Towards Collaboration Literacy Development through Multimodal Learning Analytics

Marcelo Worsley
Northwestern University, Evanston, IL, USA
marcelo.worsley@northwestern.edu

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

ABSTRACT: The last ten years has involved significant growth and development in the learning analytics community. One of the developments to recently emerge as a recognized special interest group in Learning Analytics is the sub-field of Multimodal Learning Analytics (MmLA). In this paper we consider a future trajectory for MmLA that intersects with the cross-cutting 21st century skill of collaboration. Teaching collaboration is seldom the focus of formal, or informal learning experiences, as students and teachers rarely receive feedback on their collaboration process. Instead, feedback is normally reduced to an outcome measure, or requires a level of human analysis that is intractable at scale. We see a unique opportunity for MmLA to promote collaboration literacy, and for collaboration literacy to be a common space in which to grow MmLA. Concretly, MmLA can provide the theoretical and technological innovations needed to create tools that support the evaluation, assessment and development of collaborative skills. As a first step in this direction, this paper presents a framework for collaboration literacy that consists of four levels of increasing complexity. We describe examples of current work in the first three levels of the framework, and situate the fourth level as an aspirational goal for the field of MmLA. We also discuss some of the key challenges that need to be solved to facilitate increased adoption of a collaboration literacy feedback tool, and MmLA more broadly. Ultimately, we argue that the development of such a tool could be instrumental in introducing new ways for building collaboration literacy.

Keywords: multimodal feedback, data fusion, framework, data capture

1 INTRODUCTION

Being able to effectively practice collaborative problem solving has been widely identified as one of the key skills that is needed to succeed, learn, and work in the 21st-century (Dede, 2010). Despite its importance, the development of collaboration skills usually consists of exposing students to a series of collaborative experiences with limited scaffolding. Furthermore, most of the feedback is about the final product of the collaboration, with almost no feedback about student in-situ collaboration skills (Lai, 2011). These “activity-as-instruction” approaches for the development of collaboration skills usually overlooks individual performance (Lai, 2011) and severely restrict the potential learning opportunities from otherwise carefully designed activities.

This focus on evaluation of collaboration artifacts instead of evaluation of the collaborative process is not an oversight or pedagogically-justified preference. It is a predictable result of both a lack of literacy and capacity to provide such feedback. It is easier to provide and receive feedback about a concrete artifact that has a predefined physical or digital form and can be easily shared between participants, than to give feedback about a sequence of remembered actions that are not always shared (or remembered) by students or teachers. While analyzing the collaboration process is possible, and it is routinely done for research purposes (e.g.: Kleinsmann, Deken, Dong, and Lauche (2012); Berland, Davis, and Smith (2015); Shaffer, Collier, and Ruis (2016)), the level of observation, coding and analysis currently required is scarcely practical for a teacher (or students) during a normal collaborative activity. Providing feedback about the collaboration behaviour of each student is a desirable goal. However, the current practice is too laborious and time consuming to be routinely used.

There have been several proposals on how to improve the feasibility of teaching collaboration skills in regular educational contexts. These proposals could be summarized in what Griffin and Care (2014) note when talking about how to improve the teaching and assessment of 21st century skills: ``New forms of data collection needed to be devised, and methods of analysing those new forms of data need to be identified and tested``. This challenge perfectly aligns with the goal of Multimodal Learning Analytics (Ochoa & Worsley, 2016) (MmLA): capturing, analyzing and fusing several streams of data to better understand and improve learning processes. MmLA is uniquely positioned to bridge the gap between what is a desirable pedagogical approach (providing detailed feedback about student collaboration practices) and what is practical (a pedagogical/technical tool that can be easily used in general collaborative activities in the classroom). MmLA provides the technical tools to easily capture human behaviour using low-cost, synchronized, multimodal sensors (Domínguez, Chiluiza, Echeverria, & Ochoa, 2015; Martinez-Maldonado, Echeverria, Santos, Santos, & Yacef, 2018) and to estimate learning-relevant constructs based on the analysis and fusion of that multimodal data (Di Mitri, Schneider, Specht, & Drachsler, 2018; Worsley, 2018a)

As we consider the design of learning analytics tools that can facilitate the development of collaboration, we argue that a primary focus should be to rapidly provide feedback that is clear and actionable. To do this, it should be able to fulfil the following requirements:

- It should be able to automatically provide a detailed multimodal transcript (Ochoa, Chiluiza, et al., 2018) that summarizes the relevant actions that occurred during the collaboration activity. This information can support reflection by the teacher or the students. Teacher-facing dashboard for students collaboratively building database diagrams (Granda, Echeverría, Chiluiza, & Wong-Villacrés, 2015) and an Emergency Room behaviour reflection tool for nurses in training (Martinez-Maldonado et al., 2019) are examples that satisfy this requirement.

- It should also provide estimates of collaboration-relevant constructs. These estimates should be based on objectively measured quantities that augment an individual's capability to understand their collaboration behaviour. An example in another domain is the analytical report provided by an oral presentation feedback tool (Ochoa, Domínguez, et al., 2018).

- It should integrate seamlessly into the collaboration activity. For example, an instrumented table that displays the percentage of conversation-time used by each participant around the table (Bachour, Kaplan, & Dillenbourg, 2010) would satisfy this type of requirement.

- It should support different types of classroom orchestration (Dillenbourg, Prieto, & Olsen, 2018) and not be designed for just one type of collaborative activity. Alternatively, it should have utility for different learning context.

This list of requirements currently lies in the fuzzy frontier of what is pedagogically beneficial and technically possible. The main contribution of this paper is detailing the guidelines and a possible road-map to build a tool that is both theoretically-grounded and technologically feasible. Section 2 will introduce the concept of collaboration literacy and a novel multi-level framework to connect collaboration constructs with existing MmLA research. Section 3 will present the technical challenges to convert the framework into a functioning tool that could be easily deployed in real-world scenarios. Finally, we conclude with remarks on a potential path forward for MmLA and collaboration literacy.

2 DEFINING COLLABORATION LITERACY

We conceptualize collaboration literacy as the ability to ascertain and respond to changes in the quality of a collaborative experience. From the student perspective this amounts to being conscious of one's own contribution to a group, as well as the awareness and ability to intervene in order to ensure a strong collaboration. From the teacher perspective this includes awareness of how different groups are progressing, being able to respond to those groups in a timely fashion, and developing prompts and activities that afford good collaboration.

It is easy to adopt a simplistic perspective around collaboration quality that consists of generic labels for “good” and “bad” collaboration. Without question, there are practices that can promote more or less effective collaboration. At the same time, however, there are a number of more complex practices and behaviours that differentially contribute to the nature of a collaboration. Within this paper, we provide a framework for thinking about the complexity of different constructs. We position the framework as being important to guiding on-going research at the intersection of MmLA and collaboration. Concretely, it provides a framework that researchers can use to position their work, and also provides aspirational goals for where their work might go. Table 1 identifies collaboration related constructs, the research that supports their salience and where we place each construct relative to the multi-level system that we describe below. This list is not exhaustive. Advancing collaboration literacy helps participants learn to recognize the different levels at which one might quantify collaboration quality.

Note: Though we use literal versus semantic in thinking about the different levels, these are approximations, as these terms are not true binaries. Instead, there is a continuum between literal and semantic that is difficult to represent in a strict set of four categories. Nonetheless, we make every effort to draw clear distinctions between the different levels of the framework, and propose that multimodal data can afford a more semantic representation of a student's actions or perceived state.

Table 1. Levels of collaboration analysis, their associated constructs and example references

Level	Constructs	Example Reference
1 Unimodal Literal Individual	Text Contributions	Leeuwen van Leeuwen, Janssen, Erkens, and Brekelmans (2015)
	Head Position	Worsley, Scherer, Morency, and Blikstein (2015)
	Affective State	Richey, D'Angelo, Alozie, Bratt, and Shriberg (2016)
	Tactile Engagement	Worsley (2018a)
2 Unimodal Literal Multi-party	Entrainment	Lubold and Pon-Barry (2014)
	Body Synchrony	Cukurova, Luckin, Millán, and Mavrikis (2018)
	Joint Visual Attention	B. Schneider et al. (2018)
	Turn Taking	Devault, Mell, and Gratch (2015)
3 Multimodal Semantic Individual	Turn Management	Emma M. Mercier, Higgins, and da Costa (2014)
	Questioning	Hmelo-Silver and Barrows (2008)
	Monitoring	Hmelo-Silver and Barrows (2008)
	Summarizing	Wise and Chiu (2011)
4 Multimodal Semantic Multiparty	Rapport	Gratch, Wang, Gerten, Fast, and Duffy (2007)
	Negotiating Interactive	Emma Mary Mercier, Higgins, Burd, and Joyce-Gibbons (2012)
	Convergent Conceptual Change	Roschelle (1992)
	Struggling	Bassiou et al. (2016)

2.1 Level 1: Unimodal, Literal, Individual

The first level for examining collaboration quality involves looking at how a given individual is contributing or engaging through a single modality. Within research on collaboration, the modalities of speech and text tend to be privileged, as verbal, or textual, engagement is often a precursor to an effective collaboration (Hmelo-Silver & Barrows, 2008; Emma M. Mercier et al., 2014; Richey et al., 2016; van Leeuwen et al., 2015). Within this paradigm, researchers might look at the quantity and quality of speech and/or text generation. For example, researchers might look at the speech fraction, or the amount of time someone is talking, relative to the total time of the activity. Speech fraction values approaching 0 or 1 are likely to be indicative of interactions that were not very collaborative. Text also tends to be a rich modality for gleaning insights about the nature of an interaction. With the assistance of computational tools (i.e., Lightside (Mayfield & Rosé, 2013) or Natural Language Toolkit (Loper & Bird, 2002)) or through human annotation, researchers can begin to develop a better understanding of an individual's cognitive, social, or emotional state during a given collaborative activity. Inferences about participant state can also be inferred from the use of video or voice technology. These unimodal data points can be informative for characterizing the relative success of the interaction, and for enabling easy identification of noticeable changes in individual participation. Early work in MmLA demonstrated how looking at a single modality, could help predict collaboration among students completing math problems. Specifically, Ochoa et al. (2013) found that using various simple features for how fast someone writes, or draws, the percentage of time they use the calculator and how much they mention numbers or mathematical terms, are good proxies for predicting their level of expertise within a group. Hence, in certain situations, a unimodal, individual approach can provide a reasonable starting point for ascertaining the nature of a group collaboration.

2.2 Level 2: Unimodal, Literal, Multi-party

Level 1 looked at how a given individual may be engaging with a specific task through a single modality. Level 2 extends these measures to being multi-party. This has been a primary paradigm utilized by collaboration researchers (Hmelo-Silver & Barrows, 2008; Emma Mary Mercier et al., 2012; Emma M.

Mercier et al., 2014; Roschelle, 1992; Wise & Chiu, 2011). For example, in the case of verbal contribution, instead of looking at how much someone talked, researchers consider the nature of turn-taking within a given group. In addition to looking at how speaking turns are distributed across a group, researchers might take a more qualitative approach, and code participant utterances for ways that turns are managed, and the ways that a given speaker's idea is taken up by the other participants. For instance, researchers might label when someone is responding to a previous utterance, or examine when a given utterance signals student agreement with a given idea (Richey et al., 2016; Wise & Chiu, 2011). In these cases, a given utterance only becomes relevant in context of the surrounding utterances, and in the context of the other individuals within the space. This is one of the important pieces added by considering multi-party collaboration. The multi-party level also allows for more direct consideration of the power relations or social dynamics of a given group. Cukurova et al. (2018) provide an informative analysis using level 2 collaboration to study social dynamics among a group of students completing a hands-on task. Specifically, they analyze hand gesticulation to look at the extent of body synchrony among participants. They also answer questions about whether or not group participants appear to be exerting the same amount of body movement at a given time, or if the amount of physical movement is unevenly distributed. Another good example of unimodal, multiparty collaboration is visual joint attention (B. Schneider et al., 2018). Joint visual attention refers to instances where two, or more, individuals are looking at the same location, or object, at, roughly, the same time. Many prior studies have highlighted the importance of joint visual attention for promoting learning and perceptions of collaboration quality. It can also be indicative of power relations, when considering who, within a given collaborative group, receives the visual attention of their peers when speaking. Broadly speaking, Level 2 moves us closer to multi-party metrics, and, adds additional challenges and opportunities. Data must be synchronized across different participants, but upon doing so, it becomes easier to identify group-level patterns that emerge.

2.3 Level 3: Multimodal, Semantic, Individual

Level 3 introduces semantics and multimodaltility. Whereas Level 1 analysis of text using tools like Lightside (Mayfield & Rosé, 2013) and Coh-metrix (Graesser, McNamara, & Kulikowich, 2011), provide some degree of utterance understanding and intent interpretation, level 3 surfaces the opportunities for using data from different modalities to represent perceived student state. For instance, we can represent user engagement based on the presence or absence of speech. If, though, a user is generating speech without addressing their peers, as detected through head pose estimation, or eye gaze, it becomes less likely that the user is engaging in the collaborative task. Returning to the example of verbal contributions, when we talk about semantics we are examining the words that are uttered, and deriving meaning from those collections of words. Taking a multimodal and semantic perspective with gesticulations refers to identifying specific gestures in the gesticulations that a student is making, and, perhaps, connecting those gestures with user spoken utterances. A simple gesture one might make in the context of a classroom is raising one's hand, or pointing. In order to ascertain these gestures, one has to rely on a semantic understanding of the student's gesticulation. In the case of a student pointing, there is an additional multimodal component of understanding what they are pointing to. Prior work in learning analytics has begun to consider collaboration at this level. For example, Worsley and Blikstein (2018), inspired by Scherr and Hammer (2009), develop multimodal representations of individuals in collaborative pairs. Scherr and Hammer (2009) describes epistemological frames, which constitute a combination of modalities (e.g., speech, head pose, gesticulation) that, when combined, provide a sense of the type of activity a student is undertaking at a given time. Each of the epistemological frames: Discussion, Lecture, Teaching Assistant and Joking; is characterized by a different combination of the aforementioned modalities. Worsley and Blikstein (2018) extend this idea by using a electro-dermal activation, speech and hand/wrist movement to identify four representative modes of collaboration during a hands-on building task. Importantly, that particular analysis was primarily done on an individual basis and did not consider the ways individuals reacted to one another, which is a key differentiator between Level 3 and Level 4.

2.4 Level 4: Multimodal, Semantic, Multi-party

Level 4 elevates the level 3 measures to multi-party inferences. Here, we consider measures like shared understanding and convergent conceptual change. Roschelle's work (1992) on convergent conceptual change highlights ways that groups negotiate the collaborative learning process through a combination of gestures and spoken turns. Specifically, convergent conceptual change is

> [C]haracterized by: (a) the production of a deep-featured situation, in relation to (b) the interplay of physical metaphors, through the constructive use of (c) interactive cycles of conversational turn-taking, constrained by (d) the application of progressively higher standards of evidence for convergence.

Convergent conceptual change represents a complex interplay of student actions around a shared task. The component constituents of the interaction can be reasonably characterized through semantic, multimodal interpretation of gestures and verbal utterances. For example, speaking turns can be labeled through speech recognition, and physical gesticulations analyzed for specific gestures. The semantics of student utterances can be interpreted for different measures of cohesion, or argumentation, and combined with the corresponding gestures. However, the actual demonstration of conceptual change requires an additional level of inference that goes beyond the individual. It necessitates that an individual's data be interpreted relative to the data of the other participants.

Broadly speaking, Level 4 measures require a semantic and multimodal interpretation of group behavior, often across time and at variable time scales. This is an area of research that has received little attention from the MmLA community, and reasonably involves the highest amount of complexity. It also requires a certain level of accuracy within the level 3 measures and data representations. Part of what we propose in this paper is that developing theories and representations of collaboration that mirror the complexity of convergent conceptual change, is one of the opportunities for the future of MmLA.

We argue that all of the levels could benefit from MmLA. Levels 1 and 2, are forms of interaction that can be reasonably approximated through current artificial intelligence technologies. Levels 3 and 4, could, at present, be researched through a combination of human-machine analyses, with the eventual goal of being incorporated into real-time tools. In consideration of these factors, the section to follow describes technical challenges that we are exploring to realize developing such a tool.

3 TECHNICAL CHALLENGES

Most of the state-of-the-art in capturing and analyzing collaboration construct is the result of lab-based prototypes in which instructors and learners are only involved during the data-capturing phase. Building a tool that can be used on a regular basis in common learning contexts to improve collaboration literacy needs not only models to convert raw recording data into high-level constructs, but also a technical infrastructure that makes it deployable, scalable and acceptable. This section will provide a discussion of some challenges and potential solutions.

3.1 Type of Sensors and Modalities

There is large range of sensors and modalities that have been used in MmLA studies (Di Mitri et al., 2018; Ochoa, 2017). There is an inherent tension between the desire to capture as many modalities as possible and the complexity and intrusiveness of the recording apparatus. Given the set of constructs defined in the previous section, the recommended trade-off between the two extremes is a combination of high-definition horizontal 360 degrees video (captured with a simple camera and a fisheye lens) and directional audio. Apart from existing log data captured by digital systems, video cameras and microphones have been the sensors of choice in MmLA research. This preference for audio and video is because they can reliably capture the primary forms of human communication, have high information density, align with the information captured by human senses, are low cost, are easy to deploy and are non-intrusive (Worsley, 2018b). Different from considerable prior work, however, is the inclusion of microphone arrays, which allow for the collection of directional audio, and 360 degree cameras, which provide for substantial coverage of a given learning environment. Based on current state-of-the-art in MmLA, these sensors enable the capture of posture, gaze, facial expression, hand gestures and actions, position, speaker identity, speech verbal and non-verbal features (Ochoa, 2017). While other sensors (e.g., biophysiological sensors) are available for instrumenting people and learning environments, we want to be careful about balancing the utility of the sensors, with concerns about data privacy and ethics and deployability.

3.2 Synchronicity

Synchronicity of the recordings allow the fusion of information from different modalities. Synchronization precision depends on the type of signals being combined and the type of analysis to be conducted on the resulting features. To establish the level of synchronization needed, we reference Newell's time scale of human actions (Newell, 1994). Newell defined different time spans for several learning-related human actions and reactions. These speeds are divided into several bands according to the type of process that generates it. The bands are biological (100 microseconds to 10 milliseconds), cognitive (100 milliseconds to 10 seconds), rational (1 minute to hours) and social (days to months). The most relevant aspects of human collaboration, and also the ones that are deliberate by the student and perceptible to a human observer, are in the cognitive, rational and social bands. The lower bound for these kinds of signals is a tenth of a second. This level of synchronization is perfectly achievable with current off-the-shelf technologies. For example, the Social Signal Interpretation (SSI) framework (Wagner et al., 2013), which allows for synchronization on the order of milliseconds even when the recording is distributed across different devices, provides a viable solution for simplifying synchronization.

3.3 Deployment

Having a recording apparatus that can be setup and operated by non-experts is a main challenge in moving from lab conditions to real learning settings. Two options have been tested to facilitate this change: fixed pre-configured setups and user-friendly mobile setups. In the first option, a complex recording system is built and configured ahead of time. With this system, users are only permitted to complete two actions: turning on the recording, and turning off the recording. This strategy has been followed by Ochoa, Domínguez, et al. (2018) in their widely deployed Oral Presentation Feedback system. A commercial example of this type of devices is the Meeting Owl, a videoconference camera for group meetings. The creation of these kinds of mobile devices requires both engineering and user-based-design efforts to create easy-to-use interfaces for minimalist hardware. Given that collaboration activities could happen in any classroom, it is recommended that systems for collaboration literacy feedback follow this second approach where the recording device is mobile and easily operable by the collaboration activity participants.

3.4 Real-time vs. Post-hoc Feedback

When the feedback is provided can have an important effect on its usefulness and actionability. The difference is exemplified by two types of multimodal oral presentation feedback tools. The first one, introduced by (J. Schneider, Börner, Rosmalen, & Specht, 2015) presents simple feedback about posture, gaze and volume in real-time to the presenter through an augmented reality visor. The second, a system proposed by (Ochoa, Domínguez, et al., 2018) provides a more detailed feedback about the same modalities through a multimodal report but only after the presentation is finished. Both system show positive learning gains. It is not clear what system is more appropriate to develop presentation skills, or if a combination of the two is the right answer. A system for collaboration literacy feedback should explore both real-time multimodal signals and post-hoc reflection reports in order to find which one has a stronger impact on learning different collaboration literacy constructs. It is also important to consider the computational requirements that real-time feedback have in the multimodal extraction and fusion component.

3.5 Individual vs. Group Feedback

Feedback can be provided privately to individuals about their individual collaboration behaviour. However, there is also an element of group dynamics that cannot be explained by individual contributions alone. Some collaboration constructs only make sense at the group level. Also, exposing individual feedback to the group has the potential to violate the right of privacy of the individual. Studies such as (Archer-Kath, Johnson, & Johnson, 1994) where the individual vs. group feedback is empirically tested should be conducted to test impact of collaboration literacy feedback interfaces.

3.6 Automated vs. Human-augmented Feedback

Even with current advances in artificial intelligence, there are certain aspects of collaboration behaviour that cannot be detected or processed by current automated systems. Human feedback has the potential to be of higher quality than automated systems. However, human feedback is not without limitations. Most importantly, it is not scalable, and is subject to bias. Combining the right proportion of both types of feedback seems to be the right approach. This determination is also an open research question that should be addressed after determining which types of collaboration constructs can be accurately and reliably estimated automatically, and which ones still need human input. Moreover, systems for collaboration literacy feedback could provide an interface for human instructors to focus their capabilities on resolving difficult-to-judge cases or constructs. For example, an automated system could provide feedback about turn management and questioning, while an annotated multimodal transcript (such as in (Echeverria et al., 2018)) could be provided to instructors to focus their attention on key moments of the collaborative activity. The combination of automated and human-augmented feedback could also support teachers, students and researchers focusing on higher level collaboration constructs.

4 CONCLUSIONS

In this paper, we argue that MmLA is uniquely poised to analyze collaborative learning environments. Moreover, we propose a framework for considering the different levels of complexity of collaborative problem solving, with the goal of supporting the development of collaboration literacy, a form of literacy that receives little formal attention within mainstream, and even progressive learning experiences. Enacting the creation of collaboration literacy feedback tools can potentially be achieved through a combination of low-cost audio and video data capture technology, in conjunction with the development of robust multimodal fusion, and multimodal feedback strategies. Determining the design of collaboration literacy feedback tools will involve research and development along several of the dimensions outlined in this paper, and likely some additional dimensions that have yet to be identified. Nonetheless, we position the ideas included in this paper as a concrete, constructive and feasible research agenda for simultaneously advancing MmLA and collaboration literacy. Levels 1 through 3 of the framework represent constructs that MmLA can address in the short-term. These constructs can be enacted through unimodal and multimodal features that are available through current artificial intelligence technologies. Level 4 represents an aspirational goal for MmLA. Such investigations have the opportunity to drive new theories and conjectures about the complexities of group collaboration, much like Roschelle's work (1992) on convergent conceptual change. Finally, this work, as a whole suggests the need to close the gap in MmLA by promoting important, real-time feedback (Bassiou et al., 2016) and to carefully consider issues of ethics and data privacy.

Our hope is that this paper will help provide direction for the field to more quickly converge towards the development of common apparatus for distributed data collection, shared measures, and consistent feature extraction and fusion algorithms.

REFERENCES

Archer-Kath, J., Johnson, D. W., & Johnson, R. T. (1994). Individual versus Group Feedback in Cooperative Groups. *The Journal of Social Psychology, 134*(5), 681-694. doi:10.1080/00224545.1994.9922999

Bachour, K., Kaplan, F., & Dillenbourg, P. (2010). An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning. *IEEE Trans. Learn. Technol., 3*(3), 203-213. doi:10.1109/tlt.2010.18

Bassiou, N., Tsiartas, A., Smith, J., Bratt, H., Richey, C., Shriberg, E., . . . Alozie, N. (2016). *Privacy-Preserving Speech Analytics for Automatic Assessment of Student Collaboration.* Paper presented at the Interspeech 2016.

Berland, M., Davis, D., & Smith, C. P. (2015). AMOEBA: Designing for collaboration in computer science classrooms through live learning analytics. *International Journal of Computer-Supported Collaborative Learning, 10*(4), 425-447.

Cukurova, M., Luckin, R., Millán, E., & Mavrikis, M. (2018). The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Computers & Education, 116*, 93-109.

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

Devault, D., Mell, J., & Gratch, J. (2015). *Toward Natural Turn-Taking in a Virtual Human Negotiation Agent.* Paper presented at the AAAI Spring Symposium on Turn-Taking and Coordination in Human-Machine Interaction.

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning, 34*(4), 338-349.

Dillenbourg, P., Prieto, L. P., & Olsen, J. K. (2018). Classroom Orchestration. In *International Handbook of the Learning Sciences* (pp. 180-190): Routledge.

Domínguez, F., Chiluiza, K., Echeverria, V., & Ochoa, X. (2015). *Multimodal selfies: Designing a multimodal recording device for students in traditional classrooms.* Paper presented at the Proceedings of the 2015 ACM on International Conference on Multimodal Interaction.

Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., & Shum, S. B. (2018). *Driving data storytelling from learning design.* Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational researcher, 40*(5), 223-234.

Granda, R. X., Echeverría, V., Chiluiza, K., & Wong-Villacrés, M. (2015). *Supporting the assessment of collaborative design activities in multi-tabletop classrooms.* Paper presented at the 2015 Asia-Pacific Conference on Computer Aided System Engineering.

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). *Creating Rapport with Virtual Agents*, Berlin, Heidelberg.

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

Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating Collaborative Knowledge Building. *Cognition and Instruction, 26*(1), 48-94. doi:10.1080/07370000701798495

Kleinsmann, M., Deken, F., Dong, A., & Lauche, K. (2012). Development of design collaboration skills. *Journal of Engineering Design, 23*(7), 485-506. doi:10.1080/09544828.2011.619499

Lai, E. R. (2011). *Collaboration: A literature review*. Retrieved from

Loper, E., & Bird, S. (2002). NLTK: the natural language toolkit. *arXiv preprint cs/0205028*.

Lubold, N., & Pon-Barry, H. (2014). *Acoustic-Prosodic Entrainment and Rapport in Collaborative Learning Dialogues.* Paper presented at the Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge, Istanbul, Turkey.

Martinez-Maldonado, R., Echeverria, V., Elliott, D., Axisa, C., Power, T., & Shum, S. B. (2019). Making the Design of CSCL Analytics Interfaces a Co-design Process: the Case of Multimodal Teamwork in Healthcare.

Martinez-Maldonado, R., Echeverria, V., Santos, O. C., Santos, A. D. P. D., & Yacef, K. (2018). *Physical learning analytics: A multimodal perspective.* Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge.

Mayfield, E., & Rosé, C. P. (2013). LightSIDE: Open source machine learning for text. In *Handbook of Automated Essay Evaluation* (pp. 146-157): Routledge.
Mercier, E. M., Higgins, S., Burd, E., & Joyce-Gibbons, A. (2012). *Multi-touch technology to support multiple levels of collaborative learning in the classroom*.
Mercier, E. M., Higgins, S. E., & da Costa, L. (2014). Different leaders: Emergent organizational and intellectual leadership in children's collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning, 9*(4), 397-432. doi:10.1007/s11412-014-9201-z
Newell, A. (1994). *Unified theories of cognition*: Harvard University Press.
Ochoa, X. (2017). Multimodal learning analytics. In G. S. Charles LANG, Alyssa WISE, and Dragan GAŠEVIĆ (Ed.), *The Handbook of Learning Analytics* (Vol. 1, pp. 129-141).
Ochoa, X., Chiluiza, K., Granda, R., Falcones, G., Castells, J., & Guamán, B. (2018). *Multimodal Transcript of Face-to-Face Group-Work Activity Around Interactive Tabletops.* Paper presented at the CrossMMLA@ LAK.
Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B., & Castells, J. (2013). *Expertise estimation based on simple multimodal features.* Paper presented at the Proceedings of the 15th ACM on International conference on multimodal interaction.
Ochoa, X., Domínguez, F., Guamán, B., Maya, R., Falcones, G., & Castells, J. (2018). *The rap system: automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors.* Paper presented at the Proceedings of the 8th international conference on learning analytics and knowledge.
Ochoa, X., & Worsley, M. (2016). Augmenting Learning Analytics with Multimodal Sensory Data. *Journal of Learning Analytics, 3*(2), 213-219.
Richey, C., D'Angelo, C., Alozie, N., Bratt, H., & Shriberg, E. (2016). *The SRI Speech-Based Collaborative Learning Corpus*. Retrieved from
Roschelle, J. (1992). Learning by Collaborating: Convergent Conceptual Change. *The Journal of the Learning Sciences, 2*(3), 235-276. Retrieved from http://www.jstor.org/stable/1466609
Scherr, R. E., & Hammer, D. (2009). Student Behavior and Epistemological Framing: Examples from Collaborative Active-Learning Activities in Physics. *Cognition and Instruction, 27*(2), 147-174. doi:10.1080/07370000902797379
Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning, 13*(3), 241-261.
Schneider, J., Börner, D., Rosmalen, P. v., & Specht, M. (2015). *Presentation Trainer, your Public Speaking Multimodal Coach.* Paper presented at the Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, Seattle, Washington, USA.
Shaffer, D. W., Collier, W., & Ruis, A. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics, 3*(3), 9-45.
van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *Computers & Education, 90*, 80-94. doi:https://doi.org/10.1016/j.compedu.2015.09.006
Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F., Andr, E., & #233. (2013). *The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time.* Paper presented at the Proceedings of the 21st ACM international conference on Multimedia, Barcelona, Spain.
Wise, A. F., & Chiu, M. M. (2011). Analyzing temporal patterns of knowledge construction in a role-based online discussion. *International Journal of Computer-Supported Collaborative Learning, 6*(3), 445-470. doi:10.1007/s11412-011-9120-1

Worsley, M. (2018a). *(Dis) engagement matters: Identifying efficacious learning practices with multimodal learning analytics.* Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge.

Worsley, M. (2018b, 1/1). *Multimodal learning analytics' past, present, and, potential futures.* Paper presented at the CEUR Workshop Proceedings.

Worsley, M., & Blikstein, P. (2018). A Multimodal Analysis of Making. *International Journal of Artificial Intelligence in Education, 28*(3), 385-419. doi:10.1007/s40593-017-0160-1

Worsley, M., Scherer, S., Morency, L.-P., & Blikstein, P. (2015). *Exploring Behavior Representation for Learning Analytics.* Paper presented at the Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, Seattle, Washington, USA.