Mapping Writing Analytics

Joseph M. Moxley
University of South Florida
4202 E. Fowler Ave
Tampa, FL 33620
1-813-974-2421
mox@usf.edu

Katie Walkup
University of South Florida
4202 E. Fowler Ave
Tampa, FL 33620
1-813-922-1492
kwalkup@mail.usf.edu

ABSTRACT

This paper extends foundational perspectives on the definition of Writing Analytics (WA) to further conceptualize this nascent field of research. The affordances of digital tools have enabled us to reimagine WA's meanings and applications. The authors use the metaphor of mapping to understand the tensions and successes navigated by researchers and practitioners and to chart new ways in which this field can benefit the domains of academia, business, and culture. Approaching WA from an interdisciplinary perspective allows the field to consider new research questions.

Keywords

foundations, visual mapping, *My Reviewers*, STEM

1. INTRODUCTION

This paper seeks to expand the work of Buckingham Shum, Knight, McNamara, Allen, Bektik, and Crossley [1] in establishing the foundations of Writing Analytics (WA). Despite a growing interest in the applications of WA, and several conferences on these applications, including LAK (Learning Analytics and Knowledge) and EDM (Educational Data Mining), there remain surprisingly few foundational pieces on WA. This paper examines this field and reimagines its research areas with the end goal of charting the main questions of WA and its move into interdisciplinary domains. As the available literature suggests, researchers sometimes disagree as to the purpose of WA, and disagreements intensify when researchers approach the subject from a disciplinary perspective. However, when WA is visualized from an interdisciplinary perspective, the tangled threads of the field sort themselves into fairly clear questions, with plenty of room for expansion. Knight and Littleton [2] write that "there is an intriguing array of opportunities within the middle space of learning sciences and computational techniques around discourse data" (203). This paper explores this middle ground.

To help illustrate this middle ground, we introduce preliminary charts that map the main corpora, orientations, implementations, and domains in WA as a research field. These research maps represent the work of graduate students in a Rhetoric and Technology course at the University of South Florida (USF). Over the semester, students discussed WA research and mind-mapped scholars, concepts, and current applications of WA. Students were also encourage to conduct their own interdisciplinary research utilizing corpus methods.

The present study examines two infographics produced by this graduate seminar. The first infographic represents WA from a disciplinary perspective; specifically, we adopt a humanities-centric viewpoint. This research map exemplifies current tensions in the field of WA, and particularly showcases the critique of WA from many scholars in the humanities [3].

The second research map represents a wider understanding of the possibilities offered by WA. The categories suggested by this map show efforts to elide disciplinary tensions, to move beyond quick conceptions like automated writing evaluation (AWE) or machine scoring. While the first research map symbolizes disciplinary concerns and critique of WA, the second research map moves the conversation into a more productive avenue, facilitating discussion of what methods like datamining and predictive analysis can do for interdisciplinary research.

2. PERSPECTIVES ON WA

These research maps draw on critical research on corpus methods, datamining, machine learning, and the implications of these digital tools on the job market. The authors are particularly interested in the Critical Perspectives on Writing Analytics workshop from the Learning Analytics and Knowledge 2016 conference. Concerns in WA include topics on Latent Semantic Analysis (LSA), Natural Language Processing (NLP), and Discourse-Centric Learning Analytics (DCLA). From the available literature on the subject, the authors drew three preliminary research questions:

1. Should developmental tools that respond to student writing with formative and summative feedback be used in classrooms before they are considered finished [4]?

2. How does our bias as teachers of English Composition inform our concerns about WA, and how can we surmount these concerns [2]?

3. If we can overcome our disciplinary perspective, how might a wider viewpoint on the practices and implementations of WA benefit the field [2]?

In keeping with Buckingham Shum et al. [1], the authors wished to "[recognize] the many wider issues that aid or obstruct

analytics adoption in educational settings, such as theoretical and pedagogical grounding, usability, user experience, stakeholder design engagement, practitioner development, organizational infrastructure, policy and ethics." These foundational concerns of WA find their places in both research maps.

Review of available literature on WA suggests that its present service promise is pedagogical in nature, in keeping with its roots in Learning Analytics. As WA expands its domains, however, we expect its service promises to expand as well. As Knight & Littleton [2], write, "our analytic techniques should not be limited by the ways in which educational dialogue has previously been operationalized; new analytic techniques afford new opportunities to theorize and reconsider what constitutes productive dialogue" (193). In keeping with Knight & Littleton [2], we created our maps using their data-driven science method. We imagined our target goals, speculated behaviors which would indicate these goals being reached, and created tasks which would cause these behaviors.

To create a data-driven map of WA as a research area also required detailed discussions of the possibilities offered by both WA and learning analytics [6]. These possibilities were tempered with research on the current limitations of AWE technologies, as well as user's current experiences with these technologies [7]. Finally, researchers and practitioners in WA, to fulfill its current service promise of improving education, must consider issues of ethics and how WA will protect the best interests of students [8].

3. SITE OF RESEARCH

The University of South Florida (USF) is a test site for My Reviewers, a web-based learning environment. English and Chemistry instructors and students at USF use My Reviewers to collaborate digitally with students, which results in actionable feedback.

This feedback is recorded in a corpus of over a billion words. Researchers have used the My Reviewers corpus to propose and complete projects involving corpus linguistics, datamining, and automated writing evaluation (AWE). My Reviewers is a vigorous site for corpus-based research, developed in response to the needs of faculty and students at USF. Presently, about 12,000 students at USF use My Reviewers features, which include document markup, flexible peer review assignments, customizable rubrics, customizable projects, team projects, peer review for team projects, reporting tools and five e-books. In addition, faculty at MIT, Dartmouth, Penn, and NCSU are using My Reviewers in undergraduate STEM courses to research the efficacy of peer review, thanks to NSF Prime grant #1544239. These researchers have published extensively on the benefits of our development, especially ways it facilitates evidence-based curriculum changes: we have explored ways My Reviewers can be leveraged to inform evidence-based curriculum changes [9], researched the development and transfer of reasoning capabilities [10], and compared students' reviews to instructors' reviews [11]. Thanks to this user base, IRB approvals, and end-user license, we now have proprietary access to a large corpus of intermediate and final drafts that includes major course projects along with instructors' comments to these texts.

Clearly, the graduate students who originally composed these research maps already had a great deal of experience with corpus methods, due to prior experiences with My Reviewers. These previous experiences with My Reviewers and corpus methods enable the graduate students to share their concerns, as well as concerns about use of data, which the first map will show.
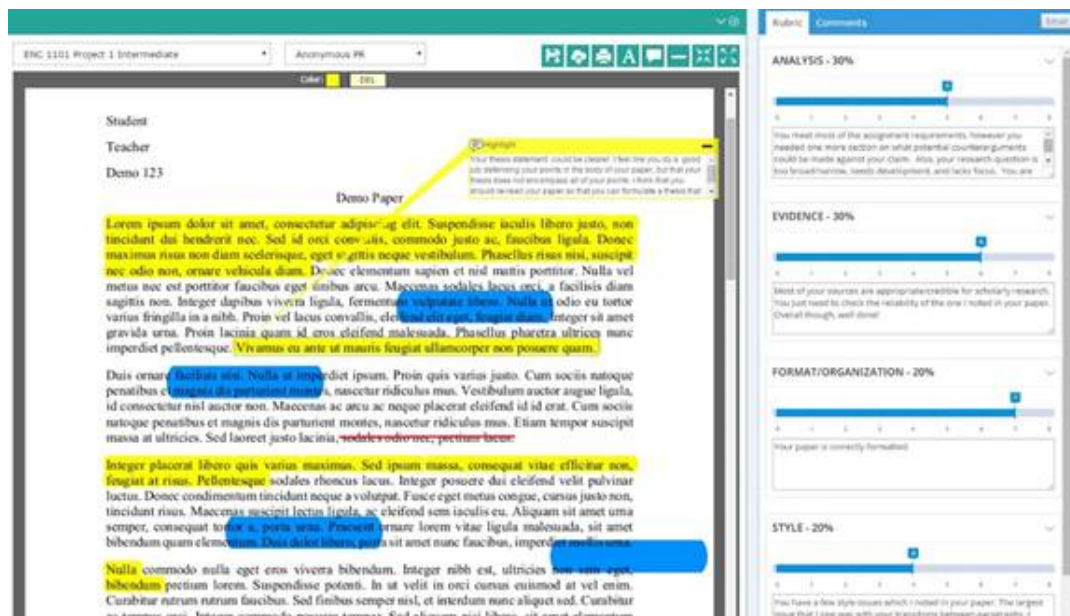

Figure 1: My Reviewers interface, site of research

4. FIRST MAP

The first map attempts to show the data WA can collect. While WA does collect data about identity, work, and culture, this map has several limitations. The first, and most obvious, is that the graduate students created a map in the shape of an eye, which highlights their perception of WA as primarily the work of surveillance. The graduate students and instructors who envisioned this map had trouble moving past their concerns about privacy, surveillance, and ethics.

These concerns are also voiced by such entities as the National Council of Teachers of English [3], who write that Machine Scoring is based on "narrowly conceived, artificial" tests that "subvert attention to other purposes and varieties of writing development in the classroom". Yet as graduate students examined work using corpus methods [12], and met with instructors using LSA and NLP to inform WA, they began to move beyond their initial construction of the research area as a field of surveillance and judgment.

To further mitigate the barriers posed by disciplinarity, Dr. Moxley presented his work with NSF I-Corps. This project involved interviewing over 100 faculty, students, and writing program administrators across the United States. These individuals' opinions toward WA made it clear that a second attempt at mapping the research field should be made.

The first map does not delineate a pathway through the middle ground. Instead, it only accounts for the many possibilities of WA, and from a humanities-centric perspective, many of these possibilities can seem negative. The second map immediately rectifies this problem with the bottom question of "Who?" To achieve an interdisciplinary perspective, researchers and practitioners must believe that WA can serve not just academia, but business and cultural interests as well.

The revised map presented by the graduate students showed addressed these needs by building a map of WA upon the idea of benefit to interdisciplinary domains and their needs. From there, researchers can select an orientation, gather data into a corpus, and implement that corpus toward a specific research goal. This map allows for a more actionable definition of WA's goals and purpose.
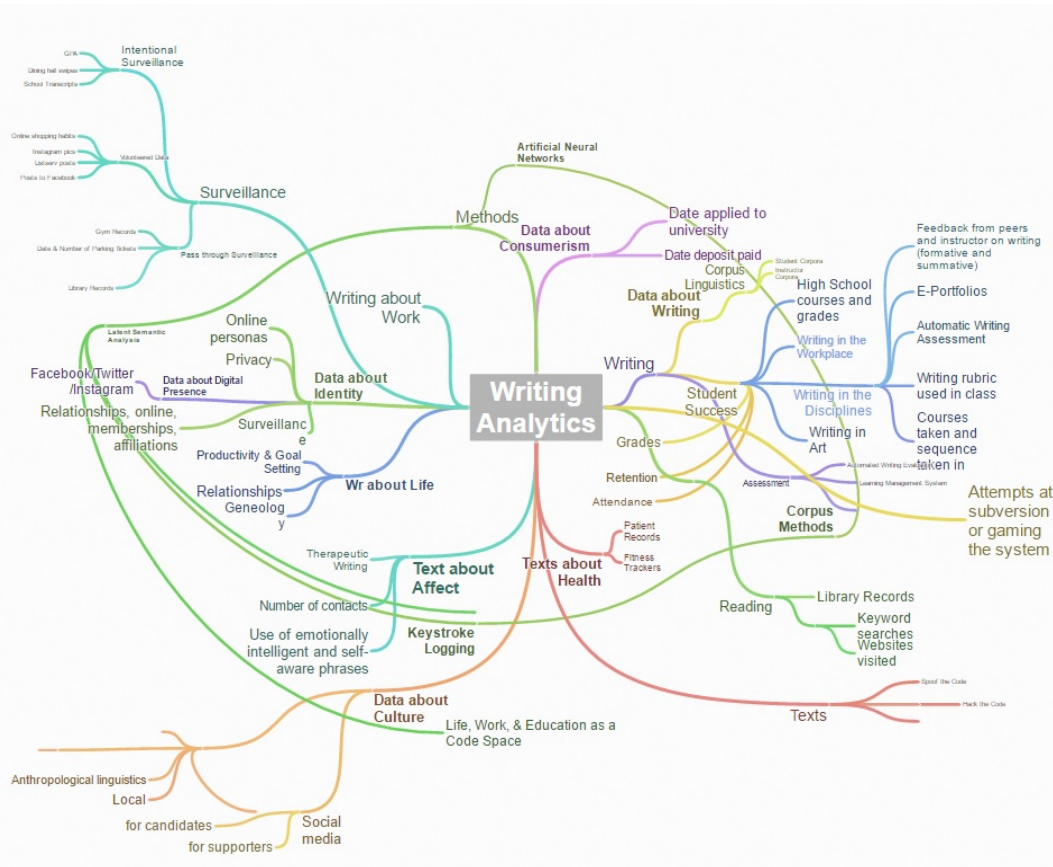

Figure 2: First attempt at mapping Writing Analytics

5. CONCLUSIONS

This use of WA to inform interdisciplinary research shows that re-conceptualizing the field helps move beyond usability discussions and stakeholder theory into questions of how WA can be used to benefit multiple domains. While usability, stakeholders, and ethical considerations are still immensely important conversations in this field, as one can see from the first map, they can create associations that distract the research or practitioner from creating a viable plan of completing research.

As shown by the graduate students in this seminar, a new visualization of the research questions in WA helped ground this new interdisciplinary perspective. By using the concept of mapping as more than an organizational tool, researchers can more effectively move through the phases of a data-driven research project.

By presenting this study, we hope to encompass and extend current attitudes and definitions of WA. By reformulating WA to perform interdisciplinary educational research, we can:

1. Structure opportunities for students to learn

2. Understand the cognitive, interpersonal, and intrapersonal constructs, as they emerge within sociocognitive and sociocultural settings, that enable students to recognize and respond to feedback

3. Gain actionable information about what practices will help students to become better writers in academic and workplace settings

When WA is reconfigured to embrace student learning, we can see that the efforts of researchers and practitioners change the learning space. With interdisciplinary collaboration, we can mediate the constructs that underlie WA as a research field.
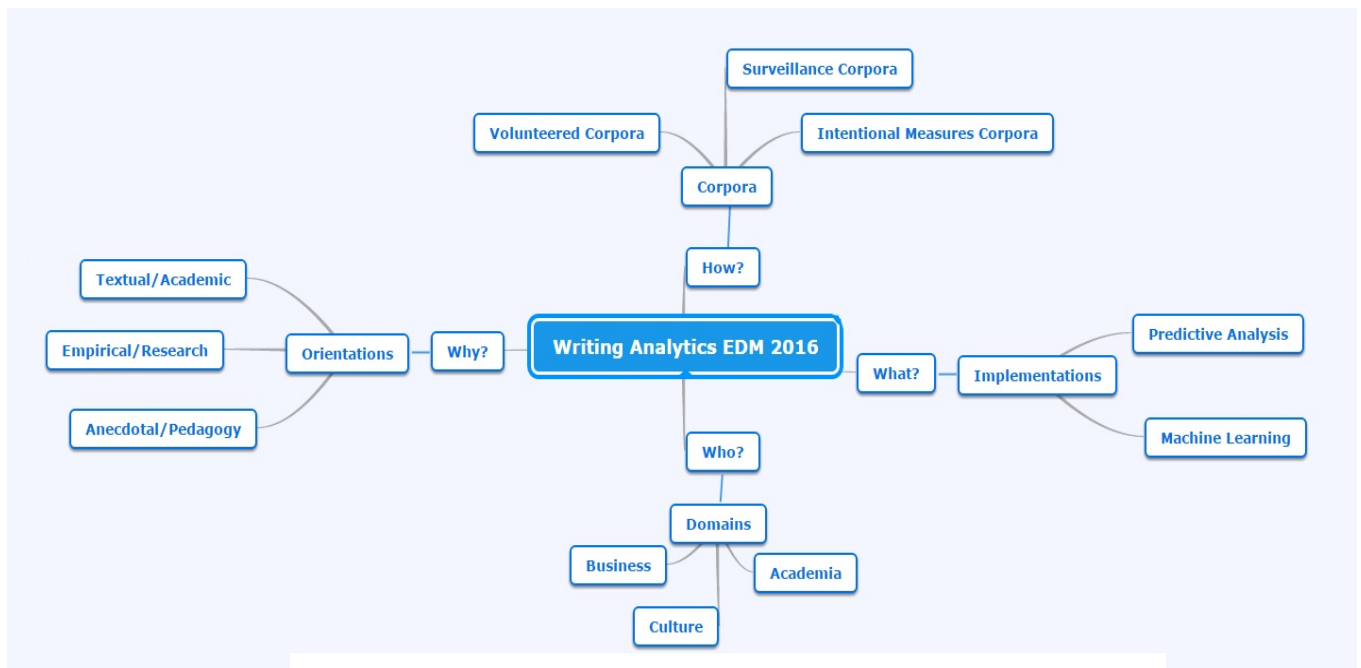

Figure 3: Reformulating WA to perform interdisciplinary educational research

6. REFERENCES

[1] Buckingham Shum, S., Knight, S., McNamara, D., Allen, L., Bektik, D., and Crossley, S. 2016. Critical perspectives on writing analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (LAK '16). ACM, New York, NY,USA,481-483.
DOI=http://dx.doi.org/10.1145/2883851.2883854

[2] Knight, S., and K. Littleton (2015). Discourse-centric learning analytics: mapping the terrain. *Journal of Learning Analytics*, 21, 185-209.

[3] NCTE (2013). Machine scoring fails the test. *NCTE Position Statement*.
http://www.ncte.org/positions/statements/machine_scoring

[4] Allen, L., Snow, E. and McNamara, D.S. 2015 Are you reading my mind? Modeling students' reading comprehension skills with Natural Language Processing techniques. In: *5th International Learning Analytics and Knowledge Conference*, (Edinburgh, UK, April 25-29 2016). LAK16. ACM, New York, NY. DOI=http://dx.doi.org/10.1145/2883851.2883955

[6] Lang, S. and C. Baehr. Data mining: A hybrid methodology for complex and dynamic research. *College Composition and Communication* 64, 1, 172-194.

[7] Elliot, N. and A. Klobucar (2013). Automated essay evaluation and the teaching of writing. Ed. M.D. Shermis and J. Burstein. *Handbook of automated essay evaluation: Current applications and new directions*. London: Routledge, 26-35.

[8] Fenwick, T. and R. Edwards (2016). Exploring the impact of digital technologies on professional responsibilities and education. *European Educational Research Journal* 15, 1, 117-131.

[9] Langbehn, K., McIntyre, M. and Moxley, J. (2013). Re-mediating writing program assessment. Ed. H.A McKee and D.N. DeVoss, *Digital Writing Assessment & Evaluation*. Logan, UT: Computers and Composition Digital Press. http://ccdigitalpress.org/dwae/13_langbehn.html

[10] Tackitt, A., Moxley, J., and Eubanks, D. (2015). Signifying scores: Instructor rating as an assessment measure. Manuscript submitted for publication.

[11] Moxley, J. and D. Eubanks. (2015). On keeping score: Instructors' vs. students' rubric ratings of 46,689 essays. Manuscript submitted for publication.

[12] Aull, L. (2015). *First-year university writing: A corpus-based study with implications for pedagogy*. London, UK: Palgrave MacMillan

7. NOTE

Joseph Moxley wishes to disclose a potential conflict of interest: while the My Reviewers software is not commercially available, it may become commercially available in the future. Because the data collection methods used in this study demonstrate the viability of My Reviewers, this research study may enhance the commercial value of My Reviewers. Ultimately, USF owns My Reviewers; however, Moxley possesses the rights to license My Reviewers. Given this potential conflict, Professor Moxley has filed the necessary USF conflict of interest paperwork. The Conflict of Interest Committee at USF has developed a management plan with which Dr. Moxley has complied prior to submitting this and similar research.