IneqDetect: A Visual Analytics System to Detect Conversational Inequality and Support Reflection during Active Learning

Stephen MacNeil, Kyle Kiefer, Brian Thompson, Dev Takle, Celine Latulipe
The University of North Carolina at Charlotte
{smacnei2,kkiefer,bthomp57,dtakle,clatulipe}@uncc.edu

ABSTRACT

A series of recent studies have shed light on the existence of sociocultural inequities in collaborative learning environments. We present IneqDetect, a system which helps students reflect on the way that they communicate as a team. Conversations during collaborative learning activities are recorded using lapel microphones, processed to determine who spoke at a given time, and then visualized. The resulting dashboard visualization provides students with a timeline of when each student was speaking, a summary of how much they spoke, and an estimate of how equitable the conversation was between team members. Students reflect on this information at the end of the class period to identify and address issues, such as conversational inequality, within their groups. IneqDetect was deployed across four CS active learning classrooms. IneqDetect led students to discuss group dynamics, change their behaviors, and gain insights about themselves and their team. However, conversational equity within groups did not improve.

KEYWORDS

collaboration; group dynamics; reflective learning; active learning

ACM Reference format:
Stephen MacNeil, Kyle Kiefer, Brian Thompson, Dev Takle, Celine Latulipe. 2019. IneqDetect: A Visual Analytics System to Detect Conversational Inequality and Support Reflection during Active Learning. In *Proceedings of ACM Global Computing Education Conference 2019, Chengdu,Sichuan, China, May 17–19, 2019 (CompEd '19),* 7 pages.
https://doi.org/10.1145/3300115.3309528

1 INTRODUCTION

Equity and inclusion are emerging as issues of critical importance in CS education research. A lack of diversity both in industry and in CS graduates has prompted many calls for broadening participation. In parallel, computing is increasingly being seen as a human literacy that should be shared by all [42]. Improving equity and inclusion in CS has been addressed with new pedagogical styles [20, 40], by changing the culture and environment through organizations and events [10, 30], or by improving the computing pipeline that leads to CS programs [7, 40]. In most of these cases, inclusion and equity are considered from a representational or structural perspective with an emphasis on race, gender, or intersectional aspects [9]. Prior research has focused mostly on improving retention [20, 29], measures of self-efficacy and motivation [4, 40], or academic performance [20, 34, 40] for under-represented groups. Improving equity and inclusivity has understandably received a lot of attention and will continue to be a focus, especially as intersectionality challenges us consider diversity more holistically and in greater detail.

In this paper, we consider another aspect of equity and inclusion that is recently gaining attention in CS education, sociocultural equity. Sociocultural inequities are disparities between students in terms of how they communicate and interact with each other socially. Sociocultural inequities were observed in pair programming [22] and in group discussions [36]. Personality types and power dynamics within teams can serve to elevate the voices of some students while silencing the voices of others. To achieve 'CS for All', these social aspects need to be considered because students who do not have a sense of belonging and community have trouble identifying with their chosen field [11]. Inequitable social learning environments may also further propagate existing negative stereotypes about CS, such as it being competitive, singularly-focused, asocial, and primarily male [21]. These problems become especially relevant as CS classrooms increasingly adopt active learning and team-based pedagogies. Existing research helps to understand some of the dynamics that occur within groups and teams. However, few tools exist to help students reflect on group dynamics.

IneqDetect is a system that helps students reflect on their group conversations that occur during active learning activities. IneqDetect uses lapel microphones to record students' conversations and uses signal processing to determine who is speaking during a given time interval. These speech segments, the total talk time per speaker, and a measure of conversational equality within the group are presented as a dashboard visualization at the end of class. Students reflect on the visualization to identify trends or areas for improvement. We found that IneqDetect led students to discuss group dynamics, change their behaviors, and gain insights about themselves and their team. Despite these benefits, IneqDetect did not improve conversational equality. We also contribute insights about how students perceive collaboration. We conclude with suggestions to improve *Research Support Tools* (RSTs) for collaboration.

2 BACKGROUND

We present the social theory and related work that inspired the creation of IneqDetect and guided its development. We provide a brief overview of reflection and describe how it can be both an individual and collaborative effort. Finally, we explain why reflection may be an appropriate vehicle for improving equity within groups that also maintains the autonomy of the group members.

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 this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CompEd '19, May 17–19, 2019, Chengdu,Sichuan, China
© 2019 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.
ACM ISBN 978-1-4503-6259-7/19/05…$15.00
https://doi.org/10.1145/3300115.3309528

2.1 Group and Team Dynamics

Understanding and evaluating small group collaboration in education is challenging for a variety of reasons. First, language consists of both verbal and non-verbal cues [12, 17]. Verbal cues can mostly be transcribed whereas non-verbal cues are harder to represent due to their variety and subtly. Similarly, verbal cues can be identified in recorded audio. However, non-verbal cues are harder to capture automatically due to video camera occlusion. Second, verbal cues can be used to convey information but also to perform actions [1, 33, 41]. These two types of verbal cues contribute to different parts of the conversation, such as what is being said and how the conversation is being regulated. Third, social interactions are heavily shaped by the task, situation, and team structure [25]. Considering task, situation, and structure is non-trivial. Joseph McGrath describes eight different settings and six different task typologies [25]. Due to all of this complexity, connecting any of these aspects to performance is challenging. Even anecdotally, a team may have great off-topic conversations which do not translate to task performance.

Given this complexity, researchers analyze many different aspects of conversations. A simple measure that is often used is the amount or distribution of talk time [22, 24, 36]. The total talk time for each speaker can also be used to compute the amount of equality present within a group as a Gini Coefficient [24, 36]. While the quantity of a conversation doesn't equate to quality, it is a good starting point for automation. Stankiewicz and Kulkarni argue that quantitative measures may also encourage students' self-disclosure better than semantic or qualitative measures [36].

IneqDetect combines these popular metrics of total talk time and the Gini Coefficient with a scrollable timeline showing when voices were detected. By providing students with this visualization dashboard to reflect on at the end of class, students can negotiate what these measures mean in their own context. This novel adaptation is a first step to determine what is most relevant for the discussants themselves. This differs from previous conversation analyses which attempted to operationalize and evaluate social theory. Additionally, IneqDetect is robust to classroom noise which enabled these studies to take place in real active learning classrooms.

2.2 Reflection

In CS education, reflection has taken many forms, including collaborative discussions [31], diaries and journaling [13, 15] or portfolios [5]. Each of these provide scaffolding to help students to develop their reflective practice. Reflection can also be scaffolded using reflection support tools [14, 23]. *Reflection Support Tools* (RSTs) provide additional data about student's learning experiences or a digital scaffold to structure students' reflections. A review of RSTs by Kim and Lee include examples as early as 1993 [18]. Many examples of RSTs exist, such as the Subtle Stone which helps students reflect on their affective states [2]. Another example is the student activity monitor (SAM) which supports both students and instructors by visualizing learner's actions [16]. In many cases, the goal of RSTs is to make the implicit explicit through visual representations. This increases awareness, and challenges students to compare their perception with additional data about an experience. Ideally, this results in critical reflection, where students think critically about their assumptions and biases related to an experience [26].

2.3 Visualizing Conversations

Conversations consist of a series of verbal and non-verbal cues occurring over a period of time. This time-series data can have micro and macro patterns as topics change and discussants become more or less involved. Visualized conversations provide *an overview* of the conversation and *details on demand* about specific interactions between discussants [35]. Visualizations also afford exploration and discoverability which can promote sense-making.

For visualizing a single audio track, TimeNotes uses zooming techniques to explore audio at multiple points in time while maintaining an overview of the audio signal [39]. The *Conversational Clock* uses a clock metaphor to show recorded audio patterns in real-time that emerge in co-located groups [3]. It features a circular table with more recent parts of the conversation radiating out from the center of the table. This highlights recent discussions while preserving the overall conversation. Although real-time visualizations provide continuous feedback, they also increase cognitive load. Participants echoed this idea in a study that compared subtle to overt real-time visual feedback during conversation by indicating that they preferred the subtle feedback [32]. In learning settings, where students are being challenged to think, communicate, and remember concepts, this additional cognitive load may be less desirable. To this end, students use IneqDetect at the end of class to reflect.

VizScribe is a visual analytics tool for analyzing verbal protocols during design sessions [8]. It presents raw data about who spoke when. This encourages designers to explore the data to find their own meaning. IneqDetect also presents raw data to encourage students to engage in sense-making and find their own meaning. We also present aggregated representations to help scaffold exploration.

In educational contexts, Toyoura et al. have explored using both video and audio to classify student interactions, including group work, lecture, private work, presentation, and movement [38]. They visualize, temporally, the type of interaction that each group is having based on clustering and classifying the video and audio data. This technique helps instructors to see which groups are interacting in ways that weren't intended by the instructor. While helpful for instructors, this approach was not intended to support students.

3 INEQDETECT SYSTEM

IneqDetect is a system that was developed to support students as they reflect on conversations that they have had within their group. Students' verbalizations during conversations are recorded and then visualized for students to reflect at the end of class. The visualization highlights conversational inequality, but it also provides raw data such as when it detected a speaker. As seen in Figure 1, the visualization consists of three parts: a barchart of overall talk time for each speaker, an equity score that represents how even the distribution of speaking was within the team as measured by the Gini Coefficient, and a timeline which shows sections where each speaker was determined to be speaking. Students review this visualization after working in groups. Stickers attached to devices help students identify themselves while preserving anonymity. This addresses previous concerns about privacy and encourages self-disclosure [36]. Justifications for many of our other design decisions have already been presented in the background sections.

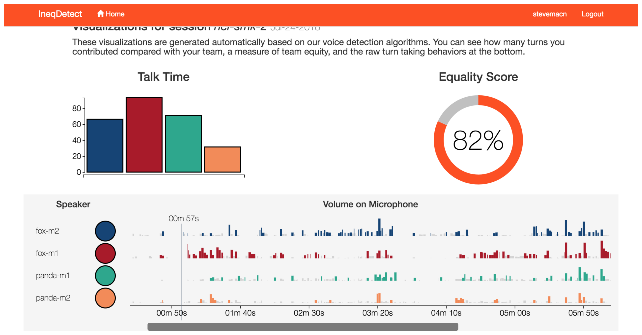

Figure 1: Visualization: A barchart summarizing each student's contribution (left), the equity score (right), and a timeline that displays when each student spoke (bottom).

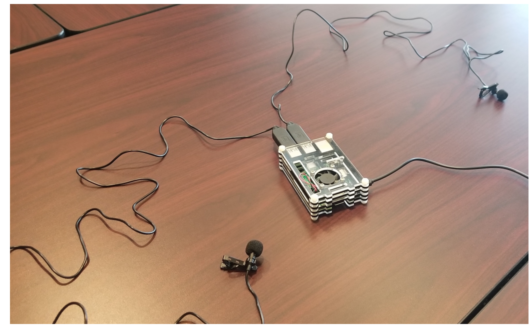

Figure 3: A Raspberry Pi Device connected to two lapel microphones to simultaneously record two speakers.

3.1 System Design and Speaker Recognition

The design of IneqDetect is shown in Figure 2. Students wear lapel microphones that are attached to Raspberry Pi devices, shown in Figure 3. Each device supported two microphones, recording two group members. In larger groups, multiple devices were used. An optional Twitter account can start and stop recordings if instructors want to use the devices without assistance. When a recording is ended, the audio is sent to the server to determine who spoke when. Students view the visualization in their web browsers.

Recording in classrooms is challenging because there is often a lot of background noise, which varies widely throughout the class period. Consequently, the audio was pre-processed before identifying the speakers. To denoise the audio signal, we used spectral whitening and a fast Fourier transform. The voice features are extracted from the audio signal as Mel-frequency cepstral coefficients (MFCC). As recommended [19], we used only the first 12 coefficients for voice activity detection (VAD). We computed the energy across these speech features during a given time interval which is a common energy-based approach for VAD [37]. The energy across all microphones was compared through a moving window and an adaptive threshold removed the lowest cluster using k-means clustering (elbow method). This effectively reduced the amount of cross-talk that was detected on the microphones. This last step was added after initial tests in classroom settings and it drastically improved our ability to distinguish between speakers even when speakers are very close (less than 3 feet apart) to each other.

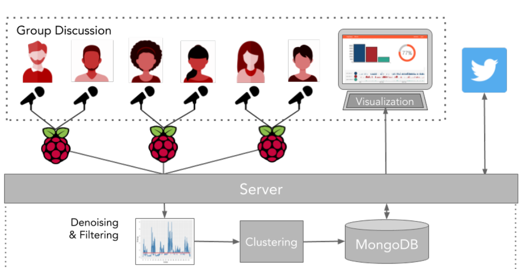

Figure 2: Raspberry pi devices are controlled by tweets and send recordings to the server for analysis.

4 IN-CLASS STUDIES

We used IneqDetect in four classes to evaluate its potential to improve equity and collaboration within teams. Consent was obtained in the first weeks of the course. For IneqDetect to be used by a group, all students in the group must have provided consent. If multiple groups were eligible to use IneqDetect then the groups were chosen at random but with a priority given to teams who had all students attend on the first day of the intervention where IneqDetect was used. Students from the groups that used IneqDetect were offered an opportunity to be interviewed about their experiences. Students received $5 each time they used IneqDetect and $15 for participating in the 30 minute interview.

4.1 Study Design and Protocol

We conducted a study that spanned four classes with two conditions: reflective writing and the IneqDetect intervention. Each class had 3-5 students per group. On the first day of class, students were introduced to IneqDetect and given a brief presentation about reflection. The presentation was tailored for each course to connect reflective practice to the course material or the instructor's intentions for reflection. The study began with a survey on the first day, four distinct reflection stages, and ended with an exit survey. In the first stage, students were given a reflective writing assignment that was related to high-level course concepts to engage them. In the second stage, students were asked to reflect on their teamwork at the end of class. In the third stage, the groups were placed into the two conditions. Some continued reflective writing and others used IneqDetect. In groups that were selected to use IneqDetect they followed the protocol displayed in Figure 4. In the fourth stage, all group members completed written reflections.

4.2 Study Context and Participants

The studies that we describe in this paper took place across four courses in Spring 2018 and both Summer Sessions I and II of 2018. The four courses were Introductory Programming II (CS1), System Integration (SI), Game Design and Development (GDD), and Human-Computer Interaction (HCI). These three courses each had opportunities for students to collaborate and work in groups. Our university (and our department in particular) has made efforts to encourage active learning across classrooms. The courses in which

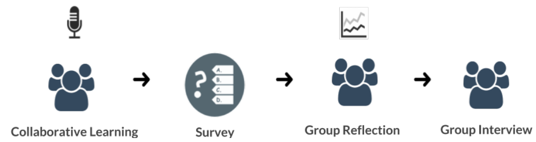

Figure 4: An overview of the in-class study.

we conducted our study were taught with a focus on active and collaborative learning experiences for the students.

The types of collaboration varied from discussion of clicker quiz questions to unstructured project work. Across classes, groups were formed randomly, and we didn't intervene with extant group formation practices. In every class except GDD, students worked in groups without formal roles. GDD had structured roles within teams. In each class, instructors rotated to help each group, but otherwise interactions were self-directed. Some groups were all-male, while others were mixed-gender. CS1 had two men and four women, all identified as white except one member who self-identified as African-American. The GDD group consisted of four men, all identified as white except one member who self-identified as Asian. The HCI group consisted of four men, all identified as white. A few groups in the SI class used IneqDetect, but none volunteered to be interviewed. The SI class was a graduate class with mostly international students with a roughly even gender distribution.

4.3 Data Collection

4.3.1 Survey Data. We collected survey data at the beginning and end of the semester. We also collected survey data each time students used the IneqDetect system, as shown in Figure 4. This data included Likert scale questions with questions such as "I find it useful to reflect on my learning process" and "IneqDetect helped me reflect on my learning." We analyzed the survey data using either inferential or descriptive statistics, depending on the amount of data and the need to generalize. Because Likert data is ordinal we applied Spearman's Rho (*Rs*). Spearman's Rank Correlation test is "appropriate when one or both variables are skewed or ordinal and it is robust when extreme values are present" [27].

4.3.2 Individual Interview Data. To gather more fine-grained information about students' experiences with IneqDetect we also conducted a series of interviews with participants in groups that used IneqDetect. Students were compensated $15 and interviews lasted about 30 minutes on average. We interviewed 6 students (5 men, 1 woman), across four teams, with at least one student from each of the four courses. For small project interviews, 6-10 people is the recommended number of participants to interview [6].

To analyze the interview data we had two independent coders review the interview recordings and take notes about interesting quotes and themes that emerged. After coding each recording, the coders met to discuss their codes. The codes that were common were kept and codes that didn't match were negotiated (followed by recoding), or they were removed if consensus wasn't reached. After coding all of the interviews, themes that appeared across sessions were merged and corresponding quotes were categorized into those themes. Quotes in the results section were taken from this analysis.

4.3.3 IneqDetect Data and Equity Scores. The visualization features two summative views of equity within the group. The first view is a bar chart which shows how much each group member spoke as aggregated by the total time that they were detected to be speaking. The second view was an equity measure based on the Gini Coefficient. The Gini Coefficient is a measure of equality that is often used in comparing economic data. In addition, we also elicited students' estimates of their own turn-taking behaviors after each learning activity in which they used IneqDetect.

5 RESULTS

To evaluate the effectiveness of IneqDetect and to evaluate our hypotheses, we conducted the study outlined above. We analyzed data generated during the study which included equity scores, survey data, and a thematic analysis of the transcribed interview recordings. This data provided some insights about how IneqDetect influenced equity and its ability to support reflection.

5.1 Triangulating Survey and Interview Data

Given the small sample size for this study, we present our results along with information from the interviews to provide some context and to provide additional evidence for our observations. Quotes used through this section were identified during the coding sessions and chosen to provide support where necessary. These quotes were not obtained after doing our statistical analysis on the survey data. While not completely unbiased, this was done to reduce the possibility of cherry-picking evidence to support our survey results.

5.1.1 Estimating Turn-Taking and Accuracy. We hypothesized that IneqDetect would improve conversational equality by challenging students' perspectives of how much they spoke. Students in our study were not able to estimate the relative amount of talk time they contributed to the conversation when compared to their peers. Spearman's Rho indicated no significant relationship between both variables. In the interviews, all but one participant indicated that they were surprised by the results of the visualization. GD-1 explained that they were "shocked at first. I didn't know that I talked that much". HCI-3 mentioned that they were "surprised ... [that] I think I talk a lot more than I do." HCI-3 described initial surprise followed by a reflection about how much he spoke, "when the equality score came up I was surprised at first how low it was [for me] but then I was like that's about right because I'm never like one of the super talkative ones." CS-1 was least surprised by the results but still indicated that on some days the results were unexpected "the amount of talking within each day you know sometimes I actually thought I was going to be an average speaker but then sometimes I notice sometimes, some days I'm talking more or less."

We didn't ask participants about how they interpreted the discrepancies between their expectations and the results as detected by IneqDetect. However, GD-2 gave one possible explanation, "There are instances where I thought I talked a lot more than others; essentially because of the high you get when you're talking, more so when you're leading the conversation." Another discrepancy is rooted in how students remember their interactions. Students appeared to prioritize some types of communication, explaining that higher quality contributions were perceived as being greater quantitatively as well.

5.1.2 Perceptions of Accuracy. As seen from the comments, students appear to estimate and value conversations differently. We've also seen initial evidence that students' estimates of conversational involvement are inaccurate. Some students described an awareness that their perception was not always accurate, and put forward interesting ideas about what accuracy might mean from a discursive standpoint. Overall, students described IneqDetect as accurate with HCI-3 describing the system as being "90-95%" accurate. GD-2 saying it was "100% accurate" at distinguishing who spoke but that "30% was irrelevant" and related to "jokes and side conversation."

Other students didn't provide an estimate but described how they knew that it was accurate. HCI-2 described its accuracy by his ability to identify events from the class period saying he was "pretty confident. That's where we all paused and we were working. That's where we were yelling about Star Wars." He went on to say, "Pretty accurate ... some points I couldn't place exactly." CS-1 explained that consistency was his way of evaluating IneqDetect's accuracy, "Because the result was consistent, constant results made me know."

5.1.3 Roles: Leaders and Non-Leaders. Across the classes in the study, only one group had defined roles provided by the instructor. Despite the lack of structured team work, there was evidence that students within groups defined roles for themselves. These roles were not always agreed upon explicitly, or even perceived in the same way by members within the group. The main roles that emerged were leadership roles. We asked students about their perception of themselves as leaders within the group. We found that the higher that students ranked themselves as a leader on a Likert scale, the more they estimated that they spoke ($Rs = 0.32$, $p < 0.05$) and the more they did speak as measured by IneqDetect ($Rs = 0.40$, $p < 0.05$). Furthermore, these self-perceived leaders strongly agreed with the statement that they spoke more than they would have liked to ($Rs = 0.55$, $p < 0.05$).

It is unclear from the survey whether these students took on the leadership role reluctantly and wanted more support from their team-members, which is why they said that they spoke more than they wanted to. Alternatively, as a leader, they may have wanted to speak less to ensure that the voices of others within the group were heard. Our interview data speaks to this aspect, GD-1 described his reluctant acceptance of the leadership role and he summarized his performance negatively saying that he didn't have enough experience with the course topics to lead the group, "As a leader, I didn't do a good job of assigning roles". His leadership role was assigned, though more generally he does prefer the role of leader, "I never say no to a leadership position." It is also interesting to note different types of leaders appeared in our groups. HCI-3 explained how two members of the group acted as leaders "HCI-2 talked most on hardworking days, HCI-1 talked more joking."

5.1.4 Motivation, Focus, and Behavioral Change. Students associated a variety of benefits with using IneqDetect. Most students cited general benefits, such as improved motivation and focus, but others provided very detailed accounts of how their teams changed over time. CS-1 talked about how IneqDetect helped him stay focused on the course material, "It motivated me to talk about the topic at hand." CS-2 also explained that it was "keeping me more focused." CS-2 indicated that students did more to explain ideas to each other so as to extend the amount they contributed. CS-2 said that previously they would just tell each other the answer. She speculated that this may have led peers to do the prep-work, "one or two of them might have even read the book even."

GD-2 indicated that wearing the lapel microphones "gave legitimacy, it made it all feel so real." He thought that IneqDetect "turned this into a fun activity to challenge ourselves to talk more." While fun, he also described it as "strangely competitive." He was in a group with four male students and described it thusly, "take four dudes give them all a microphone and you're going to find a competition." While the competition might be motivational he also described it as a potential stressor, "It was motivational and a kind of worry." In the same group, GD-1 did not experience any competition and described his experiment to get a group member more involved, saying "Our artist/designer didn't get as much time as I would like ... [and so] after the first project, I changed gears from leadership." He suggested that he took a back seat to give that partner more time "As you saw in the second time we recorded ... I took a backseat." He perceived his experiment as successful. We corroborated his perception using IneqDetect - his contributions did go down and that person's contributions went up.

An interesting change occurred in the HCI group. Initially, HCI-1 was the primary leader of the group. Both students interviewed described him as someone who cracked jokes and derailed the conversation away from the course topics. They enjoyed that about him, but they described how IneqDetect led to some changes. After reviewing the visualization, HCI-3 said his group observed that "HCI-2 talked most on hardworking days, HCI-1 talked more joking." After the group noticed this trend, HCI-3 said, "The group like unanimously decided and HCI-2 became the main worker ... HCI-2 took [the] leadership [role]", and as a result, "HCI-1 stopped talking." HCI-3 said this was motivational for him in his career and although he hadn't previously considered himself a leader, he wanted to "... take more of the [leadership] role HCI-2 did." When asking HCI-2 about whether he observed any changes within the group, he said "After? Honestly, no." It was interesting that both HCI-2 and HCI-3 had such different views of the same experience. The changes described by HCI-3 were also detected by IneqDetect with HCI-2 speaking more after the first session and HCI-1 speaking much less.

5.1.5 Equity and Changes in Equity. Based on analysis of the teams using data from IneqDetect, the inequality detected across teams with low initial inequality went up (0.12 => 0.38, 0.04 => 0.16 => 0.25, and 0.18 => 0.34) and the inequality in the one team with high initial inequality went down (0.60=>0.28). Furthermore, we observed high variability in terms of equality calculated for each team for the first time that they used IneqDetect ($n = 8$, $mean = 0.25$, $sd = 0.21$). IneqDetect was designed to improve equality within groups and so these results were surprising. These results were consistent with estimates made during group observations.

Equally surprising, equity was also not a theme that emerged when analyzing the interview data, which makes it harder to interpret these results. One possible explanation is that IneqDetect might be more helpful in teams that have a lot of inequality. HCI-2 suggested that easier classes don't need IneqDetect as much as in harder classes, saying "we were able to joke around and have a good time and still do the work that we need to do ... but there are some classes that require a lot more focus ... if I was sitting at a C or a D ... in that scenario it would be very useful for saying, dude, I need to stay on subject more cause I'm going to fail Calc 2."

5.1.6 IneqDetect as a Reflection Support Tool. The perceptions of IneqDetect as a RST were mixed. Seventy-one percent of students that used IneqDetect preferred it to reflective writing. GD-1 indicated that they preferred the structure present in reflective writing, "written reflections were a little better [than IneqDetect], they were specific." He suggested that IneqDetect could be improved "if you're given a set of tasks" adding that "reflection specifics help." CS-2 said that she received benefits from IneqDetect but explained that "writing might have had a similar effect" on improving the equity within their group. In regards to their perceived enjoyment, students that used IneqDetect indicated that they enjoyed IneqDetect more than reflective writing (4.3 > 3.7, with 5=Strongly Agree). Additionally, when comparing across both IneqDetect and non-IneqDetect users, students that used IneqDetect reported higher agreement (ID = 3.43, NO-ID = 3.13) for the statement "I enjoyed doing reflections." Neither of these two differences were statistically significant.

Qualitatively, some students discussed IneqDetect's role as a RST. HCI-2 indicated that IneqDetect resulted in "not a huge reflection. Made me think honestly that we should focus more in class." HCI-2 indicated that he wasn't a very reflective person to begin with, for example "I'm usually not overly mindful of what I'm saying" but also that the class was easy and that IneqDetect was would be more useful in harder class such as 'Calculus'. Some students were more positive about the reflective support that IneqDetect afforded for students. GD-2 described IneqDetect as "great tool to give feedback, the data it provided was great ... forced you to reflect on things."

6 DISCUSSION

Understanding the social dynamics of student groups is challenging because team formation, situation, and task each have a strong affect on collaboration [25]. We introduced IneqDetect into four CS courses, but did not make attempts to change the existing structure of those classrooms. Our data featured unstructured groups, a team with defined roles, and a mix of ages, genders, and races. This in-the-wild approach was adopted to capture the varied ways that collaboration happens in real classrooms. This makes generalization difficult, but our goal for this research was to take a first step in understanding how students interact with reflection support tools (RSTs) designed to help students reflect about collaboration.

IneqDetect was designed to make students more aware of their collaborations and inequalities present within their group. We expected that through reflection and positive reactivity [28], students would start conversing more equitably. While this didn't happen in the few groups that we studied, we did observe benefits for students. Students described instances of behavioral change that occurred within their groups. These changes included increased motivation, a renewed focus on course topics, and intentions to either take on leadership roles, speak more, or be more inclusive of team members. Not all students observed these changes and students in the same group often described very different perceptions of the same shared experiences. These profound differences in subjective interpretations were surprising.

Our results also suggest that students were unable to accurately estimate how much they contributed to a group conversation. Some students indicated an awareness that their perception may not reflect reality. They suggested that their perception changed depending on whether they were speaking or listening. Others suggested that high-quality contributions appeared to constitute more talking time. Others were confident that their estimates were accurate. Different students also ascribed value to different aspects of the conversation. In a single group, one student described silence as when the work happened, whereas another student equated more conversation to more productivity. Understanding students' perceptions of collaboration is an interesting area for future work.

Finally, we observed mixed, but mostly positive, results about IneqDetect's effectiveness as a RST. A few students created their own experiments based on the data and explored hypotheses in subsequent weeks. However, some students also indicated that they didn't know what patterns to look for or how to use the data. In particular, a few students struggled to interpret conversational inequality within their groups. Students requested additional scaffolding, such as specific tasks to perform. To improve IneqDetect as an RST, students asked for more information about conversational quality. This could be extracted from linguistic, paralinguistic, and gestural data to display conversation topics, the focus of each student's attention, and their affective states. Despite these challenges, IneqDetect improved motivation, focus, and awareness. Students also described intention to make changes, such as to take on leadership roles, participate more, or be more inclusive of others.

7 CONCLUSION AND FUTURE WORK

We present IneqDetect, a new RST for supporting reflection about collaboration and conversational equity within groups. Through our in-the-wild study across four classes, we found that students struggle to reliably estimate their contribution to the group. In addition, group members often have different, even conflicting, perceptions of the same experience and group dynamics. Despite these interesting challenges, students associated benefits with using IneqDetect and described instances of behavioral change that occurred within their groups. Results for IneqDetect as a RST were mixed. The majority of students preferred it to written reflections (71%), some students explicitly described how it supported reflection, but others said it would only be helpful in harder classes.

In this initial probe, we provided students with a mixture of raw and summarized data to reflect on. This information was based on the quantity of conversation, but students requested more information about quality of conversations. For future work, we have already extended IneqDetect to automatically transcribe the audio from conversations. From these transcripts, we plan to use natural language processing (NLP) to extract topics and sentiment from the conversations. Paralinguistic cues, such as pitch and resonance, and non-verbal communication, such as gestures and gaze, could eventually be included to track attention and emotion. In addition to supplementing the data presented, we can also provide more structure. For example, we didn't tell students how to reflect on the visualization or provide tasks. This open-endedness allowed some students to form and test hypotheses about group dynamics, but most students had difficulty knowing how to reflect on their collaboration. In the future, we plan to provide students tasks to complete while reviewing the visualization.

REFERENCES

[1] John Langshaw Austin. 1975. *How to do things with words*. Vol. 88. Oxford university press.

[2] Madeline Balaam, Geraldine Fitzpatrick, Judith Good, and Rosemary Luckin. 2010. Exploring affective technologies for the classroom with the subtle stone. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1623–1632.

[3] Tony Bergstrom and Karrie Karahalios. 2007. Conversation Clock: Visualizing audio patterns in co-located groups. In *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on*. IEEE, 78–78.

[4] Sylvia Beyer. 2014. Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades. *Computer Science Education* 24, 2-3 (2014), 153–192.

[5] Madhumita Bhattacharya and Maggie Hartnett. 2007. E-portfolio assessment in higher education. In *Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual*. IEEE, T1G–19.

[6] Virginia Braun and Victoria Clarke. 2013. *Successful qualitative research: A practical guide for beginners*. sage.

[7] Amy Bruckman, Maureen Biggers, Barbara Ericson, Tom McKlin, Jill Dimond, Betsy DiSalvo, Mike Hewner, Lijun Ni, and Sarita Yardi. 2009. Georgia computes!: Improving the computing education pipeline. In *ACM SIGCSE Bulletin*, Vol. 41. ACM, 86–90.

[8] Senthil Chandrasegaran, Sriram Karthik Badam, Lorraine Kisselburgh, Kylie Peppler, Niklas Elmqvist, and Karthik Ramani. 2017. VizScribe: A visual analytics approach to understand designer behavior. *International Journal of Human-Computer Studies* 100 (2017), 66–80.

[9] J McGrath Cohoon and William Aspray. 2006. *Women and information technology: Research on underrepresentation*. Vol. 1. The MIT Press.

[10] Teresa Dahlberg, Tiffany Barnes, Kim Buch, and Audrey Rorrer. 2011. The STARS Alliance: Viable Strategies for Broadening Participation in Computing. *Trans. Comput. Educ.* 11, 3, Article 18 (Oct. 2011), 25 pages. https://doi.org/10.1145/2037276.2037282

[11] Regina Deil-Amen. 2011. Socio-academic integrative moments: Rethinking academic and social integration among two-year college students in career-related programs. *The Journal of Higher Education* 82, 1 (2011), 54–91.

[12] Paul Drew and John Heritage. 2006. *Conversation analysis*. Vol. 1. Sage London.

[13] Alan Fekete, Judy Kay, Jeff Kingston, and Kapila Wimalaratne. 2000. Supporting Reflection in Introductory Computer Science. In *Proceedings of the Thirty-first SIGCSE Technical Symposium on Computer Science Education (SIGCSE '00)*. ACM, New York, NY, USA, 144–148. https://doi.org/10.1145/330908.331844

[14] Angela Fessl, Oliver Blunk, Michael Prilla, and Viktoria Pammer. 2017. The known universe of reflection guidance: a literature review. *International Journal of Technology Enhanced Learning* 9, 2-3 (2017), 103–125.

[15] Susan E. George. 2002. *Learning and the Reflective Journal in Computer Science*. Vol. 24. IEEE Computer Society Press, Los Alamitos, CA, USA. 77–86 pages. https://doi.org/10.1145/563857.563811

[16] Sten Govaerts, Katrien Verbert, Erik Duval, and Abelardo Pardo. 2012. The Student Activity Meter for Awareness and Self-reflection. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. ACM, New York, NY, USA, 869–884. https://doi.org/10.1145/2212776.2212860

[17] Robert A Hinde. 1972. *Non-verbal communication*. Cambridge University Press.

[18] Dongsik Kim and Seunghee Lee. 2002. Designing collaborative reflection supporting tools in e-project-based learning environments. *Journal of Interactive Learning Research* 13, 4 (2002), 375–392.

[19] Tomi Kinnunen, Evgenia Chernenko, Marko Tuononen, Pasi Fränti, and Haizhou Li. 2007. Voice activity detection using MFCC features and support vector machine. In *Int. Conf. on Speech and Computer (SPECOM07), Moscow, Russia*, Vol. 2. 556–561.

[20] Celine Latulipe, Stephen MacNeil, and Brian Thompson. 2018. Evolving a Data Structures Class Toward Inclusive Success. In *Frontiers in Education Conference (FIE 2018)*. IEEE, 1–5.

[21] Colleen M Lewis, Ruth E Anderson, and Ken Yasuhara. 2016. I Don't Code All Day: Fitting in Computer Science When the Stereotypes Don't Fit. In *Proceedings of the 2016 ACM Conference on International Computing Education Research*. ACM, 23–32.

[22] Colleen M. Lewis and Niral Shah. 2015. How Equity and Inequity Can Emerge in Pair Programming. In *Proceedings of the Eleventh Annual International Conference on International Computing Education Research (ICER '15)*. ACM, New York, NY, USA, 41–50. https://doi.org/10.1145/2787622.2787716

[23] Stephen MacNeil. 2017. Tools to Support Data-driven Reflective Learning. In *Proceedings of the 2017 ACM Conference on International Computing Education Research (ICER '17)*. ACM, New York, NY, USA, 299–300. https://doi.org/10.1145/3105726.3105745

[24] Roberto Martinez, Judy Kay, James R Wallace, and Kalina Yacef. 2011. Modelling symmetry of activity as an indicator of collocated group collaboration. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 207–218.

[25] Joseph Edward McGrath. 1984. *Groups: Interaction and performance*. Vol. 14. Prentice-Hall Englewood Cliffs, NJ.

[26] Jack Mezirow. 1998. On critical reflection. *Adult education quarterly* 48, 3 (1998), 185–198.

[27] Mavuto M Mukaka. 2012. A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal* 24, 3 (2012), 69–71.

[28] Rosemery O Nelson and Steven C Hayes. 1981. Theoretical explanations for reactivity in self-monitoring. *Behavior Modification* 5, 1 (1981), 3–14.

[29] Tia Newhall, Lisa Meeden, Andrew Danner, Ameet Soni, Frances Ruiz, and Richard Wicentowski. 2014. A support program for introductory CS courses that improves student performance and retains students from underrepresented groups. In *Proceedings of the 45th ACM technical symposium on Computer science education*. ACM, 433–438.

[30] Johanna Okerlund, Madison Dunaway, Celine Latulipe, David Wilson, and Eric Paulos. 2018. Statement Making: A Maker Fashion Show Foregrounding Feminism, Gender, and Transdisciplinarity. In *Proceedings of the 2018 Designing Interactive Systems Conference*. ACM, 187–199.

[31] Johanna Pirker, Maria Riffnaller-Schiefer, and Christian Gütl. 2014. Motivational Active Learning: Engaging University Students in Computer Science Education. In *Proceedings of the 2014 Conference on Innovation & Technology in Computer Science Education (ITiCSE '14)*. ACM, New York, NY, USA, 297–302. https://doi.org/10.1145/2591708.2591750

[32] Gianluca Schiavo, Alessandro Cappelletti, Eleonora Mencarini, Oliviero Stock, and Massimo Zancanaro. 2014. Overt or subtle? Supporting group conversations with automatically targeted directives. In *Proceedings of the 19th international conference on Intelligent User Interfaces*. ACM, 225–234.

[33] John R Searle and John Rogers Searle. 1969. *Speech acts: An essay in the philosophy of language*. Vol. 626. Cambridge university press.

[34] Robert M Sellers, Tabbye M Chavous, and Deanna Y Cooke. 1998. Racial ideology and racial centrality as predictors of African American college students' academic performance. *Journal of Black Psychology* 24, 1 (1998), 8–27.

[35] Ben Shneiderman. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages (VL '96)*. IEEE Computer Society, Washington, DC, USA, 336–. http://dl.acm.org/citation.cfm?id=832277.834354

[36] Adam Stankiewicz and Chinmay Kulkarni. 2016. $1 Conversational Turn Detector: Measuring How Video Conversations Affect Student Learning in Online Classes. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale (L@S '16)*. ACM, New York, NY, USA, 81–88. https://doi.org/10.1145/2876034.2876048

[37] Rong Tong, Bin Ma, Kong-Aik Lee, Changhuai You, Donglai Zhu, Tomi Kinnunen, Hanwu Sun, Minghui Dong, Eng Siong Chng, and Haizhou Li. 2006. The IIR NIST 2006 Speaker Recognition System: Fusion of Acoustic and Tokenization Features. In *presentation in 5th Int. Symp. on Chinese Spoken Language Processing, ISCSLP*.

[38] Masahiro Toyoura, Mayato Sakaguchi, Xiaoyang Mao, Masanori Hanawa, and Masayuki Murakami. 2016. Visualizing the lesson process in active learning classes. In *Frontiers in Education Conference (FIE), 2016 IEEE*. IEEE, 1–8.

[39] James Walker, Rita Borgo, and Mark W Jones. 2016. TimeNotes: a study on effective chart visualization and interaction techniques for time-series data. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 549–558.

[40] David C Webb, Alexander Repenning, and Kyu Han Koh. 2012. Toward an emergent theory of broadening participation in computer science education. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education*. ACM, 173–178.

[41] Victor H Yngve. 1970. On getting a word in edgewise. In *Chicago Linguistics Society, 6th Meeting, 1970*. 567–578.

[42] Madeline Zug, Hanna Hoffman, Forest Kobayashi, Miles President, and Zachary Dodds. 2018. CS for All Academic Identities. *J. Comput. Sci. Coll.* 33, 4 (April 2018), 130–137. http://dl.acm.org/citation.cfm?id=3199572.3199590