Multimodal Data Analytics for Assessing Collaborative Interactions

Yanghee Kim, Sachit Butail, Lichual Liu, Michael Tscholl, and Jaejin Hwang
ykim9@niu.edu, sbutail@niu.edu, liu@niu.edu, mtscholl@niu.edu, jhwang3@niu.edu
Northern Illinois University

Francesco Cafaro, Milka Trajkova, and Kyungbin Kwon
fcafaro@iu.edu, milktraj@iu.edu, kwonkyu@indiana.edu
Indiana University

Danielle Espino, Seung Lee, and Eric Hamilton
danielle.espino@pepperdine.edu, seung.lee@pepperdine.edu, eric.hamilton@pepperdine.edu
Pepperdine University

Cynthia D'Angelo, cdangelo@illinois.edu, University of Illinois Urbana-Champaign

Xavier Ochoa, xavier.ochoa@nyu.edu, New York University

Aaron Kline, aaron.s.kline@gmail.com, Stanford Medical School Pediatrics Research

Sungchul Lee, lees@uww.edu, University of Wisconsin-Whitewater

Abstract: This symposium will discuss the current status of the research and development of multimodal data analytics (MDA) for the observation of collaboration. Five research groups will present their current work on MDA, each with a unique focus on different data sources and different approaches to the analysis and synthesis of multimodal data sets. A few themes emerge from these studies: i) the studies seek to examine collaborative behaviors as a process in ordinary settings, both formal and informal; ii) with MDA being in its early stage, manual and computational approaches are taken complementarily, also using human annotation as the ground truth for the computational approach; and iii) several different discipline-specific research and development lines contribute integrally to generating authentic measures of collaborative interactions *in situ*, making this line of research transdisciplinary.

Overview of Symposium

In both formal and informal learning, in-person collaboration has typically been studied by human observers meticulously attending to interaction behaviors and taking notes while the collaboration is ongoing, then analyzing the observations a posteriori at the microgenetic level. Manual observations and analyses of the resulting rich and thick data demand tremendous amounts of resources and severely limit the scalability of the programs that support collaboration, and in turn the research progress. Emergent computational approaches to what is popularly referred to as "Big Data" can enable the automation of the collection, processing, and analysis of a wide range of collaborative behaviors in real time (Baker & Siemens, 2014; Martin & Sherin, 2013). Yet, it is unknown to what extent these computational approaches are being developed, tested, and used by the LS community for the authentic assessment of collaborative learning processes *in situ*. Also, theoretical and methodological perspectives and assumptions must be clarified to provide a solid grounding to the application of the new approaches.

This symposium brings together six studies (S1-S6) that employ multimodal sources of information to examine in-person or digitally connected collaborative behaviors. The studies are in different stages of development: some have the analysis results (e.g., S3, S4), and for others, data analyses are still in progress (e.g., S1, S2). Each of the studies is unique in their use of different modalities, different age groups of learners, varying task domains, and/or formal and informal learning contexts. The studies collected rich sets of individual's collaborative behaviors in ordinary settings over time and analyzed them both manually and computationally. By discussing these research cases in the symposium, we will be able to jointly build the current knowledge base on the progress in multimodal assessment of collaborative learning processes. More importantly, in many technology-enhanced learning environments, students engage in collaborative problem solving in pairs or in groups. We give special attention to advanced computational approaches to automating the collection, analysis, and synthesis of multimodal collaborative behaviors. The guiding question for the presenters and the audience is, *in what ways and to what extent does multimodal data analytics help assess on-going collaborative interactions* *over time?* Specific questions include i) *What sources of multimodal behaviors are observed and analyzed?* ii) *What kind of information is derived from each source of multimodal behavior data?* iii) *What theoretical assumptions, conceptual frameworks, and/or methodological tools have been used and need to be further explored to analyze and synthesize the multimodal sources of information for assessing collaborative engagement?*

The theme of this symposium will be a significant contribution to ICLS 2020 pinpointing the *interdisciplinarity of the learning sciences*. Multimodal data analytics of collaborative interactions requires multidisciplinary theories and expertise, inevitably involving teams of researchers from multiple disciplines, such as learning scientists, computer scientists, engineers, psychologists, and others. The symposium studies will exemplify how discipline-specific expertise works together to generate transdisciplinary knowledge and models to examine collaborative engagement authentically.

Studies (S1 - S6)

S-1. Speech-Based Multimodal Analysis of Collaboration

This section presents two case studies projects that implement a speech-based analysis of group collaboration using data from individual audio channels. Both cases build off of initial work that was developed within the Speech-Based Learning Analytics for Collaboration project (D'Angelo et al., 2019; Smith et al., 2016). This project utilized features derived from Speech Activity Detection (SAD) as well as manually coded indicators of collaborative behaviors to predict the quality of small-group collaboration among middle school students working on short mathematics problems. Individual audio channels (one for each student in a triad) were used to better understand individual and group contributions to the collaboration.

The data collection setup allowed students to speak freely, resulting in audio recordings with overlapping speech from the students in each group. SAD-based features were extracted that characterized either the individual student speech or the speech patterns of the group. These features capture information about the number, duration and location of the speech regions. Specifically, we extracted features that capture information about the overall speech duration and the "spurts" of speech. Spurts were defined as regions of speech that are at least 50ms long and were uninterrupted by pauses longer than 200ms. As students deal with the cognitive load of simultaneously solving problems and negotiating with their peers, they frequently interrupt each other or speak in short phrases.

We extracted these SAD-based features across the channels individually and in combination, taking into consideration speech activity from regions in which each individual student was the only speaker, when each student spoke, ignoring speaker overlap, when each pair of students spoke simultaneously, when all students were silent, or if all students spoke simultaneously. Duration-based features for individual and pairs of students were mapped to the group level using ratios and entropy statistics as described in Smith and colleagues (2016). These features capture information about the distribution of speech duration across the members of the group. The spurt-based features capture information about turn-taking and other features of the speaking style of the group.

Case 1: Engineering problem solving

This case is focused on small groups of undergraduate engineering students that worked collaboratively on tablet-based engineering problems. The teaching assistants (TAs) for this class are graduate and undergraduate students who do not have extensive teaching experience and need prompts and other support in order to successfully help students in their collaborative problem solving. This project involves working on analysis of speech data to aid in the prediction of problem-solving situations that need the help of a TA through a prompting system.

Data collection for this project, completed in Spring 2019, involved three groups of engineering students over the course of five weeks of instruction. Each class lasted approximately 50 minutes and students were engaged in collaborative problem solving for most of that time on synced tablets that allowed them to see each other's drawings and annotations of the problem. Individual and group audio was collected and synced with other modalities of data, including the tablet log data and video data.

Case 2: Medical problem-based learning

This case involves studying larger groups, ten medical students, engaged in collaborative problem solving with the assistance of a facilitator. The case combines multiple audio sources to look primarily at instructor versus group speech. Problem-Based Learning (PBL) is a popular student-based model of instruction in medical education and is guided by a facilitator who is expected to shape students' conversation so that it generates learning. Facilitators need to be deeply engaged in the conversations to coach students toward considering important aspects of the case they may not have identified and to promote engagement and thinking about target concepts. This role requires dynamic and real-time decisions about when to intervene and how to best do so in the context of a particular case at a particular point in time in the discussion.

Data collection for this project, completed in Summer 2019, used an individual audio channel for the facilitator and a group microphone for the students. Collecting 10 channels of audio data for individual students was not feasible from a practical standpoint, and also would have been unlikely to capture much additional information. This project involves developing a process of visualizing features extracted from speech data from the PBL sessions that will enable the identification of patterns in speech frequency, proportion, duration, and speaker turns. These visualizations, along with a professional learning community, will allow facilitators to improve their implementation of these PBL sessions.

S-2. Video Processing for the Design of Intuitive Embodied Interactions

This paper discusses approaches to the analysis of videos recorded during a series of formative user studies for a prototype interactive installation at Discovery Place, a science museum in Charlotte, NC. Our work focuses on the design of a specific class of embodied interaction: Human-Data Interaction (HDI) (Cafaro, 2012; Elmqvist, 2011). Designing HDI installations that can engage museum visitors is very challenging. HDI displays often compete with surrounding stimuli (other exhibits, people, signs, etc.). This limits the number of visitors who notice them, a phenomenon called display blindness (Cheung, Watson, Vermeulen, Hancock, & Scott, 2014). Also, museum visitors do not consult user manuals before interacting with an exhibit and often think that the system is broken if the installation does not respond to their signal quickly, a phenomenon called affordance blindness (Coenen, Claes, & Moere, 2017).

Museum visitors can use our installation to explore and compare two datasets. The visualization consists of two interactive globes, displayed on a 75" screen. Data are visualized at a country level, using color gradations (i.e., darker colors mean higher data values). Each globe visualizes one dataset expected to be thought-provoking for museum visitors and that illustrate issues such as gun violence, immigration, and unemployment.

This presentation discusses how to automatically identify the gestures and body movements that visitors typically make in front of an interactive installation. This is to develop techniques to move gesture elicitation studies (Wobbrock, Morris, & Wilson, 2009) out of research labs. The context of use has a profound impact on interactions in museum settings, and therefore researchers need to craft strategies for conducting elicitation studies *in situ*. We want to assess if observing the gestures and body movements of visitors in-situ will produce the gestures and body movements that are more intuitive and easily discovered. To achieve this goal, we need to design a tool that can track people's movements from video recordings and is able to identify the most common gestures and body movements. Also, an immediate use of this tool is to investigate if people perform different gestures and body movements when interacting alone with the screen, vs. when they collaboratively explore our data visualization in groups.

To explore different levels of engagement with our prototype, we conducted a manual analysis. Two researchers reviewed and coded the videos of the experimental sessions using a video annotation tool (Anvil). Because of the nature of our prototype (a large, interactive screen), visitors generally interacted with it in groups. The coding was structured using multiple tracks (to represent, for instance, the age range of the user, and the gestures that she/he used). In order to account for multiple users, we had to duplicate such tracks for each of the users who collaboratively interacted with our installation. This was challenging for the coding process: it forced us to decide a-priori the maximum number of visitors that we wanted to code, and it complicated our effort to ensure inter-coder reliability. We considered the segmentation agreement, i.e., whether the two coders agreed on the beginning and end time for a code, and the category agreement, i.e., whether the two coders agreed on the actual code on a segment of the video. In ongoing analysis, we are using a video analysis tool (OpenPose) that automatically labels gestures and body movements to explore collaborative interaction patterns. We will discuss the challenges that we encountered when crafting a dictionary of the gestures when tracking multiple people in collaboration.

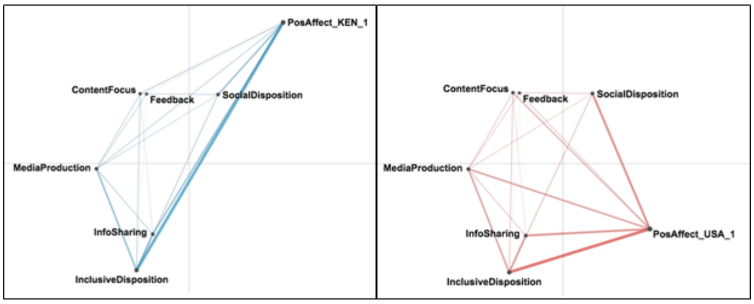
S-3. Epistemic Network Analysis to Connect Affect and Dialogue in Collaboration

Distributed environments provide a unique opportunity for learners to work together across geographical boundaries. This work examines the collaboration of middle and high school age adolescents in afterschool clubs from different countries across four continents as they develop STEM content-focused media projects. They share ideas and provide feedback asynchronously through email or Slack (a cloud-based team messaging application) or synchronously through video conferences from two or more sites.

Video conferences are a key source of data, recording the interactions of students as they collaborate. Participants' conversations in the conferences were transcribed, followed by qualitative coding for constructs that emerged from the text. While this method provides a tangible analysis of the recorded data, it does not account for the emotional interactions of participants that video data can capture. Combined use of visual and spoken data provides genuine information about the collaboration dynamics but presents a complex challenge on how to integrate them meaningfully with methodological rigor.

Epistemic network analysis (ENA) provides a useful analytical tool for examining the relationships among constructs derived from multiple types of data (Lund et al., 2017). In this study, ENA was utilized to model the connections among group discourse and the affective states of individual participants within a recent temporal context. The use of the moving window allows for the model to capture the linkages that are created as participants respond to and build upon each other's utterances and emotional expressions.

To prepare the data for ENA, a qualitative coding approach was used for the transcribed dialogue. Coding for affect involved identifying the presence of positive or negative emotions demonstrated by each participant (Pekrun, 1992). Neutral emotional states were indicated as the absence of a positive or negative affect. The coding process became complex, depending on the number of participants and the length of the video conferences because the emotional state of each participant must be accounted for the entire duration of the online meeting both when they were speaking and when not speaking. Once the affective valence for each individual was coded, it was integrated alongside the coding of dialogue into the ENA webtool to examine connections between positive and negative affect with discourse constructs. Examples of this analysis can be seen in Figure 1 (Lee et al., 2020). The data used in this analysis were collected from a 30-minute online meeting held in November 2018 with learners from Kenya and the U.S. The models visualize the relationship among constructs in the group discourse and positive affect exhibited by a Kenyan and a U.S. student for the entire duration of the meeting even when they were not speaking.



(a) (b)

Figure 1. Sample ENA models for a Kenyan learner (a) and a USA learner (b).

While it provides an opportunity to integrate two sources of data, ENA also entails several challenges that will require further attention. The first issue was harmonizing different types of data into a single model, including determination of the appropriate segmentation and moving window size. While an utterance is often used to segment discourse data, researchers will need to consider if it can be applied for analyzing affective states within video data. Similarly, incorporating multiple types of data will require analysis of how recent temporal context is conceptualized and defined for each data type. Second, further work is needed to develop new approaches to handling simultaneous data within ENA. In the example above, the simultaneity of individual affect was modeled as separate constructs, which was feasible due to the small number of participants and constructs involved. Addressing this will promote broader application of ENA.

S-4. Information Theory to Assess Engagement in Robot-Mediated Collaboration

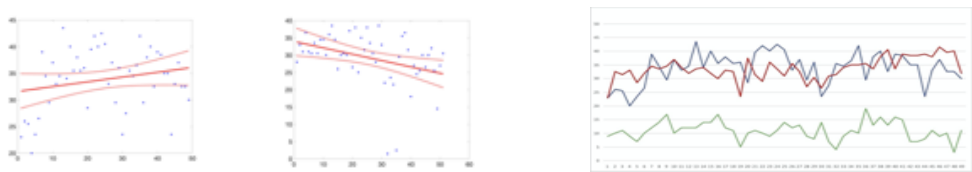
Studies 4 and 5 deal with designing robotic interactions to support equitable collaboration among young children. Collaboration is one of the key factors for young children's social and intellectual development as they start public schooling, and critical for their academic success. Leveraging social robots' appeal for young children, our project has introduced a humanoid robot as a collaboration mediator, instantiating an interaction triad of two children and a robot, where the children are encouraged to engage in collaborative conversations.

In this section, we report a study on children's engagement with each other and with the robot over time. In line with viewing engagement as a three-faceted phenomenon including behavioral, emotional, and cognitive engagement (Fredericks et al., 2004), children in our study expressed their collaborative engagement through talk, emotion, and body posture and gestures. We conducted a process analysis of these different expressions and their changes over time. Acknowledging that multimodal data analytics is in its infancy, we had four preliminary research questions: Q1) *How does a child's engagement progress over time?* Q2) *To what degree do the three multimodal data types correlate per child?* Q3) *To what degree is the robot's mediation related to the child's engagement?* And Q4) *To what degree does the engagement relationship of two children in a pair evolve over time?*

We explored a new approach to synthesizing multimodal datasets to assess children's collaborative engagement, guided by the aforementioned engagement theory and the information theoretic analysis of *mutual information (MI)*. We used data from two triads participating in three sessions, each taking 20 mins: children A&B and C&D. A total of 60-minute audio and video data were used for each triad, recorded using four directional and ambient microphones and two HD video recorders. Audio data were processed and analyzed by automatic speech recognition at 3-second intervals. Matching video data were annotated manually at the same interval for kinesics (body posture and gestures) and linguistic alignment (whether children respond to each other and/or the robot). The assessment phase came after the analyses of each of multimodal data sets were complete and also involved calculating the engagement values as a compound variable for each child. With the MI construct, we quantified the dependence among three time series of multimodal data: i) vocal pitch and intensity as evidence for emotional engagement, ii) kinesics for bodily engagement, and iii) linguistic alignment for cognitive engagement. Dependence was defined as shared information (variable *I*), with a value between 0 and 1 (perfect sharing). We computed collaborative engagement as the sum of bodily engagement (***K****inesics*), cognitive engagement (***A****lignment in utterances*), and emotional engagement (***I****ntensity* and ***P****itch)*. Denoting these values at a time *t*, engagement, *E*(*t*), was calculated as

$$E(t) = K(t) \ + \ A(t) + \frac{I(t) + P(t)}{2}$$

For Q1, Figure 2(a) depicts regression lines of child A's and C's engagement (y-axis) over time (x-axis). Similar trends are found for the children in the same group. Regression lines are best fits of all compound engagement values (blue dots) for 1-min intervals. Fit values are as high as .6, meaning that the lines adequately represent how engagement evolves. Also visible is engagement consistency of a child. Child C's values are clustered along the regression line (more consistent), but more dispersed for child A (less consistent). For Q2, we calculated means and *SDs* of normalized MI between data sources for each child over all sessions, finding that voice Intensity and Pitch correspond more with each other ($I > .75$ for all sessions) than Kinesics and Alignment ($I < 0.1$, for all sessions). Alignment conforms strongly with both Pitch and Intensity ($I > 0.6$), meaning that most of the children's speech is collaborative. We found no large variation between these values across sessions.



(a) (b)

Figure 2. Regression lines of a child's Engagement over time (a) and trends of children and robot talk (b).

For Q3, the frequencies of engagement values and robot talk per 1-minute interval are presented in Figure 2(b). Trends in a child's Engagement and the robot's mediating talk over time correspond with each other, showing that their interactions were reciprocal. Correlations between child engagement and robot speech range between 0.57 and 0.76, confirming visual analysis. There were no statistical differences in this range. For Q4, we compared the engagement relationship of the two children in the same group over time. Means of normalized MI are .63 (*SD* = .2) for A & B and .51 (*SD* = .23) for C & D, showing a higher MI relationship for children A & B.

Through both theory- and data-based modeling, we were able to compute a compound variable of children's collaborative engagement. The strength of this approach was evident when we compare it with qualitative human observations on site. For example, the progression in engagement of child C in Figure 1(a) showed that her engagement decreased (even with no statistical significance); the on-site observations noted that her interest moved from the task at hand to the robot itself, so her collaborative engagement with child D decreased. Analyzing data dispersions, it is evident that some children's engagement is more consistent over time than others'. Thus, our approach allows the individual characterization of a child to be used for identifying intrinsic and extrinsic factors for those individual variations.

S-5. Automated Video Analysis of Human vs. Robot Mediated Collaboration

This section is a continuous effort to examine collaborative engagement of children in triadic interactions. For this study, we examined if there is any difference in children's engagement when they had a human mediator compared to when they had a robot mediator. A similar interaction triad of two children and a mediator were implemented, where the children are encouraged to engage in collaborative conversations. A new group of twentyfour children participated in both of two phases, one with a human mediator (the first two weeks) and the other with a robot (the second two weeks). The study has two primary goals: exploring AI-based techniques that could lead to more automated methods of engagement measurement, and identifying the challenges involved in order to formulate structural recommendations for the community, to inform future study design and improve the likelihood of successful automated engagement measurement.

Towards these goals, the research team is now annotating the levels of overall engagement and more specific collaborative features, such as individual gaze targets and body orientation, at every three-second interval. Following this, we will leverage analysis techniques explored in a study of Superpower Glass, a wearable learning aid for children with autism. As part of the study's exploratory analysis, a number of automated techniques were applied to evaluate communication and engagement. Applying these approaches to the current study, neural network-based face detection provides locations of faces in the scene, as well as several landmark features representing salient features of each face (e.g. the shape of eyes, eyebrows and mouth).

As presented in Figure 3(a), combining this face detection (face locations shown as green boxes) with a customized object detection framework that identifies the robot's location (orange box) enables calculation of the learners' orientation towards the robot mediator. A convolutional pose machine (CMU's OpenPose) identifies key body pose features such as the head pose direction, hand and joint locations. As presented in Figure 3(b), this pose estimation is applied to further enable body orientation analysis and gauge kinetic aspects of engagement. Together these visual features will inform the construction of higher-level engagement features, such as instances of gaze directed from a learner to their peer or the robot, and body orientation relative to the group. Convolutional neural network-based emotion recognition applied to detected faces will provide prediction of emotional state. We will then explore a number of approaches (from basic logistic regression to recurrent neural networks and three-dimensional convolutional neural networks) for using these features to predict the overall engagement level of a session as it progresses.

(a) (b)

Figure 3. Examples of automated video analysis techniques.

There are significant challenges in the automated analysis of collaborative engagement from video. Even when the structure of a common mediation narrative and a generally consistent locale are provided, comparing engagement between teams of children is extremely difficult due to a number of factors, such as heterogeneity of background activity (e.g. others present in the room), differences in camera location, relative starting alignment of the students to human vs. robot mediators, and individual differences in how learners express engagement. Identifying the effects of these variations on the techniques being explored will be critical for the goal of producing structural recommendations for future studies.

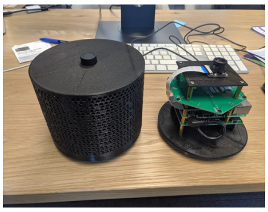
Additionally, some students may exhibit even more significantly different expressions of engagement, including those with learning challenges. Ensuring inclusion of such learners and a diverse representation of students overall in ongoing sessions will be a vital step towards developing more robust, broadly applicable and equitable methods for automated engagement analysis.

S-6. Collaboration Literacy Feedback Tool (CLiF)

Providing and receiving feedback about collaboration skills in real classrooms require a tool that is able to 1) capture synchronized multimedia signals from the environment, 2) automatically extract, analyze and fuse multimodal features from the recordings in real-time, and 3) provide multimodal feedback to students and instructors. This section will discuss our current advances in solving the technological and pedagogical challenges involved in creating this kind of tool, that we have named CLiF (Collaboration Literacy Feedback).

The first challenge was the ability to record synchronized multimedia signals of students' actions and interactions during collaborative learning activities in a way that is affordable and easy-to-use. We solved this by designing and building a low-cost sensor array consisting of a Raspberry Pi 3B microcomputer, a 6-microphone array and a 220-degree fisheye camera, as presented in Figure 4(a). A custom-made software system enables the synchronized use of several recorders at the same time with minimal training and setup. This system is able to capture video with a resolution of 1640x1232, a frame rate of 15 fps which is compressed using the high-quality profile of the H.264 codec. The captured audio is 6 channel, 16 bits, 16 Khz without compression. The system can be expanded to capture other media (digital pens, electro-dermal-activation) via external sensors through bluetooth connectivity. The approximate cost of each sensor is 150 USD.

Data extraction and analysis involves data capturing, a multi-level process, and AI techniques similar to the approaches described in various sections above. Lower-level features, such as body and head posture and speech content and speaker ID, will be used to create higher-level features that estimate relevant collaboration constructs. For example, gaze direction coupled with speaker ID and speech content provides an estimation of the level of attention that a student pays to her/his teammates and uptake of ideas across speakers. The final step in the project is to explore new and innovative ways to provide feedback to students and teachers. The main challenge in this step is to develop technologies that could help us exploit the multimodal nature of human communication to provide feedback in a way that is not disruptive of the collaborative activity. Online feedback (during the activity) will be compared with offline feedback (after the activity). Also, different combinations of automated vs. human feedback will be explored to determine their cost/benefit ratio.



(a)

(b)

Figure 4. Collaboration recorder (a) and body posture and facial features extracted (b).

One distinctive aspect of CLiF is that has been designed to respect and protect the privacy of students and instructors. Data ownership and the right to override any algorithmic decision from data capture to feedback interfaces. is embedded in the design of the feedback system. For example, each student has the capability of removing her data from the analysis or to restrict who has access to the final analysis of the data. All the human-faced systems will be evaluated according to perceived usefulness and level of intrusiveness.

Discussion

The work presented demonstrates the ways in which multiple behavioral modalities were marshalled to derive synthetic accounts of collaborative learning processes. The studies addressed the research problems by using complementarily both manual and computational approaches to collecting and analyzing behavioral data ranging from audio (S1, S4, S6) to visual (S2, S3, S4, S5, & S6), to utterances (S4, S5), and to log data (S1). Common to all was that data were collected unobtrusively and in authentic settings. Multimodal data analytics are still evolving, and it is noteworthy that human annotations were warranted as the ground-truth for the development of computational models.

Overall, the studies sought to overcome challenges with the large amount of analytical work while analyzing and interpreting of collaborative processes data, and thus served as testbeds for research programs aiming to leverage the potential of Big Data in education research. The computational approaches that the studies used enabled the automatic analysis of multimodal data sets and the production of credible outcomes in a timely manner (even in real time). This significantly increases the scalability of the research examining individuals' and groups' collaborative processes authentically at both microgenetic and global levels and for an extended time span. It is also foreseeable that the programs proven effective locally for supporting collaboration can be scaled up.

A challenge common to all is the extraction of meaningful information from the data, i.e. identifying information in the data that is relevant to answering research questions and that is grounded in theories of learning. The significance of learning theories was evident in the studies. Relevant learning theories drive choices of the appropriate computational approaches and the selection of information in the ontological and temporal perspectives. Some studies relied on keeping temporal segmentation at a minimum (e.g., every 3 sec) so as to capture changes in behaviors thoroughly, through which inferences to affective dynamics (S3) and collaborative engagement (S4, S5) are made. S1 and S2 selected the data that were meaningful within a theoretical framework, through which inferences to learning processes were made. S4 exemplified how learning theories are operationalized into formalisms that validated computational analyses. This area should be further developed.

Linking learning theory to computational models requires close collaboration among experts from multiple disciplines, and the studies presented evidence for the effectiveness of this multidisciplinary collaboration. Learning scientists, behavioral scientists, interaction design researchers, and computer scientists were able to jointly create new conceptual, theoretical and methodological innovations, breaking new ground in the study of collaboration in authentic contexts.

References

Baker, R., & Siemens, G. (2014). Educational Data Mining and Learning Analytics. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 253-272), New York: Cambridge University Press.

Blikstein, P. (2011). *Using learning analytics to assess students' behavior in open-ended programming tasks.* Proceedings of the 1st International Conference on Learning Analytics and Knowledge LAK '11, Banff, Alberta, Canada, Feb 27, 2011.

Cafaro, F. (2012). Using embodied allegories to design gesture suites for human-data interaction. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12* (p. 560). New York, New York, USA: ACM Press. https://doi.org/10.1145/2370216.2370309

Cheung, V., Watson, D., Vermeulen, J., Hancock, M., & Scott, S. (2014). Overcoming interaction barriers in large public displays using personal devices. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces* (pp. 375–380). ACM.

Coenen, J., Claes, S., & Moere, A. Vande. (2017). The concurrent use of touch and mid-air gestures or floor mat interaction on a public display. In *Proceedings of the 6th ACM International Symposium on Pervasive Displays* (p. 9). ACM.

D'Angelo, C. M., Smith, J., Alozie, N., Tsiartas, A., Richey, C., & Bratt, H. (2019). Mapping individual to group level collaboration indicators using speech data. In Lund, K., Niccolai, G., Lavoué, E., Hmelo-Silver, C., Gweon, G., & Baker, M. (Eds). A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings: 13th International Conference on Computer Supported Collaborative Learning. Lyon, France: International Society of the Learning Sciences.

Elmqvist, N. (2011). Embodied human-data interaction. In *ACM CHI 2011 Workshop "Embodied Interaction: Theory and Practice in HCI* (pp. 104–107).

Fredericks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: potential of the concept, state of the evidence. *Review of Educational Research,* 74(1), 59-109.

Grzadzinski, R., Carr, T., Colombi, C., McGuire, K., Dufek, S., Pickles, A., & Lord, C. (2016). Measuring changes in social communication behaviors: preliminary development of the brief observation of social communication change (BOSCC). *Journal of Autism and Developmental Disorders*, *46*(7), 2464-2479.

Griffin, P., & Care, E. (2014). *Assessment and teaching of 21st century skills: Methods and approach*: Springer.

Kantz, H., Kurths, J., & Mayer-Kress, G. (Eds.) (2012). *Nonlinear analysis of physiological data*. Springer Science & Business Media, Secaucus, NJ.

Lee, S. B., Espino, D. P., Van Tress, L., & Hamilton, E. R. (2020). Examining connections between positive affect and discourse in a global, STEM-focused digital makerspace community. Paper accepted for presentation at the 2020 Annual Meeting of the American Educational Research Association.

Lund, K., Quignard, M., & Shaffer, D. W. (2017). Gaining insight by transforming between temporal representations of human interaction. *Journal of Learning Analytics, 4*(3), 102-122.

Martin, T. & Sherin, B. L. (2013). Learning Analytics and Computational Techniques for Detecting and Evaluating Patterns in Learning: An Introduction to the Special Issue. *The Journal of the Learning Sciences, 22*(4), 511-520.

Ochoa, X., & Worsley, M. (2016). Augmenting Learning Analytics with Multimodal Sensory Data. *Journal of Learning Analytics, 3*(2), 213-219.

Pekrun, R. (1992). The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied Psychology*, *41*(4), 359-376.

Smith, J., Bratt, H., Richey, C., Bassiou, N., Shriberg, E., Tsiartas, A., D'Angelo, C., & Alozie, N. (2016). Spoken interaction modeling for automatic assessment of collaborative learning. Proceedings of the Speech Prosody Conference 2016, pp 277-281. Boston, MA.

Voss, C., Schwartz, J., Daniels, J., Kline, A., Haber, N., Washington, P., ... & Feinstein, C. (2019). Effect of wearable digital intervention for improving socialization in children with autism spectrum disorder: A randomized clinical trial. *JAMA Pediatrics*, *173*(5), 446-454.

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1083–1092). New York, NY, USA: ACM. https://doi.org/10.1145/1518701.1518866

Acknowledgments

The authors gratefully acknowledge funding support from the US National Science Foundation for the work included in this symposium: Awards #1848898 (S2), #1612824 (S3), #1839194 (S4), & #1623561 (S5). Views appearing in this paper do not reflect those of the funding agency.