Using Physiological Synchrony as an Indicator of Collaboration Quality, Task Performance and Learning

Yong Dich, Joseph Reilly, and Bertrand Schneider(✉)

Harvard University, Cambridge, USA
ydich@college.harvard.edu,
josephreilly@g.harvard.edu,
bertrand_schneider@gse.harvard.edu

Abstract. Over the last decade, there has been a renewed interest in capturing 21st century skills using new data collection tools. In this paper, we leverage an existing dataset where multimodal sensors (mobile eye- trackers, motion sensors, galvanic skin response wristbands) were used to identify markers of productive collaborations. The data came from 42 pairs (N = 84) of participants who had no coding experience. They were asked to program a robot to solve a variety of mazes. We explored four different measures of physiological synchrony: Signal Matching (SM), Instantaneous Derivative Matching (IDM), Directional Agreement (DA) and Pearson's Correlation (PC). Overall, we found PC to be positively associated with learning gains and DA with collaboration quality. We compare those results with prior studies and discuss implications for measuring collaborative process through physiological sensors.

Keywords: Biosensors · Collaborative learning · Physiological synchrony
Electrodermal activity · Galvanic skin response wristbands · Motion sensors
Multimodal

1 Introduction

There has been a renewed interest in leveraging new data collection tools for capturing students' learning processes that go beyond the acquisition of conceptual knowledge. With an ever-increasing ease of access to information, educational researchers are more and more interested in teaching "21st century skills [1]. Those skills include (but are not limited to) students' curiosity, critical thinking, collaborative skills, grit, persistence or creativity. Having accurate and reliable tools for capturing them can pave the way for new kinds of instruction (for example by displaying levels of mastery to teachers through dashboards; by designing awareness tools for students [2]; or by providing real-time, just-in-time, personalized feedback). To reach this goal, educational researchers are starting to use multimodal sensors and learning analytics to richly capture students' behavior (i.e., through Multimodal Learning Analytics, MMLA; [3]). The goal of this project is to make a first step in this direction by exploring how various kinds of sensors, such as eye-trackers, motion sensors, galvanic skin response wristbands, can capture proxies of 21st skills during learning activities. In this paper, we focus on capturing productive social interactions using

C. Penstein Rosé et al. (Eds.): AIED 2018, LNAI 10947, pp. 98–110, 2018.
https://doi.org/10.1007/978-3-319-93843-1_8

physiological measures from motion sensor and galvanic skin response wristbands data. More specifically, we computed various metrics of physiological synchronization and correlated them with a coding scheme for assessing collaboration quality in dyads.

The paper is structured as follows: first, we review prior work that has used electrodermal activity for studying collaborative processes. We then describe the study where the data was collected, our preprocessing procedure and analyses. We conclude with a discussion of our findings and future work for this line of research.

2 Literature Review

2.1 Electrodermal Activity (EDA) in Educational Research

Electrodermal Activity (EDA) is any electrical change measured at the surface of the skin whenever the skin receives innervating signals from the sympathetic nervous system. The sympathetic system is activated in case of an emotional activation like physical exertion or cognitive workload. The electrical conductance increases as the pores begin to fill below the surface with sweat. EDA is generally considered to be a reliable way of measuring sympathetic activation, because the skin is the only organ that is innervated just by the sympathetic nervous system [4].

In educational research, EDA has been used to capture students' affective state. As an example, [5] used data from four different sensors (camera, mouse, chair, and wristband) to predict students' affects in a school setting and was able to explain 60% of the variance of their emotional state when interacting with intelligent tutors. What is of interest to us, however, is the use of physiological sensors to predict students' social interactions. Prior work has identified various indicators of physiological synchrony and correlated those measure with outcome measures. [6], for example, used Signal Matching (SM), Instantaneous Derivative Matching (IDM), Directional Agreement (DA) and Pearson's correlation Coefficient (PC). We describe these measures in more details below, but the idea is that SM captures the difference between two EDA time-series; IDM the rate of change; DA the direction of those changes; and PC the linear relationship between them. The paragraph below and Table 1 summarize some results found by prior work.

Table 1. Summary of Results from prior studies (reproduced and modified from [6]).

Dependent measure	Indicators	Study
Team performance	SM, IDM, DA, PC	[7, 8]
Collaboration, task performance, learning	SM, IDM, DA, PC	[6]
Team work	PC	[10]
Interaction	PC	[9]
Completion time	PC	[11]
Conflicting interactions	PC	[12]

[7] found those indicators of physiological synchrony to be associated with task performance for pairs of participants in a multitask environment under varied task and technology conditions; more specifically, [8] 's findings suggest that PC and DA were the most useful indicators to differentiate between low and high performers. In a collaborative problem-solving task (i.e., designing a healthy, appropriate breakfast for an athlete training for a marathon), [6] found that IDM best predicted collaborative interactions and DA was positively associated with learning. In a collaborative game, [9] collected physiological data in dyads of learners and found that PC was correlated with participants' interaction and self-reported social presence. In a continuous tracking-task simulating teleoperation, [10] reported that PC was a significant predictor of completion time in two-person teams. In a four-persons team, they found that PC was also associated with teamwork effectiveness during real planning meetings [11]. Finally, [12] compared cooperative and competitive play and found that PC was correlated with conflicting interactions.

In summary, there is ample evidence that indicators of physiological synchrony are associated with outcomes of interest to educational researchers (social interactions, learning, task performance). However, most prior work has only looked at PC and there is not a clear understanding of the difference between the four physiological indicators considered in this paper (PC, DA, IDM, SM). In the next section, we present the study where the data was collected, describe our measures of physiological synchrony and correlate them with our dependent measures (task performance, learning gains, collaboration quality).

3 The Study

In this study, participants with no prior programming knowledge were paired and given 30 min to program a robot to autonomously solve a series of increasingly difficult mazes. Two different interventions were designed and used to support collaboration: an informational explanation of the benefits of collaboration and a visualization showing relative verbal contributions of each participant. Participants were given a pre- and post-survey on computational thinking skills and were asked to self-assess their collaboration at the end of the activity. Researchers coded the quality of the collaboration, the progress of the participants, and the quality of their final code. During the study, two mobile eye-trackers captured where participants were looking, a motion sensor captured gross motor movement and position, and two wearable technology bracelets captured EDA.

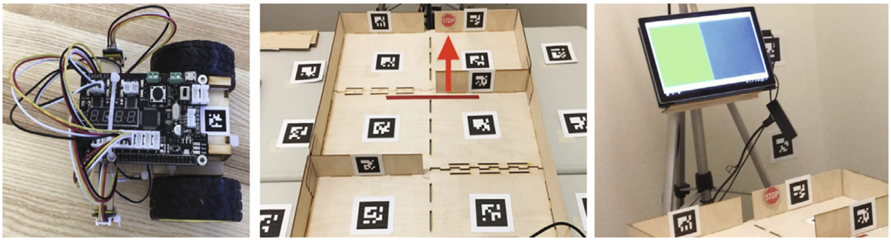


Fig. 1. The material used in the study: the robot that participants had to program (left image), one maze (middle image) and the Kinect-based speech visualization (right side).

3.1 Design

The study employed a 2 × 2 between-subjects design to measure the effects of the interventions. A quarter of the dyads received neither intervention (Condition #1, "No Explanation, No Visualization"), a quarter received solely the visualization (Condition #2; "No Explanation, Kinect Visualization"), a quarter received solely the informational intervention (Condition #3; "Explanation, No Visualization"), and the final quarter received both interventions (Condition #4; Explanation, Kinect Visualization") Assignment to conditions was done randomly prior to participant sign-up.

The informational collaboration intervention consisted of the researcher verbally informing the participants about several research findings related to collaboration such as equity of speech time predicting the quality of a collaboration. Dyads not assigned to conditions with this intervention received no such information. The visualization intervention used audio data from the motion sensor to display what proportion of total talk came from each participant over the past 30 s. The proportion of the screen filled with a certain color represented the relative contribution to total talk time (Fig. 1, right side).

The task asked participants to use a block-based programming language to program a robot through a series of mazes (Fig. 1, middle). The robot came equipped with a microcontroller, two DC motors connected to wheels, and three proximity sensors on the front, left, and right (Fig. 1, left). Participants were first shown a tutorial video to illustrate the basic concepts of how to use a block-based programming language to program the robot. Following the video, participants had five minutes to write a simple program to move the robot past a line two feet ahead of it. Data collected during this tutorial activity is not included in our analysis.

After this initial activity, a second tutorial video showing more advanced features such as using prewritten functions to turn and checking the values of the proximity sensors was shown to participants and a printed reference sheet covering the material from the video was provided. The main activity asked participants to spend 30 min attempting to get their robot through the increasingly more difficult mazes. As soon as a robot could solve a maze twice in a row, the next maze was provided. Participants did not know the layout of the mazes ahead of time and were encouraged to write code that could work for any maze. During this main activity, identical hints were given at five-minute intervals to all groups regarding common pitfalls that could lead to stuckness.

3.2 Methods

Forty-two dyads participated in the study (N = 84), and forty groups were used in the final dataset. Participants were recruited from a study pool at a university in the northeastern United States. 62% of participants self-identified as students and ages ranged from 19 to 51 years old. 60% identified as female. Participants were compensated $20 per 90-min session of the study. No participants previously knew each other.

In addition to a variety of other sensors (Fig. 2), an Empatica E4 wrist sensor tracked several physiological markers from each participant, including electrodermal activity (at 4 Hz), blood volume pulse (at 64 Hz), and XYZ acceleration (at 32 Hz).

During the 30-min session, roughly between 7,200 to 115,200 data points were generated for each participant per measure, depending on the sampling rate.

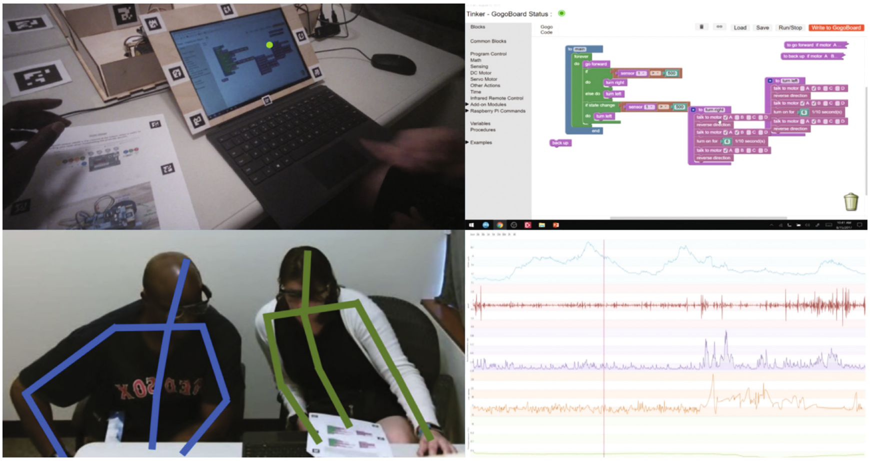


Fig. 2. The various measures collected during the study: top left shows the eye-tracking data; bottom left the motion data generated from the Kinect sensor; top right the block-based environment used to program the robot; bottom right the physiological data from the Empatica wristbands.

Learning of computational thinking skills in abstract was assessed by a pre- and post-test consisting of four questions assessing knowledge of computer science principles such as looping, conditional statements, and interpreting code (adapted from [13, 14]). These questions required near-transfer and application of skills learned in the activity. Researchers evaluated the completeness of answers and how well answers demonstrated understanding of computational thinking skills. The sum of the scores were used to generate pre, post, and gains scores for each individual.

While participants worked on the task, the researcher assessed their collaboration and task behaviors. The dyads' collaboration was assessed on nine scales adapted from [15] on a −2/+2 scale: sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation (see [15] for a definition of those dimensions). Two graduate students from the Harvard Graduate School of Education rated video recordings of the sessions using this coding scheme. The task behavior measures were task performance (number of mazes solved), task understanding (use of computational thinking concepts), and improvement over time (evidence of increased conceptual or technical understanding during the task). To calculate inter-rater reliability, a second researchers double coded 20% of the sessions from videos collected during the session and achieved an inter-rater reliability of 0.65 (75% agreement).

After the post-test, participants filled out a self-assessment of their collaboration experience during the activity (also adapted from [15]) that aligned with the researcher coding scheme as well as a demographic survey. Following the conclusion of the activity, the final block-based code each dyad created was evaluated to determine in abstract how well the code could successfully solve different types of mazes. The rubric aligned with the measures used to assess "Task Understanding" in the live coding to act as a check that no hardware or other issues had impacted the code's performance.

4 Data Preprocessing

We collected the following data from the Empatica wristbands: accelerometer, blood volume pulse (BVP), interbeat intervals (IBI), electrodermal activity (EDA), heart variability (HR), tag numbers to different sections in each session. All of this was accessible through Empatica's web portal. In this paper, we focus more specifically on electrodermal activity (EDA).

EDA is measured in a variety of ways: skin potential, resistance, conductance, admittance, and impedance. The wearable we used, Empatica E4 [16], captures electrical conductance (inverse of resistance) across the skin. It passes a minuscule amount of current between two electrodes in contact with the skin in units of microSiemans (μS). In the section below, we describe how we preprocessed the data for our analyses.

4.1 Cleaning the Data

The EDA data from the Empatica E4 comes in the following format: starting Unix time, frequency at which EDA was collected, and the EDA values. The first step was to synchronize the EDA's values. During the study, we asked participants to synchronize their sensors by pressing the button on the wristband before/after each step, which generated a tag in our dataset. By aligning those tags, we were able to synchronize the data from each participant and select one subset of the data (i.e., completing the maze task).

Before calculating our indicators of physiological synchrony, we had to clean the data by removing any extra noise. This noise or "artifacts" can be introduced whenever an individual adjusts the sensor, knocks the wearable against something or place pressure on the device. We used EDA Explorer [17], which is a machine learning classifier, focusing mainly on using support vector machines, that detects noise with 95% accuracy to remove artifacts.

In the paragraph below, we describe the four physiological coupling indices we explored in this paper: Pearson Correlation, Directional Agreement, Signal Matching, and Instantaneous Derivative Matching.

4.2 Computing Indicators of Physiological Synchrony

Each physiological compliance index was coded in the Python language following the mathematical descriptions below this paragraph. To render the data in the Empatica CSV files, Python's Pandas library was utilized to preprocess/clean, format, and analyze the data via dataframes. Other common libraries for mathematical analysis and visualization were used (seaborn, matplotlib, numpy, scipy).

Pearson's Correlation (PC): Pearson's correlation looks for a linear relationship between the EDA level of each participant. For example, a strong correlation means that both participants were likely to be physiologically activated, if positive, (or not, if negative) at similar times. As a sanity check, we looked at the scatter plots, correlations, and line graphs between two dyads, namely group 7 and group 8. According to our coding of collaboration and task performance, group 7 was judged to be a group that demonstrated poor performance and collaboration and success while group 8 was judged to be a good performing group. Figure 3 shows a slight positive correlation in EDA signals for group 8 individuals and a negative correlation in EDA signals for group 7, which is was we would expect. Directional agreement (DA) is identifying whether within each dyad the individuals' signal data points increase or decrease at the same time steps. In Python, each individual's data point at a time step was subtracted with a data point that occurred right before it [8] to determine the change in signal. We then compared both individuals' data points' change: increasing or decreasing. If any points are null, the change in signal calculated would also become null, automatically excluding both individuals' EDA data points at that specific time step. If both data points were indicated as increasing or both were decreasing, then this pair of points would mean both individuals were in "directional agreement" and then the counter variable named "tracking" would increase by 1. DA would be the ratio of the total directionally agreeing pairs of points out of the total number of pairs of data points compared.

Signal matching (SM) was used to look at the differences in area between the data curves of team members [8]. There is an inverse relation with greater area between the curves meaning less synchrony between the signals while less area between the curves meant closer synchrony or higher physiological synchrony. Thus, a negative correlation between a SM value and a qualitative measure would indicate that a small SM value means higher synchrony (and vice versa). Since individuals have different characteristics affecting their EDA signals, their signals need to be normalized to be compared on an equal scale. We normalized those values using z-scores. Once the absolute differences were calculated between the data points of each individual, the overall mean difference of each team was recorded.

Instantaneous Derivative Matching (IDM) also used the same normalized z-scored data. IDM calculates the level of matching between slopes of the two individual's physiological signal curves [8]. The slopes are calculated as the difference between the current point and the one ahead of it. Then these slopes or derivatives calculated for each individual's EDA signals were calculated by subtracting a data point at a time from a data point that's a time step ahead. These differences between the individuals' slopes were summed up and divided by the total time range observed. The following equation was used to compute IDM:

$$\frac{1}{T}\sum_{t=0}^{T-1}|(a_{t+1}-a_t)-(b_{t+1}-b_t)|$$

4.3 Filtering Outliers

Before correlating our four measures of physiological synchrony (DA, SM, IDM, PC) with our dependent measures, we looked for outliers in our dataset. Three groups were missing EDA data due to Empatica wearable malfunctions. The left side of Fig. 3 shows that each measure (except SM) has an outlier that was beyond two standard deviations of the mean. The right side of Fig. 3 shows the percentage of data that was removed after removing noisy artifacts. For our analyses, we removed those outliers because the data was either missing, too noisy or drastically different from other participants (which likely indicates that the wristband did not function properly).

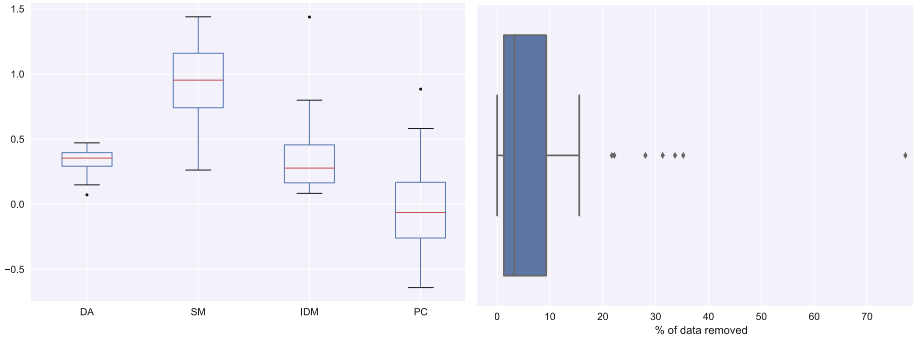
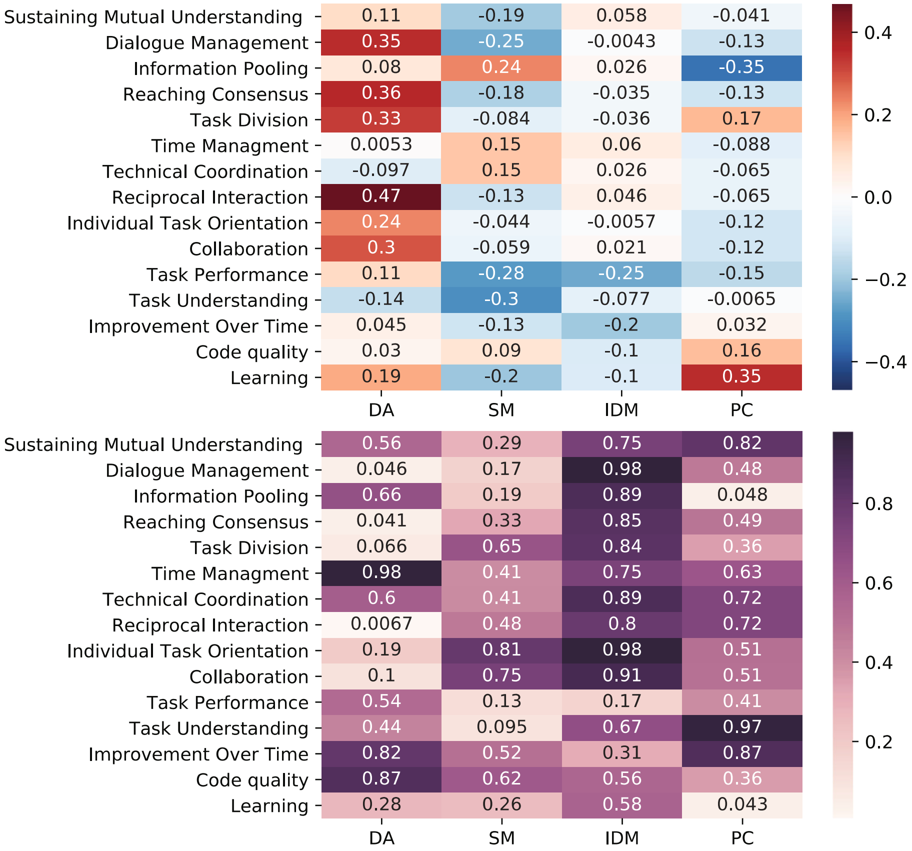


Fig. 3. Left side: boxplots for our 4 measures of physiological synchrony. We can see three outliers (one for DA, one for IDM and one for PC) and one group for which more than 50% of the data was removed.

5 Results

We first briefly summarize the main results of the study (described in detail in [18]). The quality of dyad collaboration (as coded by researchers) revealed significant differences between the two conditions that received the informational intervention to support collaboration (3&4) versus those that did not (1&2): dyads in condition 3 scored 7.1 points higher than those in condition 1 (p < 0.001), both of which did not receive the Kinect-based visualization intervention; dyads in condition 4 scored 4.8 points higher than those in condition 2 (p = 0.03), both of which did receive the Kinect intervention. Our coding of collaboration was significantly positively correlated with the quality of produced code (r = 0.52, p < 0.001) as well as all three performance metrics: task performance (r = 0.35, p < 0.001), task understanding (r = 0.53, p < 0.001), and improvement over time (r = 0.54, p < 0.001). Participants gained an average of 19.8% points on the survey of computational thinking principles (t = 6.18, p < 0.001). Learning gains did not differ significantly by condition, individual gender, the gender makeup of the group, or level of education. The quality of the final block-based code dyads produced was significantly correlated with the number of mazes completed (r = 0.45, p < 0.001), task understanding (r = 0.45, p < 0.001), and improvement over time (r = 0.54, p < 0.001).

The main contribution of this paper are the findings from the EDA analysis (not included in [18]). Results are presented visually in Fig. 4. We found that PC was positively correlated with learning gains: r(30) = 0.35, $p < 0.05$; DA was positively correlated with Dialogue Management: r(30) = 0.35, p = 0.063, Reaching Consensus r (30) = 0.36, $p < 0.05$ and Reciprocal Interaction r(30) = 0.470, $p < 0.001$. PC was also negatively correlated with Information Pooling: r(30) = −0.35, $p < 0.05$.



	DA	SM	IDM	PC
Sustaining Mutual Understanding	0.11	-0.19	0.058	-0.041
Dialogue Management	0.35	-0.25	-0.0043	-0.13
Information Pooling	0.08	0.24	0.026	-0.35
Reaching Consensus	0.36	-0.18	-0.035	-0.13
Task Division	0.33	-0.084	-0.036	0.17
Time Managment	0.0053	0.15	0.06	-0.088
Technical Coordination	-0.097	0.15	0.026	-0.065
Reciprocal Interaction	0.47	-0.13	0.046	-0.065
Individual Task Orientation	0.24	-0.044	-0.0057	-0.12
Collaboration	0.3	-0.059	0.021	-0.12
Task Performance	0.11	-0.28	-0.25	-0.15
Task Understanding	-0.14	-0.3	-0.077	-0.0065
Improvement Over Time	0.045	-0.13	-0.2	0.032
Code quality	0.03	0.09	-0.1	0.16
Learning	0.19	-0.2	-0.1	0.35

	DA	SM	IDM	PC
Sustaining Mutual Understanding	0.56	0.29	0.75	0.82
Dialogue Management	0.046	0.17	0.98	0.48
Information Pooling	0.66	0.19	0.89	0.048
Reaching Consensus	0.041	0.33	0.85	0.49
Task Division	0.066	0.65	0.84	0.36
Time Managment	0.98	0.41	0.75	0.63
Technical Coordination	0.6	0.41	0.89	0.72
Reciprocal Interaction	0.0067	0.48	0.8	0.72
Individual Task Orientation	0.19	0.81	0.98	0.51
Collaboration	0.1	0.75	0.91	0.51
Task Performance	0.54	0.13	0.17	0.41
Task Understanding	0.44	0.095	0.67	0.97
Improvement Over Time	0.82	0.52	0.31	0.87
Code quality	0.87	0.62	0.56	0.36
Learning	0.28	0.26	0.58	0.043

Fig. 4. Correlations between our dependent measures (collaboration, task performance and learning) and the four indicators of physiological synchrony. The heatmap on the top shows Pearson's correlation coefficients and the heatmap on the bottom shows p values.

Because we saw an effect of our intervention on participants' collaboration (but not on learning gains), we hypothesized that it also impacted their physiological synchrony. To check this assumption, we broke down the correlation matrix by condition (Fig. 5). By visually inspecting the heatmaps, we found that the groups in the "No Explanation" condition exhibited negative correlations with our indicators of physiological synchrony (e.g., DA and PC for the Kinect Visualization group and SM and PC for No Visualization). This is represented by dark blue columns in Fig. 5 below. We focus on

DA in our next analyses (first column of Fig. 5), because this measure was significantly correlated with participants' quality of collaboration.

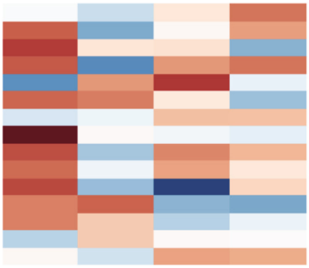
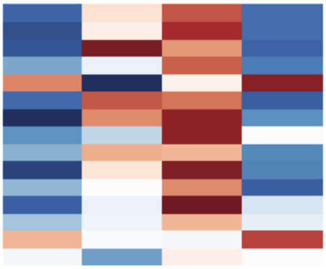
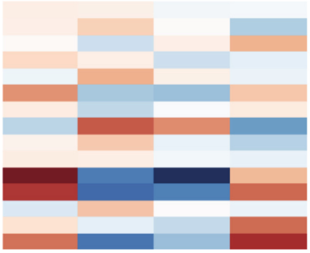
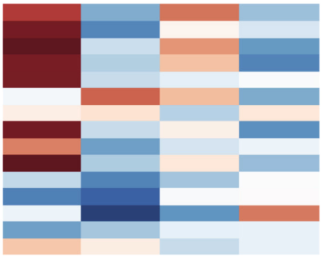
Independent Variables (2x2)	Explanation	No Explanation
Kinect Visualization		
No Visualization		

Fig. 5. Correlations matrices between our dependent measures (collaboration, task performance and learning) and the four indicators of physiological synchrony for each experimental condition. See Fig. 4 for the horizontal/vertical labels.

For DA, the experimental condition that saw the Kinect Visualization but received No Explanation was found to have negative correlations with most dependent measures (1st row, 2nd column of the heatmaps). Because this group of participants seemed to behave differently from the other experimental conditions (at least for DA), we explored whether removing it from our sample would produce different results. We found the following significant correlations with those three conditions grouped together (Kinect with Explanation, No Kinect with Explanation, No Kinect with No Explanation – 1st column and 1st row of Fig. 5):

- DA - Sustaining Mutual Understanding: r(25) = 0.436, p = 0.023
- DA - Dialogue Management: r(25) = 0.597, p = 0.001
- DA - Information Pooling: r(25) = 0.551, p = 0.003
- DA - Reaching Consensus: r(25) = 0.555, p = 0.003
- DA - Task Division: r(25) = 0.381, p = 0.050
- DA - Reciprocal Interaction: r(25) = 0.605, p = 0.001
- DA - Individual Task Orientation: r(25) = 0.395, p = 0.042
- DA - Collaboration: r(25) = 0.582, p = 0.001

Here, DA is significantly and positively correlated with most dimensions of our rating scheme for assessing collaboration quality. PC is significantly correlated with Learning. Interestingly, SM and IDM are negatively correlated with task performance:

- SM - Task Performance: r(25) = −0.387, p = 0.046
- IDM - Task Performance: r(25) = −0.491, p = 0.009
- PC - Learning: r(26) = 0.410, p = 0.030

By removing participants in the first condition (No Kinect, No Explanation), all correlations become non-significant - except Learning gains with PC (r(24) = 0.398, $p < 0.05$).

6 Discussion

In summary, we found that for our entire sample PC seems to be positively correlated with learning gains. When looking at task performance, it was surprising to see that our indicators were generally negatively correlated with how well participants completed the task. This is at odd with prior research. We plan to reexamine how our metrics of performance were coded, and how they relate to our other dependent measures. Finally, DA was associated with our dyads' quality of collaboration, especially when removing the experimental condition where participants saw the Kinect Visualization (but did not receive an explanation about the importance of working together). This suggests that seeing how much each person was talking - but no knowing how to use this information - had a distracting effect on our participants. DA, for example, became negatively correlated with participants' quality of collaboration. This suggests that groups that worked well together were more likely to be unsynchronized; it might be that in those groups, participants paid attention to the visualization (which increased their physiological activation) but each participant might have done it at different times. Those "spikes" lowered participants' scores on our measures of physiological synchronization - when in fact it meant that they were potentially aware of unbalanced levels of participation. This interpretation will be checked in future work through video analyses and log data from the Kinect sensor (i.e., the data can tell us if the participants were looking above the maze, where the visualization was presented).

Finally, it should be noted that our correlations did not agree with prior research. In other studies [6], team performance was found to be positively correlated with SM, IDM, DA and PC. Teamwork and interaction were also positively correlated with PC, and learning gains with DA. In this paper, we found collaboration quality to be associated with DA, task performance to be negatively correlated with IDM, and SM and PC to be positively associated with learning gains. Some of those differences are likely caused by how the constructs were operationalized. [6], for example, used self-report scales for capturing social interactions while we applied a validated rating scheme in the learning sciences [15]. Task performance and learning gains also depend on the task that participants have to complete and can vary widely in their measures (e.g., completion time, success, factual knowledge, transfer questions, etc.). But it is striking to see that our four measures of physiological synchronization seem to be sensitive to different outcomes measures compared to prior work.

7 Conclusion

In conclusion, this paper successfully identified predictors of task performance, collaboration quality and learning gains from physiological sensors. Those results are encouraging, especially in the context of developing real-time, just-in-time, personalized feedback to students. We also plan to leverage those measures to develop dashboards for teacher and awareness tools for students [2].

There are few limitations that are worth mentioning. In order to control the quality of the EDA signal, the study followed recommended recording conditions based on maximum signal-to-noise ratio: subjects were seated at a table which limited their movement. However, since they were manipulating the robot it generated some noise in the data. This required us to remove some participants from our sample. Additionally, new findings show too that EDA should not just be measured on the non-dominant side. There were also 3 groups removed from the dyad data set due to an Empatica wearable or user malfunction because not all of the EDA data was collected, inhibiting proper clean up and analysis.

In terms of future work, we want to consider the more specific characteristics of EDA: tonic versus phasic changes. Tonic skin conductance level is the smooth underlying slowly changing levels in the absence of external stimuli. This is also known as skin conductance level (SCL). Phasic skin conductance response is identified by the rapidly changing peaks which are associated with short term events and occur in the presence of external stimuli (sight, sound). These peaks or abrupt increases in the skin conductance are referred to as Skin Conductance Responses (SCRs).

In conclusion, our next step is to compute those indicators in real-time and test out the effectiveness of displaying this information to learners and teachers in order to promote self-regulation, real-time monitoring, data analysis, and provide the opportunity to give formative feedback.

References

1. Dede, C.: Comparing frameworks for 21st century skills. In: 21st Century Skills: Rethinking How Students Learn, vol. 20, pp. 51–76 (2010)
2. Buder, J.: Group awareness tools for learning: current and future directions. Comput. Hum. Behav. **27**, 1114–1117 (2011)
3. Blikstein, P., Worsley, M.: Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks. J. Learn. Anal. **3**, 220–238 (2016)
4. Dawson, M.E., Schell, A.M., Filion, D.L.: The electrodermal system. In: Handbook of Psychophysiology, vol. 2, pp. 200–223 (2007)
5. Arroyo, I., Cooper, D.G., Burleson, W., Woolf, B.P., Muldner, K., Christopherson, R.: Emotion sensors go to school. In: AIED (2009)
6. Pijeira-Díaz, H.J., Drachsler, H., Järvelä, S., Kirschner, P.A.: Investigating collaborative learning success with physiological coupling indices based on electrodermal activity. In: Proceedings of the Sixth International Conference on Learning Analytics and Knowledge (2016)

7. Montague, E., Xu, J., Chiou, E.: Shared experiences of technology and trust: an experimental study of physiological compliance between active and passive users in technology-mediated collaborative encounters. IEEE Trans. Hum. Mach. Syst. **44**, 614–624 (2014)
8. Elkins, A.N., Muth, E.R., Hoover, A.W., Walker, A.D., Carpenter, T.L., Switzer, F.S.: Physiological compliance and team performance. Appl. Ergon. **40**, 997–1003 (2009)
9. Järvelä, S., Kivikangas, J.M., Kätsyri, J., Ravaja, N.: Physiological linkage of dyadic gaming experience. Simul. Gaming **45**, 24–40 (2014)
10. Henning, R.A., Boucsein, W., Gil, M.C.: Social–physiological compliance as a determinant of team performance. Int. J. Psychophysiol. **40**, 221–232 (2001)
11. Henning, R., Armstead, A., Ferris, J.: Social psychophysiological compliance in a four-person research team. Appl. Ergon. **40**, 1004–1010 (2009)
12. Chanel, G., Kivikangas, J.M., Ravaja, N.: Physiological compliance for social gaming analysis: cooperative versus competitive play. Interact. Comput. **24**, 306–316 (2012)
13. Brennan, K., Resnick, M.: New frameworks for studying and assessing the development of computational thinking. In: Proceedings of the 2012 Annual Meeting of the American Educational Research Association, Vancouver, Canada (2012)
14. Weintrop, D., Wilensky, U.: Using commutative assessments to compare conceptual understanding in blocks-based and text-based programs. In: 11th Annual ACM Conference on International Computing Education Research, ICER 2015 (2015)
15. Meier, A., Spada, H., Rummel, N.: A rating scheme for assessing the quality of computer-supported collaboration processes. Comput. Support. Learn. **2**, 63–86 (2007)
16. Garbarino, M., Lai, M., Bender, D., Picard, R.W., Tognetti, S.: Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In: 2014 EAI 4th International Conference on Wireless Mobile Communication and Healthcare (Mobihealth) (2014)
17. Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., Picard, R.: Automatic identification of artifacts in electrodermal activity data. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2015)
18. Starr, E., Reilly, J., Schneider, B.: Using multi-modal learning analytics to support and measure collaboration in co-located dyads. In: The 13th International Conference on the Learning Sciences