A Multi-theoretic Analysis of Collaborative Discourse: A Step Towards AI-Facilitated Student Collaborations

Jason G. Reitman(✉), Charis Clevenger, Quinton Beck-White, Amanda Howard, Sierra Rose, Jacob Elick, Julianna Harris, Peter Foltz, and Sidney K. D'Mello

University of Colorado Boulder, Boulder, CO 80309, USA
jason.reitman@colorado.edu

Abstract. Collaboration analytics are a necessary step toward implementing intelligent systems that can provide feedback for teaching and supporting collaborative skills. However, the wide variety of theoretical perspectives on collaboration emphasize assessment of different behaviors toward different goals. Our work demonstrates rigorous measurement of collaboration in small group discourse that combines coding schemes from three different theoretical backgrounds: Collaborative Problem Solving, Academically Productive Talk, and Team Cognition. Each scheme measured occurrence of unique collaborative behaviors. Correlations between schemes were low to moderate, indicating both some convergence and unique information surfaced by each approach. Factor analysis drives discussion of the dimensions of collaboration informed by all three. The two factors that explain the most variance point to how participants stay on task and ask for relevant information to find common ground. These results demonstrate that combining analytical tools from different perspectives offers researchers and intelligent systems a more complete understanding of the collaborative skills assessable in verbal communication.

Keywords: Collaboration Analytics · Collaborative Problem Solving · Team Communication · Academically Productive Talk · Talk Moves

1 Introduction

Collaboration as a skill is increasingly emphasized at work and in school as both technology- and pandemic-related changes highlight the importance of developing collaboration skills. Researchers have studied collaboration – and how to measure it – for decades [1], but efforts to develop intelligent systems to support collaboration in human teams have only emerged more recently [2]. Crucially, much of that work has focused on the collaborative practices and AI supports

N. Wang et al. (Eds.): AIED 2023, LNAI 13916, pp. 577–589, 2023.
https://doi.org/10.1007/978-3-031-36272-9_47

that predict content learning and task performance outcomes. There is comparatively little work on intelligent systems that emphasize the development of collaboration skills as valued outcomes in their own right. Accordingly, our long-term focus is on developing AIED systems that monitor and facilitate effective small group classroom collaboration where outcomes of interest are both collaboration skills and learning measures.

Measurement is essential for any intelligent system that aims to provide feedback support for collaborative processes. However, a major challenge of collaboration assessment is the complex nature of the construct. Researchers working from different perspectives have developed a variety of approaches to assess collaboration based on different assumptions, foci, and analytic approaches [3]. Whereas this multidisciplinary approach affords an expansive view on collaboration it risks the jingle-jangle problem, where the same term can refer to different items and different terms can refer to the same item. Conversely, the use of a single framework to investigate collaboration can result in the opposite problem of construct deficiency, where the approach is too narrow to encompass all relevant phenomena. Researchers, therefore resort to defining their own frameworks to analyze collaboration as it unfolds in a particular context, which limits generalizability [4] – a major goal for developing intelligent collaborative measures and supports.

Taking a somewhat different approach, the present study focuses on analyzing classroom collaboration data using a multi-theoretical approach spanning three different perspectives towards measuring and modeling collaborative processes in small groups: Collaborative Problem Solving (CPS) [5], Academically Productive Talk (APT) [6], and communication in Team Cognition (TC) [7]. Our goal is to examine incidence of collaborative behaviors across these different frameworks, examine shared variance across frameworks, and identify underlying latent dimensions of collaboration common to these three schemes. Taken together, our results suggest certain communication practices that can be measured and encouraged to assess and support collaborative learning. They provide a stepping stone toward implementing intelligent systems that can give monitor and facilitate collaborations in classrooms.

1.1 Background and Related Work

We focus on collaboration frameworks which emphasize communication, a central component of how groups learn and work together. Whereas communication can also occur non-verbally, we limit our scope to verbal communications.

Collaborative Problem Solving. CPS describes a set of skills and tasks related to problem solving [8]. It addresses with both the cognitive abilities required to problem-solve as a team and the social dynamics between group members as they coordinate their knowledge and actions [8,9]. Over the last decade, researchers have proposed several frameworks to measure CPS [8]. For example, Sun et al.'s CPS framework identified three core facets of CPS measurable in verbal discourse: "constructing shared knowledge, negotiation/coordination, and maintaining team function" [10]. This framework blends both the cognitive and social dimensions of collaboration whereas other frameworks, such as Andrews-Todd's CPS ontology delineates among these dimensions [11].

Accountable Talk Theory & Academically Productive Talk. With a focus on how essential certain kinds of talk are to many students' learning, education research has focused on how teachers and students use and react to certain kinds of verbal communication. Talk moves, or accountable talk theory, are families of utterances that are derived from ethnographic studies of teachers who were successful in promoting rich, student-driven, academically productive talk in their classrooms. Those talk moves serve three broad instructional goals of ensuring (1) accountability to the learning community (LC), (2) accountability to content knowledge (CK), and (3) accountability to rigorous thinking (RT) [12]. Research has consistently documented how implementing such talk moves promotes student learning [13] and contributes to educational equity [14]. By using these talk moves, students not only contribute their ideas but also attend to and build on their classmates' ideas, helping to ensure they are collaboratively engaging in challenging academic work.

Team Cognition & Communication. Team cognition describes how teams gather, retain, and use information through interactions and communication to work toward their goals [15]. Marlow, Lacerenza, and Salas's [7] framework describes the role of communication in team cognition. They note that how much a team communicates can have little to do with how well they function together, but communication quality impacts how teams perform cognitive work to a greater degree [7]. Communication quality is broad, though, and qualities of communication with relationships to team performance range from clarity of speech to timeliness of information [16]. The specific qualities that are most relevant to the present work are informed by peer mentoring styles identified by Leidenfrost et al. [17], who identified key aspects of communication styles in student mentor-mentee relationships that related to mentee performance: motivating, informative, and timely mentor communication. These are also aspects of communication quality emphasized in Marlow, Lacerenza, and Salas's [7] framework of team communication.

Related work. Whereas most studies analyze collaboration from a unitary perspective, some researchers have utilized multiple perspectives. Jordan and Henderson's [18] foundational work emphasized that the complex nature of interaction data benefits from analysis by theoretically diverse working groups, as it "reveals and challenges idiosyncratic biases" (p. 43). Suthers, Lund, Rosé, and Teplovs' [19] productive multivocality is a guiding example of how to effectively integrate theory and practice from multiple disciplines. Their work presenting shared data sets to groups of analysts from diverse traditions yielded insights into how challenging "individual operationalizations of complex constructs" (p. 588) can reveal the limited scope of how different literatures deal with broad constructs. They argue that analyzing the same data from different perspectives can offer deeper understanding. Analysts from diverse theoretical backgrounds participated in the Productive Multivocality Project over the course of five years, analyzing shared data corpora using their own methods and juxtaposing their analyses with those of the other participants, producing insights into the difficulties and advantages of combining their disparate perspectives. One of these data sets consisted of video data student groups working on chemistry problems in an undergraduate chemistry class. The methods brought to these data included ethnographic analysis , coding and counting analysis , and social network analysis . The researchers found that although they had all studied leadership in the student groups, their interpretations of the construct differed in ways that made clear that each perspective on its own lacked the depth of understanding of the complexities of leadership that they were able to achieve with a comparative analysis of all three approaches [20]. For example, the qualitative and social network analyses both identified a difference between groups in terms of whether they primarily moved procedurally through the steps of a problem or spent more time discussing concepts relevant to the problem at hand. The analysts inferred based on this difference that the procedural approach lost out on meaningful group interactions that could support collaborative learning. The code and count analysis yielded the same difference between groups, but focusing on transactivity yielded findings that throughout the step-by-step procedure, group members were reasoning together and solving steps collaboratively [20]. The ability of disparate analytical approaches to challenge each other in this way argues for inclusion of diverse perspectives when measuring complex constructs. For our work, comparing results from multiple analyses is essential for rigorous measurement so that each perspective can inform our understanding of collaboration.

Current study. We applied three different measures for analyzing collaboration data to investigate how they relate to each other and what can be learned from a multi-theoretical approach. We have three guiding research questions: (RQ1) how often do the behaviors of interest for each measure occur, (RQ2) to what extent do these measures align, and (RQ3) what are the underlying factors of collaboration common to these measures?

Our work contributes to rigorous measurement of collaboration as a step toward developing AIED systems that can assess and support collaborative learning in classrooms. We provide insights into collaboration measurement from diverse theoretical backgrounds and detail how the resultant frameworks can work together to provide a more complete understanding of collaboration in classrooms. We focus on the process of collaboration, as opposed to content outcomes, because the development of collaboration skills is the outcome of interest. This research, which aims to identify underlying dimensions of collaborative discourse that are common across diverse theoretical frameworks, is an important step towards the development of generalizable computational models of collaborative processes for feedback and intervention.

2 Methods

Our data are 31 transcribed videos from dyads and triads of middle school students working on collaborative tasks as part of their normal classroom instruc-

Table 1. Excerpt of student dialog during sensor programming

Speaker	Utterance
S1	So, so far we have the "forever" button and it's if the sound intensity is over 1000 then it will show the icon and play this sound for four beats [TC: Inf, PresentFuture, Plural; APT: Relating, Providing; CPS: MonitoringExecution]
S2	So like four times that length. [TC: Inf, PresentFuture; APT: Relating, Making; CPS: MonitoringExecution]
S1	And if it's not over that, then it'll plot a bar graph. [TC: Inf, PresentFuture; APT: Relating, Providing; CPS: SharingUnderstanding]
S2	Yeah. So the "forever" makes it sense the sound intensity, and if it goes over a thousand then it's gonna show this icon and then play this [TC: Inf, PresentFuture; APT: Providing; CPS: Responding, MonitoringExecution]

tion. Video and audio were collected with iPads and Yeti Blue microphones set up at tables with participating groups during normal class time. Researchers transcribed and analyzed five minutes of audio from each group. These five-minute samples were randomly selected from longer videos with the constraints that the audio and video were of good quality and included the midpoint of the period of small-group work.

Participants. Participants were 34 middle school students from four classes taught by one teacher in a school district with 48.8% female students and a ethnic demography of 62.3% White, 30.0% Hispanic, 3.3% Asian, 3.0% two or more races, 0.9% Black, 0.3% American Indian or Alaska Native, and 0.1% Hawaiian/Pacific Islander students in the 2020–2021 school year. 22.0% of students in this district are receiving free or reduced price lunch. These participants each gave assent and provided consent forms signed by their parents, and the data collection procedure was approved by the appropriate Institutional Review Board.

Collaborative Task. The tasks are part of a Sensor Immersion (SI) curriculum that focuses on developing scientific models and computational thinking. Through wiring and programming sensors and data displays using Micro:Bit hardware and MakeCode block programming software, students practice computational thinking to design and describe their systems. Tasks build from using specific sensors to answer specific questions to students postulating and answering their own questions about their environment by wiring and programming a multi-sensor system.

Table 1 contains a sample of collaborative conversation between two students during a task in the SI curriculum annotated with the three coding schemes (see below). The students in this group are working to program the data display that they wired to a sound sensor. Their goal is to have the setup perform certain actions when the sound sensor registers specific volume levels.

Measures. We analyzed this data set with coding schemes based on each CPS, APT, and TC. Separate teams of annotators worked simultaneously to code the data with their respective scheme. The CPS team was comprised of four coders including one master coder who had previously worked with the CPS scheme. This team worked to reach consensus with the master coder, who checked all annotated transcripts. The APT and TC teams were each comprised of two coders who worked to consensus.

Collaborative Problem Solving - CPS. The CPS framework and coding scheme, as detailed by Sun et al. [10] is built around three facets of collaborative problem solving: constructing shared knowledge, negotiation/coordination, and maintaining team function. Each of these facets is comprised of sub-facets (see Table 2), which are in turn comprised of verbal and non-verbal indicators. These indicators were originally defined in terms specific to a different collaborative task. The CPS coding team, therefore, adapted them to SI data before analyzing this corpus. A team of three coders and one master coder annotated SI transcripts, recorded discrepancies they found in applying the coding scheme to the new data, supplemented the coding scheme with notes and examples from coding the SI data, and discussed utterances of high disagreement until they reached consensus. To ensure the reliability of the adapted coding scheme's application to the SI data, each annotated transcript was reviewed by the master coder until the original coder and the master coder reached consensus. The data can be analyzed at all three levels(facet, sub-facet, and indicator), and we chose to focus on the intermediate sub-facet level here. For this, the indicators were aggregated to the sub-facet level so that if an utterance was coded with any of the indicators that comprise a sub-facet, then it was coded with that sub-facet.

Academically Productive Talk - APT. The APT coding scheme draws on a set of student talk moves detailed by Suresh et al. [21], originally defined to analyze student speech during math lessons. Four talk moves were identified based on suggestions from experts and required relatively minor adaption for the SI curriculum (see Table 2).

Team Communication - TC. To measure communication quality and style as defined by Marlow et al.'s [7] team communication framework and Leidenfrost et al.'s [17] styles of peer mentoring, the TC coding scheme includes indicators of motivational, informative, timely, and collective speech (see Table 2). This coding scheme was originally developed to measure communication styles in competitive video game teams with the explicit purpose of measuring communication quality with a domain-agnostic method [22]. Because it is quite general, it did not require much adaptation to SI.

3 Results and Discussion

RQ1: Occurrence of Collaborative Behaviors Within Schemes. There was considerable variability in the different codes as noted in Table 2, which provides insights into their collaborative processes. For example, APT coding

Table 2. CPS, APT, and TC coding scheme definitions, examples, and mean frequencies

CPS Facet	Sub-facet	Occurrence Mean (SD)	Definition
Constructing shared knowledge	Sharing understanding of problems/solutions	0.26 (0.13)	Contributing expertise and ideas regarding particular problems and toward specific solutions
	Establishing common ground	0.06 (0.05)	Acknowledging ideas, confirming understanding, and clarifying misunderstanding
Negotiation/ coordination	Responding to others' questions/ideas	0.14 (0.06)	Providing feedback, offering reasons for/against claims, and implementing agreed on solutions
	Monitoring execution	0.09 (0.08)	Talking about strategy, progress, and results
Maintaining team function	Taking initiative to advance the collaboration process	0.06 (0.05)	Asking questions, acknowledging others' contributions, and helping to maintain team organization
	Fulfilling individual roles on the team	0.06 (0.05)	Performing own roles and responsibilities
	Participating in off-topic conversation	0.11 (0.16)	Initiating or joining off-topic conversation
APT Facet	Code	Occurrence Mean (SD)	Definition
LC	Relating to another student	0.52 (0.21)	Using, commenting on, or asking questions about a classmate's ideas
LC	Asking for more information	0.05 (0.04)	Requests more information or says they are confused or need help
CK	Making a claim	0.38 (0.15)	Makes a math claim, factual statement, or lists a step in their answer
RT	Providing evidence/ reasoning	0.02 (0.03)	Explains their thinking, provides evidence, or talks about their reasoning
TC Facet	Code	Occurrence Mean (SD)	Definition
Mot	Positive	0.04 (0.05)	Positive emotional valence
	Negative	0.07 (0.08)	Negative emotional valence
Inf	Informative	0.51 (0.20)	Contains or requests useful info
	Uninformative		Lacks substance
Tim	Present/Future	0.85 (0.11)	Present or future tense or referring to present or future
	Past		Past tense or referring to past
Col	Plural	0.11 (0.07)	Collective references to group and group members
	Singular	0.25 (0.10)	Singular references to group members
	No Reference		No reference to group members

indicated that students make a lot of claims (m = 0.38), but do not appear to explain their reasoning behind those claims (m = 0.02). Their CPS discourse which mainly consists of shared knowledge construction (m = 0.33), followed by negotiation/coordination (m = 0.23) and lastly maintaining team function (m = 0.12 – not counting off-topic conversations), mirrors distributions on other tasks [10]. Some indicators require more context to understand what their occurrence rates suggest about collaboration. Are students referring to individual team members (TC: Singular, m = 0.25) more than twice as much as they refer to their team as a collective (TC: Plural, m = 0.11), because they are not working as a team or because they are working together to assign aspects of the task to specific teammates? Are they encouraging off-topic conversations (m = 0.11, SD = 0.16) as much as almost every other CPS behavior, because they are not working together on the task or because they want to relate to their teammates to promote team cohesion?

RQ2: Relationships Between Schemes. To understand relationships across schemes, we aggregated the data by observations and computed Pearson correlation coefficients between CPS sub-facets and APT indicators, CPS sub-facets and TC indicators, and APT indicators and TC indicators. To aggregate from the coded utterances to the observation level, we computed the frequency of each code applied in relation to the total number of utterances in each observation. We found low to moderate mean correlations across all three schemes. CPS and APT yielded a mean r(29) of 0.29 (SD = 0.22, median = 0.30, range = <0.01–0.77). CPS and TC yielded a mean r(29) of 0.20 (SD = 0.18, median=0.18, range = <0.01-0.80). APT and TC yielded a mean r(29) of 0.28 (SD = 0.26, median = 0.20, range = <0.01–0.83). Overall, this suggests a modicum of overlap across the three different coding approaches, confirmation their different foci according to their respective theoretical foundations, while also demonstrating shared variance across schemes.

RQ3: Factors of Collaboration. We conducted a principle components factor analysis with a varimax rotation to identify latent factors comprised of the three different coding schemes. Measures that implied these data were factorable included at least moderate correlation between codes, an overall Kaiser-Meyer-Olkin measure of sampling adequacy of 0.59 (which approaches the recommended threshold of 0.60), and a significant Bartlett's test of sphericity ($\chi^2(78) = 475$, $p<0.001$). We selected six factors with eigenvalues > 1 that explained a cumulative 82% of the variance in our data (see Table 3).

We interpret the first factor as a measure of task-related communication. It includes codes from each scheme that indicate on-topic, informative discourse. The excerpt in Table 1 exemplifies this dimension. The students stay on task, share information, relate to each other, and make claims about their work with almost every utterance. This type of discourse is important for collaboration.

The second factor pertains to finding common ground. APT's "Asking for more information" and CPS's "Establishing common ground" load onto this factor strongly. This dimension is evident when group members prompt each other for relevant information that the group can orient their thinking around. For example, when a triad was struggling to get their data display to respond to sensor input, one student started asking the rest of the group for more information on how to proceed: "Okay, show strings? What should we tell it to do?" Thus, When a group is having difficulty, they center the collaborative effort around prompts for specific information.

The third factor is most strongly informed by CPS's "Monitoring execution" and is most visible in utterances where a group member is assessing the current state of the group's work by making a claim about how their sensor setup is functioning. For example, consider the following excerpt when a team was trying to make their data display react at a certain noise level and were struggling to find the right threshold: "So even my click was louder than this." The implied assessment the student is making is that they currently have the threshold set too low, because even their clicking the computer mouse was louder than their current threshold. CPS's "Sharing understanding of problems/solutions" loads negatively onto this dimension, meaning that the problem has been established, and now the focus is on finding a solution and debugging faulty solutions (i.e., monitoring execution).

Groups display the fourth factor when they focus on what is happening in the present or coming in the future, without communicating negative emotions about it. If something goes wrong in a collaborative task, it can be easy for a group to harp on that past event. Utterances that plan, strategize, or consider possible group achievements can be a vital part of what keeps a collaboration productive, as long as they are not presented with demotivating emotions. Many examples of this dimension occur when participants offer ideas to guide their work toward exciting designs, like a student suggesting fun ways to program the data display, like "We could make it look like it's exploding," and "Let's graph ASMR." These utterances spur on collaborative work toward a planned goal.

The fifth factor is where CPS's "Taking initiative to advance the collaboration process" and TC's "Positive" converge. This dimension assesses motivational discourse and expressly positive emotional valence. These can be simple feedback remarks, like "That is very cool," or personal compliments, like "We're pure geniuses." Discourse in this dimension encourages collaborators to work together and rewards them when they do.

The final factor is informed by only one code. CPS's "Fulfilling individual roles on the team" measures talk that indicates individual group members are owning their responsibilities and doing their part for the collaboration as a whole. This includes offering instructional support when the group is stuck (e.g. "Alright, you gotta do 'on button B press', so add another one of these.") and owning up to mistakes by apologizing for them (e.g. "No, no no no. Sorry. My bad. Scroll over to where you can see the whole thing.").

4 General Discussion

Measuring a complex construct like collaboration benefits from the nuance gained by combining different analytical perspectives [20]. Each coding scheme

Table 3. Factors and loadings yielded by factor analysis

Factor	Qualitative Description	Proportion of Variance	Codes	Loadings
Factor 1	Task content	0.26	TC:Informative	0.84
			APT:RelatingOther	0.81
			APT:MakingClaim	0.82
			CPS:SharingUnderstanding	0.74
			CPS:OffTopic	-0.82
Factor 2	Using information to find common ground	0.15	APT:AskingInfo	0.92
			CPS:EstablishesCG	0.88
Factor 3	Monitoring teamwork	0.11	CPS:MonitorsExecution	0.91
			APT:MakingClaim	0.33
			CPS:SharingUnderstanding	-0.47
			TC:Positive	-0.35
Factor 4	Strategizing and planning	0.11	TC:PresentFuture	0.83
			TC:Negative	-0.80
Factor 5	Motivating	0.10	TC:Positive	0.72
			CPS:Initiative	0.80
Factor 6	Fulfilling role	0.09	CPS:FulfillRole	0.92

used in these analyses focuses on different aspects of collaboration. They each have their own strengths and weaknesses for use as a foundation on which to develop computational models for AIED interventions to measure and support effective collaboration.

Regarding our first research question about occurrence of behaviors of interest, student groups in this sample show similar distribution of the CPS facets in the SI tasks as they do in previously researched tasks. They participate most in shared knowledge construction and least in maintaining team function. APT analysis revealed that these groups almost never provide their reasoning and infrequently ask for more information, despite frequently relating to each other and making claims about their group work. TC points to mostly informative talk, without strong emotional valence, that does not linger on events that have already passed. A system equipped with these assessments could, for example, elicit explicit reasoning and encourage maintenance of team function for the average group in these data.

For our second research question asking how related these measures are, correlations across schemes indicate that they are identifying behaviors that do relate to each other, but only weakly to moderately. We infer that these measures are in fact approaching the same construct from different perspectives and therefore contain some overlap with each other while also focusing on unique aspects of collaboration. Simply put, they are complimentary and not redundant.

The underlying dimensions of collaboration informed by these schemes surfaced by factor analysis answer our third research question. The measures, when taken together, do in fact point to underlying factors of collaboration that none of the schemes can fully assess on its own. Four of the six factors include codes from multiple schemes (factors 1, 2, 3, & 5), indicating dimensions of collaboration that no one scheme accounts for on its own. Two of these include input from all three schemes (factors 1 & 3). In other words, these factor analysis results support the inference that these schemes are complementary, yet distinct, approaches to measuring collaborative processes in student groups.

The overall implication of these findings is that it might be beneficial to incorporate discourse from multiple coding schemes and theoretical frameworks to obtain a more nuanced, contextual analysis of complex classroom collaboration. An initial hypothesis is that computational models might focus on predicting the factors which are largely composed of multiple schemes rather than scheme-specific indicators. An open question is also whether this will result in more accurate and generalizable models because they are focused on more abstract dimensions of collaboration.

Limitations and Future Work. While the data were collected in a real-world classroom setting, the sample was small and occurred in a single teacher's classes in a single school district. We are continuing to collect data in multiple classrooms in multiple schools for subsequent analysis with larger, more diverse samples. The small sample size precluded an analysis of the effects of varying curricula and group composition on the unfolding collaboration dynamics, so replication and generalizability across curricula, domains, and group compositions is an important item for future work.

Conclusion. In order to develop systems that can provide feedback and support collaboration, we require rigorous and reliable tools for measuring collaboration in the first place. This work is a necessary step toward implementing an AIED system that puts these collaboration analytics to work in live measurement and feedback for small groups in classrooms. Effective interventions will also need to consider how, when, and what metrics are presented to groups as they learn collaboratively. Work developing, implementing, and studying such a system is underway with the intent to inform design of interventions in collaborative learning situations and collaborative systems more broadly.

Acknowledgements. This research was supported by the NSF National AI Institute for Student-AI Teaming (iSAT) under grant DRL 2019805. The opinions expressed are those of the authors and do not represent views of the NSF.

References

1. Gedney, J.: Development of an instrument to measure collaboration and satisfaction about care decisions. J. Adv. Nurs. **20**(1), 176–182 (1994)
2. Stewart, A., D'Mello, S.K.: Connecting the dots towards collaborative AIED: linking group makeup to process to learning. In: Penstein Rosé, C., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10947, pp. 545–556. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93843-1_40

3. Forsell, J., Forslund Frykedal, K., Hammar Chiriac, E.: Group work assessment: assessing social skills at group level. Small Group Res. **51**(2), 87–124 (2020)
4. Walters, S.J., Stern, C., Robertson-Malt, S.: The measurement of collaboration within healthcare settings: a systematic review of measurement properties of instruments. JBI Evid. Synth. **14**(4), 128–197 (2016)
5. Fiore, S.M., Graesser, A., Greiff, S.: Collaborative problem-solving education for the twenty-first-century workforce. Nat. Hum. Behav. **2**(6), 367–369 (2018)
6. O'Connor, C., Michaels, S., Chapin, S., Harbaugh, A.G.: The silent and the vocal: participation and learning in whole-class discussion. Learn. Instr. **48**, 5–13 (2017)
7. Marlow, S.L., Lacerenza, C.N., Salas, E.: Communication in virtual teams: a conceptual framework and research agenda. Hum. Res. Manag. Rev. **27**(4), 575–589 (2017)
8. Graesser, A.C., Fiore, S.M., Greiff, S., Andrews-Todd, J., Foltz, P.W., Hesse, F.W.: Advancing the science of collaborative problem solving. Psychol. Sci. Public Interest **19**(2), 59–92 (2018)
9. Fiore, S.M., Rosen, M.A., Smith-Jentsch, K.A., Salas, E., Letsky, M., Warner, N.: Toward an understanding of macrocognition in teams: predicting processes in complex collaborative contexts. Hum. Factors **52**(2), 203–224 (2010)
10. Sun, C., Shute, V.J., Stewart, A., Yonehiro, J., Duran, N., D'Mello, S.: Towards a generalized competency model of collaborative problem solving. Comput. Educ. **143**, 103672 (2020)
11. Andrews-Todd, J., Kerr, D.: Application of ontologies for assessing collaborative problem solving skills. Int. J. Test. **19**(2), 172–187 (2019)
12. Resnick, L.B., Asterhan, C.S.C., Clarke, S.N.: Accountable talk: instructional dialogue that builds the mind. In: Marope, M., Vosniadou, S. (eds.) Educational Practices Series 29, pp. 14–32. The International Academy of Education (IAE) and the International Bureau of Education (IBE) of the United Nations Educational, Scientific and Cultural Organization (UNESCO), Geneva (2018)
13. Webb, N.M., Franke, M.L., Ing, M., Turrou, A.C., Johnson, N.C., Zimmerman, J.: Teacher practices that promote productive dialogue and learning in mathematics classrooms. Int. J. Educ. Res. **97**, 176–186 (2019)
14. O'Connor, C., Michaels, S.: Supporting teachers in taking up productive talk moves: the long road to professional learning at scale. Int. J. Educ. Res. **97**, 166–175 (2019)
15. Cooke, N.J., Gorman, J.C., Myers, C.W., Duran, J.L.: Interactive team cognition. Cogn. Sci. **37**(2), 255–285 (2013)
16. González-Romá, V., Hernández, A.: Climate uniformity: its influence on team communication quality, task conflict, and team performance. J. Appl. Psychol. **99**(6), 1042 (2014)
17. Leidenfrost, B., Strassnig, B., Schabmann, A., Spiel, C., Carbon, C.-C.: Peer mentoring styles and their contribution to academic success among mentees: a person-oriented study in higher education. Mentoring Tutoring: Partnership Learn. **19**(3), 347–364 (2011)
18. Jordan, B., Henderson, A.: Interaction analysis: foundations and practice. J. Learn. Sci. **4**(1), 39–103 (1995)
19. Suthers, D.D., Lund, K., Rosé, C.P., Teplovs, C.: Achieving productive multivocality in the analysis of group interactions. In: Suthers, D.D., Lund, K., Rosé, C.P., Teplovs, C., Law, N. (eds.) Productive Multivocality in the Analysis of Group Interactions. CCLS, vol. 15, pp. 577–612. Springer, Boston, MA (2013). https://doi.org/10.1007/978-1-4614-8960-3_31

20. Rosé, C.P.: A Multivocal Analysis of the Emergence of Leadership in Chemistry Study Groups. In: Suthers, D.D., Lund, K., Rosé, C.P., Teplovs, C., Law, N. (eds.) Productive Multivocality in the Analysis of Group Interactions, pp. 243–254. Springer, New York. (2013). https://doi.org/10.1007/978-1-4614-8960-3_13
21. Suresh, A., et al.: Using transformers to provide teachers with personalized feedback on their classroom discourse: The TalkMoves application. Paper presented to the 2021 AAAI Conference on Artificial Intelligence in K-12 Education (2021)
22. Reitman, J. G.: Generalizable Communication Styles in Novice and Expert Team Performance. Doctoral Dissertation. UC Irvine, (2022)