Toward Using Multi-Modal Learning Analytics to Support and Measure Collaboration in Co-Located Dyads

Emma L. Starr, Joseph M. Reilly, and Bertrand Schneider
ems519@mail.harvard.edu, josephreilly@g.harvard.edu, bertrand_schneider@gse.harvard.edu
Harvard Graduate School of Education

Abstract: This paper describes an empirical study where the productive interactions of small collaborative learning groups in response to two collaboration interventions were evaluated through traditional and multi-modal data collection methods. We asked 42 pairs (N=84) of participants to program a robot to solve a series of mazes. Participants had no prior programming experience, and we used a block-based environment with pre-made functions as well as video tutorials to scaffold the activity. We explored 2 interventions to support their collaboration: a real-time visualization of their verbal contribution and a short verbal explanation of the benefits of collaboration for learning. This paper describes our experimental design, the effect of the interventions, preliminary results from the Kinect sensor, and our future plans to analyze additional sensor data. We conclude by highlighting the importance of capturing and supporting 21st century skills (i.e., collaboration and effective communication) in small groups of students.

Introduction

For decades, the field of CSCL (Computer-Supported Collaborative Learning) has been concerned with promoting socio-constructivist outcomes. The idea that students learn particularly well in social settings is not new (e.g., Vygotsky, 1980; Piaget, 1998). What is new, however, is that we now have unprecedented means to measure and support students' interaction. On the measurement side, in particular, we can now have access to massive datasets characterizing students' collaboration and learning processes. Recently, sensing technologies have become increasingly affordable and easy to use, allowing researchers to collect large datasets on students' interactions in real-world settings, such as classrooms, makerspaces, museums, or other informal learning environments. In the near future, we are envisioning that a combination of qualitative and computational measures will provide us with rich information about the different facets of productive collaborative learning groups. Being able to accurately measure collaborative skills is of primary importance, because our current educational system tends to *teach what it can measure*. If we can develop innovative and automated ways of capturing those skills, we can pave the way to new forms of formative assessment and new ways of teaching those skills in traditional school curricula.

Theoretical framework

While there is a wealth of theories of collaborative learning, we focus on Roschelle's (1992) framework of convergent conceptual change. In this framework, collaboration is seen as the process of constructing shared meanings for conversations, concepts, and experiences. It gradually leads to the construction of new meanings and results in conceptual change. From this perspective, markers of collaborative learning are captured through iterative cycles of interactions that converge toward a shared understanding of the task and the concepts taught. For example, researchers in Computer-Supported Collaborative Learning (CSCL) have developed tools to capture the level of transactivity of a dialogue over time (i.e., the extent to which students build on each other's ideas in small groups; Ward & Litman, 2007). So far, the main tools for capturing these processes are limited to time-consuming qualitative analyses or quantitative methods applied to verbal exchanges (transcripts) and self-reports (surveys, questionnaires). We suggest that new technologies, sensors in particular, can provide researchers with a richer and complementary way of capturing collaborative processes.

Capturing collaboration through Multi-Modal Learning Analytics (MMLA)

Over the past decade, high frequency sensors (such as eye-trackers, motion sensors, wearables) have become affordable and reliable, which opens new doors for capturing students' multi-modal interactions. They allow educational researchers to collect significantly larger datasets: a sensor typically runs at 30-120Hz and collects various streams of information. The Microsoft Xbox Kinect sensor, for example, can collect information about a person's body joints (x,y,z coordinates), their facial expressions, and their speech at 30 Hz (i.e., 30 times per second). One can easily define ~100 variables that can be captured from the Kinect sensor. This means 3000 data points per second for one person, which translates to roughly 10 million data points for an hour of data collection. Multiply this figure by the number of sensors (eye-trackers, GSR sensors, emotion detection tools, speech features), participants, and studies to get a sense of the possibilities of combining sensors with data mining techniques. The field of Multi-Modal Learning Analytics (MMLA; Blikstein & Worsley, 2016) is about exploiting this new development. Sensors and data mining techniques have both reached a level of maturity that allows researchers to tackle new research questions and develop new educational interventions. MMLA can be used not only to study collaborative learning, but also to support it in innovative ways.

Supporting collaboration through MMLA

The effect of technology-based collaboration interventions on collaboration quality has been examined previously. Bachour, Kaplan, and Dillenbourg (2010) developed an interactive table system known as *Reflect*, which supported participation for groups of four by presenting a visualization of how much each individual talked during a learning activity. Disparity in collaborator participation can be detrimental to both individual and group experience during a collaborative activity (Salomon & Globerson, 1989). Bachour et al. found that their interactive table intervention helped participants be more aware of their participation levels during the activity, and encouraged groups to be more balanced in participation levels when the participants in the group felt that balanced participation was important to them. This research demonstrates that technological interventions that target participation balance can change the behavior of collaborators during a task. However, their study did not explore the relationship between that behavioral change and other constructs, such as experimenter and participant ratings of collaboration quality, task understanding and completion, nor the learning gains of individual participants. The present study implements a similar visualization as an intervention, while collecting data about these other critical constructs using traditional and multi-modal measurement techniques.

General description of the experiment

In this study, 42 pairs of participants had 30 minutes to program a robot to navigate through a series of increasingly difficult mazes. Two interventions were used to support collaboration: a visualization representing verbal contributions and a short verbal explanation of the benefits of collaboration for learning. Dependent measures included 1) how well participants programmed the robot (task performance), 2) differences from pre- to post-test (learning gains), and 3) the quality of their collaboration (using a validated rating scheme described below). In terms of process variables, we used three kinds of sensors: two mobile eye-trackers, a motion sensor, and two bracelets capturing electrodermal activity. In this paper, we focus on the impact of our two interventions on our dependent measures. More specifically, we predict that each intervention should positively impact task performance, collaboration and learning gains.

Methods

Subjects

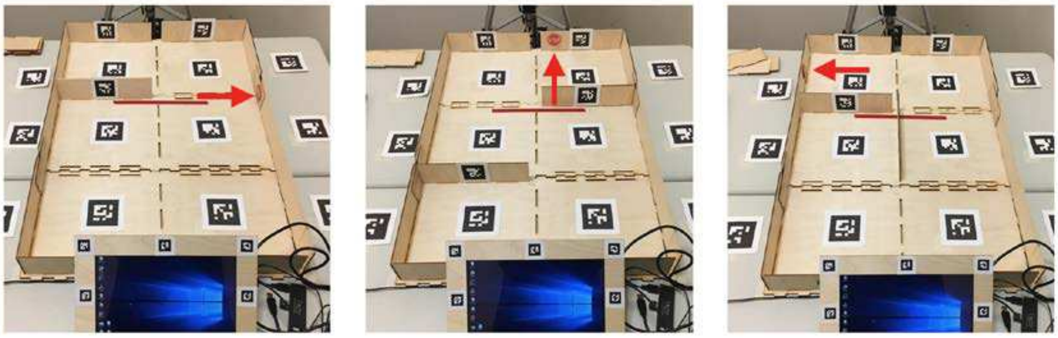
Forty-two dyads (N=84) participated in the study, although two groups were dropped from analyses due to experimenter error during data collection. Participants were recruited from the study pool of a laboratory at a university in the northeastern United States. Of the participants in the study, 62.2% were full or part-time students while the rest were members from the surrounding community. Participants ranged in age from 19 to 51 years old with a mean age of 26.7 years, and 60% identified as female. Participants were compensated $20 for the 90-minute session.

Materials

Participant learning was measured with a pre- and post-test learning assessment. The learning assessment consisted of four short answer or fill-in-the-blank questions that assessed their understanding of basic computer science competencies, such as calculating outcomes from a loop and interpreting the purpose of example code (adapted from Brennan & Resnick, 2012; Weintrop & Wilensky, 2015). There were two versions of the learning test with slight differences in question wording and figures.

In addition to the four learning test questions, during the posttest participants filled out a self-assessment of their collaboration experience, adapted from Meier, Spada, and Rummel (2007), and a demographic survey. The self-assessment included six Likert scale questions that asked about features of collaboration, such as task division, time management, and partner respect. The demographic survey asked participants to report their age, gender identity, student status, and educational level.

The task asked participants to use a block-based programming language called Tinker to navigate a simple robot through a series of mazes. The robot was constructed out of a GoGo Board, an open-sourced educational hardware device (Sipitakiat, Blikstein, & Cavallo, 2004). The robot came equipped with proximity sensors on the front, left, and right of the robot and DC motors on each side (Figure 2a). Participants used Tinker and the GoGo Board's app interface, the GoGo Widget, to learn about and control the robot. Participants were first given a training task to code the robot to go straight across a line approximately two feet in front of them so they could learn about the robot and coding environment. Behavior during this first activity was not included in the analysis. The main activity asked participants to spend thirty minutes attempting to get the robot through three mazes (Figure 1). The first maze consisted of one wall, requiring participants to make one right turn. The second maze consisted of two walls, requiring participants to make seven left and right turns. The final maze required participants to navigate a dead end. Participants were told that the overall goal was to create flexible code so that the robot would be able to solve any maze it encountered.



Maze 1 **Maze 2** **Maze 3**

Figure 1. Experimental mazes used in task. Participants were told to start the robot in the bottom left corner of the maze. The red arrow indicates the end position.

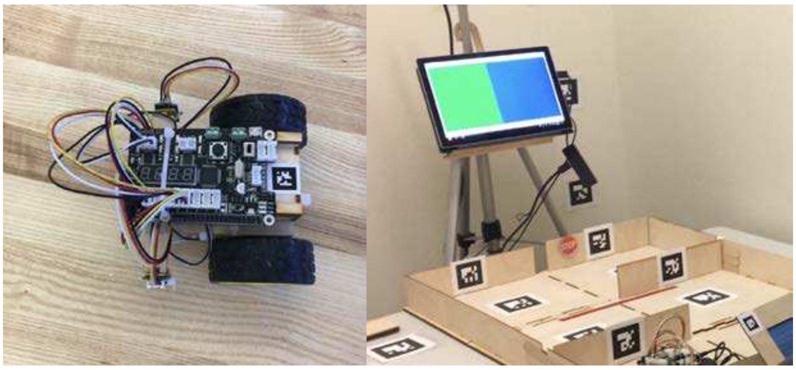
While participants attempted the task, their collaboration and task behaviors were assessed by the researcher using scales that ranged from negative two to positive two. The dyads' collaboration was assessed on nine scales, adapted from Meier, Spada, & Rummel (2007): sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation. The task behavior measures included task performance, task understanding, and improvement over time.

Design

This study utilized a two by two between-subjects design. The two independent variables were a speech equity visualization intervention (Figure. 2b; inspired from Bachour et al. 2010) and an informational collaboration intervention. The speech equity intervention utilized data from an Xbox Kinect sensor to track how much each participant spoke during the activity. Dyads assigned to receive this visualization saw a tablet that displayed how much each of them had spoken in relation to the other's speech in the past 30 seconds by presenting rectangles proportional to their speech time. As one participant spoke more than the other, their rectangle grew larger and took up more of the screen. Dyads not assigned to this condition received no visual intervention.

The informational collaboration intervention consisted of a researcher giving participants instructions around collaboration before the study activity began. Dyads assigned to this intervention were reminded that they were expected to collaborate during the learning activity, and were invited to think about how they were collaborating throughout the session. They were told that previous research has found that factors like equity of each partner's speech time is predictive of the quality of collaboration and learning gains. Dyads not assigned to this condition received no informational intervention.

A quarter of dyads (10 of 40) received no visualization nor the collaboration information (Condition #1), a quarter received only the visualization (#2), a quarter received only the collaboration information (#3), and the final quarter received both the visualization and the collaboration information (#4). These conditions assignments were generated randomly.



(a) (b)

Figure 2. (a) The robot used in experimental task. (b) The Kinect-based speech visualization.

Procedure

Each participant first signed an informed consent, and then individually completed the pretest learning assessment. Participants had five minutes to complete the pretest assessment. The two participants worked on different versions of the assessment on separate computers. Participants were then informed that they would be collaborating with one another, and were instructed to introduce themselves to their partner.

A tutorial video was shown that illustrated basic concepts of how to use Tinker to program the robot, such as how to find, insert, and delete code blocks. After the video, participants had five minutes to complete the training task to code the robot to move forward across a line roughly two feet in front of them. If after five minutes they were unable to complete this task, the researcher demonstrated how to accomplish it and explained the rationale behind the code used. If participants did successfully complete the task, the researcher showed them the same code solution that they showed to unsuccessful participants if any differences existed between that solution and the participant's code. All participants received an explanation of the rationale for this code solution, so that all dyads received the same explanation and code example before progressing.

Participants then saw a second tutorial video, which showed more advanced features of the GoGo Board, including how to use prewritten functions for going forward, left, or right. Participants were also exposed to more complicated code examples, including how to use conditional statements to trigger actions when certain conditions were met, and how to utilize the GoGo Widget. After this, dyads were given 30 minutes to try to code the robot to solve a series of three increasingly complex mazes, as described above and depicted in Figure 1.

While they were completing the series of three mazes, the participants had access to a printed reference sheet that reviewed some of the material presented in the two tutorial videos. Participants also received hints every five minutes from the researcher throughout the thirty-minute session. These hints were the same across all groups, and consisted of reminders of the available tools and suggestions for what code blocks to try to use.

After the 30-minute session, participants had 10 minutes to complete a posttest learning assessment. Participants received the version of the pretest that they did not complete previously as the first section of their posttest. They also responded to questions that asked them to reflect on their experience with the task and about their collaboration experience, and filled out a demographic survey. Finally, participants were debriefed, thanked for their participation, and compensated with $20.

Coding

Dyads' collaboration behavior and task performance were live-coded by the researcher conducting the session using the scales described above. During the main 30-minute work session, researchers looked for evidence of behaviors that corresponded to one of five levels of each scale, which ranged from good behaviors at positive two and poor behaviors or the absence of good behaviors at negative two. Multiple researchers conducted sessions of the study and thus coded dyads' behavior. Researchers double coded 20% of the sessions from videos collected during the session, and had an inter-rater reliability of 0.65 (75% agreement).

The four items on the learning test were graded on a zero to three rubric scale to evaluate completeness of answers and understanding of computational thinking skills. These scores were added together to generate total pre- and post-test scores for each participant as well as learning gains. The final code each dyad created was evaluated on a zero to four scale to determine how well the code in abstract could perform the maze solving task. This rubric aligns with the live coding of "Task Understanding" done during the session, acting as a post-hoc assessment to ensure dyads' final products were fully evaluated.

Multi-modal data

A number of multi-modal sensors were also used to collect data from both participants in each session. Tobii Pro Glasses 2 eye-tracking glasses (https://www.tobiipro.com/product-listing/tobii-pro-glasses-2) were used to follow where each participant looked throughout the session. The eye-tracking glasses sampled at a rate of 50 Hz, thus generating roughly 90,000 data points per person during the 30-minute session.

An Empatica E4 wrist sensor (https://www.empatica.com/e4-wristband) was used to track several physiological markers from each participant, including electrodermal activity (at 4 Hz), blood volume pulse (at 64 Hz), and XYZ acceleration (at 32 Hz). During the 30-minute session, roughly between 7,200 to 115,200 data points were generated for each participant per measure, depending on the physiological measure's sampling rate.

Finally, a Kinect was used to track the motions of the dyads. The sampling rate for the Kinect was 30 Hz, generating roughly 54,000 data points during the main session for around 100 different variables. Data from these sensors will be analyzed and presented more thoroughly in future publications.

Results

Assessment of collaboration

Analysis of the coding of dyad collaboration revealed significant differences between the two conditions that received the informational intervention to support collaboration (3&4) versus those that did not (1&2). Dyads in condition 3 scored 7.1 points higher than those in condition 1 ($p < 0.001$), both of which did not receive the Kinect-based visualization intervention. Dyads in condition 4 scored 4.8 points higher than those in condition 2 ($p = 0.03$), both of which did receive the Kinect intervention. Differences in collaboration between the conditions that received the Kinect-based visualization intervention (2&4) and those that did not (1&3) were not significant when controlling for the verbal intervention on collaboration.

Participants' self-reported collaboration scores at the individual level differed significantly from the researchers' assessment of their collaboration at the dyad level ($F = 15.21$, $p < 0.001$) but they are significantly positively correlated ($r = 0.43$, $p = 0.001$). Self-reported scores were on average higher for measures of task division, time management, and reciprocal interaction while being lower for reaching consensus, dialog management, and sustaining mutual understanding. Further qualitative analysis and analysis of multi-modal data sources may reveal additional details regarding participants' self-perceptions of effective collaboration and why they differ from the researchers' coding. Neither scale differed significantly by individual gender, the gender makeup of the group, or level of education of participants.

Researcher coding of collaboration was significantly positively correlated with the quality of produced tinker code ($r = 0.52$, $p < 0.001$) as well as all three performance metrics: task performance ($r = 0.35$, $p <0.001$), task understanding ($r = 0.53$, $p < 0.001$), and improvement over time ($r = 0.54$, $p < 0.001$) (See Figure 3a.)

Learning test

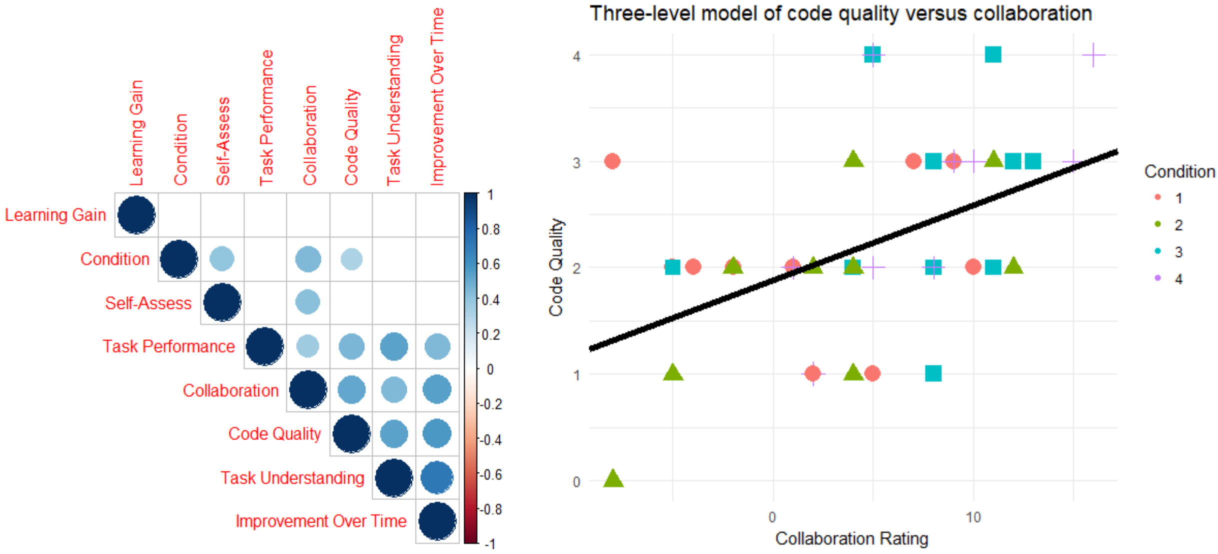
Pairwise comparisons of treatment group means with a Bonferroni correction for multiple comparisons revealed no significant differences between groups' pretest, posttest, or gains scores by condition on the learning test. All pairwise comparisons were confirmed with a Tukey's honest significance post-hoc test. The two interventions did not target content knowledge gains directly but aimed to increase collaboration within the dyads. While not differing significantly by condition, participants on average gained 19.8 percentage points between the pre- and post-survey ($t = 6.18$, $p < 0.001$). This indicates the efficacy of even a short interactive programming lesson targeting computational thinking fundamentals. Learning gains do not differ significantly by individual gender, the gender makeup of the group, or level of education of participants.

Posttest scores on a question that asked participants to explain how to solve a maze were significantly correlated with the quality of the code the dyads wrote ($r = 0.26$, $p = 0.04$). A question on interpreting code that used nested conditional statements was significantly correlated with our coding of dyad collaboration ($r = 0.33$, $p < 0.01$) as well as the number of mazes dyads completed ($r = 0.35$, $p < 0.005$).

Performance metrics and quality of produced code

The mean number of mazes completed by each group (task performance) and improvement over time did not differ significantly by condition. Mean scores on task understanding only significantly differed between Conditions 2 and 3 ($p < 0.05$) but both interventions differed between those two groups. The quality of the final block-based code dyads produced is significantly correlated with our assessment of the quality of their collaboration ($r = 0.52$, $p < 0.001$), the number of mazes completed ($r = 0.45$, $p < 0.001$), their task understanding ($r = 0.45$, $p < 0.001$), and their improvement over time ($r = 0.54$, $p < 0.001$).

To estimate the relationship between collaboration and code quality, a three-level (participants in dyads in conditions) linear mixed-effects model was fit by residual maximum likelihood methods. On average in the population, a one-unit increase on our rating of collaboration was associated with a 0.071-point increase in the grand mean of the quality of their code ($p < 0.001$) when controlling for gender and education (Figure 3b.) At a collaboration rating of 0, the expected mean code quality is 1.88 ($p < 0.001$). To aid interpretation, a 14-point increase in our rating of collaboration corresponded to roughly a one-point increase on our 0-4 code quality scale. Additional model building and discussion of the random effects of the model will be explored in future work.



(a) (b)

Figure 3. (a) Correlogram of different performance metrics and ratings of collaboration. All correlations are positive, only significant ones plotted. (b). Linear fit of fixed effects of code quality versus collaboration rating.

Preliminary results: Movement and talking of dyads

The total amount of movement across all upper body joints and body parts was calculated for each session as well as the total amount of talking per dyad. The amount of talking was significantly positively correlated with our assessment of the quality of collaboration ($r = 0.45$, $p = 0.007$) as was movement of the participants' left elbows ($r = 0.36$, $p = 0.03$) and left hands ($r = 0.36$, $p = 0.04$). From observer notes, most participants were right handed and used that hand to manipulate the keyboard and mouse to program the robot. This would leave the left hand more free to gesture. While the intervention related to visualizing participant verbalization did not appear to have a significant effect on quality of collaboration, the strength of the relationship between amount of talking and quality of collaboration suggests this is fruitful area for additional exploration. Future exploration of this data will examine differences in movement and talking within groups and how this affects collaboration as well as determining prototypical postures or gestures that may indicate quality of collaboration.

Plans to analyze physiological markers, eye tracking, and motion data

Our main goal is to conduct an in-depth analysis of the sensor data collected during this study. More specifically, our first step will be to design measures of convergence using physiological, gestural, and visual data. For example, it is well-known that Joint Visual Attention (JVA) is a prerequisite for effective collaborations (Schneider et al., 2016). There is also some initial evidence that physiological synchrony is indicative of productive groups (Pijeira-Díaz, Drachsler, Järvelä, & Kirschner, 2016). Finally, body postures and gestures can be used to identify leadership behaviors - though bodily synchronization has not been found to correlate with collaboration quality (Schneider & Blikstein, 2015). We plan to replicate those findings on our datasets, and combine them to develop multi-modal measures of synchronicity in dyads.

Discussion

The purpose of this paper was to explore the effect of two collaboration interventions and the relationship between collaboration quality, task performance, learning gains. The main hypothesis was that the two collaboration interventions, one visual and the other informational, would improve dyads' collaboration quality. Administration of the interventions was crossed such that an equal number of groups (10 dyads) received neither, one, or both of the interventions. Analysis of researcher's coding of participants' collaboration quality found that while collaboration quality improved for both groups that received the informational intervention, there was no significant improvement of collaboration quality for the groups that received the visualization intervention and no significant advantage conferred to the group that experienced both interventions.

The failure of the visualization intervention to influence the collaboration behavior is somewhat surprising. Bachour et al. (2010) implemented a similar visualization intervention that targeted participants' speech balance as a way to alter collaboration behavior, and found that groups who experienced the visualization intervention had more balanced participation levels when participants reported that they believed balanced participations levels were important. Based on those findings, it might have been predicted that dyads' who experienced both the informational and visualization interventions would have benefitted the most, as they had the external feedback tool and the importance of participation balance made salient. However, there was no evidence to support this. Possible explanations for this could include the design of the visualization tool. The visualization tool was presented on a relatively small tablet screen on the other side of a table from the participants. It could be that the visualization was not salient enough for participants to be motivated to attend to during the challenging task. Further analysis of the data could examine the eye-tracking of participants to assess how frequently they looked at the visualization and whether that correlates to collaboration quality. Additionally, the tool used in the present study operated on a different timescale than Bachour et al.'s. Rather than showing total participation throughout the session, the visualization was based on the past thirty seconds of speech. Perhaps this timescale was not optimal for altering participant behaviors. Lastly, Bachour et al.'s study did not include any measure of collaboration quality other than speech time balance among participants. It could be that participation balance simply is not an effective predictor of collaboration quality more broadly, and the effect of the informational intervention stems more from the mere reminder of the importance of collaboration than from encouragement of speech equity in particular.

Additional metrics were used to evaluate the efficacy of the two interventions, including task performance and learning gains on the pre- to posttest assessment. However, analyses showed that neither of these metrics were significantly affected by collaboration condition assignment. This is likely due to reasons similar to those listed above. Other reasons could include the challenging nature of the activity that resulted in a performance ceiling effect and that the interventions did not directly addressed the learning goals of the activity.

The present study also aimed to explore the relationship between collaboration quality, learning gains, and task performance. There was evidence that stronger task performance, as measured by final code quality, was significantly associated with researcher's coding of collaboration quality. Learning gains, as measured by the changes in performance from the pre- to posttest, revealed only modest evidence that certain questions on the test correlated with collaboration quality and task performance. This finding is not surprising, as the research on the relationship between collaboration quality and learning outcomes is mixed (Dillenbourg, Baker, Blaye, & O'Malley, 1996). There was a significant improvement in participant scores from pre- to posttest overall, suggesting that even this relatively short learning activity led to increase in computational thinking skills.

Conclusion

While this study was not able to show a clear effect of providing a real-time visualization to support collaboration, it made many other valuable contributions. First, it showed that simple verbal interventions can help participants pay attention to particular aspects of their collaborative behavior (i.e., how much they are talking and how much space they are providing to their partner). Second, it suggested that awareness tools such as the one developed for this study have to be designed differently to impact social interactions (e.g., by being more salient or be used in a setting where users have the mental bandwidth to reflect on their collaborative style). Third, we collected a rich multi-modal dataset that can be used to build proxies for measuring effective collaborations. As a preliminary analysis, we found that various indicators captured by the Kinect sensor were correlated with participants' quality of collaboration (e.g., amount of talking and movements). Finally, we are showing that well-designed learning activities can teach beginning computational thinking skills to a variety of participants - even those with no programming experience.

In the future, in addition to developing more multi-modal measures of collaboration, we are planning to improve the activity by making it longer and by providing more scaffolding. We will also design alternative ways of displaying the amount of talking, for example by making it more salient in the environment. Finally, we are interested in studying longer-term activities, for instance when students are working together over the span of several days or weeks. This will address an important limitation of most studies where collaborative episodes take place in a short time frame (i.e., 1-2 hours).

References

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*(3), 203-213.

Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. Journal of Learning Analytics, 3(2), 220-238.

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. *Proceedings of the 2012 annual meeting of the American Educational Research Association, Vancouver, Canada,* 1-25.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In *P. Reimann & H. Spada (Eds.) Learning in Humans and Machine: Towards an interdisciplinary learning science*, pp. 189–211. Oxford: Elsevier.

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 2(*1), 63–86.

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., & Kirschner, P. A. (2016). Investigating collaborative learning success with physiological coupling indices based on electrodermal activity. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 64-73). ACM.

Piaget, J. (1998). *The language and thought of the child*. Psychology Press.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *The journal of the learning sciences, 2*(3), 235-276.

Salomon G., & Globerson, T. (1989). When teams do not function the way they ought to. *International journal of Educational Research, 13*(1), 89-99.

Schneider, B., & Blikstein, P. (2015). Unraveling Students' Interaction Around a Tangible Interface using Multimodal Learning Analytics. International Journal of Educational Data Mining, 7(3), 89-116.

Schneider, B., Sharma., K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2016). Unpacking The Perceptual Benefits of a Tangible Interface. *ACM Transactions on Computer-Human Interactions (TOCHI), 23(6),* 39.

Sipitakiat, A., Blikstein, P., & Cavallo, D. P. (2004). GoGo board: Augmenting programmable bricks for economically challenged audiences. *Proceedings of the 6th International Conference on Learning Sciences*, 481-488.

Tschan, F. (2002). Ideal Cycles of Communication (or Cognitions) in Triads, Dyads, and Individuals. *Small Group Research, 33*(6), 615–643.

Vygotsky, L. S. (1980). Mind in society: The development of higher psychological processes. Harvard university press.

Ward, A., & Litman, D. (2007). Dialog convergence and learning. *Frontiers in Artificial Intelligence and Applications, 158*, 262.

Weintrop, D., & Wilensky, U. (2015). Using commutative assessments to compare conceptual understanding in blocks-based and text-based programs. *ICER, 15,* 101-110.