Unpacking Collaborative Learning Processes During Hands-on Activities Using Mobile Eye-Trackers

Bertrand Schneider, Harvard University, bertrand_schneider@gse.harvard.edu

Abstract: This paper describes a multimodal dataset captured during a collaborative learning activity typical of makerspaces. Participants were introduced to computational thinking concepts using a block-based environment: they had to program a robot to solve a variety of mazes. Mobile eye-trackers, physiological wristbands and motion sensors captured their behavior and social interactions. In this paper, I leverage prior work on joint visual attention (Tomasello, 1995) and analyze the eye-tracking data collected during the study. This paper provides three contributions: 1) I use an emerging methodology to capture joint visual attention in a co-located setting using mobile eye-trackers (Schneider & al., 2018); 2) I then replicate findings showing that levels of joint visual attention are positively correlated with collaboration quality; 3) finally, I present a new measure that captures cycles of collaborative / individual work, which is positively associated with learning gains (but not with collaboration quality). I discuss these results and conclude with implications for capturing students' interactions in co-located spaces using Multimodal Learning Analytics.

Introduction

In the last decade there has been a growing interest in cultivating skills that are not traditionally taught in traditional school settings. Those skills are often referred to as "21st century skills" (Dede, 2010; e.g., Collaboration, Communication, Creativity, Critical Thinking) because they are deemed essential for jobs that do not yet exist. New learning environments, such as digital fabrication labs and makerspaces, are ideal spaces for their development. They allow students to learn complex concepts in STEM (Science, Technology, Engineering, Mathematics) through hands-on learning and applied projects. Measuring the development of those skills and providing formative assessment, however, remains a challenge (Berland, Baker, Blikstein, 2014), because each student is unique, and the development of those 21st century skills takes different forms depending on interacting factors (e.g. learners' personalities, prior knowledge, SES background). The CSCL community has long been studying those skills before they gained a renewed attention from researchers and the general public.

For the scope of this paper, I focus on students' collaboration and communication by leveraging a new field of research called Multimodal Learning Analytics (Blikstein & Worsley, 2016; MMLA) to capture the quality of learners' interactions. MMLA uses multiple high-frequency sensors to capture users' behavior and applies data mining techniques to find trends and predictors in large datasets. Joint visual attention has been extensively studied by social and developmental psychologists and has been shown to be critical to many social interactions. Based on prior literature (e.g., Richard & Dale, 2015; Schneider & Pea, 2013), the main hypothesis of this paper is that productive groups exhibit higher levels of joint attention compared to less productive groups. This construct was captured using multiple mobile eye-trackers in co-located spaces (Schneider & al., 2018). More specifically, I designed a hands-on task typical of makerspaces (i.e., learning to program a robot to solve a variety of mazes) and computed measures of joint visual attention. correlated them with three outcomes measures: the quality of their collaboration (coded with a validated rating scheme in the learning sciences), their task performance (i.e., how successful they were) and their learning gains (computed from a pre and post-test). Finally, because collaboration can be a powerful way to support learning, I analyzed the eye-tracking data to find behaviors that were not just related to collaboration quality, but also learning outcomes.

This paper is structured as follows: the first part reviews the literature on dual eye-tracking and the various measures that researchers have developed over the years to capture joint visual attention. The second part describes the study, participants and data collection protocol. The third part discusses the steps to pre-process the data and compute metrics of joint visual attention to correlate them with outcomes of interests. Finally, I discuss our results and conclude with future steps for capturing students' 21st skills in makerspaces using MMLA.

Literature review

This section provides a succinct review of foundational work in developmental and social psychology, as well as in Computer-Supported Collaborative Learning (CSCL) and Computer-Supported Collaborative Work (CSCW) where multiple eye-trackers are used to look at participants' visual alignment.

There are currently three strands of research studying collaborative processes through dual eye-tracking. The first strand uses remote eye-trackers to capture joint visual attention, where users are each looking at a different computer displays (for example through video conferencing). In an early study, Richardson & Dale (2005) explored the coupling between speakers' and listeners eye movements and its relationship with discourse comprehension. They found a positive correlation between discourse comprehension and dynamic coupling between conversants' eye movement. In a subsequent study, they replicated those results for a live conversation (Richardson, Dale & Kirkham, 2007). In CSCL, researchers have used this methodology to study pair programming tasks (Jermann, Mullins, Nüssli & Dillenbourg, 2011) and found that collaboration quality was characterized by higher levels of "gaze cross-recurrence" (i.e., joint visual attention). A second strand of research has started to study more ecological settings using mobile eye-trackers. Yu and Smith (2013), for example, used mobile eye-trackers to explore infant cross-situational word learning through eye-hand coordination. In education, Schneider & al. (2018) studied apprentices in logistics interacting with a tangible user interface. They found that levels of joint visual attention (as captured by mobile eye-trackers) were correlated with their quality of collaboration. Additionally, they developed a methodology to capture leadership behaviors from dual eye-tracking data by identifying who initiated and who responded to an offer of joint visual attention. Imbalances of these behaviors were negatively correlated with learning gains. Finally, a last strand of research has started to explore the benefits of novel visualization techniques to improve learning in a collaborative setting, for example by displaying participants' gaze to each other. This intervention is sometimes called a "gaze awareness tool", "shared gaze visualization" or "Bidirectional Gaze" (for a review, see D'Angelo & Schneider, 2018). D'Angelo & Begel (2017) have enhanced remote pairs' speed and success in communicating when resolving a coding problem by using eye tracking devices to show each participant where their partner is looking on the screen. In education, Schneider & Pea (2013) found that making the gaze of each partner visible promoted interactions of higher quality and consequently increased students' learning gains.

In conclusion, there is ample work showing that joint visual attention is a central mechanism by which group members coordinate their actions and establish a common ground (Clark & Brennan, 1991). Furthermore, recent research has been leveraging new sensing technology to quantify joint visual attention in dyads of users (e.g, Jermann, Mullins, Nüssli & Dillenbourg, 2011). While most studies have looked at remote collaborations, there is some nascent work in co-located settings using mobile eye-trackers (e.g., Yu & Smith, 2013). Ultimately, however, the goal from a CSCL perspective is to understand how collaborative processes *contribute* to learning. Joint visual attention (JVA), for example, is a necessary but not sufficient condition for productive social interactions. This paper is about going beyond capturing JVA and finding more precise indicators of collaborative learning. This paper builds upon prior findings (e.g., Schneider & al., 2018), replicates results, and provides new contributions by isolating collaborative learning processes from the eye-tracking data.

Methods

Summary of the study

In this study, participants with no prior programming knowledge were given 30 minutes to program a robot to autonomously solve a series of increasingly complex mazes (see Fig. 1 for the setup of the experiment). Two different interventions were developed and used to support collaboration: a visualization of relative verbal contributions of the participants shown in real time and a brief informational explanation delivered verbally summarizing literature findings on the value of collaboration for learning. While dyads completed the activity, a variety of sensors described in 2.2 collected eye gaze, movement, verbal, and electrodermal activity data on participants. Dependent measures were an assessment of the quality of the collaboration, how well the participants coded the robot to perform the assigned task, and learning gains related to computational thinking. The study is described in more detail in Starr, Reilly & Schneider (2018).

Participants

Participants were drawn from an existing study pool at a university in the northeastern United States. 42 pairs of participants (N=84) were used in the analysis. 62% of participants identified as students, with ages ranging from 19 to 51 years old (mean age = 26.7 years). 60% of participants identified as female. Participants were paid $20 for the 90-minute session and did not know each other prior to the study.

Experimental design

The study utilized a two-by-two between-subjects design where dyads were assigned to one of four conditions that would receive different interventions. 25% of dyads received neither intervention (Condition #1), 25% received solely the visualization intervention (#2), 25% received solely the informational intervention (#3) while the remaining quarter received both interventions (#4). The speech equity visualization utilized speech collected by the sensors in the experiment to display how much each participant spoke as a proportion of total talk during the activity. Dyads with this intervention saw a tablet display representing this data over the past 30 seconds by presenting colored rectangles that grew to take up more of the screen as relative contribution increased. The informational intervention involved a researcher reading a short passage that reminded dyads that they were expected to collaborate and invited dyads to think about how they were collaborating during the activity. They were also told that research has found that equity of each partner's speech time is predictive of the quality of collaboration and learning gains. For an analysis of the differences between each experimental condition, please see Starr, Reilly & Schneider (2018).

Procedure

After taking the pre-survey and calibrating all sensors, participants were shown a short tutorial video that introduced the basics of writing a program in Tinker, a block-based programming language designed for use with the microcontroller of the robot. Participants were then given five minutes to write code that would move the robot forward across a red line roughly two feet directly in front of it. The robot consisted of a microcontroller, two DC motors, and three proximity sensors. Following this tutorial activity, a second tutorial video was shown that highlighted more advanced features of Tinker such as using prewritten functions to turn the robot and using sensor values to trigger conditional statements. Dyads were also given a reference sheet summarizing the content covered in the tutorial video. Dyads then had 30 minutes to write code to navigate a robot through a series of mazes. Once the robot successfully completed a maze twice, a more challenging maze was provided. Dyads did not know the layout of the mazes ahead of time and were encouraged to write code that would allow the robot to solve any simple maze. During this portion of the study, the researcher provided standard hints at 5-minute intervals to all dyads regarding common pitfalls researchers identified in pilot testing of the activity.

Independent, dependent measures and process data

The quality of the dyad's collaboration and task performance was assessed during the task by the researcher running the session. The quality of collaboration was measured by aggregating the 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 (refer to Meier, Spada, & Rummel, 2007 for a definition of those terms). Researchers double-coded 20% of the sessions and had a Cronbach's alpha of .65 (75% agreement). Task behaviors evaluated included task performance (how many mazes were completed by the robot in 30 minutes), task understanding (how much major coding concepts were included in the design such as using sensors with appropriate thresholds in conditional statements), and improvement over time (how much a team's understanding of requisite technical skills and conceptual understanding of the task changed over time). The final written code of the dyad's was also assessed to determine theoretically how well it could have performed the assigned task barring technical issues. To assess learning of computational thinking principles, participants individually completed pre- and post-surveys consisting of four questions related to conditional statements, looping, and interpreting the output of given code (adapted from Brennan & Resnick, 2012; Weintrop & Wilenski, 2015). After the activity, participants also self-assessed their collaboration and wrote a brief reflection regarding how their thinking changed over the course of the activity. Table 1 presents a summary of the measures described in this section.

Table 1: Independent, process and dependent measures looked at in this paper (described above)

Independent Measures (2x2)	**Process Measures**	**Dependent Measures**
Speech visualization (on/off) Verbal intervention (yes/no)	Eye-tracking data: Individual gaze points on AOIs, Joint Visual Attention	Collaboration quality (9 sub-scores; 1 overall score) Task performance: code quality, improvement, # of mazes solved) Learning gains (computational thinking)

Dual eye-tracking measures and hypotheses

Prior work has explored multiple ways of capturing joint visual attention and collaborative processes from dual eye-tracking data. In this paper, I first follow a methodology described by Schneider & al. (2018) to compute joint visual attention from mobile eye-trackers and attempt to replicate previous results showing that JVA is associated with collaboration quality. Second, I was inspired by previous results showing that collaborative problem-solving is a cycle between moments of understanding and non-understanding (Miyake, 1986), and that ideal cycles of communication are related to group performance (Tschan, 2002). In this paper, I hypothesize that collaborative learning interactions are characterized by more frequent cycles of individual work and group interactions – which are captured from the eye-tracking dataset. In short, the hypotheses of this paper are as follows:

1. JVA is associated with higher quality of collaboration; more specifically, JVA is associated with participants' ability to sustain mutual understanding (Schneider & Pea, 2013).
2. The number of cycles of individual work (no-JVA) and collaborative interactions (JVA) is positively associated with the three outcome measures (collaboration, task performance, learning gains).

In the next section, I describe the data, preprocessing steps and measures.

Data collection (multimodal sensors)

Several sensors were used to collect data from both participants in each session. Tobii Pro Glasses 2 eye-tracking glasses were worn by each participant to follow eye gaze relative to a set of fiducial markers placed around the study environment. An Empatica E4 wrist sensor tracked participant electrodermal activity, blood volume pulse, and acceleration. Finally, a Kinect motion sensor was used to track the movement and position of the participants in space. This sensor collects approximately 100 variables related to a person's body joints and skeleton (24 different points with columns for x, y, z coordinates), their facial expressions, and their amount of speech. Typically collected at 30 Hz, this results in roughly 5.4 million observations per individual during a 30-minute session.

This paper focuses on the Tobii eye-tracker and the data it generates. The glasses include multiple cameras (two infrared cameras recording eye movements and one scene camera recording the participant's field of view), an accelerometer, gyroscope, microphone, a wearable recording unit running and associated controller software running on Windows. The Tobii eye-Tracker outputs data of multiple kinds including an audio recording of the session, a video recording from the point of view of the user, the x and y coordinate of the user's eye-gaze relative to its point of view. These glasses sampled at 50 Hz, generating roughly 90,000 observations per person during the main 30-minute activity. No participant reported being bothered by the glasses. Anecdotally, a few participants forgot that they were wearing them and attempted to leave the room without removing the glasses at the end of the study.

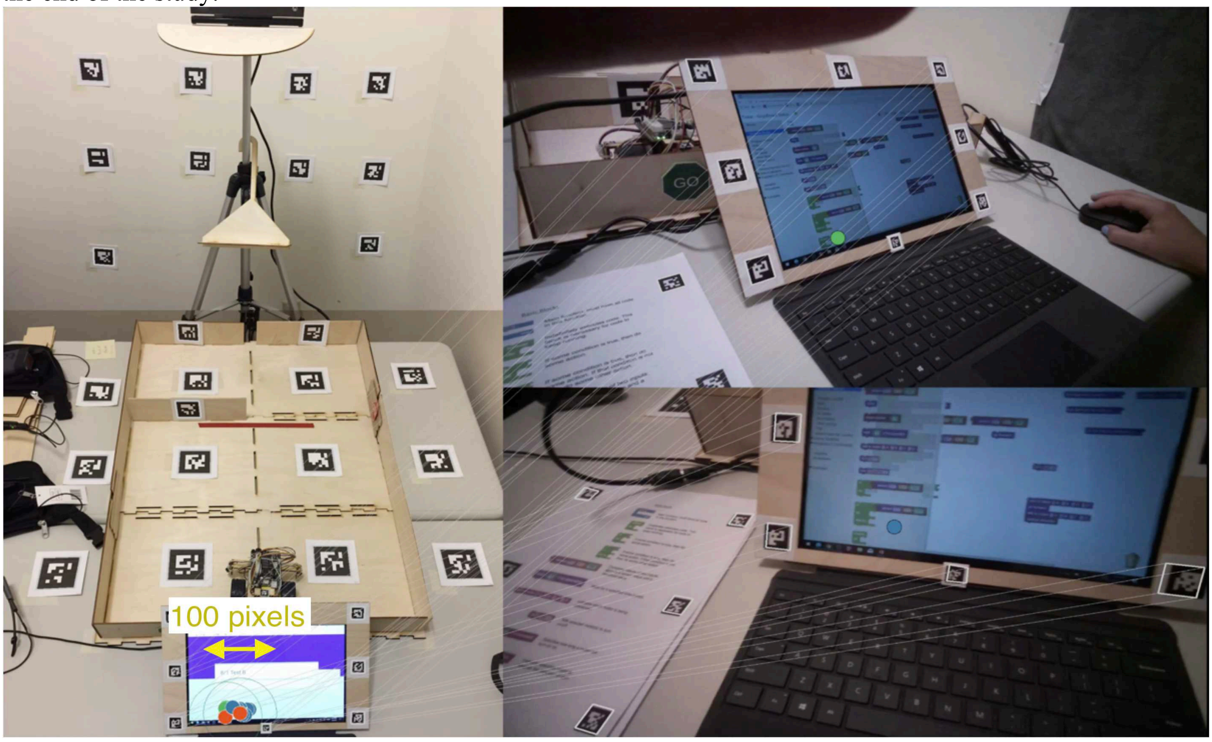

Figure 1. Example of a video frame generated for sanity-checking purposes where a homography was used to remap participants' gaze (shown in blue and green on the right side of the image) onto a ground truth (left side). The white lines represent the points detected from the fiducial markers to do the homography. On the left, the gaze points turn red if there are within a certain radius (e.g., 100px), which signifies some joint visual attention.

Data preprocessing – Temporal and spatial alignment

Temporal alignment

In order to clearly mark when transitions between different portions of the study took place across all sensors and recording devices, several fiducial markers with accompanying audio cues were placed in a PowerPoint presentation used by researchers and participants to guide the flow of the study. Whenever specific points in the study were reached, participants would simultaneously see the fiducial marker, hear the sound, and press a button on the EDA bracelet. In this way, all of the sensors on both participants as well as the video recorder would have some tagged record of the event and therefore a way to synchronize all of the data. The eye-tracking data analyzed in this paper is solely from the main 30-minute portion of the study and was synchronized via these tags.

Spatial alignment

One challenge of using mobile eye-trackers is that users can freely move - they can walk around, stand, sit, and change the orientation of their head. Their eye gaze is calibrated on the frames provided by the scene camera, which changes in its content depending on where users are looking. This kind of data is significantly more challenging to analyze compared to traditional (i.e. remote) eye-trackers, where the main area of interest is the screen of a computer. Thus, when using a mobile eye-tracker, we need to identify which part of the environment users are looking at. The solution used in this paper is to add fiducial markers to the environment (they look like QR codes on Figure 1). Detecting those markers is relatively easy for computer vision algorithms, and since they each have a unique ID they provide common coordinates across different perspectives. More specifically, a panoramic picture of the workspace (Fig. 1, left side) and the markers detected from the scene camera of the mobile eye-tracker (Fig. 1, right side) were associated to the markers of the workspace (referred to as “ground truth” below). Knowing this common set of points allowed to infer the location of users’ gaze points on a common plane using a *homography*. The left side of Figure 1 shows the last 5 gaze points for each user (shown as a gaze plot; additionally, the dots turn red if they are within 100 pixels of each other). Finally, for each group a video recording was generated for sanity checking purposes and to make sure that the homography was accurate.

Results

Areas of Interest (AOIs)

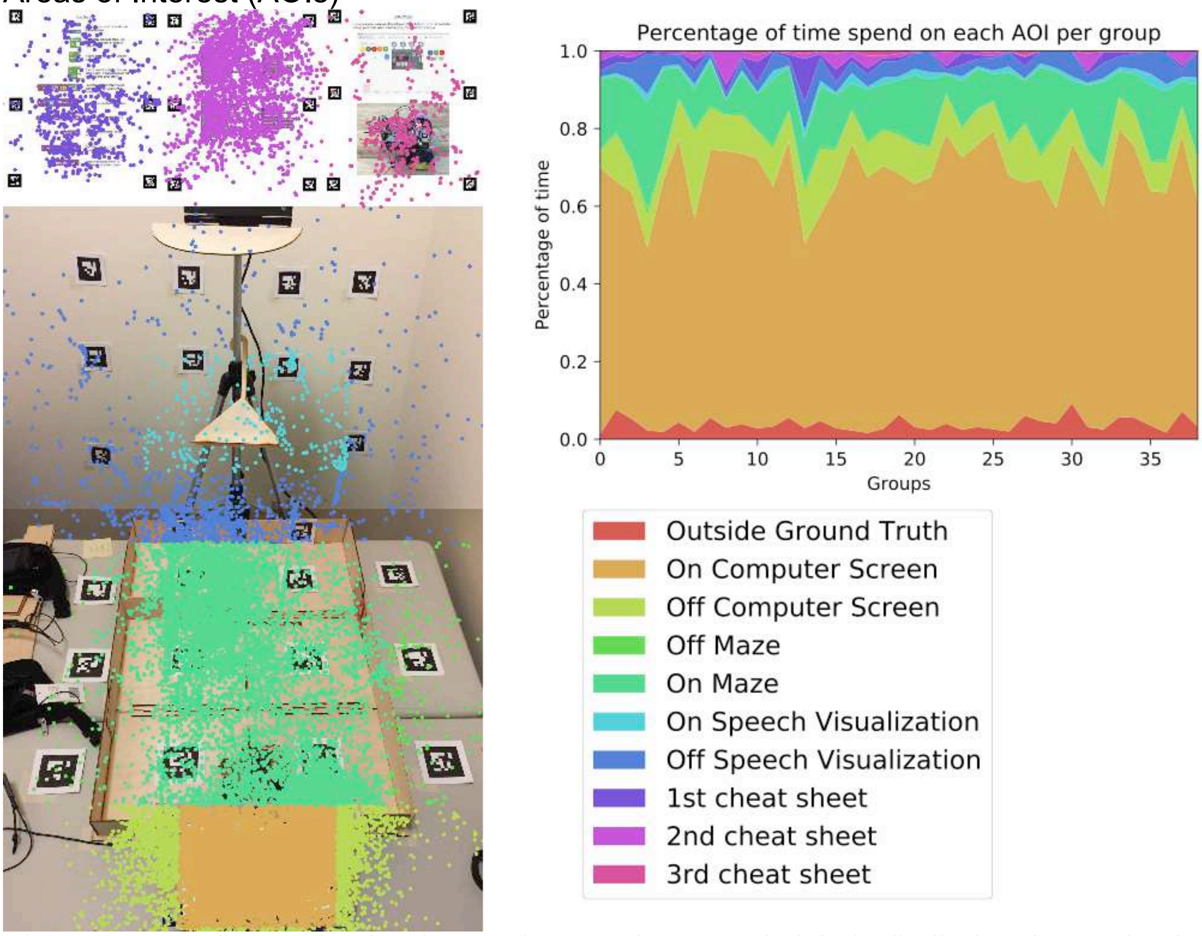

Figure 2. AOIs on the ground truth and Gaze points on each AOI. On the left: the distribution of gaze points for one group (42). On the right: the percentage of eye-tracking data for each AOI (y-axis) and for each group (x-axis). “Outside Ground Truth” refers to the gaze points outside the image shown on the left side of Fig. 1.

Areas of Interests (AOIs) divide the participants view into 7 different regions, at three different height levels. At the lower level, they differentiate between looking at the computer screen where participants wrote code and looking around it. At the level of the maze, they differentiate gaze points within the maze and outside the maze. Finally, at the level of the wall, they separate the area corresponding to the speech visualization (only relevant to groups having access to it, i.e. condition #2 and #4) from the one around it.

Running Pearson's correlations between the number of gaze points on each AOI at the individual level generated the following results: looking at the maze and code quality ($r(37) = 0.331$, $p = 0.040$) / learning gains ($r(37) = 0.360$, $p = 0.025$). Additionally, there were negative correlations between looking at the computer screen and code quality ($r(37) = -0.320$, $p = 0.047$); looking at the first cheat sheet and sustaining mutual understanding ($r(36) = -0.364$, $p = 0.025$) / quality of collaboration ($r(36) = -0.323$, $p = 0.048$); looking at the second cheat sheet and task performance ($r(36) = -0.608$, $p < 0.001$) / code quality ($r(37) = -0.350$, $p = 0.029$). In summary, looking at the number of times that individual participants looked at different AOIs seemed to be mostly associated with negative outcomes.

Cross-recurrence graphs

Before computing measures of joint visual attention, it is recommended to generate cross-recurrence graphs to sanity check the data. A Cross-Recurrence Graph (Jermann, Mullins, Nüssli & Dillenbourg, 2011) is a plot representing the eye-tracking data of the dyad. One axis is the time for one person and the other axis is the time for the other participant. If the two people are looking at the same location at the same time, we plot a black dot along the diagonal. If there is a delay, we plot this point above and below the diagonal. The distance from the diagonal is proportional to the delay. Therefore, by looking at the points on the diagonal we can estimate visual coupling within the pair. Gray dots represent no joint visual attention and white dots represent missing data.

Cross-recurrence graphs provided a visual representation of the groups' attentional alignment. I generated one for each group and used them as a sanity check for the JVA measures (Fig. 3): for example, it confirmed that group 42 had high levels of JVA, which is represented by more black pixels. Group 38 had low levels of JVA which is represented by more white pixels. Color-coded cross-recurrence graphs (i.e., using the colors from Fig. 3) also helped us observe patterns of interaction: groups spent most of their time looking at the computer screen (gold), and the maze (green).

Group 42 (high JVA and learning gains)		**Group 38** (low JVA and learning gains)	
B&W	Colored with AOIs	B&W	Colored with AOIs

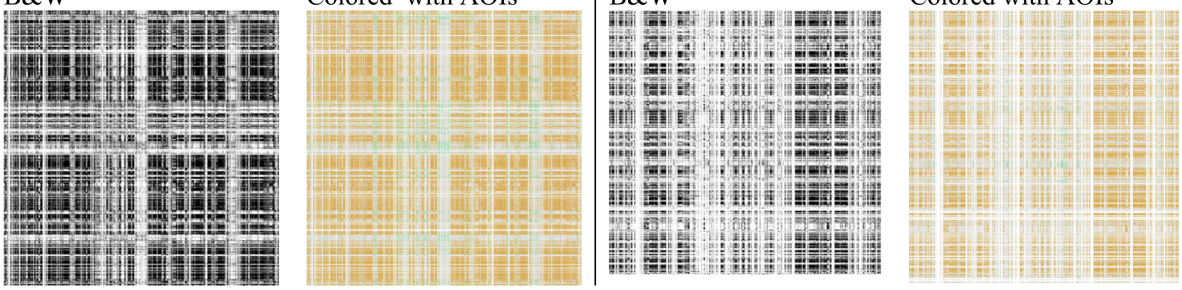

Figure 3. Cross-recurrence graphs for two dyads. The left side shows moments of joint visual attention (black), no joint visual attention (gray) and missing data (white). The right side shows join attention on particular AOIs (gold = computer screen, green = maze). The two graphs on the left show a productive group with high learning gains (42), and the two graphs on the right shows a group with low learning gains (38).

Joint Visual Attention

Joint Visual Attention (JVA) was computed according to prior research (Richard & Dale, 2015; Schneider & al., 2018; Schneider & Pea, 2013; Gergle & Clark, 2011). As a first pass, I used a radius of 100 pixels for two gaze points to be considered as JVA (this radius is shown on Figure 1). For each gaze point, I also checked whether the other group member looked at the same area within +/- 2 seconds (which has been shown to be the amount of time necessary for someone to disengage from what they are doing and pay attention to a partner's actions; Richardson). The results are summarized below.

I found significant correlations between JVA and: "sustaining mutual understanding": $r(36) = 0.397$, $p = 0.014$; "task division": $r(36) = 0.351$, $p = 0.031$; and their overall quality of collaboration: $r(36) = 0.341$, $p = 0.036$ (see Meier, Spada, & Rummel. 2007 for a definition of those constructs). There was no significant correlation with task performance or learning gains. It should be noted that I also looked at other radius sizes in addition to 100 pixels (50, 150, and 200 pixels) which defined the distance between two gaze points to be considered as a moment of joint visual attention. In these analyses, the size of the radius did not influence the correlations reported above.

Cycles of collaboration (JVA) and individual work (no-JVA)

For those analyses, I look at the number of times participants shifted between collaborative work (i.e., with increased levels of joint visual attention) and individual work (i.e., with lower levels of joint visual attention). I tried several approaches and found that the following steps provided the most conclusive measure: 1) I summed the number of moments of JVA for different time windows (a 60 second time window is shown on Fig. 4); 2) I compared each observation with the previous point and looked at whether JVA was going up or down; 3) I counted the number of times the group shifted from increasing to decreasing (and vice versa) their levels of JVA. In other words, this measure offers an estimate for cycles of individual and collaborative work.

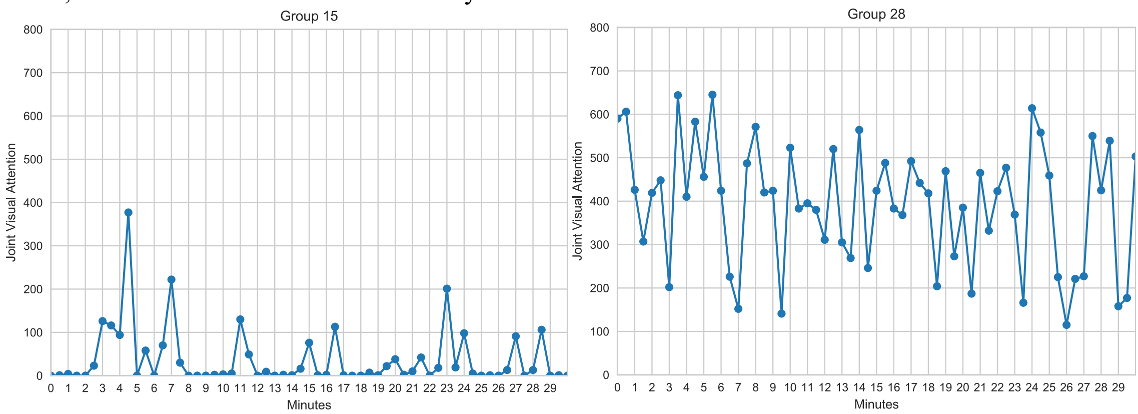

Figure 4. Levels of joint visual attention over time for two groups. Group 15 had the lowest number of cycles according to the measure above (12) and group 28 had the larger number of cycles (24).

This measure was correlated with learning gains using 30 seconds increment (i.e., 30sec., 60sec., 90sec., 120sec.). I found significant correlations with a 30 second window: $r(35) = 0.349$, $p = 0.035$, 90 sec. window: $r(35) = 0.355$, $p = 0.031$, 120 sec. window: $r(35) = 0.515$, $p = 0.001$ but not with a 60 sec. window: $r(35) = 0.001$, $p = 0.99$. Finally, by looking at smaller time windows between 10sec and 60sec., there was a time window (40 sec.) that was significantly correlated with all three dependent measures: overall quality of collaboration $r(34) = 0.347$, $p = 0.038$, Task Performance $r(34) = 0.355$, $p = 0.034$ and Learning gains $r(35) = 0.398$, $p = 0.015$. There was no significantly correlation between the aggregated JVA measure reported in the section above (i.e., the total amount of JVA) and the measure described in this section - which suggests that they are capturing two different constructs. I discuss these results below.

Discussion

This paper replicates prior results showing that JVA is positively correlated with high quality collaborative interactions (Jermann, Mullins, Nüssli & Dillenbourg, 2011; Schneider & Pea, 2013; Schneider, Sharma, Cuendet, Zufferey, Dillenbourg & Pea, 2018). It also provides further evidence that JVA can be computed in a co-located setting using fiducial markers disseminated in the environment. Since most of the prior work was done in remote settings, it is timely that new approaches allow researchers to study collaborative learning in more ecological ways. The final and main contribution of this paper is a new measure that captures cycles of collaboration and individual work in dyads. This measure provides a complementary lens into collaborative processes: I found JVA to be positively associated with collaboration quality, and this new measure with learning gains (as well as task performance and collaboration quality, depending on the threshold used). This suggests that an important feature of successful collaborative learning groups is to balance individual cognition with group work. While this is beyond the scope of this paper, future work will study this effect in more detail by qualitatively analyzing dyads that are driving this effect (i.e., groups with low learning gains and low scores on this measure, and dyad with high learning gains and high scores on this measure). Additionally, I am planning to replicate the findings above on a different dataset, which would provide further evidence that cycles of collaboration and individual work positively contribute to learning.

Conclusions

This paper presents a study where dyads of participants worked on programming a robot to solve a variety of mazes. I found that Joint Visual Attention can be captured using dual eye-trackers in co-located settings, and that this measure is positively correlated with collaboration quality. Additionally, I designed a new measure intended to capture cycles of individual work and group collaboration. This measure shed a new light on what constitutes productive interaction in dyads of students.

It should be acknowledged that this paper has several limitations (for a discussion of the limitations related to the task and the dependent measures used, please refer to Starr, Reilly & Schneider, 2018). First, this paper mostly relied on correlations. While this provides intuitive results, future work should use more comprehensive statistical tests to model participants' interactions and control for collinearity. Second, the new measure presented in this paper relies on several parameters (minimum distance between two gazes to qualify as joint visual attention, different time windows) that were arbitrary defined. A better understanding of how those parameters need to be fine-tuned is important for generalizing this measure to other settings. Finally, dual eye-tracking only offers a limited view of collaborative processes. Future work will integrate sensor data from multiple modalities (e.g., electrodermal, motion and speech data) to get a more complete picture of what constitutes productive interactions in co-located settings.

In conclusion, this study shows that it is possible to develop new ways of capturing 21st century skills in hands-on tasks typical of makerspaces. Even with the limitations mentioned above, this work makes a first step in this direction and opens the way to more rigorously studying collaborative processes in open-ended learning environments using dual mobile eye-trackers.

References

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, *19*(1–2), 205–220.

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.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC, US: American Psychological Association.

D'Angelo, S., & Begel, A. (2017). Improving Communication Between Pair Programmers Using Shared Gaze Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 6245–6290). New York, NY, USA: ACM.

Dede, C. (2010). Comparing frameworks for 21st century skills. *21st Century Skills: Rethinking How Students Learn*, *20*, 51–76.

Gergle, D., & Clark, A. T. (2011). See What I'M Saying?: Using Dyadic Mobile Eye Tracking to Study Collaborative Reference. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (pp. 435–444). New York, NY, USA: ACM.

Jermann, P., Mullins, D., Nüssli, M.-A., & Dillenbourg, P. (2011). Collaborative Gaze Footprints: Correlates of Interaction Quality. *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings.*, *Volume I-Long Papers*, 184–191.

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.

Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science, 10* （2), 151-177.

Richardson, D. C., & Dale, R. (2005a). Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science*, *29*(6), 1045–1060.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The Art of Conversation Is Coordination Common Ground and the Coupling of Eye Movements During Dialogue. *Psychological Science*, *18*(5), 407–413.

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, *8*(4), 375–397.

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2018). Leveraging Mobile Eye-Trackers to Capture Joint Visual Attention in Co-Located Collaborative Learning Groups. *International Journal of Computer-Supported Collaborative learning*.

Starr, E., Reilly, J., & Schneider, B. (2018). Toward Using Multi-Modal Learning Analytics to Support and Measure Collaboration in Co-Located Dyads. *ICLS2018 Conference Proceedings., Volume I-Long Papers,* 448–455.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

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

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS One*, *8*(11), e79659.