International Society of the Learning Sciences

Exploring the Process of Group-Based Collaboration: A Validation Argument for a Collaboration Model and Observation Rubric for Training Explainable Machine Learning Models

Nonye Alozie, Audra Wingard, Ron Fried, Bladimir Lopez-Prado, Arif Rachmatullah, Anirudh Som, Mindy Hsiao, Hui Yang and Sujeong Kim

Maggie.alozie@sri.com, audra.wingard@sri.com, ron.fried@sri.com, bladimir.lopez-prado@sri.com, arif.rachmatullah@sri.com, Anirudh.som@sri.com, mindy.hsiao@sri.com, hui.yang@sri.com, and sujeong.kim@sri.com

SRI International

Collaboration is an important learning process. During collaborative learning, students engage in group activities where they converge on goals, solve problems and make joint decisions. To understand the process of collaboration, we focused on how behavior and interaction patterns contribute to the social-relational space of collaboration. We have designed a multilayered conceptual model for the collaboration process and an observation rubric that identifies behaviors and interactions during collaboration that serves as the foundation for machine learning models that can provide behavioral insight into the process of collaboration. This study reports results on several validation studies performed to establish a validation argument for our collaboration conceptual model and collaboration rubric. Through disconfirming evidence, interrater reliability testing, expert reviews, and focus group interviews, we found that our stratified architecture of collaboration and rubric provide valid accounts and descriptions of human behavior and interactions that can be used to substantiate the collaboration process.

Introduction

Collaborative learning experiences involve interactive peer-to-peer interactions as students solve problems and converge on decisions (Roschelle, 1992) and is comprised of speech-based and behavior-based interactions that intersect for successful product outcomes. Collaboration is a central component of how students learn but presents challenges when determining its role in the learning process. Teachers use speech, namely, discourse and the quality of group products to determine how students negotiate solutions to problems as part of the collaboration process (Roschelle & Teasely, 1995; Guo, Saab, Post, & Admiraal, 2020). For this work, we use machine learning to delve deeper into the behavior-based interactions that students exhibit when collaborating because it is important for students to be mindful of how speech *and* behavior work together for the creation of high-quality group products (Steinbock, 2008). To explore the feasibility and challenges of using behavior to explain the process of collaboration, we have designed a stratified conceptual model for collaboration and rubric that uses nonverbal behaviors and interactions to identify patterns of collaboration to train machine learning models that can provide behavioral analytics for collaboration.

We performed various validation studies of our collaboration conceptual model and rubric to derive a validation argument for valid, reliable, and objective constructs of collaboration. To demonstrate a construct validity argument, we present evidence from disconfirming approaches, interrater reliability testing, expert reviews, and focus groups interviews. Our validation research questions are as follows: (1) How does our concept of collaboration accurately represent theoretically based constructs of collaboration? (2) To what extent do individual key behaviors demonstrate collaboration competencies needed in group settings? (3) To what extent does nonverbal behavioral information from each level of collaboration provide teachers with adequate information for understanding how well students collaborate?

Literature Review

Collaboration As a Process

We describe collaboration as a complex process through which a group of students interdependently and constructively explore ideas to search for solutions, construct new knowledge, and create a socio-cognitive space to generate and share goals, ideas, plans and tools (Roschelle, 1992; Grosz & Kraus, 1996; Child & Shaw, 2015). Our work focuses on two components of collaboration: 1) the process of collaboration (Roschelle & Teasely, 1995), and 2) the nonverbal gestures and behavioral interactions exhibited during collaborative group work.

Our collaboration conceptual model (Alozie et al., 2020) is based on theoretical models that bring together research on social factors (i.e., behaviors, social cohesion), cognitive science (i.e., social-cognitive systems), and education research (i.e., problem-solving strategies). We broke collaboration down into four building blocks. (1) *Gestures*, where students move their bodies to express an idea or meaning. (2) *Individual awareness,* where students build a community through their understanding of self. (3) *Social and cognitive awareness,* where students engage in cooperative interactions (Johnson & Johnson, 2008), interdependance, and build an understanding of "other" in a group space. (4) *Shared interactive culture*, where students share plans, intensions, goals and tools and work as a team.

Automating Collaboration Assessment Using Speech-based and Behavior-based Analytics.

The use of multimodal learning analytics offers insights into how students learn when performing unstructured tasks like problem solving in the digital and physical world (Blikstein, 2013). Emerging work in speech and textual analytics has shown promise in using machine learning to observe group interactions. Online collaboration assessment has been done through the analysis of peer discourse (Erkens & Janssen, 2008). Other research has collected speech-based data of students working face-to-face to assess collaboration quality (D'Angelo et al., 2015). Behavior analytics for collaboration has made great gains in isolating gestures, sounds, and interactions that can indicate the quality of collaboration (e.g., Subburaj et al., 2020). Visual behavior analytics have focused on evaluating individual facial expressions, group interactions, affect, and task success (D'Mello, 2020). Some studies combined speech and behavior analytics to assess collaboration (Martinez-Maldonado et al., 2013).

Addressing Bias in Artificial Intelligence and Human Interactions

Biased data will create biased machine learning models. If the training data is more representative of some behaviors than others, then the model's predictions would systematically be biased against the under-represented behaviors, and vice-versa (Roselli, Matthews, & Talagala, 2019). We understand that eradicating all bias is difficult and use vetting of the training data to decrease the occurrence of biased annotations (Roselli, Matthews, & Talagala, 2019). To address potential bias due to unequal amounts of data, we perform controlled mix-up data augmentation (Som et al., 2020) which creates additional training samples that use features from existing samples. To expand our sample variation, we recruit students that are diverse in race and gender identification and provide anti-bias training for human observers and analysts. We also use a modular annotation approach to avoid propagating bias across samples and to increase the explainability of each machine learning model.

Methods

Validating the Collaboration Conceptual Model and Rubric Using Disconfirming Evidence and Interrater Reliability Testing. We used discomfirming evidence (Booth et al., 2013) to establish validity of our collaboration conceptual model (RQ1); a procedure whereby the researcher looks for areas of disagreement and for divergent data (Roselli, Matthews, & Talagala, 2019). Two education researchers with expertise in learning science, classroom studies, and STEM education examined the collaboration conceptual model, looking for areas of disagreement. Six different researchers also watched and annotated videos (Barron & Pea, 2013) to establish inter-rater reliability (Cohen's Kappa above 0.6; Landis & Koch, 1977) and look for areas of disagreement. When disconfirming evidence was found, the researchers discussed their responses, came to a consensus, and refined the collaboration rubric.

Validating the Collaboration Rubric Using Participant Focus Groups. To determine the authenticity of our collaboration rubric (RQ2), we convened 9 focus groups that demonstrated diversity in age, race, gender, and occupation. Participants were asked several questions about collaboration. For example, *How do you define collaboration?* and *What do you think are the characteristics of a good collaborator?* We used open coding and selective coding processes to analyze the transcriptions of focus group interviews.

Validating the Collaboration Conceptual Model and Rubric Through Expert Reviews. We conducted expert reviews of the collaboration conceptual model and rubric using a survey and open-ended questions (RQ2 and 3). The expert review was a questionnaire consisting of Likert scale questions about the alignment of the collaboration building blocks to theory-based descriptions and constructs. For example, experts were asked questions like "*To what extent do the building blocks [of collaboration] align with the descriptions [of collaboration]?*" and given choices ranging from *Does not align* to *Aligns very well*. Using qualitative thematic analyses of the open-ended responses and calculating reliability scores of the ratings, we checked for consistency across reviews and then determined a set of revisions based on recommendations.

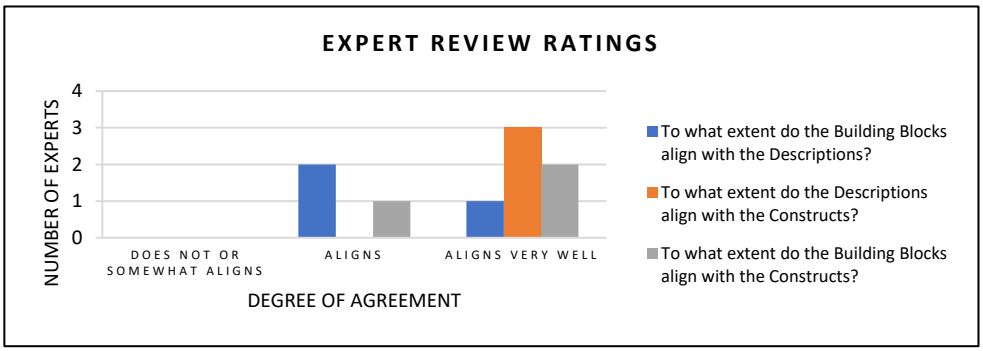
Findings

Definition of collaboration. Most participants stated that collaboration can be defined as shared goals, diversity, inclusivity of behaviors and interactions, distributed responsibilities and roles, an intersection of talents, and opportunities to participate. For example, one participant from Group 7 said, "The bare bones definition of collaboration is just a group working towards a common goal, and everyone bring their own ideas, background and experience in achieving that goal, whether it be through for school or for a personal project or for work." Another participant in said, "Inherency of the intersection of multiple people towards a common goal," and added, "Yeah, and it has to include progress." Similarly, a participant from Group 11 said, "In an ideal world, I would say that a collaborative task is one where you get meaningful efficiencies from having multiple people work on something," then added, "One opportunity for that is differentiated skills, knowledge, expertise, but not necessarily." Common themes among participants about the components of effective collaboration were 1) clear communication, 2) well assigned roles and division of labor, 3) trust and respect among team members, and 4) having shared goals and norms.

Collaboration Conceptual Model. According to the expert responses, our collaboration conceptual model showed alignment with theory and had internal consistency (see Figure 1). However, they commented that the conceptual model should be less linear to show that there is a network of movement between levels.

Figure 1

Construct alignment results from expert review



Collaboration Rubrics. The human annotators achieved above 0.6 Cohen's Kappa inter-rater reliability, showing a convergence in judgement while maintaining variation. However, the experts suggested that a wide range of behaviors would help reduce bias and suggested that the use of descriptions, rather than ratings would be more informative. For example, one expert wrote, "The demeanor codes are unlikely to have all students fitting the description. The codes are fine, but the descriptions are not detailed enough." Our experts pointed to differences in the kinds of interactions that are exhibited among neurodiverse individuals. For instance, one reviewer wrote, "Looking at screen: 'looking at speaker or activity' places significance on one's social interactions. Those who are neurodivergent may not be completing these activities with the same regularity as the neurotypical learners. This is true also of "sustained fidgeting" as well as "sitting still." The experts also recommended measuring heterogeneity of the group using percentage, to allow for group size variation.

Behaviors and interactions of a good collaborator: Emergent themes showed that participants identified a good collaborator as someone that is open-minded and driven, open to critiques, accepting of different perspectives, and welcoming of challenges. Themes also indicated that participants thought that an effective collaborator is considerate and attentive. For instance, a participant in Group 9 said, "[A good collaborator] doesn't just decide on, like, what they want to happen and how they want it to happen and just go for it. With the project, I think people who, like, work to help others express their ideas." Participants also highlighted good communication as important for collaboration. A participant in Group 2 said, "I think good communication from each team member; open mindedness and willingness to share ideas and to discuss other ideas and not force your own opinion based on, like, what you know. So, it's about open mindedness and communication for me." Additional interaction patterns identified by the participants were the ability to provide and follow directions and manage the group dynamics to achieve the articulated shared goal.

Conclusions and Next Steps

Our validation studies provided an effective validation argument for our collaboration conceptual model and rubric. Findings from our studies showed that (1) our concept of collaboration accurately represents theoretically based constructs of collaboration, (2) individual behaviors identified in our rubric effectively measure the

collaboration competencies needed in group settings, and (3) nonverbal behavioral information from each level of collaboration can provide teachers with insightful information for understanding how students collaborate. We found alignment to our working definition of collaboration (e.g., shared goals, co-construction of knowledge, and sharing of tools) and used the recommendations to revise our collaboration conceptual model.

Ultimately, our collaboration conceptual model and rubrics will be used to inform the development of machine learning models that can be used to capture behaviors and interactions in group settings and provide behavioral insights about the collaboration process. Our next steps will be to examine the social structures using epistemic and social network analysis and explore the relationship between talk and gesture. We will also perform machine learning validation tests to explore the consistency, accuracy, and explainability of the machine learning models. Behavior analytics is intended to inform the teacher's exploration of group dynamics- not replace it.

References

Alozie, N., Dhamija, S., McBride, E., & Tamrakar, A. (2020, June) Automated collaboration assessment using behavioral analytics. *International Conference of the Learning Sciences (ICLS),* Nashville, TN.

Barron, B. J., & Pea, R. (2013). Advancing Understanding of Collaborative Learning with Data Derived from Video Records. Randi A. Engle (University of California, Berkeley). In The international handbook of collaborative learning (pp. 215-231). Routledge.

Blikstein, P. (2013, April). Multimodal learning analytics. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 102-106).

Booth, A., Carroll, C., Ilott, I., Low, L. L., & Cooper, K. (2013). Desperately seeking dissonance: identifying the disconfirming case in qualitative evidence synthesis. *Qualitative health research*, *23*(1), 126-141.

Child, S., & Shaw, S. (2015). Collaboration in the 21st century: Implications for assessment. *Economics*, *21*, 2008.

D'Angelo, C. M., Roschelle, J., Bratt, H., Shriberg, L., Richey, C., Tsiartas, A., & Alozie, N. (2015). Using students' speech to characterize group collaboration quality. *11th International Conference on Computer Supported Collaborative Learning: Exploring the Material Conditions of Learning*, 819–820.

D'Mello, S. K., Craig, S. D., Witherspoon, A., Mcdaniel, B., & Graesser, A. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction, 18*(1–2), 45–80.

Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols. *International Journal of Computer Supported Collaborative Learning, (3),* 447–470.

Grosz, B. J., & Kraus, S. (1996). Collaborative plans for complex group action. *Artificial Intelligence*, *86*(2), 269-357.

Guo, P., Saab, N., Post, L. S., & Admiraal, W. (2020). A review of project-based learning in higher education: Student outcomes and measures. *International Journal of Educational Research*, *102*, 101586.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363-374

Martinez-Maldonado, R., Dimitriadis, Y., Martinez-Monés, A., Kay, J., & Yacef, K. (2013). Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop. *International Journal of Computer-Supported Collaborative Learning*, *8*(4), 455–485.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *The journal of the learning sciences*, *2*(3), 235-276.

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning* (pp. 69-97). Springer, Berlin, Heidelberg.

Roselli, D., Matthews, J., & Talagala, N. (2019, May). Managing bias in AI. In *Companion Proceedings of The 2019 World Wide Web Conference* (pp. 539-544).

Som, A., Kim, S., Lopez-Prado, B., Dhamija, S., Alozie, N., & Tamrakar, A. (2020). A machine learning approach to assess student group collaboration using individual level behavioral cues. *European Conference on Computer Vision*, 79–94.

Steinbock, D. (2008). Mindful of process: Scaffolds for collaboration discourse in design education.

Subburaj, S. K., Stewart, A. E., Ramesh Rao, A., & D'Mello, S. K. (2020, October). Multimodal, multiparty modeling of collaborative problem solving performance. *Proceedings of the 2020 International Conference on Multimodal Interaction*, 423–432.

Acknowledgements

Thank you to the expert reviewers (Carol Tate, Sara Rutherford-Quach, Erika Buckley, Camille Erskine and Charu Gulati) and the participants that made this validation study possible. This work is funded by National Science Foundation grant #2016849 and an internal research and development grant at SRI International.