, research interest, and knowledge (their personal resources), as well as their professional network and access to other experts (their social resources) (Lin, 2001).

Newcomers usually have only little knowledge about a new community and need to acquire it to become able to contribute more and in a proper way (Levine & Moreland, 2013). Compared to mono-disciplinary communities, this might be even more complicated in multi- and interdisciplinary communities because of the variety of research lines. Although knowledge about researchers seems easy to acquire as most of them present their bios and publications on their websites, it is probably hard for a newcomer to identify the 'important' people in a community or those who could be relevant for their own research in face-to-face settings. In friendship networks, students with little information about their peers are assumed to be less active in initiating new friendships. They seem also more likely to use rather passive selection strategies, such as transitivity (forming new friendships with friends of their friends), instead of initiating new friendships with peers who might be a good fit to them regarding for example, norms and values (Baerveldt et al., 2010). Consequently, newcomers in scientific communities may be disadvantaged in finding new collaboration partners as they may not only use less active strategies for initiating new collaborations with other researchers but they can also benefit less from network-based selection patterns as they are not linked to many other researchers.

Newcomers in scientific communities can be found at all career levels because researchers tend to be involved in several scientific communities at the same time and, therefore, are used to switching roles, often from expert in one scientific community to newcomer in another scientific community (Kienle & Wessner, 2005). However, PhD students are the most common form of newcomers in scientific communities and they can be expected to suffer from the most disadvantages in finding collaboration partners. In contrast to more experienced researchers, PhD students do not only lack group awareness in the new scientific community but also lack knowledge on scientific collaboration in general. They are moreover often seen as having less expertise and being less available to collaborations not involving their supervisor and, therefore, as less attractive collaboration partners for other researchers. Therefore, it seems promising to support PhD students at scientific events. Results from group awareness research in smaller groups suggest that enhancing group awareness may also be a helpful means of support in scientific communities.

## Research question

Summing up the previous line of argumentation, it is an open question if certain types of members are more advantaged or disadvantaged in finding interaction and collaboration partners and to what extent the interaction and the initiation of collaborations at meetings of scientific communities follows similar patterns as the development of other forms of relationships. Assuming that PhD students are an especially relevant group of newcomers for scientific communities but disadvantaged in finding collaboration partners, a further question is whether they can benefit from cognitive group awareness support. Finally, the connection between present interaction and later collaboration is still unclear. These research gaps lead to the following question:

To what extent do career level, disciplinary background, and homophily regarding these two attributes, as well as group awareness support, reciprocity and transitivity predict the selection of face-to-face interaction partners and collaboration partners?

## Method

### Study context and participants

The study was conducted at the Alpine Rendez-Vous' 2011 and 2013 in France, two scientific events that aimed at bringing together researchers from multiple disciplines working on technology-enhanced learning to foster community building and scientific progress in the field.

Both Alpine Rendez-Vous' were structured in a similar way: each event consisted of several workshops on specific topics of technology-enhanced learning, and each workshop lasted one and a half days. Half of the workshops took place in the first part of the event, followed by a community event for all event participants in the evening of the second day. The second group of workshops took place after the community event. Both Alpine Rendez-Vous' were deliberately located in a large hotel at a remote place in the French Alps

to avoid external influences and to provide many opportunities for networking among the participants. Almost all participants stayed in the same hotel.

While each workshop had been selected in a competitive process and was organized independently, there was a general schedule for all workshops at the event to synchronize starting time, breaks, and end time. Between the workshop time slots, all present participants had breakfast, lunch, dinner, and coffee breaks together, as well as a long afternoon break which allowed networking also across workshops in independent social activities.

The Alpine Rendez-Vous 2011 consisted of four workshops in the first half and four workshops in the second half of the event. Additionally, a winter school for doctoral students was held across the whole event. However, winter school data will not be reported in the following analyses because of its unique design and composition compared to other workshops. The Alpine Rendez-Vous 2013 comprised five workshops in the first half and five workshops in the second half of the event. Altogether, 136 persons participated in workshops at the Alpine Rendez-Vous 2011 and 151 individual persons participated in workshops at the Alpine Rendez-Vous 2013, leading to a sample of  $N = 287$  individual participants. The majority of the participants was from European countries.

## Study design, data collection, and instruments

The study had an experimental design in which the factor *group awareness support* (with vs. without) was varied across different workshops in a randomized way. Additional quasi-experimental variables *career level* (doctoral student vs. experienced researcher) and disciplinary background (Information Technology vs. Social Sciences) varied naturally among participants within the workshops.

The data collection procedure was the same in both scientific events and for both experimental condition and control group: After being informed about their participation in a study and signing a form of consent during conference registration, participants were equipped with an RFID device, which immediately started tracking their face-to-face proximity with other participants during the conference. Tracking was deactivated when participants checked out of the hotel and returned their RFID device. Additionally, a social network questionnaire was handed out to each participant at the end of each workshop. Participants who had to leave earlier were asked to fill in an online version of the questionnaire. Personal data about the participants (career level and disciplinary background) were collected together with the social network questionnaire and within the registration form for the event.

## Independent variables

### Group awareness support

In the experimental condition, workshop participants received a brochure with information about other workshop participants. The brochure contained profiles of all workshop participants, which we compiled based on information from participants' personal websites. Each profile included basic information about each person (name, picture, and contact information), information about personal resources (research interests and exemplary publications), and information pointing to their social resources (affiliations and background). We handed the brochure to the participants at the beginning of the workshop without further instructions. The control group did not receive a brochure.

### Career level and disciplinary background

Data on both career level and disciplinary background were extracted from a questionnaire. While three different career levels were originally specified on the questionnaire (PhD/doctoral student, Early/Mid career (postdoc), and full professor), we collapsed the two latter categories. This resulted in a variable distinguishing only between PhD/doctoral students and experienced researchers (Early/Mid career and full professors forming a unique category). This allowed us to include career level as a single dummy variable into the model and made the results easier to interpret.

Information about the disciplinary background of the participants was handled similarly, resulting in a variable separating researchers with a background in Information Technology from those with a background in a social science (e.g. psychology, education, learning sciences etc.). This classification was chosen because the Alpine Rendez-Vous' had the specific focus on bringing researchers from those two types of disciplines together and to foster their collaboration.

## Data sources and measures of dependent variables

### Number of interaction partners and duration of interaction measured by RFID devices for tracking face-to-face proximity

The RFID devices, developed by the SocioPatterns collaboration (<http://www.sociopatterns.org>) were integrated into the name badge of the participants. The devices engage in bidirectional low-power radio communication. As the human body acts as a shield for the used radio frequency, and as the badges are worn on the chest, badges can exchange radio packets only when the individuals wearing them face each other at close range (about 1 to 1.5 m). The measuring infrastructure captured close face-to-face proximity between two individuals with a temporal resolution of 20 seconds, and therefore gives access to the amount of time that two participants spent together (see Cattuto et al., 2010) for a detailed description of the infrastructure). The RFID devices only tracked face-to-face proximity within the range of antennas, which were located in public spaces of the hotel only, so spare time activities taking place outside of the hotel were not tracked.

In order to exclude noise and very brief, insignificant contacts, we considered only pairs of individuals with a total measured interaction time of at least 100 seconds during the total event. The set of pairs of individuals with such interactions gives us the interaction network of the event. For each participant, we extracted from this network his/her number of distinct interaction partners.

### Number of previous and potential new collaboration partners, measured by social network questionnaires

Social network questionnaires were individually adapted to each workshop and contained a list of all workshop participants' names. Participants were asked to indicate with whom they had collaborated already before the event and with whom they had found potential for future collaboration.

We computed the number of previous collaboration partners and potential new collaboration partners using the social network questionnaire data. For each participant, we computed the number of previous collaboration partners ('Freeman degree') as the number of participants in the workshop with whom the participant declared to have had collaborated before the workshop. Likewise, we summed up the number of participants in the workshop with whom the participant had indicated to see potential for future collaboration and this yielded their number of potential new collaboration partners. As many of these include previous collaboration partners, we also consider specifically within these potential future collaboration partners the potential *new* collaboration partners (i.e., the ones who were not declared as previous partners).

### Reciprocity, transitivity, and homophily

Reciprocity, transitivity, and homophily in the selection of interaction partners and potential new collaboration partners were computed from both data sources using the analysis tool RSiena. The tool compares the expected number of reciprocal, transitive, and homophile interaction relations and new collaboration relations with the actual numbers (Ripley, Snijders, Boda, Vörös, & Preciado, 2014).

## **Analyses**

Two different analysis methods were used: Linear mixed models were created using the R package lme4 version 1.1 – 11 to predict each person's number of interaction partners / potential new collaboration partners. For each dependent variable, interactions and new collaborations, two models were computed. The first model was computed without interaction effects whereas the second model included interaction effects of career level, disciplinary background, and group awareness support. The models contained random effects for the different workshops and were run on the whole dataset, which included the data from both ARVs. To investigate selection patterns, we used the RSiena package version 1.1-232 in R for simulation investigation for empirical network analysis. We computed a model in which the development from the network of previous collaborations to the network of potential future collaborations was predicted for each workshop. In a meta-analysis of the estimates in the individual workshop models, we calculated the overall estimate across 1) all workshops, 2) all workshops without cognitive group awareness support, and 3) all workshops with cognitive group awareness support.

## **Results**

Table 1 gives an overview of four models aiming at explaining the participants' number of interaction partners during the workshop and their number of potential new collaboration partners at the end of the workshop. Having previous collaboration partners lead to more interaction partners and more potential new collaboration partners. Career level is not a significant predictor in any of the models. Disciplinary background, in contrast, predicts the number of potential new collaboration partners, showing that participants with a background in Information Technology have less potential new collaboration partners at the end of the workshop. We do not

see similar effects regarding the number of interaction partners. Moreover, group awareness support does not show an influence on the number of interaction partners nor on the number of potential new collaboration partners. No interaction effect of group awareness support with career level or disciplinary background is observed. The models aimed at predicting the number of interaction partners do not fit well, as they actually create more variance than a model with random effects only. The models explaining the number of potential new collaboration partners, in contrast, explain a large amount of variance.

Table 1: Models for the number of interaction partner and for the number of potential new collaboration partners

	<i>Number of interaction partners</i>		<i>Number of potential new collaboration partners</i>	
	<i>Main effects model Estimate (se)</i>	<i>Interaction effects model Estimate (se)</i>	<i>Main effects model Estimate (se)</i>	<i>Interaction effects model Estimate (se)</i>
<b><i>Fixed Effects</i></b>				
Intercept	<b>4.497**</b> (0.939)	<b>4.695**</b> (0.984)	<b>9.006**</b> (0.669)	<b>9.196**</b> (0.774)
Previous collaboration partners	<b>0.138*</b> (0.061)	<b>0.135*</b> (0.061)	<b>0.283**</b> (0.082)	<b>0.281**</b> (0.083)
career level (PhD student)	0.571 (0.486)	0.415 (0.680)	-0.498 (0.720)	-0.759 (1.004)
Discipline (Information Technology)	0.348 (0.505)	0.005 (0.769)	<b>-1.778*</b> (0.680)	<b>-2.044*</b> (0.975)
Group awareness support	0.965 (1.344)	0.611 (1.449)	-0.387 (0.831)	-0.757 (1.130)
Career level * group awareness support	-	0.256 (0.952)	-	0.484 (1.421)
Discipline * group awareness support	-	0.577 (1.024)	-	0.481 (1.375)
<b><i>Random Effects</i></b>				
$\sigma^2$	9.338	9.426	21.601	21.791
$\tau_{00}$	7.215	7.069	1.284	1.264
<b><i>Model fit</i></b>				
Pseudo-R <sup>2</sup> (variance between workshops explained)	-13%	-10%	73%	74,2%

\*p < .05; \*\*p < .01

We now look at the RSiena results that aimed at explaining how dyads that had not previously collaborated identified the potential for a new collaboration after the workshop and which factors are related to this change. Table 2 shows the results of the meta-analyses across the workshops for the different predictors. In the overall model across all workshops, we find that both reciprocity and transitivity seem to be related to the selection of potential new collaboration partners. When the workshops without cognitive group awareness support are contrasted to the workshops with cognitive group awareness support, we find that the reciprocity pattern disappears in both cases, while transitivity is still a significant selection pattern. This highlights the important role of previous collaboration partners, who can introduce different participants to each other. The duration of a face-to-face interaction is also an important positive predictor for selecting a potential new collaboration partner, with a more important role if no cognitive group awareness support is provided.

Regarding the role of career level, we find in the overall model that PhD students reach out to other researchers as potential new collaboration partners to the same extent as experienced researchers. However, PhD students are chosen significantly less than experienced researchers as potential new collaboration partners. In the separated models, we do not find any career level related effects. Finally, we do not find any homophilous selection behavior among career levels. Looking at disciplinary backgrounds of the participants, we find in the overall model and in the workshops without cognitive group awareness support that researchers with an Information Technology background select significantly less potential new collaboration partners. However, this effect cannot be found in the workshops with cognitive group awareness support. We do not find differences between disciplines in being chosen as potential new collaboration partners, nor do we find homophile behavior in relation to disciplines.



Table 2: Models for the selection of potential new collaboration partners

	<i>Overall model</i>		<i>Without group awareness support</i>		<i>With group awareness support</i>	
	<i>Estimate (sd)</i>	<i>p</i>	<i>Estimate (sd)</i>	<i>p</i>	<i>Estimate (sd)</i>	<i>p</i>
reciprocity	<b>0.562 (0.541)</b>	<b>.023</b>	0.397 (0.462)	.212	0.719 (0.534)	.062
transitivity	<b>0.261 (0.087)</b>	<b>&lt; .001</b>	<b>0.269 (0.108)</b>	<b>.006</b>	<b>0.226 (&lt; 0.001)</b>	<b>.001</b>
duration of interaction	<b>0.0003 (&lt; 0.001)</b>	<b>.010</b>	<b>0.0004 (&lt; 0.001)</b>	<b>.001</b>	<b>0.0002 (&lt; 0.001)</b>	<b>.014</b>
ego career level (PhD student)	0.860 (1.1219)	.150	-0.270 (< 0.001)	.654	-1.414 (1.379)	.174
alter career level (PhD student)	<b>-0.500 (0.479)</b>	<b>.028</b>	-0.412 (< 0.001)	.066	-0.670 (1.074)	.242
homophily career level	-0.194 (0.703)	.452	-0.140 (< 0.001)	.469	-0.358 (1.235)	.552
ego discipline (Information Technology)	<b>-1.278 (1.344)</b>	<b>.036</b>	<b>-2.266 (&lt; 0.001)</b>	<b>.006</b>	-0.397 (1.162)	.571
alter discipline (Information Technology)	-0.012 (0.341)	.949	-0.832 (1.173)	.251	0.105 (< 0.001)	.495
homophily discipline	-0.039 (0.258)	.800	-0.220 (.0591)	.547	-0.044 (0.125)	.785

## Discussion

Our results have contrasting aspects. On the one hand, we have not been successful in explaining with how many other researchers participants interacted during the scientific events considered. It might be that the variables considered are not the relevant ones for understanding the initiation of face-to-face interaction but others e.g. specific features of the workshop design or personality traits are more relevant. On the other hand, we were more successful in explaining the number of potential new collaboration partners the participants have identified after a workshop and in understanding the patterns that drive the selection process. The number of previous collaboration partners seems to play a major role, which is in line with the finding that transitivity is a constant and relevant selection pattern. Lacking connections to other researchers seem to be the only disadvantage of PhD students, who join a meeting of a scientific community for the first time. The lack of homophily in the interaction patterns is in contrast with the results of Barrat, Cattuto, Szomszor, van den Broeck, and Alani (2010), who found clear signs of homophily with respect to career level in the face-to-face interactions. This discrepancy might be due to the number of participants in the individual workshops, which potentially makes it easier for researchers with different levels of experience to mingle or makes it more difficult for the statistical analysis to uncover such detailed patterns. Moreover, we found that the longer researchers interact with each other, the more likely it is that they will select each other as potential collaboration partners later on. This finding is complementary to the fact, found both here and by Barrat et al. (2010) that previous collaboration partners interacted for longer time on average than pairs of individuals who had not collaborated prior to the workshop. An especially interesting finding for multi- / interdisciplinary scientific communities is that participants' background is a relevant factor for the number of potential new collaboration partners and in the selection patterns and that cognitive group awareness was associated with less differences in selection patterns between disciplines.

However, it has to be noted that this study has several limitations, with the small sample size of workshops being the most important one. Also, the identification of potential new collaboration partners cannot be assumed to equal actual new collaborations and further data about joint publications after the scientific events needs to be analyzed. So far, we can only say that our results indicate that scientific communities seem to share several selection patterns with other kinds of social networks, especially transitivity. However, other expected selection patterns, especially homophily, seem rather to depend on the group setting (larger vs. smaller scientific events). Furthermore, the literature on cognitive group awareness support has been controversial and our findings add new questions about who benefits from this kind of support.

From a practical point of view, the findings indicate that tandems of a new member and an experienced member - as they have been initiated at ICLS 2016 - could be a very valuable endeavor for community building and newcomer integration. Furthermore, supporting cognitive group awareness may be especially helpful when

different disciplines come together. However, these findings need further replication and validation to justify these conclusions and can so far only be seen as a first hint in these directions.

## References

- Aiello, L. M., Barrat, A., Cattuto, C., Ruffo, G., & Schifanella, R. Link Creation and Profile Alignment in the aNobii Social Network. In *2010 IEEE Second International Conference on Social Computing (SocialCom)* (pp. 249–256). doi:10.1109/SocialCom.2010.42
- Baerveldt, C., van de Bunt, G. G., & de Federico de la Rua, A. (2010). Why and how selection patterns in classroom networks differ between students: The potential influence of network size preferences, level of information, and group membership. *REDES - Revista hispana para el análisis de redes sociales*, *19*(12), 273–298.
- Barrat, A., Cattuto, C., Szomszor, M., van den Broeck, W., & Alani, H. (2010). Social Dynamics in Conferences: Analyses of Data from the Live Social Semantics Application. In P. F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Z. Pan, . . . B. Glimm (Eds.), *Lecture Notes in Computer Science. The Semantic Web – ISWC 2010* (Vol. 6497, pp. 17–33). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-17749-1\_2
- Byrne, D., & Clore, G. L. (1970). A reinforcement model of evaluative responses. *Personality*, *1*, 103–128.
- Cannon-Bowers, J. A., & Salas, E. (2001). Reflections on shared cognition. *Journal of Organizational Behavior*, *22*, 195–202.
- Cattuto, C., van den Broeck, W., Barrat, A., Colizza, V., Pinton, J.-F., & Vespignani, A. (2010). Dynamics of person-to-person interactions from distributed RFID sensor networks. *PloS ONE*, *5*(7), e11596.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *The American Journal of Sociology*, *94*.
- Fischer, G. (2000). Symmetry of ignorance, social creativity, and meta-design. *Knowledge-Based Systems*, *13*(7–8), 527–537. doi:10.1016/S0950-7051(00)00065-4
- Goodwin, C. (2000). Action and Embodiment Within Situated Human Interaction. *Journal of Pragmatics*, *32*, 1489–1522.
- Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J. F., & van den Broeck, W. (2011). What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of theoretical biology*, *271*(1), 166–180.
- Janssen, J., & Bodemer, D. (2013). Coordinated Computer-Supported Collaborative Learning: Awareness and Awareness Tools. *Educational Psychologist*, *48*(1), 40–55. doi:10.1080/00461520.2012.749153
- Kienle, A., & Wessner, M. (2005). Principles for Cultivating Scientific Communities of Practice. In P. Besselaar, G. Michelis, J. Preece, & C. Simone (Eds.), *Communities and Technologies 2005. Proceedings of the Second Communities and Technologies Conference, Milano 2005* (pp. 283–299). Dordrecht: Springer.
- Kienle, A., & Wessner, M. (2006). The CSCL community in its first decade: Development, continuity, connectivity. *International Journal of Computer-Supported Collaborative Learning*, *1*(1), 9–33. doi:10.1007/s11412-006-6843-5
- Lee, V. R., Ye, L., & Recker, M. (2012). What a long strange trip it's been: A comparison of authors, abstracts, and references in the 1991 and 2010 ICLS Proceedings. In J. van Aalst, K. Thompson, M. J. Jacobson, & P. Reimann (Eds.), *The Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012)* (pp. 172–176). Sydney, NSW, Australia: ISLS.
- Levine, J. M., & Moreland, R. L. (2013). Knowledge Transmission in Work Groups: Helping Newcomers to Succeed. In J. M. Levine, L. L. Thompson, & D. M. Messick (Eds.), *Series in Organization and Management. Shared Cognition in Organizations. The Management of Knowledge* (pp. 267–296). Hoboken: Taylor and Francis.
- Lin, N. (2001). *Social capital: A theory of social structure and action. Structural analysis in the social sciences: Vol. 19*. Cambridge, New York: Cambridge University Press. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=74320>
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*, 415–444.
- Ripley, R. M., Snijders, T. A. B., Boda, Z., Vörös, A., & Preciado, P. (2014). *Manual for RSiena version 4.0 (version May 10, 2014)*. Oxford. Retrieved from University of Oxford, Department of Statistics; Nuffield College website: <http://www.stats.ox.ac.uk/siena/>
- van den Besselaar, P., & Heimeriks, G. (2001). Disciplinary, Multidisciplinary, Interdisciplinary: Concepts and Indicators. In M. Davis & C. S. Wilson (Eds.), *ISSI 2001* (pp. 705–716). Sydney: UNSW 2001.
- Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). New York: Springer.

# Indexes



# Author Index

Pages 1-526: Volume 1  
Pages 527-1012: Volume 2

- Abrahamson, Dor, 909  
Acosta, Alisa, 343  
Adams, Jennifer D., 976  
Adleberg, Barrie, 735  
Agrawal, Harshit, 609  
Aleven, Vincent, 95, 633  
Almatrafi, Omaima, 605  
AlQahtani, Mona, 573  
Anderson, Emma, 199, 952  
Anderson, Richard C., 537  
Andrade, Alejandro, 151  
Andrews, Jessica J., 743, 827, 913  
Antle, Alissa Nicole, 471  
Applebaum, Lauren R., 41, 351, 664, 936  
Arastoopour, Golnaz Irgens, 529  
Arukovic, Selma, 877  
Ashley, R. William, 891  
Asterhan, Christa, 49  
Aurovindh, Sree Viswanathan, 335  
Avry, Sunny, 789  
Ayer, Tugba, 359  
Baker, Ryan, 103  
Baltés, Jennifer, 311  
Banerjee, Amartya, 851  
Bang, Megan, 735  
Bannan, Brenda, 119  
Barany, Amanda, 924  
Barrat, Alain, 519  
Barron, Brigid, 735  
Beauchamp, Gary, 819  
Beck, Stephan, 719  
Bell, Philip, 735  
Belland, Brian R., 255  
Ben-Horin, Hava, 833  
Benichou, Maya, 897  
Bennett, Cynthia L., 239  
Benson, Stephanie, 920  
Ben-Zvi, Dani, 787  
Bereiter, Carl, 703, 871  
Bétrancourt, Mireille, 789  
Bielaczyc, Katerine, 703  
Biemans, Harm, 533  
Bientzle, Martina, 771  
Birk, Gurpreet, 857  
Bishara, Sarah, 311  
Biswas, Gautam, 319  
Blikstein, Paulo, 175  
Bodemer, Daniel, 127, 287, 997  
Bodnar, Stephen, 857  
Bogouslavsky, Maria, 407  
Borge, Marcela, 573, 652, 916  
Bouton, Edith, 49  
Boyer, Kristy Elizabeth, 207, 295, 991  
Brady, Corey, 656  
Brami, Uzi Zevik, 553  
Brand, Charleen, 633  
Breuleux, Alain, 613  
Briskin, Jessica, 863  
Buessing, Shawn, 719  
Burns, Mary, 537  
Calandra, Brendan, 359  
Carnes, Molly, 549  
Carvalho, Renato, 773  
Cassell, Justine, 383  
Cattaneo, Alberto A. P., 557  
Cearrtto-Pargman, Teresa, 905  
Celepkolu, Mehmet, 295  
Chai, Shaoming, 703  
Chamberlain, John, 827  
Chan, Carol K.K., 703, 711, 867, 873, 881, 883, 901, 993  
Chan, Melvin, 696  
Chanel, Guillaume, 789  
Charles, Elizabeth S., 811, 897  
Chen, Bodong, 711, 916  
Chen, Feng, 711, 867  
Chen, Mei-Hwa, 703, 711  
Chen, Yuxin, 857  
Cheng, Kenneth, 431  
Cheng, Teng-Yao, 641, 855  
Chounta, Irene-Angelica, 589  
Christmann, Nina, 625  
Close, Kevin, 887  
Cobb, Sue, 719  
Cohen, Jonathan, 359  
Cook, Claire, 719  
Corning, Wiley, 719, 879  
Costa, Stacy A., 797  
Coulter, Bob, 783  
Cress, Ulrike, 751, 771, 897  
Crick, Tom, 819  
Crossley, Scott, 103  
Csanadi, Andras, 215  
Cukurova, Mutlu, 263, 565, 839, 934  
Czerniewicz, Laura, 3  
D'Angelo, Cynthia, 916  
D'Angelo, Sarah, 727  
Dado, Marielle, 287, 997

Dai, Zhaihuan, 777  
 Damsa, Crina, 905  
 Danish, Joshua A., 853  
 Daradoumis, Thanasis, 803  
 Dascalu, Mihai, 103, 223  
 Davis, Bria, 853  
 Davis, Katie, 795  
 Davis, Richard Lee, 175, 735  
 Dawes, Les, 621  
 de Jong, Frank, 703, 711  
 De Wever, Bram, 557,  
 del Castillo, Fernando Diaz, 703  
 Dellatola, Eirini, 803  
 DeSchryver, Michael, 111  
 Di Eugenio, Barbara, 629  
 Dillenbourg, Pierre, 727  
 Dimitriadis, Yannis, 803  
 Dishon, Gideon, 815  
 Donaldson, Jonan Phillip, 924  
 Dong, Yi, 319  
 Dornfeld, Catherine, 191, 1005  
 Doyle, Tanya, 621  
 Dudek, Jaclyn, 775  
 Dugdale, Michael, 811  
 Dutta, Nikita, 247  
 Eagan, Brendan R., 215, 529  
 Eberle, Julia, 877, 897  
 Elby, Andrew, 65  
 Elming, Anna Lindenhoff, 271  
 Elson, Malte, 625  
 Emar, Mona, 319,  
 Emmanuel, Navo, 439, 767  
 Enyedy, Noel, 853  
 Erdmann, Julia, 625  
 Erkens, Melanie, 127  
 Espino, Danielle, 861  
 Fake, Helen, 119  
 Faulkner, Roosevelt, 439  
 Fefferman, Nina, 865  
 Feng, Xueqi, 873, 993  
 Fields, Deborah A., 865  
 Fields, Gabriel, 719  
 Finkelstein, Samantha, 511  
 Fischer, Frank, 215, 519, 751, 847, 897, 922, 971  
 Fleming, Mary, 779  
 Flynn, Paul, 779  
 Ford, Cecilia E., 549  
 Fossati, Davide, 629  
 Foster, Aroutis, 821  
 Frensley, Troy, 399,  
 Fricke, Kyle W., 664  
 Friend, Michelle, 71  
 Fröhlich, Bernd, 719  
 Funaoi, Hideo, 593  
 Fuqua, Aeslya, 359  
 G, Nagarjuna, 609  
 Galoyan, Tamara, 821  
 Gao, Dandan, 644  
 Gebre, Engida, 684  
 Gemmell, Jonathan, 311  
 Georgen, Chris, 151  
 Georgiou, Yiannis, 423  
 Gergle, Darren, 727  
 Giang, Michael T., 865  
 Giannakos, Michail, 183  
 Giri, Sagun, 415  
 Gleasman, Cory, 617  
 Gnesdilow, Dana, 327  
 Gomez, Kimberly, 680  
 Goodyear, Peter, 897  
 Gouveia, Christine, 719  
 Graesser, Art, 743, 913  
 Gravel, Brian, 447  
 Gray, Steven, 399  
 Green, Nick, 629  
 Greenwald, Scott W., 719, 879  
 Guo, Yu, 656  
 Gutu, Gabriel, 223  
 Hakkarainen, Kai, 703, 751, 831, 981  
 Hall, Allison H., 773  
 Hall, Rogers, 33, 735  
 Halverson, Erica, 901  
 Hämäläinen, Raija, 557  
 Hamid, Taha, 311  
 Hamilton, Eric, 861  
 Hansen, Jennifer, 735  
 Hao, Jiangang, 743, 913  
 Harrell, D. Fox, 4  
 Harrell, Maralee, 589  
 Harsley, Rachel, 629  
 Hartley, Kylie, 439, 767  
 Hartmann, Christian, 633  
 Hatami, Javad, 533  
 Healion, Donal, 565  
 Hecking, Tobias, 287, 625  
 Henderson, J. Bryan, 887  
 Herrmann, Thomas, 625  
 Hickey, Timothy J., 463  
 Hill, Roger B., 617  
 Hinojosa, Leighanna, 648  
 Hirayama, Ryoya, 593  
 Hmelo-Silver, Cindy E., 399, 439, 711, 767, 857,  
 916  
 Hod, Yotam, 87, 787, 897, 922  
 Hoffman, Kelly H., 795  
 Holland, Jennifer, 719  
 Hollett, Ty, 601  
 Hong, Helen, 672, 1009  
 Hoppe, H. Ulrich, 287, 625  
 Horn, Michael, 656, 851  
 Horwitz, Paul, 827  
 Houlihan, Barry, 779  
 Howard, Sarah K., 801  
 Hsu, Pi-Sui, 781

Huang, Joey, 399  
 Jackson, Julia, 537  
 Jahnke, Isa, 905  
 Jeong, Heisawn, 439, 767, 771  
 Jermann, Patrick, 727  
 Jimenez Pazmino, Priscilla, 843  
 Johri, Aditya, 479, 605  
 Jordan, Rebecca, 399  
 Joyce-Gibbons, Andrew, 819  
 Jung, Yong Ju, 652  
 Kaatz, Anna, 549  
 Kafai, Yasmin B., 569, 735, 815, 865, 981  
 Kahn, Jennifer, 735  
 Kali, Yael, 833, 897  
 Kamsan, Muhamad Ansar B., 455  
 Kanasa, Harry, 621, 801  
 Kapur, Manu, 931  
 Kassab, Dima, 1001  
 Kato, Hiroshi, 593  
 Kawas, Saba, 795  
 Kazemitabar, Maedeh, 857  
 Ke, Fengfeng, 777  
 Keith, P. Kevin, 841  
 Kelly, Susan B., 581  
 Kerr, Deirdre, 743  
 Kessler, Aaron, 807  
 Keune, Anna, 545  
 Khanlari, Ahmad, 585, 668, 869, 871, 875  
 Kici, Derya, 1007  
 Kim, ChanMin, 617  
 Kim, David, 785  
 Kim, Nam Ju, 255  
 Kim, Soo Hyeon, 676, 863  
 Kim, Yanghee, 799  
 Kim, Yoon Jeon, 829  
 Kimmerle, Joachim, 771  
 Kirschner, Paul, 711  
 Koh, Elizabeth, 672, 962  
 Kollar, Ingo, 215  
 Koon, Al, 827  
 Korhonen, Tiina, 831  
 Koskinen, Anniina, 831  
 Krämer, Nicole C., 625  
 Kubota, Yoshihiko, 593  
 Kulik, Alexander, 719  
 Kunert, André, 719  
 Kyewski, Elias, 625  
 Kyllonen, Patrick, 743  
 Kyza, Eleni A., 423  
 Laferrière, Thérèse, 613, 703  
 Laina, Vasiliki, 447  
 Lajoie, Susanne P., 688, 857  
 Land, Susan M., 863  
 Landy, David, 909  
 Lasry, Nathaniel, 811  
 Lavonen, Jari, 831  
 Law, Nancy, 431  
 Lawrence, LuEttaMae, 581  
 Lee, Alwyn Vwen Yen, 985  
 Lee, Eric Monsu, 781  
 Lee, Hee-Sun, 135, 577  
 Lee, Jin Ha, 167  
 Lee, Teo Chew, 5, 455, 692, 696, 703  
 Lee, Ung-Sang, 680  
 Lee, Victor R., 735  
 Lee, Victoria, 719  
 Leftheriotis, Ioannis, 183  
 Lenton, Kevin, 811  
 Levy, Sharona, 656, 909  
 Lewis, Whitney, 920  
 Li, Taihua, 311  
 Liang, Leming  
 Liao, Jian, 775  
 Lin, Feng, 867, 940  
 Lin, Lijia, 765  
 Lindgren, Robb, 813, 909  
 Ling, Lee Yu, 692  
 Linn, Marcia C., 41, 351, 375, 664  
 Litts, Breanne K., 569, 920  
 Liu, Kun, 881  
 Liu, Lei, 743, 913  
 Lord, Trudi, 135  
 Luckin, Rosemary, 839  
 Ludvigsen, Sten, 901  
 Lui, Debora A., 569  
 Lui, Michelle, 495, 948  
 Lund, Kristine, 231, 519, 922, 971  
 Lynch, Collin F., 991  
 Lyons, Leilah, 487, 843  
 Ma, Jun, 801  
 Ma, Leanne, 455  
 MacArthur, Chris, 311  
 Madaio, Michael, 383  
 Maes, Pattie, 719, 879  
 Maggiore, Chrystal, 863  
 Magnussen, Rikke, 271  
 Margulieux, Lauren, 932  
 Martin, Fred, 143  
 Martin, Caitlin K., 311  
 Martin, Nicole, 327  
 Marx, Sherry, 799  
 Matcuk, Matt, 851  
 Matsuzawa, Yoshiaki, 703, 711  
 Matuk, Camillia, 375, 974  
 Mavrikis, Manolis, 839  
 McAuley, Alexander, 703  
 McBride, Elizabeth, 41, 983  
 McGee, Steven, 851  
 McIntyre, Cynthia, 827  
 McKeown, Jessica, 439, 767  
 McKinley, Zachary, 891  
 McLaren, Bruce M., 589  
 McMullen, Kyla, 295  
 McNamara, Danielle S., 103

McNaughton, James, 819  
 McSweeney, Niall, 779  
 Mercier, Emma, 581, 823, 825  
 Millan, Eva, 839  
 Mislevy, Robert J., 743  
 Mochizuki, Toshio, 593  
 Moher, Tom, 495, 503  
 Molinari, Gaëlle, 789  
 Montané, Mireia, 703  
 Moon, Jewoong, 777  
 Mortensen, Chase, 920  
 Mulder, Martin, 533  
 Nachankar, Mrunal, 609  
 Nacu, Denise C., 311  
 Nardi, Nicholas, 861  
 Nathan, Mitchell J., 549, 901, 909  
 Natsis, Antonis, 791  
 Ndegemo, Joyce, 861  
 Newbutt, Nigel, 719  
 Newman, Greg, 399  
 Ng, Meixi, 735  
 Nguyen, Tung, 799  
 Ni, Lijun, 143  
 Nickels, Katherine, 621  
 Nistor, Nicolae, 561  
 Noor, Jama, 183  
 Noroozi, Omid, 533  
 Nouri, Jalal, 837  
 Noushad, Noora F., 597  
 Nunes, Cesar, 703  
 Nussbaum, Miguel, 905  
 Nutchey, David, 621  
 Obwegeser, Nikolaus, 791  
 Ogan, Amy, 383  
 Ohsaki, Ayano, 57  
 Olsen, Jennifer K., 95, 511, 633  
 Oshima, Jun, 57, 703, 711, 763, 981  
 Oshima, Ritsuko, 57  
 Ottinger, Sarah, 847  
 Ouyang, Xueying, 573  
 Owho-Ovuakporie, Kesiena, 829  
 Owoo, Mama Adobea Nii, 859  
 Oztok, Murat, 957  
 Pal, Joyjeet, 479  
 Pallant, Amy, 135, 577  
 Pan, Yanjun, 777  
 Papadopoulos, Pantelis, 791  
 Pardo, Abelardo, 801  
 Pardos, Zach, 727  
 Park, Miyoung, 199  
 Parsons, Sarah, 719  
 Patchan, Melissa, 955  
 Payne, Scott, 719  
 Peppler, Kylie, 545, 981  
 Perera, Nishan, 159  
 Phillips, Abigail, 735  
 Pier, Elizabeth L., 549  
 Piland, Jacob, 255  
 Pinkard, Nichole, 311  
 Pion, Carmit, 833  
 Plevinski, Justin, 111  
 Polman, Joseph L., 648, 849  
 Popov, Vitaliy, 938  
 Poza, Ricardo, 841  
 Pratt, Wanda, 239  
 Price, Kimberly Michelle, 207  
 Price, Nancy, 735  
 Prieto, Luis P., 727  
 Pryputniewicz, Sarah, 135  
 Puhl, Thomas, 247  
 Pun, Thierry, 789  
 Puntambekar, Sadhana, 191, 327, 1005  
 Quigley, David, 809  
 Quignard, Matthieu, 231  
 Quintana, Rebecca, 503  
 Raclaw, Joshua, 549  
 Raes, Annelies, 942  
 Raicu, Daniela, 311  
 Ratté, Sylvie, 711  
 Rau, Martina, 79, 279, 727  
 Rayner, Mason, 769, 835  
 Rebedea, Traian, 223  
 Recker, Mimi, 735  
 Reeve, Richard, 703  
 Reich, Justin, 829  
 Reiss, Kristina, 847  
 Renken, Maggie, 359  
 Resendes, Monica, 585, 869, 875  
 Ricarte, Thales, 743  
 Richard, Gabriela T., 415, 891  
 Roberts, Jessica, 487, 851, 959  
 Robinson, Kevin, 829  
 Rodriguez, Kimberly, 769, 835  
 Rodríguez, Fernando J., 207  
 Rogers, Bradley, 529  
 Roschelle, Jeremy, 901  
 Rosé, Carolyn, 25, 711  
 Rosson, Mary Beth, 303  
 Ruffaldi, Emanuele, 263  
 Ruis, Andrew R., 529  
 Rummel, Nikol, 95, 625, 633, 727, 805, 877  
 Russell, Sam, 565  
 Sagy, Ornit, 87  
 Säljö, Roger, 905  
 Sasaki, Hiroshi, 593  
 Scaff, Ligaya, 795  
 Scardamalia, Marlene, 585, 668, 703, 751, 797,  
 869, 871, 875, 1007  
 Schmitt, Lara Johanna, 9  
 Schneider, Bertrand, 175, 727  
 Schönfeld, Tim, 877  
 Schwaighofer, Matthias, 847  
 Schwendimann, Beat A., 557  
 Sedik, Zahira Mohd, 692



Seitamaa-Hakkarainen, Pirita, 703, 831, 981  
 Seo, JooYoung, 573  
 Serafin, Yvonne, 561  
 Serlin, Ronald, 529  
 Seufert, Tina, 817  
 Shaban, Yara, 447  
 Shaenfield, David, 761  
 Shaffer, David Williamson, 215, 529, 769, 835, 1003  
 Shah, Mamta, 821  
 Shaikh, Rafikh, 609  
 Shapiro, Ben Rydal, 33  
 Shareff, Becca, 447  
 Sharma, Kshitij, 183, 727  
 Shehab, Saadeddine, 823, 825  
 Shim, Jooeun, 597  
 Shin, Minyoung, 617  
 Shinohara, Kristen, 239  
 Shirouzu, Hajime, 703  
 Siebert-Evenstone, Amanda, 1003  
 Sierschynski, Jarek, 637  
 Silvis, Deborah, 391  
 Slattey, Brian, 843  
 Slotta, James D., 343, 503, 897  
 Smith, Brian K., 924  
 Smith, Thomas J., 781  
 Snyder, Anne, 719  
 Sommer, Stephen R., 648, 849  
 Sommerhoff, Daniel, 922  
 Sormunen, Kati, 831  
 Spaulding, Scott, 637  
 Spikol, Daniel, 263, 565  
 Steele, K-Fai, 735  
 Steffen, Wiebke, 771  
 Stegmann, Karsten, 519  
 Steier, Rolf, 966  
 Stoddard, Jeremy, 769, 835  
 Strohmaier, Anselm, 847  
 Stromholt, Shelley, 735  
 Stucker, Michael, 151  
 Subramaniam, Mega, 795  
 Sullivan, Florence R., 841  
 Sumner, Tamara, 809  
 Sun, Jiangshan, 765  
 Sun, Jingjing, 537  
 Sun, Na, 303  
 Suzuki, Hideyuki, 593  
 Swiecki, Zachari, 769, 835  
 Tabak, Iris, 553  
 Tan, Elaine, 819  
 Tan, Samuel, 455  
 Tan, Seng Chee, 703, 985  
 Tao, Dan, 644, 660, 995  
 Tarimo, William T., 463  
 Taylor, Katie Headrick, 391  
 Telhan, Orkan, 735  
 Terwedow, Ilka, 847,

Thompson, Kate, 621, 801, 897  
 Tjietjen, Phil, 897  
 Tong, Yuyao, 881, 883  
 Toprani, Dhvani, 652  
 Tran, Kelly M., 845  
 Trausan-Matu, Stefan, 103, 223  
 Traut, Hilary, 648  
 Tsan, Jennifer, 991  
 Tscholl, Michael, 319  
 Tsovaltzi, Dimitra, 247  
 Tsunakawa, Takashi, 763  
 Tu, Xintian, 853  
 Tzou, Carrie, 735  
 Ufer, Stefan, 847  
 van Aalst, Jan, 703, 711, 873, 881, 883, 901, 993  
 van den Ende, Joan, 711  
 Van Dyke, Margot, 781  
 van Heijst, Hennie, 711  
 van Leeuwen, Anouschka, 805, 944  
 Vanlehn, Kurt, 335  
 VanMeerten, Nicolaas, 541  
 Varma, Keisha, 541  
 Velazquez Godinez, Erick, 711  
 Vinha, Telma, 703  
 Viswanathan, Anandhi, 335, 479  
 Vitale, Jonathan M., 41, 351, 664  
 Vogel, Freydis, 751, 847, 922  
 von Davier, Alina A., 743, 827, 913  
 Wagh, Aditi, 656  
 Wakimoto, Takehiro, 593  
 Walker, Justice T., 569  
 Walker, Susan, 367  
 Walkington, Candace, 909  
 Walkoe, Janet, 65  
 Wallon, Robert C., 813, 989  
 Wan, Hai-Peng, 17  
 Wang, Li-Jen, 641, 855  
 Wang, Qi, 17  
 Wang, Xu, 25, 987  
 Warren, Jillian, 471  
 Weible, Jennifer, 111  
 Weidler-Lewis, Joanna, 648  
 Weinberger, Armin, 9, 247, 817, 931  
 Weiss, David M., 255  
 Wen, Miaomiao, 25  
 Wen, Yun, 946  
 Whiteman, Jennifer, 119  
 Whittaker, Chris, 811  
 Wichmann, Astrid, 625  
 Widman, Sari A., 569  
 Wiggins, Joseph B., 295  
 Wilensky, Uri, 656  
 Wilkerson, Michelle, 65, 447  
 Williams, Christopher A., 817  
 Williams-Pierce, Caro, 909, 1001  
 Windleharth, Travis, 167  
 Wise, Alyssa Friend, 159, 471, 711, 916

Wobbrock, Jacob O., 239  
Wong, Jacqueline, 865  
Wozniak, Kathryn, 807  
Wright, Kenneth, 577  
Wu, Sally P. W., 79, 279  
Wu, Ying-Tien, 641, 855, 875  
Xia, Lei, 719  
Xing, Wanli, 950  
Xu, Xinhao, 777  
Xu, Yang, 793  
Xuan, Andy Ng Ding, 692  
Yamada, Yuki, 57  
Yan, Shulong, 652, 999  
Yang, Christine, 711  
Yang, Jie, 801  
Yang, Yuqin, 873  
Yeoman, Pippa, 897  
Yin, Jia, 765  
Yip, Jason, 167  
Yoon, Susan A., 199, 597, 901, 931  
Young, Nick, 819  
Yu, Sheng-Quan, 17  
Yuan, Guangji, 407  
Yuan, Jiangmei, 617  
Zapata, Diego, 913  
Zhang, Jianwei, 407, 644, 660, 703, 711, 995  
Zhang, Jiayuan, 375  
Zhao, Jian, 765  
Zhao, Naxin, 191  
Zheng, Xudong, 765  
Zhu, Gaoxia, 585, 668, 869, 875  
Zhu, Mengxiao, 743  
Zimmerman, Heather Toomey, 676, 863

# Keyword Index

Pages 1-526: Volume 1  
Pages 527-1012: Volume 2

access.....	95, 247, 423, 601, 672, 827
action research.....	239, 815, 825, 881
activity theory.....	151, 672, 795, 815, 841
adoption of technologies.....	71, 87, 247, 263, 463, 471, 565, 641, 644, 791, 811, 825, 827, 863
affective dimensions.....	135, 431, 815, 851
anthropology.....	672
architectures.....	135, 775
argumentation.....	57, 581, 668
artifact analysis.....	17, 668
artificial intelligence.....	383, 644
assistive technology.....	57, 95, 247, 573, 644, 773
augmented/virtual reality.....	487, 664, 767
authoring tool.....	207, 553, 644, 668, 815, 829
behaviorism.....	343, 644
best practices.....	247, 519
broadening participation.....	111, 247, 343, 399, 569, 827, 863
case studies.....	65, 151, 263, 311, 463, 511, 519, 533, 565, 597, 641, 648, 672, 688, 821, 833, 837
case-based instruction.....	41, 773
cognitive dimensions.....	644, 837, 869
cognitive psychology.....	57, 343, 644, 648, 775, 839
cognitive studies.....	57, 637, 644, 785
collaboration skills.....	65, 79, 231, 263, 303, 447, 463, 553, 565, 660, 688, 775, 777, 779, 795, 811, 827, 857
collaborative problem solving skills.....	9, 135, 167, 287, 303, 431, 519, 593, 629, 652, 660, 779, 811, 817, 827, 863, 875, 883
collective regulation/socio-metacognition.....	279, 648, 680, 775, 813
communication studies.....	263, 335, 463, 545
community building.....	223, 247, 255, 263, 455, 463, 821, 827, 849, 863, 867
community settings.....	255, 263, 463, 676, 773, 815
competence/expertise.....	41, 57, 191, 479, 565, 773
computational models.....	415, 829
computer science.....	57, 119, 127, 135, 199, 215, 247, 311, 625, 644, 672, 793, 801, 827, 863
computing education.....	33, 57, 79, 127, 135, 191, 215, 247, 255, 311, 487, 519, 617, 625, 644, 793, 827, 829, 839, 845, 871
conceptual change.....	585, 648, 795, 805, 877
connected learning.....	815
constructivism/constructionism.....	33, 57, 183, 247, 255, 455, 581, 621, 644, 668, 672, 773, 789, 829, 837, 865, 877
content analysis.....	17, 57, 343, 545, 817, 821, 861
contextual design.....	247, 255, 533, 633, 668, 811
conversation analysis.....	303, 335, 545, 773, 779, 827
core competencies.....	41, 319, 519
creativity.....	247, 319, 519, 785, 827, 839
cultural psychology.....	672, 839, 863
Cultural-Historical Activity Theory.....	151, 672, 795, 815
culturally relevant pedagogy/design.....	247, 399, 533, 668, 827, 863
culture.....	247, 463, 827, 863
descriptive studies.....	151, 263, 303, 423, 519, 553, 573, 597, 609, 617, 633, 672, 797, 833, 837
design experiments.....	49, 479, 644, 785
design thinking.....	359, 621, 644, 664, 672, 769, 803, 821, 827, 863
design-based research.....	9, 41, 79, 103, 111, 255, 327, 367, 399, 439, 471, 511, 519, 549, 668,

	672, 680, 684, 769, 773, 783, 813, 821, 827, 829, 837, 847, 865, 879
didactics.....	391
discourse analysis.....	431, 561, 577, 581, 629, 680, 761, 779, 841, 875
discovery learning.....	343, 644
distance learning.....	431
distributed cognition.....	247, 672, 797, 821, 827, 839, 875
diversity and inclusion.....	95, 247, 533, 541, 553, 573, 601, 609, 641, 672, 815, 821, 827, 863
early childhood education.....	17, 95, 553, 644
education technology.....	231, 247, 255, 263, 303, 375, 391, 399, 407, 415, 431, 463, 471, 511, 519, 533, 537, 545, 549, 589, 601, 605, 644, 652, 668, 763, 765, 769, 777, 789, 797, 801, 807, 809, 811, 819, 821, 827, 853, 857, 879
educational psychology.....	49, 65, 343, 447, 519, 553, 637, 641, 644, 672, 688, 805, 807, 815, 819
engagement.....	17, 239, 255, 263, 343, 431, 463, 511, 668, 801, 815, 851, 859, 875
engagement and transfer.....	644, 811, 827
English as a Second Language.....	601, 609
equity and gender.....	135, 247, 311, 533, 827, 863
ethnography.....	71, 151, 239, 617, 641, 672, 799, 863
experimental studies.....	135, 167, 343, 805
game-based media.....	359, 519, 533, 644, 765, 821, 827, 831, 835, 861
gaming.....	375, 519, 549, 644, 652, 761, 765, 785, 811, 821, 827, 831, 835, 839
group cognition.....	135, 553, 597, 787, 839
handheld/mobile devices.....	41, 151, 255, 271, 367, 511, 553, 811, 835, 847
history.....	255, 789, 821, 881
human-computer interaction.....	143, 151, 247, 367, 557, 641, 644, 672, 773, 815, 821, 827, 837, 863
identity and community.....	143, 247, 529, 533, 815, 827, 867
informal learning settings.....	175, 247, 255, 311, 359, 367, 399, 503, 519, 569, 577
information sharing.....	95, 399, 809, 827, 839
innovative algorithms.....	801
innovative design approach.....	49, 103, 207, 343, 407, 511, 519, 529, 633, 644, 803, 819, 821, 827
innovative software/artifacts.....	41, 231, 391, 511, 644, 773, 883
intelligent systems.....	135, 215, 303, 383, 644, 807, 819
interaction analysis.....	135, 279, 343, 359, 431, 503, 511, 553, 581, 763, 767, 771, 795, 855
interaction tool.....	41, 151, 399, 511, 549, 644, 664, 761, 815, 831, 833
interactive whiteboard.....	431, 791, 803
intersubjectivity.....	839
knowledge building.....	57, 95, 103, 263, 271, 439, 463, 537, 541, 561, 585, 613, 629, 680, 684, 692, 696, 763, 769, 773, 783, 823, 827, 833, 869, 879
large-scale studies.....	601, 791
learning analytics/educational data mining.....	183, 199, 223, 303, 541, 644, 773, 781, 811, 829, 841, 851
learning analytics and monitoring tools.....	119, 279, 303, 327, 343, 537, 644, 791, 841, 867
learning management systems.....	79, 223, 849, 853, 859, 883
learning outcomes.....	239, 343, 549, 644, 660, 811, 859, 871
life-long learning.....	57, 247, 263, 431, 463, 537, 644, 815, 827
linguistics.....	383
literacy.....	151, 255, 375, 519, 537, 553, 668, 672, 799, 815, 825, 845, 863
longitudinal studies.....	263, 849
mathematics education.....	231, 303, 343, 431, 447, 471, 672, 797, 831, 853, 855
MOOCs.....	87, 423, 439, 601, 609, 867
motion capture/tracking.....	815
motivation.....	239, 247, 529, 601, 644, 807, 819, 821, 827
multi-touch devices.....	95, 167, 503, 553, 811
online learning.....	57, 87, 143, 263, 351, 415, 463, 519, 545, 561, 593, 644, 801, 817, 819, 821, 859, 873
participatory design.....	247, 455, 519, 549, 565, 633, 641, 773, 815, 827, 849
participatory evaluation.....	785, 827, 839
participatory simulations.....	41, 191, 789, 815
partnering across sectors/organizations.....	773, 849
personalized learning environments.....	247, 343, 644, 807, 819, 827

portable/ubiquitous collaboration .....	263, 271, 391, 463, 839, 847
primary school .....	49, 95, 151, 279, 303, 407, 519, 553, 625, 629, 644, 672, 783, 797, 823, 833, 849, 869, 879
problem solving tool .....	65, 319, 431, 519, 688, 815, 821, 883
problem-based learning .....	199, 519, 815, 827, 837, 875
process support .....	71, 135, 191, 407, 565, 597, 644, 656, 763, 773, 817
project-based instruction .....	498, 530, 926, 998, 1090
reflection .....	57, 263, 431, 463, 487, 648, 656, 787, 815, 873
regulation .....	279, 644, 648, 680, 775, 813
representational affordances .....	247, 343, 495, 511, 593, 597, 668, 815
representational tools .....	343, 495, 511, 597, 637, 668, 787, 815, 817, 829, 865, 881
research methodologies .....	255, 351, 577, 815
scaffolding .....	57, 207, 239, 255, 287, 295, 343, 447, 495, 519, 589, 593, 648, 656, 668, 773, 787, 811, 819, 827, 835, 839, 849
scenario-based design .....	111, 431
science education .....	33, 57, 159, 255, 279, 287, 295, 327, 399, 415, 479, 495, 511, 519, 541, 549, 589, 605, 629, 668, 761, 763, 771, 783, 795, 805, 807, 811, 817, 819, 829, 833, 839, 853, 865, 869, 871
secondary school .....	199, 255, 303, 343, 399, 533, 549, 565, 613, 625, 641, 644, 648, 668, 672, 771, 773, 777, 789, 795, 827, 829, 839, 849, 875
self-regulated learning .....	529, 644
shared knowledge .....	167, 247, 271, 431, 597, 605, 641, 775, 797, 823, 827, 839, 875
simulations .....	9, 33, 41, 111, 183, 191, 287, 511, 593, 789, 815, 829, 857
situated learning .....	41, 549, 597, 637, 668, 672, 781, 795, 811, 813, 815, 821, 827, 873, 877
skill development .....	17, 57, 191, 239, 247, 295, 407, 519, 545, 549, 773, 827, 873
social constructivism .....	65, 135, 255, 263, 431, 439, 463, 511, 553, 565, 648, 688, 775, 803, 831
social network analysis .....	119, 263, 383, 613, 652, 676, 684, 781, 791, 839, 861
social networking (Facebook, etc) .....	175, 597, 799, 825, 843
social psychology .....	335
stakeholder analysis .....	25, 191
student achievement .....	263, 463, 827
surveys and questionnaires .....	191, 399, 423, 533, 573, 601, 801, 807, 811
tangible computing .....	247, 557, 827
teacher education .....	17, 25, 41, 255, 263, 407, 463, 471, 537, 565, 637, 773, 793, 839, 873
technical proof of concept .....	407, 664, 781
transfer .....	519, 644, 811
university .....	17, 41, 57, 71, 127, 239, 255, 303, 407, 431, 511, 621, 644, 763, 809, 819, 821, 881
usability test .....	9, 95, 819
video analysis .....	199, 511, 533, 557, 593, 795, 829, 837
virtual communities .....	95, 135, 143, 375, 644, 825, 839, 845, 851
virtual/3D environments .....	351, 511, 519, 761, 839
visual/performance arts .....	613, 855
workplace .....	191, 263, 407, 463, 644, 773