Identifying Collaborative Learning States Using Unsupervised Machine Learning on Eye-Tracking, Physiological and Motion Sensor Data

Karina Huang
Harvard University
qhuang@g.harvard.edu

Tonya Bryant
Harvard University
tonya_bryant@gse.harvard.edu

Bertrand Schneider
Harvard University
bertrand_schneider@gse.harvard.edu

ABSTRACT

With the advent of new data collection techniques, there has been a growing interest in studying co-located groups of students using Multimodal Learning Analytics [3] to automatically identify collaborative learning states. In this paper, we analyze a multimodal dataset (N=84) made of eye-tracking, physiological and motion sensing data. We leverage unsupervised machine learning algorithms to find (un)productive collaborative states. We found a three-states solution where different states (and transitions between them) were significantly correlated with task performance, collaboration quality and learning gains. We interpret these findings in light of collaborative learning theories and discuss their implications for studying groups of students using MMLA.

Keywords

Multimodal Learning Analytics; Eye-tracking; Physiological Sensing; Motion Sensing; Unsupervised Machine Learning.

1. INTRODUCTION

The last decade has seen educational researchers go beyond the study of conceptual learning to understand non-cognitive skills. These skills are considered central for preparing students for the challenges of the 21st century and turning them into resilient, creative, curious and collaborative individuals. Additionally, new learning environments are becoming popular to foster those skills, such as makerspace and digital fabrication labs. While there has been some important progress made in the study of 21st century skills, measuring and assessing them remains a challenge. Most educational researchers and practitioners still rely on traditional collection tools such as participant observations and in-depth interviewing. While research strategies provide valuable insight into learning and development, they are no longer the most efficient way of collecting data.

With the advent of new data collection techniques, however, there has been a growing interest in capturing 21st century skills using Multimodal Learning Analytics (MMLA; [3]). MMLA is about using high frequency sensors, such as eye-trackers, motion sensors, physiological devices and brain sensors to capture students' learning trajectories. Additionally, by combining multiple sensors it is possible to study collaborative learning groups and capture various aspects of productive collaborations. Traditionally, these sensors have been studied in isolation. The promise of MMLA is to combine multimodal data sources to capture a more holistic picture of students' learning. Being able to capture 21st century skills [6] (such as collaboration) in real time opens new opportunities for providing feedback and designing new kinds of interventions to teach these skills.

The paper is organized as follows. First, we review the literature on several constructs related to collaborative skills that can be captured using high frequency sensors (e.g., Joint Visual Attention, Physiological Synchrony, body postures). We then describe the study that generated our dataset and detail how these constructs were measured. Finally, we present findings where we identified collaborative states using unsupervised machine learning algorithms and discuss their implications.

Karina Huang, Tonya Bryant and Bertrand Schneider "Identifying Collaborative Learning States Using Unsupervised Machine Learning on Eye-Tracking, Physiological and Motion Sensor Data" In: *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*, Collin F. Lynch, Agathe Merceron, Michel Desmarais, & Roger Nkambou (eds.) 2019, pp. 318 - 323

2. Literature Review

For decades, socio-constructivist theories have emphasized the importance of social interactions for learning (e.g., [12]). Among other things, collaborative learning can help students develop critical thinking skills, increase their motivation, provide a support system and facilitate assessment by making learning visible [10]. Capturing collaborative processes, however, remains a challenge – even though researchers have argued for almost a century that we need more rigorous ways to capture learning processes [23]. In the study of collaborative learning, Dillenbourg [8] argues that "empirical studies have started to focus less on establishing parameters for effective collaboration and more on trying to understand the role which such variables play in mediating interaction. This shift to a more process-oriented account requires new tools for analyzing and modelling interactions". Below we review how Multimodal Learning Analytics (MMLA; [3]) can help us make a first step in this direction. More specifically, we describe how dual eye-tracking, motion sensors and physiological sensors can provide fine-grained indicators of collaboration.

2.1 Joint Visual Attention and Dual Eye-tracking

Joint Visual Attention (JVA; [4]) is the most fundamental building block by which human beings coordinate their actions, establish a common ground, advance toward a common goal, solve problems, and learn together. It is a construct that encompasses numerous visual processes, and is observed as important for learning to socialize [4], engaging collaboratively [22], and developing social motivation [16] for diverse populations in varying collaborative conditions. The last decade has seen a small but growing number of researchers take advantage of synchronized eye-trackers to quantitatively measure gaze alignment in various collaborative situations interpersonal communication [14]. With the emergence of MMLA, quantifying gaze synchronization in remote learning and problem-solving environments has similarly popularized. In video lectures, projecting the professor's gaze onto the screen (as a substitute for the use of deictic gestures in co-located teaching environments) while making explicit references to information on slides can be useful for students and increase learning gains [20]. In co-located collaborative problem-solving situations, students' level of JVA has been found to be positively correlated to behaviors such as managing group dialogue, reaching consensus, and equally dividing work between members of the group [18]. Regardless of the context, JVA measurement and visualization tools are providing new ways to allow for objective inferences to be drawn about gaze synchronization as it relates to various collaborative states.

2.2. Body Postures and Motion Sensing

Students' use of their bodies has received a great deal of attention from learning scientists over the last decades. Numerous studies have unraveled links between students' understanding of various topics [1, 5] and specific gestures [15]. More generally, there has been a plethora of studies linking people's intuitive representations of everyday situations and bodily language (e.g., embodied cognition [2]). Recently, researchers have started using motion sensors to provide more fine-grained analyses of body postures in collaborative learning settings. For example, [19] found that hand movement could distinguish between students who were more dominant (called "drivers") and those who were more passive (called "passengers").

2.3 Physiological Sensors and Group Synchrony

Researchers have recently started to use EDA sensors (Electrodermal Activity) to look at collaborative learning interactions. [13] describes four measures of physiological synchrony in small groups of students: Signal Matching (SM), Instantaneous Derivative Matching (IDM), Directional Agreement (DA), and Pearson's correlation coefficient (PC). They found that IDM was related to collaboration quality and task performance, and DA with learning gains. In a separate publication, we applied the same methodology to the dataset described in this paper and found that those indices provided significant predictors for collaborative learning [7]. DA was significantly correlated with collaboration quality, IDM with task performance, and PC with learning gains. In a different study, [9] found that DA was the best predictor for task performance. It is interesting to note the discrepancy between the findings above, which is likely due to the nature of the task and the way we operationalized our constructs. But overall, researchers have found that physiological synchrony seems to be sensitive to social interactions in a variety of contexts.

In summary, there is significant evidence that collaborative learning processes can be captured using multimodal sensors. This paper goes one step further by combining modalities together, instead of studying them in isolation.

3. METHODS

3.1. The Study

The dataset used in this paper was collected as a part of a Multi-Modal Learning Analytics study [21]. 84 participants (Male = 40%; Female = 60%) with no prior programming experiences were randomly assigned to dyads ($N_{dyad} = 42$) and programmed a robot to solve a series of mazes during 30-minute sessions (Fig. 1). Each dyad was randomly assigned to one, both, or neither of two designed interventions: 1) a verbal explanation of the benefits of collaboration (e.g. past research findings using equity of speech time as indication of collaboration quality), and 2) real-time visualizations showing relative verbal contributions of each participant. The number of dyads was evenly distributed among four experimental conditions. For the analyses reported below, we analyze the aggregated data and did not consider the four experimental conditions.

During each session, we used two Empatica E4 wrist sensors to track participants' physiological activities, two Tobii Pro Glasses 2 eye-trackers to capture eye gaze, and one Kinect sensor to record movement as well as facial expressions. Participants were given a survey before and after the study for assessment of their computational thinking and collaboration experiences. A dyad's collaboration, task performance, and learning outcomes were assessed by the researcher responsible for running the session; ratings were given on nine scales based on prior work by Meier, Spada, and Rummel [11] (inter-rater reliability of 0.65 – i.e., 75% agreement). Analyses of learning gains, coding schemes, inter-judge reliability scores and individual sensor data have been reported in [21]. The current analysis aimed to combine all sensor data in order to identify collaborative states.

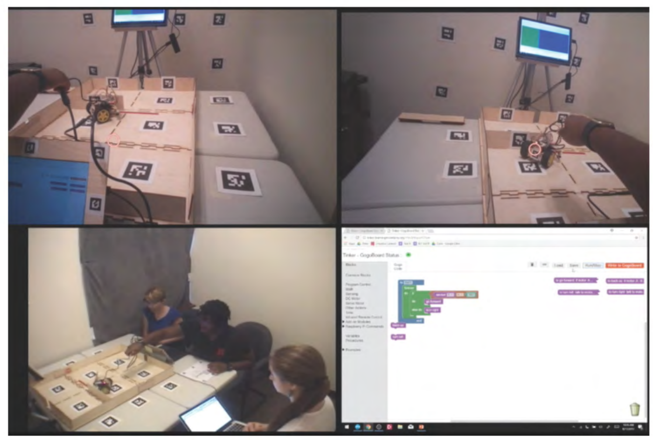

Figure 1. Two participants from the study. The top images show the video feed from the mobile eye-trackers (with a participant's gaze shown on the top right image). The bottom left image shows a 3rd person perspective. The bottom right image shows the programming environment.

3.2. Data Collection

3.2.1. Empatica

The Empatica E4 wristbands collected participants' accelerometer, blood volume pulse (BVP), interbeat intervals (IBI), electrodermal activities (EDA), and heart variability (HR). The current study focused on the EDA data, which is a measure of skin potential, resistance, conductance, admittance and impedance. Participants were asked to tag the wristbands before and after each step during the sessions (i.e. before and after completing the maze task). We synchronized the dyadic data according to the tags and timestamps of the sessions. The resulting data frame per dyad contained timestamp, four aggregated measures of physiological synchrony described in the Measures section below, and EDA values of each participant. Five dyads were removed from the current analysis because the sensor data was missing, too noisy, or identified as outliers (see [21]).

3.2.2. Tobii Pro Glasses 2

The Tobii eye-trackers generated data at 50Hz per second and recorded the x and y coordinates of each participant's eye-gaze relative to its point of view. The resulting data frame per dyad contained time indicated by second (ranging from roughly 1s to 1800s), the x and y coordinates of each participant's eye gaze and counts of joint visual attention (JVA) by pixel distance. The eye-tracking data was synchronized with the EDA data by briefly presenting a fiducial marker on the computer screen between each step of the session; participants were asked to tag this event on their wristband as accurately as possible. While we were able to clean and synchronize most dyadic data by seconds, two groups were excluded from the current study due to missing data.

3.2.3. Kinect

The Kinect motion sensor captured around 100 variables related to a participant's body joints and skeleton. The sensor generated data at 30 Hz per second, resulting in about 3,000 observations per second per participant. Noisy data (e.g. when the session facilitator entered the Kinect frame) were removed for each group, after which data were aggregated by second according to timestamps generated by the Kinect sensor. Researchers manually trimmed the Kinect data for each 30-minute session based on video records and aligned them with the eye-tracking and EDA data. Nine dyads were removed from the current study due to missing data.

3.2.4. Synchronizing All Data

The Empatica and Kinect data were synchronized by trimming each session to exactly 30 minutes and outer-joining sessions' data on the timestamp column. Per-dyad eye-tracking data were synchronized by matching the "second" column (i.e. from 1s to 1800s during a 30-minute session; see Fig. 2) generated based on timestamps of the EDA data. For analysis purpose, we concatenated all per-dyad data into a master data frame, with an additional column indicating to which session the data belonged to. Due to an unequal amount of data loss between sensors, two datasets were created for analysis: 1) Combined EDA and JVA ($N_{dyad} = 35$), and 2) Combined EDA, JVA and Kinect ($N_{dyad} = 31$). The current analysis used the second dataset, including 67,656 rows and 19 columns of by-second original and scaled data from all sessions investigated (Fig. 2).

	session	second	DA	PC	IDM	SM	jva100	moveDiff	headDiff	shoulderDiff
0	2	1	0.233333	-0.447563	0.177397	0.208709	24.0	1.941329	0.047057	0.658437
1	2	2	0.232708	-0.445837	0.177548	0.207804	7.0	2.013969	0.040868	0.654975
2	2	3	0.233125	-0.442000	0.177468	0.206289	18.0	1.825878	0.048828	0.642805
3	2	4	0.234583	-0.439776	0.176532	0.205172	23.0	1.675585	0.037022	0.648898
4	2	5	0.234375	-0.438995	0.175948	0.204015	6.0	1.967405	0.042931	0.646596

Figure 2. A snapshot of the final data frame. Note that the scaled column measures are excluded for legible visualization.

3.3. Data Processing

3.3.1. Electrodermal Activities (EDA)

Four measures of physiological synchrony were computed based on participants' electrodermal activities (refer to [7] for an exhaustive description of these measures and related analyses): 1) Pearsons's Correlation (PC) represented the linear relationship between the EDA level of each participant in a dyad; a strong, positive correlation indicated that the dyad was physiologically activated at similar times. 2) Directional Agreement (DA) captured whether the EDA level of each participant in a dyad increased or decreased at the same time steps; an increase in DA value in the positive direction indicated higher physiological synchrony. 3) Signal Matching (SM) was computed as the area between data curves of each dyad. A greater SM value indicated lower physiological synchrony. 4) Instantaneous Derivative Matching (IDM) computes, for each dyad, the level of signal matching between slopes of participants' signal curves. A higher IDM value indicated lower physiological synchrony between participants.

3.3.2. Joint Visual Attention (JVA)

JVA was qualified by looking at participants' location of eye gaze after mapping these coordinates into a common plane (Fig. 3; see [17] for the complete procedure of computing JVA). The current analysis looked at the number of JVA per second where a dyad's eye gazes were within 100 pixels of each other.

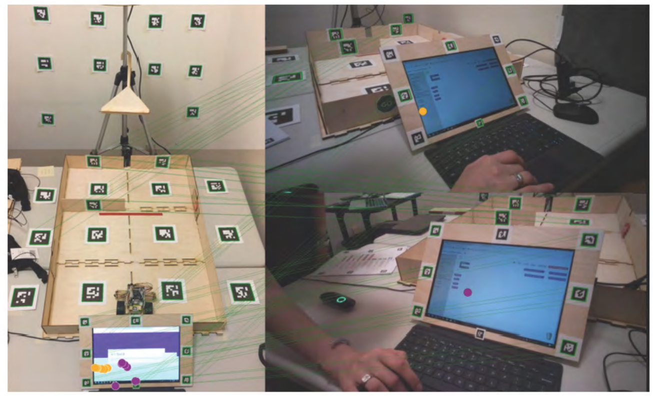

Figure 3. The procedure used to compute Joint Visual Attention (the left side shows the data from the mobile eye-trackers, and the right side shows how the participants' gaze were mapped onto a common plane using a homography).

3.3.3. Kinect

We explored various collaborative measures based on prior work [18, 23]. The current analysis aggregated three measures of movement differences per dyad: 1) Total difference in movement (MoveDiff), 2) Vertical difference in head orientation (HeadDiff), and 3) Horizontal difference in shoulder orientation (ShoulderDiff). Total movement was computed by taking the Euclidean distance of all joint coordinates; difference in movement within dyad was the absolute difference in the participants' total movements. Vertical difference in head orientation was the absolute difference in the y coordinates of participants' heads. Horizontal difference in shoulder orientation was calculated by taking the absolute value of the difference between 1) the absolute difference in x coordinate of the left shoulder of the left participant and the x coordinate of the right shoulder of the right participant, and 2) the absolute difference in the x coordinate of the right shoulder of the left participant and the x coordinate of the left shoulder of the right participant.

3.3.4. Outcomes Measures

The study generated three types of outcome measures: collaboration quality [11] (sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, individual task orientation, and Collaboration – the sum of those scores), overall task performance (task performance, task understanding, improvement over time) and learning gains. Collaboration and Task performance of each dyad was hand-coded by the experimenter at the end of the session. The dyad's learning gain was assessed through a pre-test and post-test. For more information, please refer to [21].

3.4. Analysis Strategy

We used K-Means Clustering with Euclidean distance to identify different collaborative states. In particular we attempted clustering using all sensor data simultaneously. All data were transformed into z-scores before clustering; the scaled values reported in results refer to the z-scores. Note that the clustering assignment was performed on the aggregated data with per-second data from all dyads. The collaborative states were identified regardless of group and time. We used the elbow curve to identify the optimal number of clusters for each clustering strategy; the current analysis used within-cluster sum of squared as the indication of distortion. We proceeded our analysis using K = 3.

Upon assigning per-second data to clusters, we computed 1) time spent in each cluster, and 2) transition probabilities between clusters for each session. Correlations between time in cluster, transition probabilities, and each qualitative outcome measure were then investigated and visualized. The results section below summarizes our findings by outcome measures.

4. RESULTS

4.1. Correlation Check

To check for underlying relationships between our sensor data aggregated at the second level and the qualitative outcomes, we first checked for correlations between each sensor and qualitative measure. Significant correlations were observed between SM and Learning (r = -0.4, p = 0.025), and between JVA and sustaining mutual understanding (r = 0.41, p = 0.027). In accordance with previous analysis [7], no other significant correlation was observed.

4.2. Cluster Centroids

Centroid 1 values were the highest in all movement variables, suggesting cluster 1 as a state where dyads exhibited the most total movement difference, vertical difference in head orientation (e.g. a person standing up versus the other seated), and horizontal difference in shoulder orientation (when dyads were far apart from each other). Centroid 2 values were the highest in DA, PC, JVA, and the lowest in IDM, and SM; cluster 2 indicated a state where dyads were physiologically synchronized and actively sharing eye gaze. In contrast, cluster 3 appeared to be a state where dyads were the most desynchronized. Centroid 3 had the highest SM, IDM, and the lowest DA, PC values. Table 1 provides a summary of the 3 clusters identified by K-means Clustering, we will use the identified states to address the clusters in the following sections.

Cluster	Physiological Synchrony	Joint Visual Attention	Movement Difference	State
1	Average	Average	Highest	*Neutral*
2	Highest	Highest	Low	*Collaborative*
3	Lowest	Lowest	Low	*Non-Collaborative*

Table 1. Summary of clusters by sensor data.

4.3. Collaboration

Figure 5 represents the scaled and unscaled sensor data values by cluster centroid. The overall collaboration measure (r = 0.50, p = 0.009) was significantly correlated with time spent in the collaborative state. Time spent in the collaborative state was significantly correlated with sustaining mutual understanding (r = 0.42, p = 0.031), dialogue management (r = 0.51, p = 0.008), reaching consensus (r = 0.48, p = 0.013), task division (r = 0.49, p = 0.012), and reciprocal interaction (r = 0.47, p = 0.015). By interpretation of the EDA measure, dyads were highly, physiologically synchronized in the collaborative state. In contrast, dyads were highly desynchronized in the non-collaborative state, as the state corresponds to the lowest DA, PC and highest SM, IDM values by clustering. Time spent in the non-collaborative state was significantly, negatively correlated with dialogue management (r = -0.49, p = 0.008), task division (r = -0.45, p = 0.015) and collaboration (r = -0.42, p = 0.030).

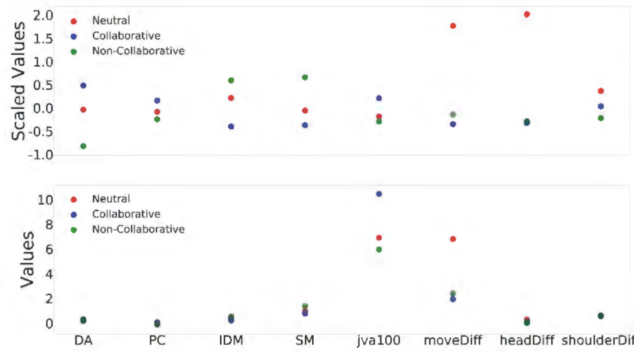

Figure 4. Sensor data mean values by cluster.

Correlations between state transition probabilities and qualitative outcomes rendered the same interpretation: the higher the probability that a dyad would transition into a desynchronized, or non-collaborative state, the lower the rating in collaboration (r = -0.41, p = 0.027).

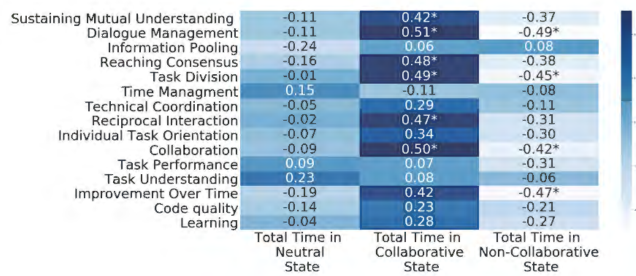

Figure 5. Correlations (left) and p-values (right) between time spent in cluster and qualitative outcome variables (* p < 0.05).

4.4. Task Performance

As shown in Fig. 6, time spent in the non-collaborative state was significantly correlated with improvement over time (r = -0.47, p = 0.018). The more likely a dyad were to transition into the desynchronized state, the lower the rating for task understanding (r = -0.47, p = 0.01, See Figure 6), and improvement over time (r = -0.53, p = 0.005). Furthermore, there was a marginally significant correlation between the probability of remaining in the neutral state and code quality (r = -0.37, p = 0.042); the more a dyad was different in movement, the lower the code quality as evaluated by the session facilitator.

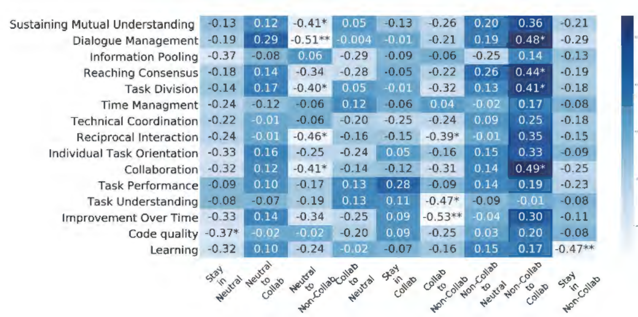

Figure 6. Correlations (top) and p-values (down) between transition probabilities and qualitative outcome variables (* p < 0.05, ** p < 0.01).

4.5. Learning Gain

We did not observe any significant correlations between learning and time spent in any of the collaborative state identified. However, learning was significantly and negatively correlated with the probability of remaining in the non-collaborative state ($r = -0.47$, $p = 0.008$). The more likely a dyad were to stay desynchronized, the lower the rating in learning gain.

5. DISCUSSION

Overall, our results suggest that K-Means Clustering is an effective method for identifying collaborative states. In accordance with previous findings, higher JVA in the current analysis significantly correlated with higher ratings of collaboration, more specifically in sustaining mutual understanding. This implies that sharing gaze facilitates collaboration by making aware the object, or the intent, of communication. In comparison with previous correlation findings [7], we were able to draw connections between the EDA measures and collaborative outcomes using clustering analysis. Specifically, physiological synchrony within dyads correlated with higher ratings in quality of collaboration, including sustaining mutual understanding, dialogue management, reaching consensus, and task division. This indicates, intuitively, that when participants in a dyad were in sync with each other, they were more likely to agree, understand, and coordinate in task with each other. Moreover, the larger the difference in movement and position within a dyad, the lower the code quality. One interpretation could be that the longer a dyad spent apart from each other, the less collaborative they were, or were deemed to be, and therefore, the lower the resulting code quality.

In a case study that compared the most collaborative (Group 11) and most non-collaborative (Group 5) groups, we observed that desirable collaborative qualities support the narrative, and the disaggregated graphs (Fig. 7) we created to depict them, that collaborative states are closely associated with levels of JVA and DA value.

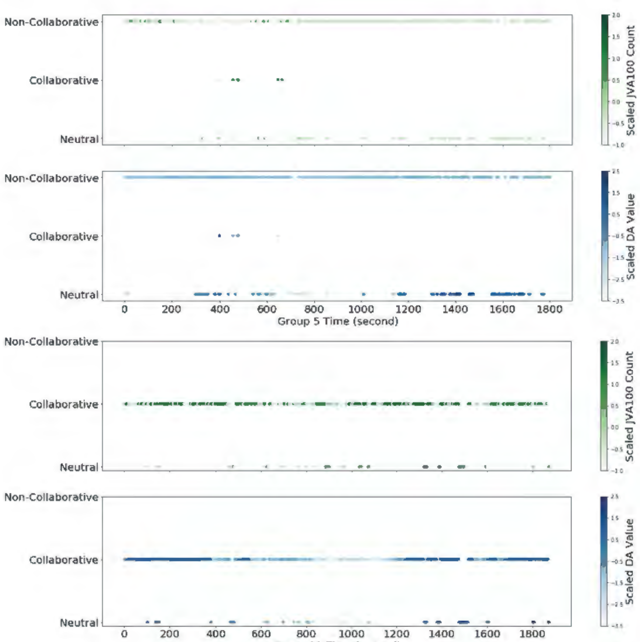

Figure 7. Progression of collaborative states by DA and JVA value in group 5 and group 11.

Though Groups 11 and 5 displayed collaborative (Group 11) and non-collaborative (Group 5) qualities rather consistently across the activity, each group mirrored qualities of the other. For example, while thinking aloud was a consistent quality exhibited by Group 11 across the activity, the participant on the left almost always remained in an observational role, which could have led to low learning gains. However, Group 11 achieved the highest learning gains of all groups in the study. One reason for this outcome could be the intent of the observer. High learning gains and consistent observation seemed to be a mode of learning for the participant on the left. This explanation supports an interpretation of observation not as a culprit of poor collaboration, but an assistant to learning, given a particular learning context. On the other hand, Group 5 also showed behaviors that were contrary to main themes that arose from their interactions. Take for instance their dialog that occurred in the middle of the activity. There, they exhibited seemingly desirable qualities in collaborative learning such as asking for help, asking clarifying questions, using demonstrative actions such as re-running code to convey conception of problem, and a continuous interactive dialog; however, demonstration of these qualities was more of an abnormality than a change in behavior. Furthermore, each participant steadfastly performed their roles. One interpretation of why this temporary change in mode of operation did not stick goes back to the very definition of collaboration: a result of continuous attempts to construct and maintain a shared problem space. This means that establishing a collaborative state two thirds of the way through the activity is perhaps too difficult of a cognitive shift when working modes have been established.

Main Results	Interpretation	Implication
Time spent in cluster 2 is sig. correlated with collaboration quality	Productive collaboration is characterized by increased levels of JVA and physiological synchrony	combining JVA values with PCIs provide better predictors for collaboration quality than just individual JVA or PCI
Staying in cluster 1 or 3 is neg. correlated with learning	The more participants stay in a state where they are not looking at the same place and are physiologically desynchronized prevents them from learning	we could potentially detect when people are in this state in real-time and offer suggestions for getting participants back on track.
Time spent in cluster 3 is neg. correlated with task performance	Non-collaborative states are associated with less task understanding and improvement over time.	*we didn't observe a positive significance with cluster 1, perhaps that while being highly synchronized and actively sharing eye gaze doesn't necessarily leads to more task understanding / improvement overtime; there's a reverse effect when participants were not at all collaborative by EDA/JVA measures.
Staying in cluster 1 is neg. correlated with code quality	The more time participants spent apart from one another, the lower the code quality.	Spending time apart from one another may indicate individual exploration, which in learning settings, may imply that participants were in need of guidance in order to proceed with task.

Table 2. Summary of results.

Finally, we found learning gains to be significantly associated with EDA measures. Particularly, the more likely a dyad were to remain desynchronized physiologically, the less likely that they had learned, or were evaluated to have learned from the task. Overall, we found that combining multimodal measures of collaboration together (e.g., eye-tracking, physiological, motion data) provides us with richer results: we found more (and stronger) significant correlations with our dependent measures. Table 2 summarizes the main results of this paper.

Nonetheless, it is important to note the limitations of the current analysis. For one, the aggregated Kinect measures utilized in the current analysis might not have best captured motor differences between collaborative and non-collaborative dyads. The current analysis only examined dyad's motor differences on single dimensions, future work should aggregate movement measures based on multiple dimensions (e.g. movement angle) in order to better capture (non-)collaboration in motion. As we identified the number of collaborative states (K = 3) using distortions computed with the current dataset, it is possible that this number is not definitive and is unique to the current study. We concluded from exploratory analysis that the implications differed as we increased the number of clusters, and/or reduced our variable dimensions. Moving forward, it is of our interest to find the optimal combination of measures, and the optimal number of states that best characterize the (un)productive collaboration.

5. CONCLUSION

The current study used unsupervised machine learning algorithms to effectively identify different states of collaboration. Combining eye-tracking and physiological-activity data better predicted collaboration quality than the two types of sensor data apart. The longer two partners were not sharing gaze, and were desynchronized from one another, the worse their task performance and the less they learned. Identification of collaborative states and their characteristics through sensor data potentially allows us to monitor collaboration in real-time, detect ineffective cooperation, and keep partnership intact. Future work should explore movement measures of various dimensions to best capture participants' postures and motions.

In summary, this paper contributes to the application of MMLA in open-ended learning environments for capturing 21st century skills. We argue that multimodal sensors can capture different aspect of productive collaboration, and that combining them can provide us with a more complete picture of productive social interactions. Because we tend to "teach what we can measure", developing tools that can capture 21st century skills is a crucial step toward studying and fostering them. This paper makes a first step in this direction by leveraging multimodal sensor data.

6. REFERENCES

[1] Abrahamson, D. 2009. Embodied design: constructing means for constructing meaning. *Educational Studies in Mathematics*. 70, 1 (Jan. 2009), 27–47.

[2] Anderson, M.L. 2003. Embodied cognition: A field guide. *Artificial intelligence*. 149, 1 (2003), 91–130.

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

[4] Bruinsma, Y. et al. 2004. Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews*. 10, 3 (2004), 169–175.

[5] Church, R.B. 1999. Using Gesture and Speech to Capture Transitions in Learning. *Cognitive Development*. 14, 2 (Apr. 1999), 313–342.

[6] Dede, C. 2010. Comparing frameworks for 21st century skills. *21st century skills: Rethinking how students learn*. 20, (2010), 51–76.

[7] Dich, Y., Reilly, J. & Schneider, B. 2018. Using Physiological Synchrony as an Indicator of Collaboration Quality, Task Performance and Learning. In *International Conference on Artificial Intelligence in Education* (pp. 98-110). Springer.

[8] Dillenbourg, P. et al. 1996. The evolution of research on collaborative learning. *Learning in Humans and Machine: Towards an interdisciplinary learning science*. 189–211.

[9] Elkins, A.N. et al. 2009. Physiological compliance and team performance. *Applied ergonomics*. 40, 6 (2009), 997–1003.

[10] Laal, M. and Ghodsi, S.M. 2012. Benefits of collaborative learning. *Procedia - Social and Behavioral Sciences*. 31, (Jan. 2012), 486–490.

[11] Meier, A. et al. 2007. A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*. 2, 1 (Feb. 2007), 63–86.

[12] Palincsar, A.S. 1998. Social constructivist perspectives on teaching and learning. *Annual review of psychology*. 49, (1998), 345.

[13] Pijeira-Díaz, H.J. et al. 2016. Investigating collaborative learning success with physiological coupling indices based on electrodermal activity. (2016), 64–73.

[14] van Rheden, V. et al. 2017. LaserViz: Shared Gaze in the Co-Located Physical World. *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction* (New York, USA, 2017), 191–196.

[15] Roth, W.-M. 2001. Gestures: Their Role in Teaching and Learning. *Review of Educational Research*. 71, 3 (Sep. 2001), 365–392.

[16] Salley, B. and Colombo, J. 2016. Conceptualizing Social Attention in Developmental Research. *Social Development*. 25, 4 (2016), 687–703.

[17] Schneider, B. (accepted). Unpacking Collaborative Learning Processes during Hands-on Activities using Mobile Eye-Tracking. In *the 13th International Conference on Computer Supported Collaborative Learning*.

[18] Schneider, B. et al. 2016. Using Mobile Eye-Trackers to Unpack the Perceptual Benefits of a Tangible User Interface for Collaborative Learning. *ACM Transactions on Computer-Human Interaction*. 23, 6 (Dec. 2016), 1–23.

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

[20] Sharma, K. et al. 2016. Visual Augmentation of Deictic Gestures in MOOC Videos. (Jul. 2016).

[21] Starr, E.L., Reilly, J.M. and Schneider, B., 2018. Toward Using Multi-Modal Learning Analytics to Support and Measure Collaboration in Co-Located Dyads. In Proceedings of the 13th International Conference of the Learning Sciences, London, UK, p. 448-455.

[22] Tomasello, M. et al. 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*. 28, 5 (Oct. 2005), 675–691.

[23] Werner, H. 1937. Process and achievement—a basic problem of education and developmental psychology. *Harvard Educational Review*. (1937).