Augmenting Qualitative Analyses of Collaborative Learning Groups Through Multi-Modal Sensing

Bin Xie, Joseph Reilly, Yong Li Dich, and Bertrand Schneider
bix175@student.hks.harvard.edu, josephreilly@g.harvard.edu, ydich@college.harvard.edu, bertrand_schneider@gse.harvard.edu
Harvard University

Abstract: In a previous study (N=84), we collected information about dyads who worked on an engineering task typical of makerspaces: programming a robot to solve mazes of increasing difficulty. We collected multimodal data using a variety of sensors, including mobile eye-trackers, galvanic skin response, motion sensors and audio / video streams. In this paper, we contrast two pairs that exhibited positive and negative learning gains. We first detail multimodal measures to compare differences and similarities across those groups, and then dive deeper into a qualitative analysis of their exchanges. We then describe how those measures could be used over the entire sample to capture productive interactions in small groups. We conclude by discussing how process data from sensors can augment traditional qualitative observations, and how it can create powerful synergies for better understanding collaborative interactions among learners in settings such as makerspaces.

Introduction

Supporting STEM learning (Science, Technology, Engineering, Mathematics) has become a primary focus of the Learning Sciences over the past decade. There is also a growing interest to understand how we can teach 21st century skills within those domains (e.g., Collaboration, Communication, Critical thinking, Creativity). The combination of those factors has contributed to the popularity of makerspaces. Makerspaces are informal learning environments where students learn complex concepts in STEM by building their own artifacts using digital fabrication tools (e.g., laser cutters, 3D printers, robotics). We are interested in understanding what promotes learning in those spaces - especially from a socio-constructivist perspective (Palincsar, 1998). We presuppose that social interactions are among the main drivers of learning, because students spend a significant amount of time interacting with their peers and facilitators. This paper is about conducting a multimodal analysis of a typical makerspace activity, and isolating factors that contribute to productive collaborations. More specifically, we isolated two pairs from a larger study (Starr, Reilly & Schneider, 2018) and are qualitatively analyzing their interactions. The main contribution of this paper is that we are leveraging methods from the field of Multi-Modal Learning Analytics (MMLA; Blikstein & Worsley, 2016) to support our qualitative observations and helps us generate measures of productive interactions in small groups.

The paper is structured as follows: first, we conduct a literature review of indicators of collaboration from a multimodal perspective (physical, physiological and visual synchronization). Second, we summarize the study and present the two pairs that we are contrasting. Third, we analyze those two groups using a variety of qualitative and quantitative methods. We leverage sensor data to augment our two case studies using data from eye-trackers, a motion sensor and wristbands capturing electrodermal activity. Fourth, based on those analyses we present measures of synchrony that we plan to extend to our entire sample of 42 pairs. We conclude by summarizing our results and discussing next steps.

Literature review

As a first step and for the scope of this paper, we are focusing on measures of synchronization in small groups. We review three kinds of synchronization that could characterize productive interactions: physical, visual and physiological. Because of space limitations, we only discuss the main contributions of each domain.

Physical synchronization

The synchrony of physical movements within groups using multimodal learning analytics is an emerging aspect of research on collaboration. Behavioral coordination between group members is generally indicative of positive outcomes and has been studied extensively (Pentland & Heibeck, 2008.) This type of qualitative analysis, however, is time-consuming and requires expert knowledge of gestures to code correctly. Using sensor data and computational methods, Worsley and Blikstein (2013) have pioneered new ways to study embodied learning and found that experts in a construction task are more likely to use both hands in a synchronized fashion and that this bimanual coordination predicted expertise. Similarly, studies have shown that learning gains can be predicted by the amount of time students spend in certain postures (Schneider & Blikstein, 2015) and that the most productive posture involves both hands being synchronously engaged in the activity. This work was extended to look at dyad interactions, with the "driver" consistently using both hands more frequently while the "passenger" asynchronously moved their hands. Other MMLA work has shown that body posture during computer-supported activities can be predictive of learning (Grafsgaard et al., 2014) and theory suggests that increased body synchronization is associated with higher quality collaboration (Chartrand & Bargh, 1999). In summary, there is some emerging evidence that physical synchronization can be indicative of productive social interactions.

Visual synchronization
Eye-trackers have been used to study joint attention in collaborative learning situations. Richardson et al. (2007) showed that building upon a mutual source of understanding --"mutual grounding"-- (i.e., hearing the same background information before the task) positively influenced the visual attention coordination in spontaneous discussions. More related to this particular study, Jermann et al. (2001) used synchronized eye-trackers to assess the degree of collaboration as programmers worked together on a segment of code. By a comparison of a 'good' and a 'bad' dyad, the study suggested that high joint visual recurrence is strongly related with collaboration. Nüssli (2009) showed that models of group behavior can be built with a combination of eye-tracking and other data: the combination of gaze and raw speech data (voice pitch and speed) afforded predictions of participants' success with an accuracy rate of up to 91 %. Lastly, Brennan et al. (2008) conducted a spatial search task and studied the effect of shared gaze and speech during the experiments; they concluded that the shared gaze condition surpassed solitary search by twofold in terms of speed and efficiency and was the most optimal of all the conditions. Consolidating the results from the above studies, we can see that joint attention and in turn synchronization between individuals are crucial for high-quality collaborations. The results suggest eye-tracking as a salient method for understanding factors that contribute to effective collaborations.

Physiological synchronization
Recently, researchers have started to study collaborative groups using electrodermal sensors. Electrodermal activity (EDA; also referred to as galvanic skin response, GSR) measures the amount of sweat produced by the sympathetic nervous system and is an indication of physiological arousal. By using synchronized EDA sensors, one can measure whether group members are aroused at the same time, or exhibit some levels of desynchronization. Early work by Pijeira-Díaz, Drachsler, Järvelä and Kirschner (2016) looked at different measures of physiological coupling indices (PCIs), and found that Directional Agreement (DA) predicted learning gains while Instantaneous Derivative Matching (IDM) was related to the quality of the produced artifact. In summary, there is some preliminary evidence that physiological synchronization can capture a facet of productive collaborations.

Summary
In sum, educational researchers are starting to use various kinds of sensors to capture facets of a productive collaboration. There is data suggesting that physical, visual and physiological synchrony can be used a proxy to collaboration quality. In the next section, we describe our study where we measured those states using a Kinect sensor, two mobile eye-trackers and two wristbands capturing participants' electrodermal activity.

General description of the study
42 pairs of participants (N=84) programmed a robot to navigate a series of increasingly complex mazes (see (for more details on the study, see Starr, Reilly & Schneider, 2018). Participants were shown two tutorial videos to acquaint them with the basics of block-based programming and how to use the values from sensors on the robot in their code. Groups were told to come up with a general solution that could solve any simple maze and then had 30 minutes to complete as many of the mazes as possible. Two different interventions were implemented to support collaboration in a two-by-two between-subjects design resulting in four different conditions (presence / absence of): 1) a visualization of the amount of individual verbal contributions as a proportion of total verbalization (Fig. 1, top left corner of the right picture; referred to as VISUALIZATION henceforth); 2) a short verbal explanation of the benefits of collaboration for learning (referred to as EXPLANATION henceforth). Outcomes of interest included the quality of the code groups produced (evaluated on a zero to four scale to determine how well the code in abstract could perform the maze solving task), the number of mazes solved, gains on a learning test administered before and after the sessions, and the quality of their collaboration. Multimodal data was collected via two mobile eye-trackers, two bracelets tracking electrodermal activity, a motion sensor, and video recording (Fig. 1).

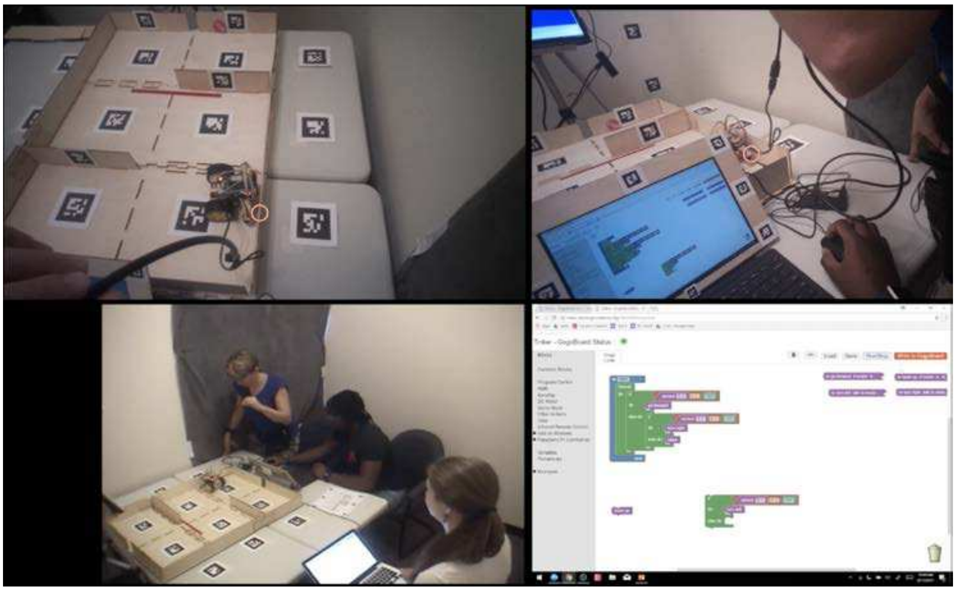


Figure 1. A frame from the video used for the qualitative analyses of this paper. The two top images show the perspective of the participants (captured by the mobile eye-trackers) and the location of their gazes (red circle). The bottom left view shows the main video feed of the session, and the bottom right view displays a screen capture of the laptop. In this frame, group 8 was programming the robot to navigate an S-shaped maze.

Data analysis: Contrasting group 7 and 8

The goal of this paper is to analyze two dyads in more depth and design measures that will allow us to contrast good versus poor collaborative styles across the entire sample. We chose to focus on groups 7 and 8 because of the stark differences in their behaviors. Group 8 (EXPLANATION, VISUALIZATION) was among the top groups in our sample: participants had a productive collaboration where group members built on each other's ideas and exhibited positive learning gains. The participants in group 7 (EXPLANATION, NO VISUALIZATION), on the other hand, exhibited lower scores on all our metrics, resulting in periods of silence and miscommunications as well as negative learning gains.

Traditional quantitative measures

Group 7 was comprised of a 50-year old male (7L) and a 26-year-old female (7R). Both self-reported "Some College" for level of educational attainment and 7R indicated she was currently a student. 7L scored 8.3 percentage points worse on the post-test for computational thinking skills, indicating some confusion about the concepts required for the task. 7R gained 20.9 percentage points between pre and post, suggesting a much better grasp of the material after completing the activity. Participants in group 7 were able to direct the robot successfully through one maze but their final code failed to nest conditional statements and did not use the pre-written functions correctly (Fig. 2, left side). In a written reflection section on the post test, 7L had the following to say about his time with the activity: "…I have no talent for programming. I did not have any breakthrough moments. I would not do this type of study again. My ideas did not change over time and I felt that I did not learn much about computers."

Group 8 was comprised of a 25-year old female (8L) and a 35-year-old female (8R). Both reported completing college and identified as no longer being students. 8L scored 16.7 percentage points higher on the post test of computational knowledge, indicating a modest improvement. 8R scored no better on the posttest compared to the pre, although she did make different errors. This suggests a level of confusion related to certain topics in computational thinking. They were also only able to complete the simplest maze but their code made use of nested conditional statements and correctly employed the prewritten functions they were given (Fig. 2 - right side). In a written reflection section on the post test, 8L had the following to say about his time with the activity: "We tried playing around with the different sensors. We started trying sensors 1 and 2, but then realized that using sensor 4 was necessary to complete the task. That was our "a-ha" moment. We tried changing the order of the if/else/do functions to get different results that helped advance our knowledge of the task."

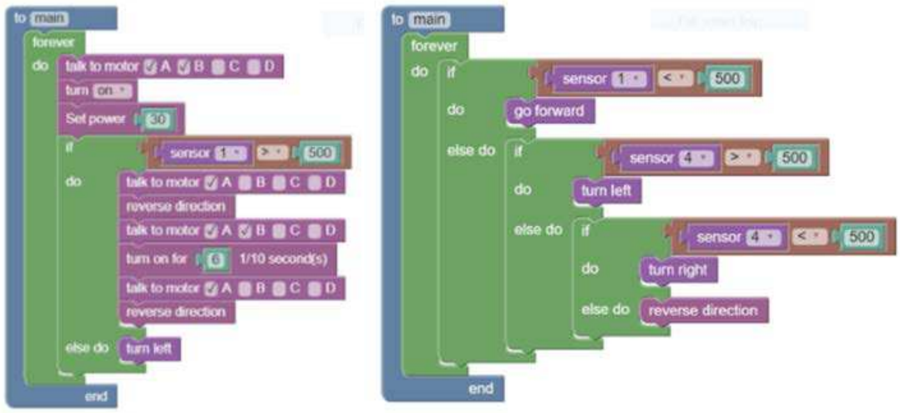


Figure 2. Final code produced by Group 7 (left) and Group 8 (right).

These two groups were selected due to their similarities of having one more knowledgeable participant, both completing the same number of mazes, and both having complete multimodal data to analyze. Exploring the difference in amount of collaboration observed and how that relates to the quality of the written code is a main goal of this work as well as identifying multimodal markers to signal quality of collaboration.

Qualitative data

Before delving into the overlay between qualitative data, eye-tracking data, and physiological data, we will overview the key themes seen between Groups 7 and 8's qualitative data. To obtain this data, subjects' experiments were recorded on video and an iterative process was used to note the various data points for qualitative data.

Group 7 engaged much less with each other than Group 8. There was little rapport built between the Group 7 members, whereas Group 8 laughed as early as the initial calibration color-reading exercise and continued throughout the experiment (17 times for Group 8 versus once for Group 7), especially after trials and during coding discussions. The length of dialogue in Group 7 was noticeably shorter than that of Group 8, and the content exchanged within Group 7 was not as detailed as Group 8. Even when Group 7 discussed more detailed code, it was a leader-follower response with the subject on the right saying most of the language and the subject on the left saying "mhmm" or "ok" and followed by heavy sighing. This type of dialogue impacted Group 7 negatively as the team was not able to understand the task at hand and dialogue came to a pause frequently through the experiment.

By contrast, Group 8 was almost the exact opposite of Group 7. Building on each other's rapport minutes into the experiment, the two subjects were able to speak specifically about each section of the code. Each subject switched between making suggestions and verifying the assertions. The pair spoke at equal lengths throughout the experience, and often asked each other questions directly related to the task at hand. Unlike Group 7, which consistently displayed confusion via repeating statements like "I don't know," Group 8 displayed encouraging enthusiasm with short exclamations to relieve stress even as it recognized rising task difficulty throughout the experiment. For instance, about halfway through task 3, right subject suggested adding code to tell the robot to reverse direction, which is responded by an enthused left subject, "right!" In another instance, the robot did not make an intended turn, but turned in an opposite direction. The dyad exclaimed in surprise, but the right subject said, "well we got it to turn left, so that's promising!" Lastly, Group 8 had many instances of "mhmm," "ok," and "does that make sense" language that were spoken by the dyad with friendly tone, versus the resigned tone of Group 7 for the same words. It is worthy to note that Group 7 did not have an equal split of reciprocal filler words, as the left subject were the one who said most of such language.

Group 8 also visibly and often celebrated for their successes, which did not happen for Group 7. Forms of celebration was most commonly displayed via loud exclamations like "whoo!", high-fives, laughing and clapping. These observable signs provide grounds to believe that reinforcing signals of goodwill such as frequent check-ins and friendly body language and enthusiastic tones build rapport that help the team sustain collaboration as the task difficulty increases. Another key difference that set Group 8 apart is the frequency of the iterations. Building upon the alternating role of suggesting and verifying, the dyad was willing to make adjustments to the code, test it, and come back to the coding platform to improve the code if the trial failed. Combined with frequent acknowledgement of each other--including looking at each other--the dyad was able to repeat the iterations numerous times, become increasingly familiar with the interface and be more hands-on than Group 7.

Eye-tracking data

Our first pass at analyzing visual synchronization involved applying the taxonomy developed by Kaplan and Hafner (2006), where they define joint visual attention as: simultaneous looking triggered by a salient event, simultaneous looking triggered by a “pop-out” effect, coincidental simultaneous looking, gaze following, or coordinated gaze on same object (Table 1). We stayed largely consistent with the five-category approach from Kaplan and Hafner (2006)’s hierarchy for generating our eye-tracking data. We decided to use this hierarchy because it was pertinent for gauging joint attention between the two agents--the partners in our study--and in turn the synchronization between the dyad. We used the ELAN software to track the eye-tracking data through annotating the occurrence and duration of the various categories of eye-tracking data as relevant data points appear in the experiment videos (Fig. 3). In the future, we will use automated ways of capturing joint visual attention using the fiducial markers tapped at various locations in the room.

Table 1: Categories used in the eye-tracking data

Category	Hierarchy Shorthand	Example	Group 7	Group 8
Simultaneous looking triggered by a salient event	1 - Simultaneous looking_Salient	Agent 1 points finger at the screen, Agent 2 looks at the screen	19	52
Simultaneous looking triggered by a “pop-out” effect	2 - Simultaneous looking_PopOut	Agent 1 looks at a different color of code block because its color stands out	1	0
Coincidental simultaneous looking	3 - Simultaneous looking_Coincidental	Agents 1 and 2 both looks for the robot, sees the robot at the same time but has no interaction with one another	7	13
Gaze following	4 - Gaze Following	Agent 1's gaze follows that of Agent 2	44	59
Coordinated gaze on same object	5 - Coordinated Gaze_SameObj	Both agents look at the same object knowing the other agent is looking at it as well	62	142

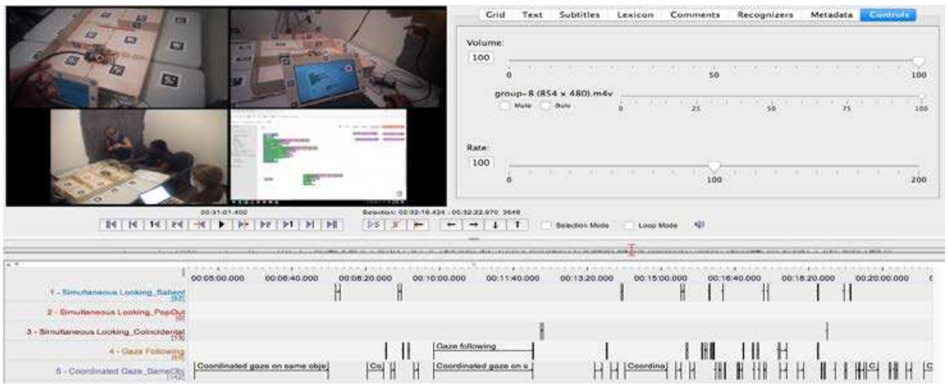


Figure 3. Screenshot of eye-tracking data for group 8 as seen in ELAN annotations, on the bottom half of the screen by category (i.e., second column of Table 1). Lengths of the annotated simultaneous gazes can be seen.

Overall, there were more coordinated gaze on the same object than other categories, followed by gaze following and then salient event. This is consistent with expectation for the study, as the dyad conducted most of its interactions via sedentary coding. The salient events were mainly caused by one subject pointing at the screen or the guide sheet provided. Gaze following usually occurred after salient events and before coordinated gazes. When gaze following overlaps with another category in sequence, the total combined gaze duration is shorter for the salient event than for the coordinated gaze.

For both dyads, most coordinated gazes longer than 10 seconds happened during task 3 (69% for Group 7 and 75% for Group 8), indicating the relative scale of attention required to complete a more difficult task. The salient event gazes did not last long for either group due to the nature of the gaze. What was different was that Group 8 had a mostly even distribution of salient event gazes through the different task periods, while Group 7's salient events were fewer and more clumped than of Group 8 (19 for Group 7 versus 32 for Group 8). Given that most of the salient events occurred because of finger pointing to attract attention of the subject's partner, this suggests that Group 8 may have had more interactions with one another.

Additionally, eye-tracking results indicate further differentiation of Group 8 from Group 7. Group 8 had the longest joint eye gazing, and more often overall than Group 7 (142 coordinated gazes for Group 8 versus Group 7's 62). Coordinated gaze on the same object composed of 53% of Group 8's total eye-tracking data, and of that category 55% of the gaze durations surpassed 10 seconds long, compared to Group 7's 11%. Finally, Group 8 had more overall eye-tracking appearances, with 266 tracked points to Group 7's 133.

Comparing the relative frequency of appearance of each category across groups, Group 8 dominated across each category except gaze following. Group 7 had a great proportion of its eye-tracking events as gaze following when compared against Group 8. This suggests that simple gaze following may not contribute to the effectiveness of team collaboration--more active feedback (as measured by various types of eye movements and other data) had to be exchanged between the partners to create more meaningful interaction.

The above data suggests that Group 7 is not as synchronized as Group 8, with fewer coordinated gazes, fewer overall gazes, and shorter gazes than Group 8. In a task like pair programming, coordinated gazes are increasingly important as tasks become more complex. The drastic difference between Group 7 and Group 8 lead us to seek further validation in qualitative data to see if further patterns can be seen.

Physiological data

Tying the themes described above to EDA spikes, Group 7's physiological arousals did not sync across the dyad and generally the state of physical arousal decreased throughout the experiment for one subject while for the other subject the EDA levels remained about the same (Fig. 4, left side). The spike in the EDA of the subject towards the right-hand side of the screen observed when the subjects sat and watched an instructional video. 40 seconds after the start of the activity, the left-hand subject had an EDA spike as he looked at the moderator when she was providing the pair directions. Neither of these events were related with the tasks at hand. The only relevant arousal happened 10 minutes after the beginning of the activity, when the right subject entered into an agreement period in which she was narrating the logic of the code to the left subject. The pair also looked at each other for the first time. Despite this, it's clear that the left subject did not match the EDA state as right.

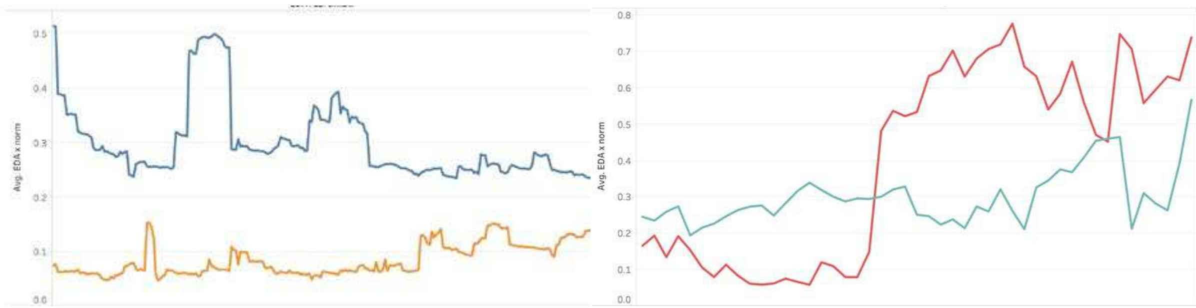


Figure 4. EDA graph, normalized, for group 7 (left side) and group 8 (right side). Indices of synchronization can be computed using measures described by Pijeira-Díaz, Drachsler, Järvelä, & Kirschner (2016).

Group 8's physiological data showed that the dyad had a higher level of synchrony (Fig. 4, right side). In particular, about five minutes after the 5th tagging procedure, the dyad ran their robot through the first maze (task 2), and upon the robot's success the right subject celebrated with a "whoo" and raised arms. This was picked up by the EDA as the intersection of the EDA measures of left and right subjects. The pair continued onwards through the remainder of the experiment in active EDA syncing, showing at least three visible intersection points starting from 20 minutes after the 5th tagging procedure, per Figure 4. The overlaps increased in frequency towards the end of the session, further supporting earlier eye-tracking and qualitative data showing the quality collaboration within the dyad. During this period, there were numerous iterations of coding and running the robot, often ending in laughter. The spikes in EDAs are mainly explained by these as well as the frequent standing by the dyad for their repeated robot retrials.

Preliminary quantitative analyses

In this section, we describe a first attempt at capturing physical and physiological synchrony between group 7 and group 8. Our strategy was to synchronize the data between group members and produce a scatter plot, where values on the x-axis are shown for the first participant and values on the y-axis are shown for the second participant. We can then roughly subdivide a graph into four quadrants: the bottom left represents when both participants were exhibiting low levels of physiological activation or movement; the top right shows when they were both aroused or moving. The last two quadrants (top left or bottom right) indicate some levels of desynchronization: one group member has high values while the other participant has low values. Figure 5 shows that group 7 exhibits a pattern that is L-shaped as well as negative correlations (Fig. 5, left side), while group 8 tends to have points that are more evenly distributed - which is captured by positive correlations. We are planning to apply those measures to our entire sample to confirm those results.

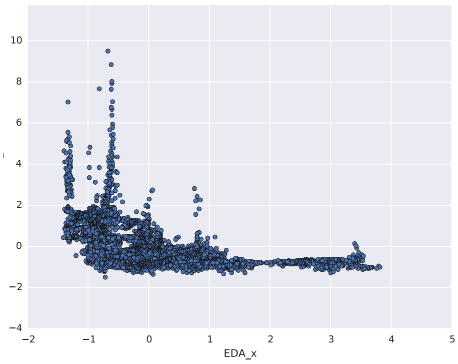
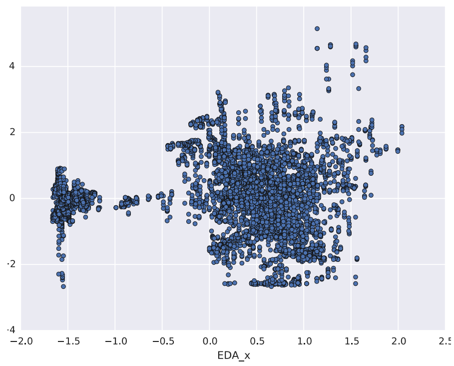
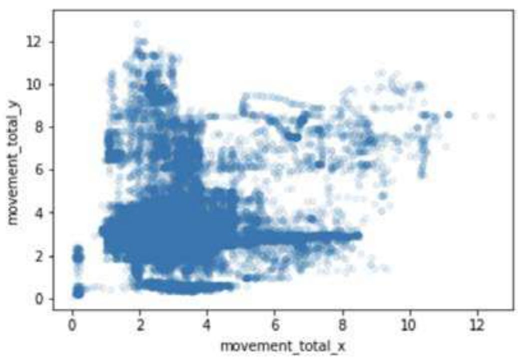
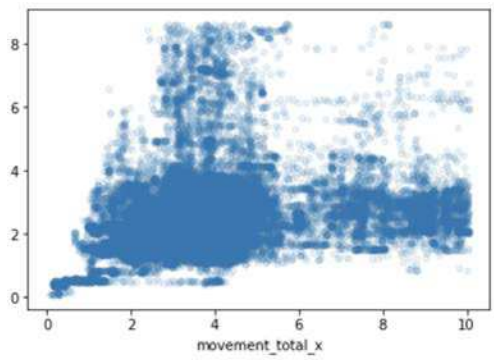
	Group 7	**Group 8**
EDA	 r = -0.047, $p < 0.05$	 r = 0.077, $p < 0.05$
Kinect	 r = -0.046, $p < 0.05$	 r = 0.179, $p < 0.05$

Figure 5. Group 7 is on the left side, group 8 is on the right side. First row shows synchronized EDA data (the first participant is on the x-axis; the second participant is on the y-axis). Second row shows the amount of movement generated by each participant on each axis.

Discussion

In this paper, we contrasted two groups sampled from a larger study (N=84) using qualitative and quantitative methods. Qualitative analyses suggested that group 7 had more issues working together and accomplishing the task, while group 8 was more successfully and enjoyed the task more. The more collaborative dyad had much more detailed language and frequent, specific interactions, but also developed rapport through body language, mutual gazing, and frequent acknowledgement of each other. Combined with the use of keywords and tones that signaled positive intention, the more collaborative dyad was able to weather the stresses of completing difficult task and maintain task engagement as one unit. Physiological and eye-tracking data further validated our observational data, providing the multimodal view of team collaboration. Eye-tracking data showed that frequent simultaneous gazes and longer gaze length were characteristic of the more collaborative group, which eventually accomplished the tasks given. This provides support that eye-tracking data can provide a good gauge for high-quality group collaboration (as previously shown by Jermann, Mullins, Nuessli, & Dillenbourg, 2001). Gaze-following is also important in indicating reciprocity of the dyad in collaborative spaces. However, simple gaze following is insufficient in contributing to collaboration. Additionally, electrodermal data confirmed those observations by showing elevated activation from the more successful group and helped us identify events of interest (e.g., by analyzing "spikes" in the data). Finally, those qualitative analyses helped us design measures of synchrony in small collaborative groups: we found that bodily and physiological synchronization should be further studied by extending the results found in Fig. 5 to the entire sample and confirming whether it can be used as a proxy for identifying successful groups.

Conclusion

In this paper, we have presented a new way of studying collaborative learning groups by using a combination of qualitative observations and data from high-frequency sensors. Our preliminary analyses suggest that Multi-Modal Learning Analytics (Blikstein & Worsley, 2016) can help us shed new lights in collaborative learning processes, especially in open-ended learning environments such as makerspaces. In the future, we are planning to further explore differences between groups in our study and develop multimodal proxies of collaborative interactions using high-frequency sensors. Those proxies could then be used by teachers and practitioners in informal learning environments to support the development of 21st century skills, especially in terms of students' ability to work effectively in small groups.

References

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, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*(3), 1465–1477.

Chartrand, T.L. & Bargh, J.A. 1999. The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910.

Grafsgaard, J. F., Wiggins, J. B., Vail, A. K., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2014, November). The additive value of multimodal features for predicting engagement, frustration, and learning during tutoring. In Proceedings of the 16th International Conference on Multimodal Interaction (pp. 42-49). ACM.

Jermann, P., Mullins, D., Nuessli, M.-A., & Dillenbourg, P. (2001). Collaborative gaze footprints: correlates of interaction quality. In Spada, H., Stahl, G., Miyake, N., & Law, N. (Eds.), *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings*, Hong Kong, July 4–8, 2011, Volume I - Long Papers, (pp. 184–191).

Kaplan, F., & Hafner, V. V. (2006). The challenges of joint attention. *Interaction Studies*, *7*(2), 135-169.

Nüssli, M. A., Jermann, P., Sangin, M., & Dillenbourg, P. (2009). Collaboration and abstract representations: Towards predictive models based on raw speech and eye-tracking data. In *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning* (pp. 78–82).

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual review of psychology, 49*(1), 345-375.

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.

Pentland, A., & Heibeck, T. 2008. Honest signals. Cambridge, MA: MIT press.

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.

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.

Roth, W. M. (2001). Gestures: Their role in teaching and learning. *Review of educational research*, *71*(3), 365-392.

Schneider, B., & Blikstein, P. (2015). Unraveling students' interaction around a tangible interface using multimodal learning analytics. *Journal of Educational Data Mining*, *7*(3), 89-116.

Starr, E., Reilly, J., & Schneider, B. (2018). Using Multi-Modal Learning Analytics to Support and Measure Collaboration in Co-Located Dyads. *12th International Conference of the Learning Sciences*.

Worsley, M., & Blikstein, P. (2013, April). Towards the development of multimodal action based assessment. In Proceedings of the third international conference on learning analytics and knowledge (pp. 94-101). ACM.

