The Relationship between Collaborative Problem-Solving Skills and Group-to-Individual Learning Transfer in a Game-based Learning Environment

Chen Sun
University of Manchester
Manchester, United Kingdom
chen.sun@manchester.ac.uk

Valerie Shute
Florida State University
Tallahassee, USA
vshute@admin.fsu.edu

Angela E.B. Stewart
University of Pittsburgh
Pittsburgh, USA
ANGELAS@pitt.edu

Sidney K. D'Mello
University of Colorado Boulder
Boulder, USA
Sidney.Dmello@Colorado.EDU

Abstract

Collaborative problem solving (CPS) is viewed as an essential 21st century skill for the modern workforce. Accordingly, researchers have been investigating how to conceptualize, assess, and develop pedagogical approaches to improve CPS. These efforts require theoretically-grounded and empirically-validated frameworks of CPS which have been emerging over the past decade with various levels of validity data. The present paper focuses on validating the generalized competency model (GCM) of CPS with respect to predicting individual learning outcomes following CPS among triads. The GCM consists of three main facets–constructing shared knowledge, negotiation/coordination, and maintaining team function–mapped to behavioral indicators (i.e., observable evidence). It hypothesizes that scores on all three facets should positively predict CPS outcomes, including group-to-individual learning transfer. We tested this hypothesis in a study where 249 students who comprised 83 triads engaged in collaborative gameplay with the Physics Playground game environment remotely via videoconferencing. We found that the only CPS facet predicting individual physics learning was maintaining team function, after accounting for pretest scores, students' perceptions of team collaboration, and their perceived physics self-efficacy. This facet was also the only significant predictor of individual learning regardless of how facet scores were computed (i.e., reverse coding of negative indicators, separating the sums of positive and negative indicators, and no reverse coding of negative indicators). Implications for the GCM and other CPS frameworks are discussed.

CCS Concepts

• **Applied computing**; • **Education**; • **Collaborative learning**;

LAK 2025, March 03–07, 2025, Dublin, Ireland

ACM ISBN 979-8-4007-0701-8/25/03
https://doi.org/10.1145/3706468.3706560

Keywords

collaborative problem solving, framework validation, physics learning outcome, triads, game-based learning

ACM Reference Format:
Chen Sun, Valerie Shute, Angela E.B. Stewart, and Sidney K. D'Mello. 2025. The Relationship between Collaborative Problem-Solving Skills and Group-to-Individual Learning Transfer in a Game-based Learning Environment. In *LAK25: The 15th International Learning Analytics and Knowledge Conference (LAK 2025), March 03–07, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3706468.3706560

1 Introduction

Collaborative problem solving (CPS) refers to the collective efforts of two or more individuals to achieve a solution to a problem [26]. CPS encompasses a collection of social-cognitive skills such as proposing ideas, sharing knowledge, and building common ground [2, 20]. It is hypothesized to be a 21st century skill for the modern workforce where people are increasingly required to come together to solve complex, nonroutine problems such as fighting forest fires, designing a rocket for a mission to Mars, or forecasting the weather. Even more mundane tasks like figuring out where to eat dinner with fussy eaters, coordinating travel schedules, and organizing a workshop for an academic conference require CPS. The inclusion of AI in the workforce of the future is expected to result in a large degree of automation of routine cognitive tasks, which has the potential to free-up human resources to tackle more complex problems if they can work together cohesively and productively.

Unfortunately, humans are not very effective at CPS skills, presumably because they do not receive formal training on developing these skills. In 2015, the Organization for Economic Co-operation and Development (OECD) conducted an assessment of CPS across 15-year-old students from 72 global economies. The results revealed that a mere 8% of the students were characterized at the highest proficiency level (level 4), with 28% being scored as low performers (level 1 and below) [27]. This indicates a general insufficiency of CPS skills among young people to perform effectively in collaborative environments. It highlights the need to support young people to develop CPS skills in order to excel in school and future workplace.

Despite the growing importance of CPS, there remains a need to understand the complexity in assessing and training CPS skills [16]. This requires rigorous quantification of CPS competencies (or skills) via validated frameworks, which serve as the basis for developing effective assessments and interventions. In educational settings, beyond predicting performance on CPS tasks [3, 38], it is imperative that engaging in CPS transfers to individual learning gains, called group-to-individual transfer of learning [28]. Whereas past work has shown evidence for this type of learning transfer (see meta-analysis [29]), research has yet to do so in the context of CPS. Accordingly, the present paper focuses on investigating the predictive validity of a CPS framework on *individual* student learning following CPS in a digital game-based learning environment.

1.1 CPS frameworks

There are a number of different frameworks in the literature designed to assess collaborative problem solving (CPS) skills [20, 26]. For instance, the OECD CPS framework comprises a 3x4 matrix illustrating the interaction between individual problem-solving behaviors (e.g., planning and executing) and collaborative processes (e.g., maintaining team organization), resulting in 12 CPS skills (e.g., establishing a common understanding of the problem, and monitoring and evaluating the results of actions). The purpose of this framework was to develop a large-scale CPS assessment for the 2015 PISA assessment that was used to evaluate CPS in 15-year-old students across the OECD economies as discussed above.

The ATC21S [17, 20] is another popular framework which dissected CPS into cognitive processes and social skills. The social skills consist of three main skills including: participation, perspective taking and social regulation. There are also sub-skills under the main skills (e.g., adapt responses to individuals). The cognitive skills are comprised of two main skills of task regulation and learning and knowledge building. Associated sub-skills include achieving a shared solution via negotiation, setting up goals, and the like. The framework aims to inform learning and teaching of CPS, as well as standardized assessments.

CPS was similarly characterized in terms of cognitive and social skills in an ontology by Andrews-Todd et al. [2] which targeted computerized simulation-based tasks. This ontology comprises four main social skills and five main cognitive skills, such as maintaining communication, negotiation, planning, and executing. Each main skill contains specific measurable behaviors related to communication and log file actions that occur within the context of CPS tasks. Andrews-Todd et al. [3] investigated domain-generalizability of their framework by testing it with the same students across two domains, finding moderate correlations across domains. To this point, the generalized competency model (GCM) of CPS [38] was specifically designed to be applicable across CPS domains because its three main facets (constructing shared knowledge, negotiation/coordination, and maintaining team function) and behavioral indicators therein were derived in a cross-domain fashion. This framework has since been applied to one additional domain [21].

It is notable that some frameworks clearly differentiate cognitive and social skills as two main constructs underlying CPS skills whereas others consider joint socio-cognitive skills. The number and type of sub-skills also vary across frameworks but some overlap such as negotiation and monitoring. Despite similarities and differences across frameworks, a fundamental issue relates to establishing their validity relative to assessing CPS skills [6]. Further, in addition to diagnosing CPS skills and predicting success on CPS tasks, a valid assessment framework should also potentially foster understanding of learning processes related to CPS activities by linking specific CPS skills to individual (i.e., student) learning outcomes, which is the focus of this paper.

1.2 Factors influencing CPS assessment

Beyond the choice of framework(s), there are several other factors that influence CPS assessment. One key factor is whether the collaboration occurs among humans (Human-to-Human; HH) or among humans and agents (Human-to-Agent; HA) [16, 19]. In HA communication, students communicate using text-based chats (having some freedom in communication) or by selecting pre-defined messages [12]. Some research has shown that HA collaboration is not as effective as HH collaboration due to the lack of coordination, communication, and shared cognition [25]. In HH communication, students can chat verbally or via a chat box. When students chat verbally, another factor relates to whether students collaborate in-person or virtually through videoconferencing [37].

Students can additionally be assigned to work in differently-sized groups (e.g., dyads or triads or even larger) although dyads are more common in the CPS literature. Triads, however, can generate different team dynamics than dyads as there is only one way for two people to interact with each other in a dyad, but four ways for a triad to interconnect (i.e., three dyads and the triad) [1]. Another phenomenon that may arise with a larger group size is the appearance of social loafers who rely on others' contributions [15]. Zhan et al. [40] compared the effects of dyads with triads in terms of their CPS learning in a high-school introductory AI course. The findings suggested that although group size did not influence learning outcomes, dyads reported significantly higher cognitive load, whereas triads had lower quality collaborative behaviors, presumably due to a social loafing phenomenon [15]. However, in general, there is little data of triadic CPS in the current literature since most studies focus on dyads. Accordingly, the present study focused on examining learning in the context of triadic CPS with open-ended human-human communications. We conducted our research in the context of a game-based learning environment because it provided the opportunity for engaging CPS skills as a team along with measurement of student learning gains.

1.3 The relationship between CPS and learning outcomes in game-based learning

Research on the connection between CPS and pertinent outcomes in game-based learning is fairly sparse. Moreover, CPS outcomes can be categorized in various ways, such as subjective perceptions [10, 34], objective task performance [2, 3, 38], and pre-post learning on individual learning transfer [7, 28, 32]. Some studies have focused specifically on CPS behaviors that occur during game-based learning (e.g., [5]), but have not further investigated the link between those specific behaviors and learning outcomes. For example, studies have attempted to detect disengagement behaviors (e.g., talking about irrelevant topics), by leveraging multiple data sources (such as group chat messages and video recordings) and machine learning techniques [5, 14], based on the premise that disengagement can negatively impact learning outcomes [4]. In a recent study, Gupta et al., [18] used machine learning techniques to predict learning outcomes from in-game actions and group chats, but more research is needed to reveal what specific behaviors contribute to learning.

Research has examined how CPS behaviors and skills derived from validated frameworks influences in-task performance. For instance, Sun et al. [38] examined CPS behaviors that predicted success in a game-based learning environment, finding that clarifying understanding among team members, actively proposing potential ideas for a solution, responding to ideas/suggestions from teammates, discussing results of problem-solving attempts, and complimenting and encouraging others were all positively predictive of task performance. This study also investigated temporal patterns in CPS behaviors that distinguished successful from unsuccessful task performance, which was further extended by Zhou et al. [44]. Andrews-Todd et al. [3] found that social regulation, which encompasses negotiating among team members and monitoring the results and progress, to be successful of CPS performance. Surprisingly, the two other key skills of communicative participation and task regulation were not predictive of task performance.

Despite the initial evidence linking CPS behaviors/skills and task performance, it is difficult to generalize findings due to the variation of assessment frameworks and CPS tasks. For example, whereas social regulation predicted task performance on one task in Andrews-Todd et al. [3], it did not on a second task in the same study. It is also unclear whether the improved task performance can transfer to individual domain learning, where the research is much sparser. Sun et al. [37] found that the socio-cognitive skill of shared knowledge construction – which entailed suggesting ideas/solutions and building on the ideas of others – predicted posttest scores in the context of CPS in a Minecraft environment. This general pattern of the importance of sharing knowledge and building on the knowledge of others has been generally shown to be beneficial to collaborative learning in traditional classroom settings [42, 43].

Lastly, the literature suggests that game design can potentially influence the relationship between collaboration and learning. For example, Sung and Hwang [39] compared individual learning, collaborative learning, and scaffolded collaborative learning of natural science topics. They found no differences between individual and collaborative learning, but when compared to a version of the game that included a scaffolding feature for students to reflect and organize their knowledge collaboratively, the collaborative version performed significantly better than the individual learning.

1.4 Current Study, Novelty, and Contributions

Prior research has showcased the need for specific investigations as to which CPS behaviors relate to pre-post learning outcomes in game-based learning contexts. Specifically, studies on what constitutes CPS provide valuable insights on how to assess CPS skills in educational settings. And although research has attempted to identify the relationship between CPS skills and performance-based outcomes like task success [38], there is a need to investigate the predictive power of CPS assessment frameworks relative to specific learning outcomes. Moreover, since most studies focus on dyads, there is a need to validate CPS frameworks in more complex interactions that emerge in triads and beyond [1].

To address those gaps, we investigated whether CPS skills manifested in triadic interactions predicted individual learning in the context of remote collaborations in a physics learning game. In addition, we compared different CPS coding approaches to determine which was more predictive of student learning. Whereas all approaches utilized the same three CPS facets of constructing shared knowledge, negotiation/coordination, and maintaining team function from the generalized competency model of CPS [37, 38], they differed at the level of specific indicators used to comprise each facet. Our hypothesis was that all three CPS facets would positively predict student learning.

2 Methods

2.1 Participants

Participants were 303 university students from two large public universities in the US. Participants reported fairly equal gender distribution, with 56% being female students, 44% male students, and 0% other genders. The average age across participants was 22 years old, and the self-reported ethnicities showed diversity: Caucasian (47%), Hispanic/Latino (28%), Asian (18%), Black or African American (2%), American Indian or Alaska Native (1%), and "other" (4%). At the end of the study, participants were either compensated with a $50 Amazon gift card (96%) or course credit (4%). In the study a total of 249 students from 83 teams were eligible to be included in the analysis based on availability of data and completion of all procedures. All participants provided written informed consent and all procedures were approved by the designated IRB.

2.2 Physics Playground

Physics Playground is a 2D digital game to facilitate physics conceptual learning, such as Newton's laws (previously known as Newton's Playground) [33]. The game contains game levels with a wide range of difficulty. The levels are arranged in terms of difficulty level from the easiest to the hardest based on physics experts' ratings. Players can quit a level and start a different one if they were struggling with a certain game level. To solve each game level, players need to guide a green ball to hit a red balloon by drawing simple physics machines—such as levers and pendulums—which become dynamic following the laws of physics. Figure 1 shows a game level where a ramp is drawn in red to lead the falling green ball to the balloon. The in-game performance has three possible outcomes: gold coin, silver coin, or unsolved (no coin), with performance measured by the elegance of the solutions manifested as the number of objects drawn (i.e., a solution with fewer objects is more elegant [gold coin] than one with many objects [silver coin]).

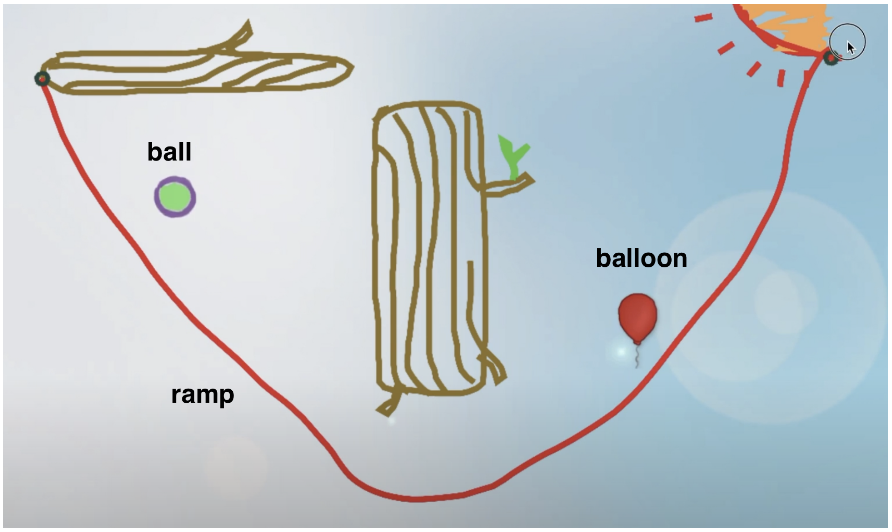


Figure 1: A level in Physics Playground – Around the tree

2.3 Measures

We collected several measures, of which we focus here on CPS skills, physics knowledge (pre-post), physics self-efficacy, and perceptions of team collaboration with the latter two measures serving as control variables.

CPS measure/coding scheme. Table 1 contains the facets and associated indicators in the validated GCM framework [38], which evolved from an earlier version [37]. We adopted an utterance-based coding strategy in our study, where utterances were transcribed and coded according to the scheme illustrated in Table 1. For example, the utterance, "*How about we draw a weight to pull the lever down*?" was coded as "suggests appropriate ideas" (drawing a weight) and "provides reasons to support a solution (to pull the lever down)".

Utterances generated by the triads were transcribed by IBM Watson speech-to-text recognition software for coding. Three coders were trained to complete the coding procedure. The coders could access the transcripts and video recordings to ensure the accuracy of utterance transcription and coding within the broader CPS context. The coders received two rounds of training to establish appropriate inter-rater reliability at the utterance-level: Gwet's AC1 (0.91 – 1.00) and percentage agreement (0.89 – 1.00). Then the coders individually coded 209 level attempts (randomly split among coders) selected using a matching procedure that differentiated task outcomes (gold, silver, none) while approximately matching on other covariates (detailed in [38]).

Physics learning assessments. The physics tests (i.e., forms A and B, used as pretest and posttest) were developed by two subject matter experts. Each of the two forms had 10 multiple-choice items focusing on two concepts: (a) energy can transfer (EcT), and (b) properties of torque (PoT). The possible score of each test ranged from 0 to 10. Figure 2 demonstrates a sample test item on energy can transfer concept. The test scores were scaled from 0 to 1. The mean pre-test score was 0.65 ($SD = 0.19$, $n = 234$) and the mean post-test score was 0.69 ($SD = 0.20$, $n = 234$). A two-tailed paired-samples t-test indicated a significant increase from pretest to posttest (t (233) = 3.06, p = .003).

Surveys on physics self-efficacy and perceptions of team collaboration. Students' self-efficacy for physics [24] was measured by three Likert-scale items, such as "I generally manage to solve difficult physics problems if I try hard enough." For each item, students rated their degrees of agreement from 1 to 7, with 1 representing *strongly disagree* and 7 representing *strongly agree*. Students' perceptions of past experiences with team collaboration [9] were also measured with similar Likert-scale items (e.g., "I can work very effectively in a group setting").

2.4 Procedure

Participants were assigned to 101 triads based on scheduling constraints. The study was conducted in controlled lab settings in two sites set up for virtual collaboration (Figure 3). The triads came to a lab at each site equipped with videoconferencing tools such as Zoom, webcams, and headphones for each team member, which they used to remotely collaborate with each other to complete the main game-based CPS activities. Students' audio, video, and screen content were recorded, and those recordings were used in the analyses.

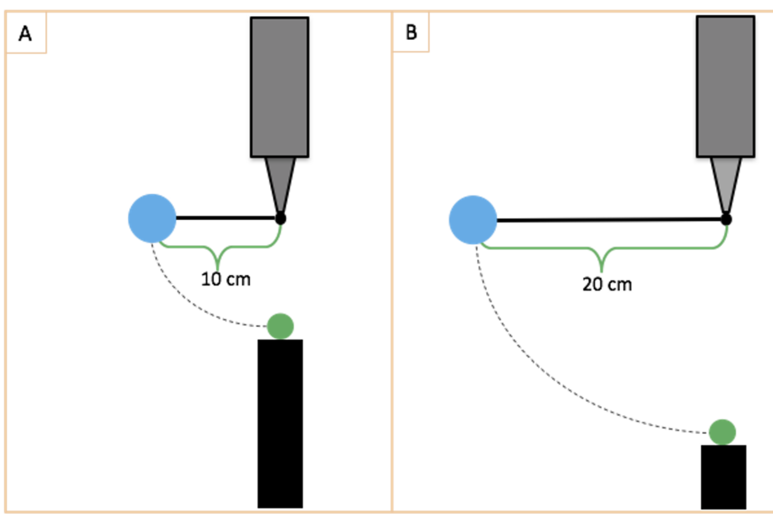
Before coming to the lab, each participant completed a demographic survey as well as an online pretest covering physics concepts. The pretest had two parallel forms (A and B) counterbalanced across participants. The participants also familiarized themselves with the collaborative physics task by going through a game tutorial and completing five easy game levels. They additionally completed a battery of individual differences measures, including physics self-efficacy and perceptions of team collaboration.

When participants were in the lab, they collaboratively played the game for three 15-minute blocks: warmup, block 1, and block 2. The warmup block served to help participants get to know each other with six easy-to-medium game levels. Block 1 and block 2 each contained either seven EcT levels or six PoT levels, with

Table 1: Details of CPS coding scheme. R indicates a reverse-coded indicator

CPS facets	Indicators [Examples]
Constructing shared knowledge	Talks about challenge situation ["What is that?"; "Can I delete it?"]
	Suggests appropriate ideas ["Draw a weight"; "Make it longer."]
	Suggests inappropriate ideas (R) ["Increase the weight" when it should be decreased]
	Confirms understanding ["Is that what you're asking?"]
	Interrupts others (R) [Jumping in when others are talking]
Negotiation/coordination	Provides reasons to support a solution ["That's because . . ."]
	Questions/Corrects others' mistakes ["It would get stuck."]
	Responds to others' ideas/questions ["yes", "I'm not sure"]
	Discusses the results ["The ball was not high enough to hit it."]
	Brings up giving up the challenge (R) ["Should we try a different one?"]
	Strategizes to achieve task goals ["Should we try again and get a gold coin?"]
	Tries to quickly save almost successful attempts ["Click the ball now."]
Maintaining team function	Asks for suggestions ["Any ideas?", "What should I do?"]
	Compliments or encourages others ["Great job!", "Good idea!"]
	Provides instructional support ["Do you see the red line? Start from there."]
	Apologizes for one's mistakes ["Sorry, my bad."]
	Criticizes, makes fun of others (R) ["That was stupid."]
	Initiates off-topic conversation (R) ["Did you have breakfast this morning?"]
	Joins off-topic conversation (R) ["Yes, but I'm still hungry."]

1. In Figures A and B, the two pendulums have different lengths, but the same mass. Which pendulum will have the greater speed *just before* it impacts with the green ball?



a. A will be faster than B.
b. A and B will move at the same speed.
c. B will be faster than A. *
d. More information is needed to answer the question.

Figure 2: A sample test item on energy can transfer.

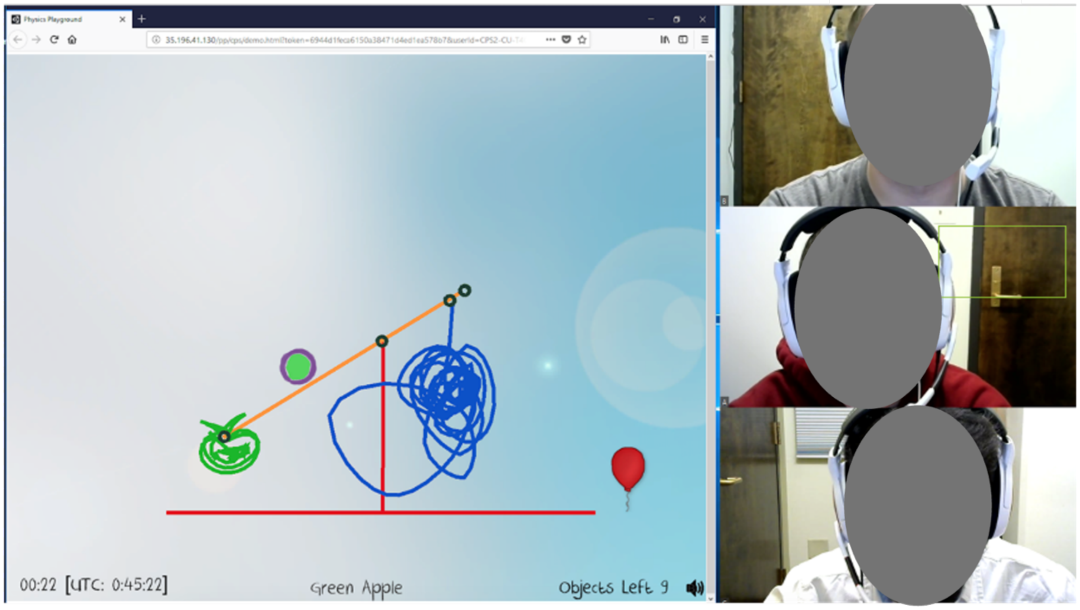


Figure 3: The set up of triadic virtual collaboration - each participant with a headset viewing a shared screen.

medium to high level difficulty. They also contained manipulated task instructions of focusing on maximizing gold coins vs. solving as many levels as possible. The assignment of task instructions (gold vs. silver), content (EcT or PoT), and order of the two blocks were counterbalanced across participants.

There was also a 15-minute transfer task after completing the above three blocks (not analyzed here). Then participants completed the posttest. The test forms were counterbalanced in the study (i.e., if a participant received Form A as a pretest, they would receive Form B as a posttest, and vice versa).

2.5 Analytic strategy

Because we aimed to investigate how CPS processes might relate to individual physics learning, the analysis was conducted at the student level (n = 249 with 234 valid data for all measures). We computed our CPS facet scores at the individual level (i.e., scores on constructing shared knowledge, negotiation/coordination, and maintaining team function) using three different methods. The first method involved *reverse coding* the negative indicators (annotated with "R" in Table 1). Specifically, the CPS facet scores were computed by taking the sum of positive indicators minus negative indicators. The second method, called *separate sums,* entailed computing the sum of positive and negative indicators separately. Thus, each CPS facet would have two sub-scores—one positive and one negative. The third method involved simply summing the raw counts of behaviors across *all* indicators within each facet without reverse coding the negative ones, referred to as *sum total* scores. The goal was to investigate whether positive and negative indicators had additive predictive power or if they canceled each other out.

Multi-level modeling was employed in predicting learning outcomes (measured by pre- and posttests), because students were nested within teams (random intercepts). The outcome was physics posttest score. The main predictors were CPS facet scores, and the three covariates were students' pretest score, their subjective perceptions of team collaboration, and their physics self-efficacy scores. We used the lme4 package in R version 4.3.0 with two-tailed tests with a $p < .05$ cutoff for significance.

3 Results

3.1 Descriptives and Correlations

Table 2 shows descriptive statistics and correlations for the three CPS facet scores (computed using reverse-coding method) and the other variables (i.e., pretest scores, posttest scores, perceptions of team collaboration, and physics self-efficacy). The CPS facets were aggregated to participant level by averaging across all utterances generated by that participant. We found that participants had slightly higher scores for negotiation/coordination (M = .18) than the other two CPS facets (M = .14 and .15), which were similar. The three facets were not significantly correlated with each other (rs from 0.06 to 0.11), suggesting they are indexing unique information. Of the three facets, only negotiation/coordination significantly correlated with both posttest (r = .14) and pretest (r = .19) scores. Maintaining team function had a non-zero correlation with posttest scores (r = .09), but constructing shared knowledge had a near zero correlation with posttest scores (r = -.05). The three facets were not significantly correlated with perceptions of team collaboration and physics self-efficacy, but this later variable was significantly correlated with both pre- and posttest scores.

Note: $^{*}\ p < .05$, $^{**}\ p < .01$, $^{***}\ p < .001$

Table 2: Means and standard deviations for three CPS facet scores (reverse coding) and the other variables in the statistical model (diagonal cells). Pairwise Pearson correlation coefficients among variables (the upper diagonal cells).

	Constructing shared knowledge	Negotiation/ Coordination	Maintaining team Function	Pretest score	Posttest score	Perceptions of team collaboration	Physics self-efficacy
Constructing shared knowledge	0.14 (0.14)	-0.11	-0.11	-0.04	-0.05	-0.11	0.05
Negotiation/ Coordination		0.18 (0.09)	0.06	0.19**	0.14*	0.00	0.12
Maintaining team function			0.15 (0.11)	-0.10	0.09	0.08	0.09
Pretest score				0.65 (0.19)	0.58***	0.08	0.29***
Posttest score					0.69 (0.20)	-0.09	0.28***
Perceptions of team collaboration						3.96 (1.23)	0.01
Physics self-efficacy							4.70 (1.29)

3.2 Predictive Models

Table 3 shows the results of three separate models for predicting posttest scores from CPS facets computed by three methods – reverse coding, separate sums, and sum total – after controlling for pretest scores, physics self-efficacy, and perceptions of team collaboration. Regardless of the method used to compute each facet score, *maintaining team function* significantly predicted individuals' physics posttest scores. In the separate sums model, only the sum of positive indicators of the maintaining team function facet significantly predicted individual posttest scores. In addition, all three covariates (i.e., pretest score, self-efficacy, and perception of team collaboration) were significant predictors across all three models (except for $p = .053$ for physics self-efficacy in the reverse coding model). The marginal R squared shows that CPS facet scores and the covariates together explained between 38% to 39% of the variance in predicting the physics posttest score. The inclusion of the random effects (conditional R squared) explained a total of 45% of the variance.

We also fit separate mixed effects regression models predicting posttest scores with the individual facets (instead of including them in the same models as above). The models only included pretest score as a covariate to address whether the lack of predictive power of constructing shared knowledge and negotiation/coordination from the above analyses might be due to the presence of the other covariates. The results in Table 4 show the coefficients and 95% confidence intervals for the facet scores only ($n = 234$). We found that the predictive results were similar to Table 3 – maintaining team function was the only statistically significant predictor of individual learning outcomes.

Note: * $p < .05$, ** $p < .01$, *** $p < .001$

4 Discussion

We explored the relationships between CPS facets from the GCM and group-to-individual transfer learning in a collaborative game-based learning environment. Of the three CPS facets tested, maintaining team function was the only significant predictor for individual physics learning, regardless of how the facet scores were computed. The findings emphasize the critical role of healthy and functional interpersonal relationships within a team. For instance, team members should proactively ask for suggestions if they are stuck, and provide necessary support when others encounter problems. The zero-order correlations indicated a significant association between negotiation/coordination and posttest scores, but this association was non-significant once the covariates were added. Surprisingly, constructing shared knowledge was not associated with posttest scores in the study, contrary to our prediction. The findings that the constructing shared knowledge and negotiation/coordination facets did *not* predict individual learning merits further investigation. For example, CPS facets might interact to predict learning or there might be nonlinear relationships with learning. It might also be the case that not all facets predict all CPS outcomes, but there is differential prediction where some predict task performance, others' perceptions of the collaboration, and others learning gains. Together, this suggests that the GCM framework may still be incomplete and further refinement is warranted.

Table 3: Multilevel modeling predicting posttest scores from CPS facet scores computed by three methods – reverse coding (upper), separate sums (middle), and sum total (bottom)

FIXED EFFECTS

Predictor	*Estimates*	*CI*	*p*
Model 1: Reverse coding			
(Intercept)	0.27	0.16 – 0.39	**<0.001**
Constructing shared knowledge	-0.03	-0.18 – 0.12	0.685
Negotiation/coordination	-0.02	-0.24 – 0.21	0.896
Maintaining team function	0.27	0.09 – 0.45	**0.004**
Pretest score	0.61	0.49 – 0.72	**<0.001**
Perceptions of team collaboration	-0.02	-0.04 – -0.01	**0.008**
Physics self-efficacy	0.02	-0.00 – 0.03	0.053
Model 2: Separate sums			
(Intercept)	0.26	0.14 – 0.39	**<0.001**
Constructing shared knowledge *Positive*	-0.01	-0.19 – 0.17	0.891
Constructing shared knowledge *Negative*	0.08	-0.23 – 0.39	0.616
Negotiation/coordination *Positive*	-0.04	-0.27 – 0.20	0.755
Negotiation/coordination *Negative*	-0.93	-2.78 – 0.92	0.321
Maintaining team function *Positive*	0.27	0.08 – 0.45	**0.005**
Maintaining team function *Negative*	-0.23	-1.22 – 0.76	0.649
Pretest score	0.62	0.50 – 0.73	**<0.001**
Perceptions of team collaboration	-0.02	-0.04 – -0.01	**0.009**
Physics self-efficacy	0.02	0.00 – 0.03	**0.046**
Model 3: Sum total			
(Intercept)	0.26	0.14 – 0.39	**<0.001**
Constructing shared knowledge	0.02	-0.13 – 0.18	0.758
Negotiation/coordination	-0.03	-0.26 – 0.20	0.811
Maintaining team function	0.26	0.08 – 0.44	**0.005**
Pretest score	0.60	0.49 – 0.72	**<0.001**
Perceptions of team collaboration	-0.02	-0.04 – -0.01	**0.008**
Physics self-efficacy	0.02	0.00 – 0.03	**0.047**

RANDOM EFFECTS

Parameter	**Model 1 Reverse coding**	**Model 2 Separate sums**	**Model 3 Sum total**
σ2	0.02	0.02	0.02
τ00 team	0.00	0.00	0.00
ICC	0.10	0.10	0.11
N team	81	81	81
Observations	234	234	234
Marginal R2 / Conditional R2	0.385 / 0.449	0.386 / 0.448	0.382 / 0.449

Nevertheless, the paper takes a further step towards developing a CPS framework given that the current literature lacks evidence of reliability and validity of guiding frameworks [6]. Further, it is important that the validation environment resembles real-world collaborative situations as much as possible. Some previous validation studies were conducted in environments devoid of collaboration with other humans as in the PISA framework [8]. Although the current study was conducted in a lab setting, it entailed free exchange of information with teammates using speech instead of typing into a chat box. The ever-increasing importance of virtual

Table 4: The coefficients of each CPS facet (computed by different methods) predicting the posttest scores with pretest as a covariate

Scoring Method	Constructing shared knowledge	Negotiation/ Coordination	Maintaining Team Function
Reverse Coding	-0.02 (-0.18, 0.13)	0.02 (-0.21, 0.26)	0.27** (0.09, 0.45)
Separate Sums Positive	-0.01 (-0.19, 0.18)	0.00 (-0.23, 0.24)	0.27** (0.09, 0.45)
Separate Sums Negative	0.08 (-0.24, 0.40)	-0.96 (-2.84, 0.93)	-0.33 (-1.33, 0.67)
Sum Total	0.01 (-0.15, 0.17)	-0.01 (-0.24, 0.23)	0.25** (0.07, 0.43)

collaboration via videoconferencing platforms such as Zoom in this globalized world makes it necessary to investigate and train effective collaborations within such contexts as in the current study. The study has limitations. For one, we focused solely on utterance analysis, but non-verbal behaviors are also likely to be important for CPS assessment [41]. Second, the study was conducted in a game-based learning setting. There is a need to examine whether the relationship between CPS and learning outcome can be replicated in other CPS contexts. The diversity of the sample is another limitation that should be addressed for broader claims of generalizability. Further, we used a human-coding method in the study because it is the gold-standard, but it is also time consuming and difficult to scale. Machine learning techniques such as natural language processing can be leveraged to automate the assessment process [30, 31, 35]. In addition to automated assessment, AI can also be utilized to improve assessment accuracy by incorporating multimodal data [23]

In parallel to framework development and validation, future research should tackle the design and development of learning analytics interfaces [22] aimed to support the enhancement of CPS skills. Some work on this front is already underway. For example, CPSCoach [36] is an intelligent system that uses natural language processing (NLP) to automatically assess students' CPS skills (based on the GCM model) from their collaborative discourse to provide feedback for reflection. This was followed by CPSCoach 2.0 which included more active learning approaches (rather than passively viewing feedback), resulting in improvement in CPS skills [11]. Research has also been leveraging validated frameworks like the GCM and automated assessment to develop collaborative reflection-support tools in classroom environments [13]. Validated CPS frameworks such as the GCM considered here serve as the foundation for automated assessment in digital learning environments and can guide the design and evaluation of interventions.

5 Conclusions

As research on CPS has been emerging, work is needed to fully understand it as a unique construct by validating existing frameworks [16]. There is little evidence on how CPS skills may be associated with individual learning outcomes post collaboration (group-to-individual learning transfer). Therefore, this research adopted an analytical approach to investigate the relationships between CPS and individual learning outcomes. Specifically, the current study assessed CPS skills in the context of a remote CPS task embedded in a digital game-based learning environment. Results indicated that only one of the three CPS skills (maintaining a positive team dynamic) predicted student learning outcomes. Findings from this study can add to the literature on effective assessment of CPS skills in dynamic and complex learning contexts. Furthermore, this work can inform researchers and educators on best practices of applying CPS skills to enhance student learning. By leveraging AI, a validated framework can not only improve the deployment of automated CPS assessments but also facilitate the development of effective interventions for CPS skill development.

Acknowledgments

This research was supported by the NSF National AI Institute for Student-AI Teaming (iSAT) under grant DRL 2019805 and by DUE 1745442/1660877. The opinions expressed are those of the authors and do not represent views of the NSF.

References

[1] Mary Jean Amon, Hana Vrzakova, and Sidney K. D'Mello. 2019. Beyond Dyadic Coordination: Multimodal Behavioral Irregularity in Triads Predicts Facets of Collaborative Problem Solving. Cognitive Science 43, 10, e12787. https://doi.org/10.1111/cogs.12787

[2] Jessica Andrews-Todd and Carol M. Forsyth. 2020. Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. Computers in Human Behavior 104, 105759. https://doi.org/10.1016/j.chb.2018.10.025

[3] Jessica Andrews-Todd, Yang Jiang, Jonathan Steinberg, Samuel L. Pugh, and Sidney K. D'Mello. 2023. Investigating collaborative problem solving skills and outcomes across computer-based tasks. Computers Education 207, 104928. https://doi.org/10.1016/j.compedu.2023.104928

[4] Ryan S.J.d. Baker, Sidney K. D'Mello, Ma.Mercedes T. Rodrigo, and Arthur C. Graesser. 2010. Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. International Journal of Human-Computer Studies 68, 4, 223–241. https://doi.org/10.1016/j.ijhcs.2009.12.003

[5] Dan Carpenter, Andrew Emerson, Bradford W. Mott, Asmalina Saleh, Krista D. Glazewski, Cindy E. Hmelo-Silver, and James C. Lester. 2020. Detecting Off-Task Behavior from Student Dialogue in Game-Based Collaborative Learning. In Artificial Intelligence in Education, Ig Ibert Bittencourt, Mutlu Cukurova, Kasia Muldner, Rose Luckin, and Eva Mill.n (Eds.). Springer International Publishing, Cham, 55–66.

[6] Huanyou Chai, Tianhui Hu, and Li Wu. 2024. Computer-based assessment of collaborative problem solving skills: A systematic review of empirical research. Educational Research Review 43, 100591. https://doi.org/10.1016/j.edurev.2023.100591

[7] Ching-Huei Chen and Victor Law. 2016. Scaffolding individual and collaborative game-based learning in learning performance and intrinsic motivation. Computers in Human Behavior 55, 1201–1212. https://doi.org/10.1016/j.chb.2015.03.010

[8] Matthias Stalder, Katharina Herborn, Maida Mustafić, and Samuel Greiff. 2020. The assessment of collaborative problem solving in PISA 2015: An investigation of the validity of the PISA 2015 CPS tasks. Computers & Education, 157, 103964. https://doi.org/10.1016/j.compedu.2020.103964

[9] Jos. M. de la Torre-Ruiz, Vera Ferr.n-V.lchez, and Natalia Ortiz de Mandojana. 2014. Team Decision Making and Individual Satisfaction With the Team. Small Group Research 45, 2, 198–216. https://doi.org/10.1177/1046496414525478

[10] Muhterem Dindar, Sanna J.rvel., and Hanna J.rvenoja. 2020. Interplay of metacognitive experiences and performance in collaborative problem solving. Computers Education 154, 103922. https://doi.org/10.1016/j.compedu.2020.103922

[11] Sidney K. D'Mello, Nicholas Duran, Amanda Michaels, and Angela E.B. Stewart. 2024. Improving collaborative problem-solving skills via automated feedback and scaffolding: a quasi-experimental study with CPSCoach 2.0. User Modeling and User-Adapted Interaction. https://doi.org/10.1007/s11257-023-09387-6

[12] Nia Dowell, Yiwen Lin, AndrewGodfrey, and Christopher Brooks. 2020. Exploring the Relationship between Emergent Sociocognitive Roles, Collaborative Problem-Solving Skills and Outcomes: A Group Communication Analysis. Journal of Learning Analytics 7, 1 (Apr. 2020), 38–57. https://doi.org/10.18608/jla.2020.71.4

[13] Thomas Breideband, Chelsea Chandler, Shriprajwal Krishnamurthy, Sachin Rathod, Tilak Singh, John Weatherly, and Sidney K. D'Mello. 2024. CoBi, the Community Builder: An AI that Helps Youth Develop Collaboration Skills. Workshop on Child-Centred AI Design (CCAI) at the ACM CHI conference on Human Factors in Computing Systems (CHI 2024).

[14] Fahmid Morshed Fahid, Seung Lee, Bradford Mott, Jessica Vandenberg, Halim Acosta, Thomas Brush, Krista Glazewski, Cindy Hmelo-Silver, and James Lester. 2023. Effects of Modalities in Detecting Behavioral Engagement in Collaborative Game-Based Learning. In LAK23: 13th International Learning Analytics and Knowledge Conference (Arlington, TX, USA) (LAK2023). Association for Computing Machinery, New York, NY, USA, 208–218. https://doi.org/10.1145/3576050.3576079

[15] Stephen M. Fiore, Art Graesser, Samuel Greiff, Patrick Griffin, Brian Gong, Patrick Kyllonen, Christine Massey, Harry O'Neil, Jim Pellegrino, Robert Rothman, Helen Soulé, and Alina von Davier. 2017. Collaborative problem solving: Considerations for the national assessment of educational progress. https://nces.ed.gov/nationsreportcard/pdf/researchcenter/collaborative_problem_solving.pdf

[16] Arthur C. Graesser, Stephen M. Fiore, Samuel Greiff, Jessica Andrews-Todd, Peter W. Foltz, and Friedrich W. Hesse. 2018. Advancing the Science of Collaborative Problem Solving. Psychological Science in the Public Interest 19, 2, 59–92. https://doi.org/10.1177/1529100618808244

[17] Patrick Griffin and Esther Care. 2015. The ATC21S Method. Springer Netherlands, Dordrecht, 3–33. https://doi.org/10.1007/978-94-017-9395-7_1

[18] Anisha Gupta, Dan Carpenter, Wookhee Min, Bradford Mott, Krista Glazewski, Cindy E. Hmelo-Silver, and James Lester. 2023. Enhancing Stealth Assessment in Collaborative Game-Based Learning with Multi-task Learning. In Artificial Intelligence in Education, Ning Wang, Genaro Rebolledo-Mendez, Noboru Matsuda, Olga C. Santos, and Vania Dimitrova (Eds.). Springer Nature Switzerland, Cham, 304–315.

[19] Katharina Herborn, Matthias Stadler, Maida Mustafić, and Samuel Greiff. 2020. The assessment of collaborative problem solving in PISA 2015: Can computer agents replace humans? Computers in Human Behavior 104, 105624. https://doi.org/10.1016/j.chb.2018.07.035

[20] Friedrich Hesse, Esther Care, Juergen Buder, Kai Sassenberg, and Patrick Griffin. 2015. A Framework for Teachable Collaborative Problem Solving Skills. Springer Netherlands, Dordrecht, 37–56. https://doi.org/10.1007/978-94-017-9395-7_2

[21] Jason G. Reitman, Charis Clevenger, Quinton Beck-White, Amanda Howard, Sierra Rose, Jacov Elick, Julianna Harris, Peter Foltz and Sidney K. D'Mello. 2023. A Multi-theoretic Analysis of Collaborative Discourse: A Step Towards AI-Facilitated Student Collaborations. In: Wang, N., Rebolledo-Mendez, G., Matsuda, N., Santos, O.C., Dimitrova, V. (eds) Artificial Intelligence in Education. AIED 2023. Lecture Notes in Computer Science, vol 13916. Springer, Cham. https://doi.org/10.1007/978-3-031-36272-9_47

[22] Rogers Kaliisa, Kamila Misiejuk, Sonsoles L.pez-Pernas, Mohammad Khalil, and Mohammed Saqr. 2024. Have Learning Analytics Dashboards Lived Up to the Hype? A Systematic Review of Impact on Students' Achievement, Motivation, Participation and Attitude. In Proceedings of the 14th Learning Analytics and Knowledge Conference (Kyoto, Japan) (LAK '24). Association for Computing Machinery, New York, NY, USA, 295–304. https://doi.org/10.1145/3636555.3636884

[23] Angela E. B. Stewart, Zachary Keirn, and Sidney K. D'Mello. 2021. Multimodal modeling of collaborative problem-solving facets in triads. User Modeling and User-Adapted Interaction 31, 4 (Sep 2021), 713–751. https://doi.org/10.1007/s11257-021-09290-y

[24] Christine Lindstrøm, and Manjula D. Sharma. 2011. Self-efficacy of first year university physics students: Do gender and prior formal instruction in physics matter? International Journal of Innovation in Science and Mathematics Education 19(2).

[25] Geoff Musick, Thomas A. O'Neill, Beau G. Schelble, Nathan J. McNeese, and Jonn B. Henke. 2021. What Happens When Humans Believe Their Teammate is an AI? An Investigation into Humans Teaming with Autonomy. Computers in Human Behavior 122, 106852. https://doi.org/10.1016/j.chb.2021.106852

[26] OECD. 2017. PISA 2015 collaborative problem-solving framework. https://doi.org/doi.org/10.1787/9789264281820-8-en

[27] OECD. 2017. PISA 2015 Results (Volume V). 308 pages. https://doi.org/10.1787/9789264285521-en

[28] Fernando Olivera and Susan G. Straus. 2004. Group-to-Individual Transfer of Learning: Cognitive and Social Factors. Small Group Research 35, 4, 440–465. https://doi.org/10.1177/1046496404263765

[29] Hui-Hua Pai, David A. Sears, and Yukiko Maeda. 2015. Effects of Small-Group Learning on Transfer: A Meta-Analysis. Educational Psychology Review, 27, 79–102. https://doi.org/10.1007/s10648-014-9260-8

[30] Samuel Pugh, Shree Krishna Subburaj, Arjun Ramesh Rao, Angela E Stewart, Jessica Andrews-Todd, and Sidney K D'Mello. 2021. Say What? Automatic Modeling of Collaborative Problem Solving Skills from Student Speech in the Wild. Proceedings of the 14th International Conference on Educational Data Mining. https://par.nsf.gov/biblio/10494306

[31] Samuel L. Pugh, Arjun Rao, Angela E.B. Stewart, and Sidney K. D'Mello. 2022. Do Speech-Based Collaboration Analytics Generalize Across Task Contexts?. In LAK22: 12th International Learning Analytics and Knowledge Conference (Online, USA) (LAK22). Association for Computing Machinery, New York, NY, USA, 208–218. https://doi.org/10.1145/3506860.3506894

[32] David A. Sears, and James Michael Reagin. 2013. Individual versus collaborative problem solving: divergent outcomes depending on task complexity. Instructional Science, 41, 1153–1172. https://doi.org/10.1007/s11251-013-9271-8

[33] Valerie Shute, and Matthew Ventura. 2013. Stealth Assessment: Measuring and Supporting Learning in Video Games. The MIT Press. https://doi.org/10.7551/mitpress/9589.001.0001

[34] Angela E.B. Stewart, Mary Jean Amon, Nicholas D. Duran, and Sidney K. D'Mello. 2020. Beyond Team Makeup: Diversity in Teams Predicts Valued Outcomes in Computer-Mediated Collaborations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376279

[35] Angela E.B. Stewart, Hana Vrzakova, Chen Sun, Jade Yonehiro, Cathlyn Adele Stone, Nicholas D. Duran, Valerie Shute, and Sidney K. D'Mello. 2019. I Say, You Say, We Say: Using Spoken Language to Model Socio-Cognitive Processes during Computer-Supported Collaborative Problem Solving. Proc. ACM HCI 3, CSCW, Article 194 (Nov. 2019), 19 pages. https://doi.org/10.1145/3359296

[36] Angela E. B. Stewart, Arjun Rao, Amanda Michaels, Chen Sun, Nicholas D. Duran, Valerie J. Shute, and Sidney K. D'Mello. 2023. CPSCoach: The Design and Implementation of Intelligent Collaborative Problem Solving Feedback. In Artificial Intelligence in Education, Ning Wang, Genaro Rebolledo-Mendez, Noboru Matsuda, Olga C. Santos, and Vania Dimitrova (Eds.). Springer Nature Switzerland, Cham, 695–700.

[37] Chen Sun, Valerie J. Shute, Angela Stewart, Jade Yonehiro, Nicholas Duran, and Sidney D'Mello. 2020. Towards a generalized competency model of collaborative problem solving. Computers Education 143, 103672. https://doi.org/10.1016/j.compedu.2019.103672

[38] Chen Sun, Valerie J. Shute, Angela E.B. Stewart, Quinton Beck-White, Caroline R. Reinhardt, Guojing Zhou, Nicholas Duran, and Sidney K. D'Mello. 2022. The relationship between collaborative problem solving behaviors and solution outcomes in a game-based learning environment. Computers in Human Behavior 128, 107120. https://doi.org/10.1016/j.chb.2021.107120

[39] Han-Yu Sung and Gwo-Jen Hwang. 2013. A collaborative game-based learning approach to improving students' learning performance in science courses. Computers & Education 63, 43–51. https://doi.org/10.1016/j.compedu.2012.11.019

[40] Zehui Zhan, Guoqing He, Tingting Li, Luyao He, and Siyu Xiang. 2022. Effect of groups size on students' learning achievement, motivation, cognitive load, collaborative problem-solving quality, and in-class interaction in an introductory AI course. Journal of Computer Assisted Learning 38, 6, 1807–1818. https://doi.org/10.1111/jcal.12722

[41] Robert G. Moulder, Nicholas D. Duran, and Sidney K. D'Mello. 2022. Assessing Multimodal Dynamics in Multi-Party Collaborative Interactions with Multi-Level Vector Autoregression. In Proceedings of the 2022 International Conference on Multimodal Interaction (ICMI '22). Association for Computing Machinery, New York, NY, USA, 615–625. https://doi.org/10.1145/3536221.3556595

[42] Noreen M. Webb, Megan L. Franke, Marsha Ing, Jacqueline Wong, Cecilia H. Fernandez, Nami Shin, and Angela C. Turrou. 2014. Engaging with others' mathematical ideas: Interrelationships among student participation, teachers' instructional practices, and learning. International Journal of Educational Research, 63, 79-93. https://doi.org/10.1016/j.ijer.2013.02.001

[43] Noreen M. Webb, Megan L. Franke, Nicholas C. Johnson, Marsha Ing, and Joy Zimmerman. 2023. Learning through explaining and engaging with others' mathematical ideas. Mathematical Thinking and Learning, 24, 4, 438-464. https://doi.org/10.1080/10986065.2021.1990744

[44] Guojing Zhou, Robert G. Moulder, Chen Sun, and Sidney K. D'Mello. 2022. Investigating Temporal Dynamics Underlying Successful Collaborative Problem Solving Behaviors with Multilevel Vector Autoregression. Proceedings of the 15th International Conference on Educational Data Mining. https://doi.org/10.5281/zenodo.6853137