Computers & Education 143 (2020) 103672

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Towards a generalized competency model of collaborative problem solving

Chen Sun[a,*], Valerie J. Shute[a], Angela Stewart[b], Jade Yonehiro[b], Nicholas Duran[c], Sidney D'Mello[b]

[a] *Florida State University, Tallahassee, FL 32306, USA*
[b] *University of Colorado Boulder, Boulder, CO 80309, USA*
[c] *Arizona State University, Phoenix, AZ 85069, USA*

ARTICLE INFO

Keywords:
Collaborative problem solving
Competency model
Validation

ABSTRACT

Collaborative problem solving (CPS) is an essential skill in the 21st century. There is a need for an appropriate framework and operationalization of CPS to guide its assessment and support and across multiple domains. Accordingly, we synthesized prior research on CPS to construct a generalized CPS competency model (i.e., skills and abilities) consisting of the following core facets: constructing shared knowledge, negotiation/coordination, and maintaining team function. Each facet has two sub-facets, which in turn, have multiple verbal and nonverbal indicators. We validated our model in two empirical studies involving triadic CPS, but in very different contexts – middle-school students playing an educational game in a 3-h, face-to-face session vs. college students engaging in a visual programming task for 20 min via videoconferencing. We used principal component analysis to investigate whether the empirical data aligned with our theorized model. Correlational analyses provided evidence on the orthogonality of the facets and their independence to individual differences in prior knowledge, intelligence and personality and regression analyses indicated that the facets predicted both subjective and objective outcome measures controlling for several covariates. Thus we provide initial evidence for the convergent, discriminant, and predictive validity of our model by using two different CPS contexts and student populations. This shows promise towards generalizing across various human-human CPS interactive environments.

1. Introduction

A defining characteristic of the 21st century is the unprecedented ability to simultaneously connect to others en masse and in real-time. Such interconnectivity promises to have, and is already producing profound changes in how people live, learn, and work (Griffin, Care, & McGaw, 2012; Nelson & Squires, 2017). To successfully navigate this changing landscape, intergovernmental economic organizations, such as the Organization for Economic Cooperation and Development (OECD), and leading educational advocacy organizations, such as the Partnership for 21st Century Learning, have identified certain skills that should be prioritized in training and education programs. That is, the Programme for International Student Assessment (PISA/OECD, 2017a) has emphasized that collaborative problem solving (CPS) is a domain-independent skill for students to succeed in group-based activities in both educational and work contexts. Similarly, according to the Partnership of 21st Century Learning (2016), CPS is a high-priority skill

* Corresponding author.
E-mail address: cs15c@my.fsu.edu (C. Sun).

https://doi.org/10.1016/j.compedu.2019.103672
Received 24 October 2018; Received in revised form 21 July 2019; Accepted 19 August 2019
Available online 22 August 2019

that should be integrated in teaching, learning, and extracurricular activities.

CPS refers to the coordinated attempt between two or more people to share their skills and knowledge for the purpose of constructing and maintaining a unified solution to a problem (OECD, 2017a; Roschelle & Teasley, 1995). Researchers often emphasize individuals' cognitive abilities and social skills required for effective collaboration with the recognition that these skills are only fully expressed in the context of individuals' interactions with each other (Griffin et al., 2012; OECD, 2017a). CPS is thus, at its core, a joint activity that requires the cooperative exchange of information to successfully transform a problem state into a desired goal state. This process hinges on how well individuals can establish common ground concerning the nature of the problem, develop a solution plan, monitor progress along the way, and accommodate multiple perspectives while respectfully managing disagreements. This requires the ability to understand task goals and constraints and consider others' perspectives and knowledge, along with the ability to communicate this understanding through negotiation, mutual regulation, and shared responsibilities.

Because CPS provides opportunities for deep learning and effective problem solving (Huang et al., 2018), students should be prepared for effective CPS in college and in workforce settings (Oliveri, Lawless, & Molloy, 2017). Recent PISA results (National Center for Education Statistics, 2017) measuring students' CPS skills across 52 countries show that U.S. students scored 520, and ranked 13th out of 52 countries based on raw scores (note that the average score was 500 with a range from 382 to 561). The overall results indicate that it is important to cultivate students' CPS skills, but there are serious challenges in effectively teaching these skills (Fiore et al., 2017). One issue is that there is no consensus on a CPS model to operationalize this construct and to measure it effectively (Andrews-Todd & Forsyth, 2018). Throughout the literature, there have been various CPS models proposed by Roschelle and Teasley (1995), Nelson (1999), PISA (OECD, 2017a), ATC21S (Griffin et al., 2012), a CPS ontology (Andrews-Todd & Forsyth, 2018), and a recently proposed model that focuses solely on hand position and head direction in a face-to-face collaborative setting (see Cukurova, Luckin, Millan, & Mavrikis, 2018).

Another challenge with effectively teaching CPS relates to accurately assessing these skills in both digital and face-to-face environments (Cukurova, Luckin, Millán, & Mavrikis, 2018; Siddiq & Scherer, 2017). In a recent review on this topic, Oliveri et al. (2017) reported that research studies use diverse measures to assess CPS (e.g., surveys, tests, observations, and think-aloud protocols) and the quality of the CPS assessments vary widely. More and more, researchers are calling for guiding principles for CPS assessment (Andrews-Todd & Forsyth, 2018; Bause, Brich, Wesselein, & Hesse, 2018). Without a consensus, systematic framework for operationalizing the construct and providing guidance for its assessment, it is difficult to teach, evaluate, and support CPS effectively. The absence of such a framework also greatly limits the generalizability of study results.

We take the first steps towards developing a competency model for CPS that can be used in both remote and face-to-face CPS settings to assess human-human interactions. Towards this end, we have synthesized prior research on CPS to construct our model of key CPS facets, sub-facets, and verbal and nonverbal indicators that directly map onto CPS skills. Our overarching research question addresses the extent to which our proposed CPS model is valid and can generalize to different CPS contexts. We provide initial evidence of its convergent, discriminant, and predictive validity using two studies in different CPS learning environments: (a) one involving physics problem solving during 3-hr face-to-face collaborative gameplay with middle school students, and (b) the second involving the use of programming skills during 20-min remote collaborations among college students. We address four specific research questions:

(1) To what degree does our theoretical model align with results of empirical data-driven principal component analysis (PCA) on the indicators?
(2) What are the inter-correlations between CPS facet and sub-facet scores as computed from our set of indicators (convergent and discriminant validity)?
(3) Are the CPS facet scores divergent from individual differences in prior knowledge, intelligence, and personality?
(4) Do the facet scores predict subjective and objective CPS outcomes?

2. Literature review—theoretical background

Our model drew on the research literature and best practices that focus specifically on the CPS construct rather than related terms, such as "computer-supported collaborative learning," "teamwork," and "problem-solving." We reviewed four widely-cited frameworks in the CPS literature, which provide the theoretical grounding for our generalized competency model presented in Section 3. One of the earliest approaches targeting critical aspects of CPS was Roschelle and Teasley's (1995) joint problem space model in which team members share knowledge of goals, understanding of problems, and possible solutions. According to this framework, establishing a joint problem space is crucial for negotiation to take place. Team members are expected to constantly monitor divergence in understanding, and cognitive convergence is essential to effective collaboration (Teasley, Fischer, Dillenbourg, Kapur, & Chi, 2008). To achieve convergence, team members should resolve conflicts by reasoning based on others' ideas/knowledge (Teasley, 1997; Weinberger & Fischer, 2006) and solve problems with coordinated efforts via negotiation and shared knowledge (Stahl, Koschmann, & Suthers, 2006).

Another early CPS framework was built from a pedagogical perspective (Nelson, 1999). The author provided guidelines for instructors to implement collaborative activities within authentic problem-solving scenarios (see Lee, Huh, & Reigeluth, 2015; Merrill, 2002). The framework consists of a set of actions that team members can take to demonstrate effective collaboration. For example, team members should identify and share relevant knowledge and resources by allocating time for individual and collective work. Moreover, team members need to acquire and apply social skills (e.g., communication, leadership, and conflict management) to engender positive interdependence. As such, all team members should understand that there is no individual success without group success and participate actively to solve problems. Further, to ensure fairness, team members should be personally accountable in their responsibilities within the CPS process.

The third conceptual framework for CPS was inspired by the Assessment and Teaching of Twenty-first Century Skills (ATC21S) project. It aimed to integrate technology for large-scale assessment targeting students between 11 and 15 years old (Griffin et al., 2012) to inform teaching and learning (Hesse, Care, Buder, Sassenberg, & Griffin, 2015). Griffin et al. (2012) noted that CPS is comprised of social skills (i.e., participation, perspective taking, and social regulation) and cognitive skills (i.e., task regulation and knowledge building). Further work elaborated on social and cognitive skills required by CPS (Care, Scoular, & Griffin, 2016; Hesse et al., 2015). The main social skills include participation (actions, interactions, and concerted task completion), perspective taking (adaptive responsiveness and audience awareness), and social regulation (negotiation, self-evaluation, transactive memory, and taking group responsibility). Cognitive skills include planning (problem analysis, goal setting, and resource management), systematically executing solutions and monitoring progress, and learning and knowledge building (domain content and skills). Thus, team members should identify the problem structure and procedures, collect and assess information required to build solutions, and engage in strategic problem solving.

Finally, the PISA framework was used to create a standardized large-scale summative assessment for CPS skills in 15-year-old students from 52 countries (OECD, 2017b). According to the latest PISA report (OECD, 2017a), CPS consists of three competencies that overlap with four problem-solving processes, resulting in 12 CPS skills (Graesser et al., 2018; Webb & Gibson, 2015). The first competency involves establishing and maintaining a shared understanding among team members, which requires establishing common ground, achieved via a free exchange of knowledge and perspectives. The second competency entails taking appropriate action to solve problems via explaining and justifying possible solution plans, negotiating with others, and monitoring solution execution. The third competency involves establishing and maintaining group organization, which relies on each team member fully understanding and fulfilling his or her roles/responsibilities in the team and providing timely feedback on progress.

In summary, the existing conceptual frameworks of CPS, though varying in focus and complexity, share much in terms of underlying constructs. Each has their own limitations. For example, Roschelle and Teasley's (1995) joint problem space model focused on one social aspect of CPS—constructing a shared knowledge structure. A joint problem space is just one of many processes (e.g., knowledge convergence and co-construction) that are essential to effective CPS (Teasley et al., 2008). Nelson's (1999) framework only targets teachers who want to design CPS activities in class. Moreover, Nelson did not specify whether such guidelines apply to contexts other than classroom education, such as computer-mediated CPS or game-based CPS (Strijbos, Martens, & Jochems, 2004).

Recently, the ATC21S (e.g., Care & Griffin, 2014; Vista, Care, & Awwal, 2017) and PISA frameworks have been used to develop CPS assessments (e.g., Hao, Chen, Flore, Liu, & von Davier, 2017; von Davier, Hao, Liu, & Kyllonen, 2017). The aim of the PISA framework is to construct summative assessment to inform education systems. It is limited in that it exclusively targets students around 15 years old. As such, it has been recommended that more empirical testing should be done across different tasks, contexts, and populations (Graesser et al., 2018). Further, the PISA framework was used to assess CPS in human-agent interactions while the ATC21S framework assessed human-human interactions (e.g., Graesser et al., 2018; Scoular, Care, & Hesse., 2017). Human-human interaction has high face validity but is difficult to control in large-scale assessment (Graesser et al., 2018; Rosen, 2015), while human-agent interaction is easy to control and standardize (Greiff, Holt, & Funke, 2013), but can feel inauthentic. That is, interacting with a computer agent using preset text chats, as in the PISA assessment, largely restrains social interaction and does not consider personalities and emotions. The ATC21S framework is intended for developing assessment tasks that can elicit CPS skills specified in the framework (Hesse et al., 2015; Scoular, Care, & Hesse, 2017). However, it exclusively focuses on dyadic communication via a chat box embedded in the CPS tasks (Care et al., 2016; Scoular et al., 2017). In sum, these frameworks do not suit our purpose. That is, our interest lies in authentic human-human interaction when small groups (triads) assume different roles, either assigned or naturally formed. Our goal was not to intentionally exclude dyads from the CPS research (see Moreland, 2010), but to examine a relatively more complex and under-researched context (i.e., collaboration among triads).

3. Proposed competency model of collaborative problem solving

Given the constraints of current frameworks, we aim to establish a CPS model that is applicable across various populations and contexts—such as collaborating during gameplay and other interactive endeavors involving human-human interaction (Funke, Fischer, & Holt, 2018). Our purpose is to understand the CPS construct and use assessment results to inform teaching and learning of CPS, not to construct a standardized summative assessment. Moreover, use of our model does not require building new tasks. Instead, any existing tasks or environments that elicit CPS skills would work.

Another key feature of our model involves the explicit integration of cognitive and social aspects of CPS, as collaboration is social in nature and the two aspects are dependent on each other (Care & Griffin, 2014). Other frameworks, as described earlier, treat the two as separate dimensions with their own processes (Funke et al., 2018; Graesser et al., 2018). Additionally, there are few validation studies of CPS frameworks, and little empirical evidence validating the construct (Care et al., 2016; Scoular et al., 2017). Instead, CPS research tends to focus on assessing collaboration processes and outcomes, and on automating the assessment process using technologies and statistical methods. Thus, we provide empirical support of our CPS model in two separate validation studies.

Culling from the literature, researchers have identified common CPS skills (see Section 2), which form the basis of our proposed generalized competency model. And while different CPS frameworks might use different terms, they all emphasize establishing shared knowledge, resolving divergence and misunderstanding, monitoring progress and results, and maintaining a functional team. Accordingly, our three main CPS facets are: constructing shared knowledge, negotiation/coordination, and maintaining team function. Our model, outlined in Table 1, consists of facets, sub-facets, and associated behavioral indicators. The facets and sub-facets

Table 1
Proposed generalized competency model of facets, sub-facets, and indicators.

Facet	Sub-facet	Indicators
Constructing shared knowledge—expresses one's own ideas and attempts to understand others' ideas	Shares understanding of problems and solutions	Talks about specific topics/concepts and ideas on problem solving ● Proposes specific solutions ● Talks about givens and constraints of a specific task ● Builds on others' ideas to improve solutions
	Establishes common ground	Recognizes and verifies understanding of others' ideas • Confirms understanding by asking questions/ paraphrasing Repairs misunderstandings ● Interrupts or talks over others as intrusion (R)
Negotiation/Coordination—achieves an agreed solution plan ready to execute	Responds to others' questions/ideas	● Does not respond when spoken to by others (R) ● Makes fun of, criticizes, or is rude to others (R) ● Provides reasons to support/refute a potential solution ● Makes an attempt after discussion
	Monitors execution	● Talks about results ● Brings up giving up the challenge (R)
Maintaining team function—sustains the team dynamics	Fulfills individual roles on the team	● Not visibly focused on tasks and assigned roles (R) ● Initiates off-topic conversation (R) ● Joins off-topic conversation (R)
	Takes initiatives to advance collaboration processes	● Asks if others have suggestions ● Asks to take action before anyone on the team asks for help ● Compliments or encourages others

Note. "R" next to an indicator means that it is reverse coded.

are latent variables while the indicators are observables.

We define the **constructing shared knowledge** facet as actively disseminating one's own ideas/knowledge and understanding others' ideas/knowledge. It contains two sub-facets—sharing understanding and establishing common ground. Sharing understanding refers to group members contributing their expertise and ideas regarding the constraints of particular problems, as well as ideas toward specific solutions. Examples of associated behavioral indicators include proposing specific solutions and talking about the givens and constraints of the task. Sharing ideas/information is a critical step towards establishing common ground among team members (Fransen, Weinberger, & Kirschner, 2013; Rummel & Spada, 2005). To further build common ground, team members should acknowledge others' ideas and expertise, confirm each other's understanding, and clarify misunderstanding when necessary (Andrews-Todd, Forsyth, Steinberg, & Rupp, 2018; OECD, 2017a). Seeking confirmation from other group members about current understanding is effective in reducing uncertainty (Jordan & McDaniel, 2014), whereas clarifying misunderstanding provides a learning opportunity, where members explicate their knowledge, which may be implicit, to other group members (Dillenbourg & Traum, 2006). We include interruption as a negative indicator because interrupting while others are speaking can impede CPS (Chiu, 2008), although some interruption can help rectify misunderstandings immediately (Roschelle & Teasley, 1995).

The second CPS facet is **negotiation and coordination**, an iterative process where team members achieve an agreed upon solution that is ready for execution. The goal of negotiation and coordination is to clearly specify shared goals, divide labor as warranted, manage synchrony among members, and produce a joint work product (Rummel & Spada, 2005). Negotiation should reduce uncertainties, resolve conflict through integrating different perspectives, and contribute to a collective solution via joint expertise (Hayashi, 2018; Jordan & McDaniel, 2014; Ke & Im, 2013). Two relevant sub-facets include responding to others' questions/ ideas and monitoring execution. Responding to others' ideas and questions is a key part of coordination (Hesse et al., 2015) and should contribute to reciprocal and balanced exchanges of ideas and knowledge (Barron, 2000). As such, team members should provide feedback regarding others' ideas, offer reasons to support or refute certain claims, negotiate when disagreements occur, and implement consensual solutions after discussion (Gu, Chen, Zhu, & Lin, 2015). Hence, providing reasons to support a potential solution is a positive indicator of this sub-facet while failing to respond when spoken to by others is a negative indicator. The second sub-facet, monitoring execution, requires team members to appraise whether the solution plan works as expected, and their progress toward task completion (Kopp, Hasenbein, & Mandl, 2014; Rosen, 2015). Members should also be able to evaluate their own and others' actions, knowledge, and skills towards task completion (Care et al., 2016; OECD, 2017a). When things go awry, team members need to be able to modify or overhaul the solution if necessary. Thus, talking about progress and results is the key indicator of monitoring execution (Andrews-Todd et al., 2018). Team members may occasionally propose to quit certain tasks, captured in the reverse coded "brings up giving up the challenge" indicator.

It is important that all team members are aware of being part of a team and realize that individual behaviors affect team success

Fig. 1. A game level of Physics Playground—Spider Web (left) and one possible solution—springboard (right).

(Fransen et al., 2013). Successful teams iteratively talk about team organization (Chang et al., 2017). Thus, the third main facet of CPS involves **maintaining team function**. Maintaining a positive and effective team requires members to take distributed responsibility to contribute to the quantity and quality of the collaborative venture (Care et al., 2016; Janssen & Bodemer, 2013). Consequently, one sub-facet of maintaining team function pertains to each member performing his or her own roles/responsibilities within the team. The roles may be assigned by an external source (e.g., a teacher), or more naturalistically evolve during collaboration. Either way, team members should stay focused on the task and on what is needed in their roles, while not distracting themselves or others. The second sub-facet – taking initiative to advance CPS processes – includes asking questions, acknowledging others' contributions, and helping to maintain team organization (Hesse et al., 2015; Rosen, 2017). Taking initiative predicts productive collaboration (Howard, Di Eugenio, Jordan, & Katz, 2017). In effective teams, team members not only regulate their own activities but also evaluate the activities of other members (Raes, Schellens, De Wever, & Benoit, 2016). Because one of the ground rules for collaboration is that group members encourage each other to speak (Mercer, Wegerif, & Dawes, 1999), relevant indicators include asking if others have suggestions and complimenting or encouraging their team members.

4. Study 1—Physics Playground

Study 1 was a pilot study, involving triads playing an educational game called *Physics Playground* (Shute & Ventura, 2013) in a face-to-face setting. *Physics Playground* is a 2D educational drawing game to teach students' physics concepts, such as Newton's laws of force and motion, potential and kinetic energy, and conservation of momentum. Players need to draw simple machines (i.e., ramp, lever, pendulum, and springboard) on the computer screen using a mouse or stylus to direct a green ball to hit a red balloon. Fig. 1 shows a level called Spider Web. One possible solution is to draw a springboard with an attached weight which, when removed, can propel the ball upwards. Players receive a gold trophy if they solve a level with an optimal solution (i.e., drawing a limited number of objects). Otherwise, they receive a silver trophy. Levels can be selected, restarted, and exited at the players' discretion. The CPS task provided conditions for collaboration, such as individual accountability and positive interdependence between peers (Szewkis et al., 2011). Our facets and sub-facets map onto these conditions. For example, players assumed different roles with individual accountability to contribute to the team success, as demonstrated by the sub-facet of "fulfills individual roles on the team."

4.1. Method

Participants. We pilot tested 33 middle school students enrolled in grades 8–9 (36% female) who volunteered to participate in this pilot study. We did not collect information about their age, race, and ethnicity, but most of the participants were 14–15 years old. The students formed 11 triads. Five of the 11 groups were heterogeneous in terms of gender. Three groups consisted of two boys and one girl, and another three groups were composed of one boy and two girls. At the end of the gameplay, each student received a 40-dollar gift card as a reward for participation.

Procedure. This study was a pretest-only design, conducted in a research lab. Students were assigned to triads based on scheduling constraints and were tested in single groups at a time. Prior to gameplay, students completed a pretest on physics knowledge which consisted of 18 multiple-choice questions developed by physics and assessment experts. Each item was scored dichotomously (correct or incorrect), with a possible score range from 0 to 18. The reliability for the physics test is acceptable (Cronbach's $\alpha = 0.72$) (Shute, Ventura & Kim, 2013).

During gameplay, students were seated in a row facing a computer in the middle. Each student assumed a particular role, assigned randomly at the beginning of the study. One student controlled the mouse (the "player"), one was responsible for asking provocative questions (the "questioner"), and the third documented their problem-solving processes, such as levels completed and solutions used (the "recorder"). The "player" sat in the middle seat and the "questioner" and the "recorder" sat on either side of the player. Though each player had a role, they all were required to contribute to the CPS process such as providing solutions and sharing knowledge/ideas. The three students switched roles each time they started a new game level, by physically moving to the seat next to them and taking the role assumed by the person who previously sat there.

The researchers videotaped students' onscreen gameplay activities, their facial expressions, and dialogue during gameplay. Each

Table 2
Descriptive statistics (n = 33) for frequencies per indicator across students—*Physics Playground*.

Indicator	Min	Max	Mean	Median	SD
Proposes specific solutions (Constructing shared knowledge)	1	112	48.24	49	27.40
Builds on others' ideas to improve solutions (Constructing shared knowledge)	2	61	31.72	31	14.47
Makes an attempt after discussion (Negotiation/Coordination)	5	54	28.39	27	12.14
Initiates off-topic conversation (Maintaining team function)	0	26	4.29	1	7.37
Joins off-topic conversation (Maintaining team function)	0	16	3.60	1	4.53
Confirms understanding by asking questions/paraphrasing (Constructing shared knowledge)	0	13	3.16	3	3.47
Asks to take action before anyone on the team asks for help (Maintaining team function)	0	21	2.98	1	4.66
Makes fun of, criticizes, or is rude to others (Negotiation/Coordination)	0	22	2.75	0	4.83
Compliments or encourages others (Maintaining team function)	0	9	2.20	1	2.67
Talks about givens and constraints of a specific task (Constructing shared knowledge)	0	7	2.13	2	1.90
Provides reasons to support a potential solution (Negotiation/Coordination)	0	8	1.94	1	2.23
Brings up leaving a challenge (Negotiation/Coordination)	0	8	1.62	1	1.87
Visibly not focused on tasks and assigned roles (Maintaining team function)	0	11	1.51	0	2.50
Talks about results (Negotiation/Coordination)	0	9	1.50	1	2.32
Asks if others have suggestions (Maintaining team function)	0	3	0.39	0	0.74
Interrupts or talks over others as intrusion (Constructing shared knowledge)	0	3	0.38	0	0.83
Does not respond when spoken to by others (Negotiation/Coordination)	0	2	0.24	0	0.56

team played the game for about 3 h, with a short break in the middle. Although this pilot study did not implement a posttest on physics knowledge, we did log the number of gold trophies obtained in the game as a proxy for learning with more gold trophies indicating greater physics understanding (Shute et al., 2013).

Data Coding. Three coders were trained to reach an agreement of 0.80 (intraclass correlation coefficient) on coding occurrences of behavioral indicators in sample videos. After reaching the inter-rater criterion, they were assigned to different 3-h videos and individually coded entire videos. The coders used one coding sheet for each 4-min segment of the video. They tallied the frequency of occurrence of relevant indicators for each team member within each 4-min segment. There were 32 indicators coded in this pilot study as the CPS model and coding scheme were being developed and refined. To make the indicators generalizable to other contexts, we recycled, re-phrased, and condensed the 32 indicators into a more manageable set of 17 indicators (see Table 1).

Following is an excerpt of an exchange between two participants (Player A and Player C) during gameplay, along with their relevant indicators.

> **Player C**: "What if you grabbed it upwards. And then drew a pendulum, knocked it out. But you drew like farther out, the pendulum" (Proposes specific solution)
> **Player A**: "I have an idea. Wait, which direction should I swing?" (Confirms understanding by asking questions/paraphrasing)
> **Player C**: "Swing from here to here." (Proposes specific solution)
> **Player A:** "Nope, then it would just fly to the spider." (Provides reason to support/refute a potential solution)

4.2. Results

Descriptive Statistics. Table 2 shows the descriptive statistics of frequencies per indicator across students (n = 33). This was done by first summing indicator occurrences across the 4-min segments per student, and then computing descriptive statistics across the students. The indicators are listed in a descending order in terms of the mean occurrence scores. We found that middle school students tended to focus on the constructing shared knowledge facet of CPS in that they frequently talked about specific solutions, and built on each other's ideas to improve existing solution attempts. They also made a solution attempt right after discussion, which is aligned with the negotiation/coordination facet. The remaining indicators were less frequently observed, although this should not be conflated with their unimportance. For example, even though compliments were rare, they can have profound effects on team morale, as would the reverse-coded indicators related to making fun of, criticizing, or being rude to other team members (which were also rarely observed).

Principal component analysis (PCA). We conducted a PCA on the raw frequency data to compare how the identified indicators were organized into factors with respect to our model's three main facets. Our unit of analysis was not at the student level; but instead a unit was a 4-min coding window of gameplay, per student. This resulted in an input of a 1,452 × 17 matrix. Each row corresponded to a 4-min segment per player, each column represented our indicators, and each cell represented the frequency of a particular indicator generated by a particular player within the segment. We used principal component analysis (PCA) with a varimax rotation in SPSS 25. A two-tailed alpha of .05 was adopted for all analyses.

The PCA followed an iterative process of checking assumptions and finding the most theoretically consistent factor structure. First, we removed two indicators (i.e., "does not respond to others" and "brings up giving up the challenge") that did not correlate with any other indicators (rs < .07). Overall, correlations among indicators were low with only three pairs having correlations greater than 0.50 and the others were all smaller than the recommended 0.30 inter-item correlation for PCA (Tabachnick & Fidell, 2007).

Next, we evaluated the Kaiser-Meyer-Olkin Measure (KMO) sampling adequacy score to determine factorability of the data.

Table 3
Factors extracted and loadings for *Physics Playground*.

% of variance explained by the factor	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
	16.63	13.55	10.00	9.30	9.13
Indicator					
Builds on ideas from others to improve solutions (Constructing shared knowledge)	0.83				
Proposes specific solutions (Constructing shared knowledge)	0.76				
Provides reasons to support a potential solution (Negotiation/Coordination)	0.57				
Initiates off-topic conversation (Maintaining team function)		0.68			
Joins off-topic conversation (Maintaining team function)		0.66			
Makes fun of, criticizes, or is rude to others (Negotiation/Coordination)		0.63			
Interrupts or talks over others (Constructing shared knowledge)			0.67		
Confirms understanding by asking questions/paraphrasing (Constructing shared knowledge)			0.66		
Not visibly focused on tasks and assigned roles (Maintaining team function)				0.69	
Compliments or encourages others (Maintaining team function)				−0.57	
Talks about results (Negotiation/Coordination)					0.93

Note. Loadings with absolute values smaller than 0.40 are not shown.

Although our KMO score of 0.39 is lower than the 0.60 threshold suggested by Kaiser (1974), the Bartlett's test of sphericity, a related statistic for determining reducibility, was significant ($\chi^2 = 4962.46$, $df = 105$, $p < .001$). We then assessed which indicators could be removed to improve the KMO score. We checked the diagonals of anti-image correlations and found three diagonal values that were lower than 0.40 (the recommended value is 0.50 (Yong & Pearce, 2013)). Iterative testing showed that deleting the indicator "makes an attempt after discussion" increased KMO the most. We also deleted three indicators whose communalities (the proportion of each indicator's variance that can be explained by the factors) were lower than the 0.40 cut-off (Costello & Osborne, 2005) (i.e., "talks about givens and constraints of a specific task," "asks if others have suggestions," and "asks to take action before player asks for help").

Proceeding with the 11 remaining indicators, we found that the assumptions of the PCA were largely supported. Specifically, the KMO increased to 0.58 (close to the cut-off value 0.60), Bartlett's test of sphericity was significant ($\chi^2 = 1033.83$, $df = 55$, $p < .001$), the communalities were larger than 0.40, and the diagonals of the anti-image matrix were all larger than 0.50. We retained five factors with eigenvalues larger than one. These factors explained 58.6% of the total variance (see Table 3).

It should be noted that perfect alignment, where indicators for a given facet only load onto the factor representing that facet, was not expected because of the 4-min unit of analysis where indicators for multiple facets can co-occur. And, in fact, we found that Factor 1 reflected both constructing shared knowledge and negotiation/coordination. An analysis of the indicators suggests that upon proposing a solution by building on others' ideas (constructing shared knowledge), the team member might follow up by providing reasons to convince other team members of its viability and quality (negotiation/coordination). Factor 2 also contained mixed indicators for negotiation/coordination and maintaining team function. The loadings suggest that when a team member is not friendly and/or patient with others, it is likely that he/she would not engage in tasks and would talk about irrelevant things. Factors 3, 4, and 5 were clearer and represented constructing shared knowledge, maintaining team function, and negotiation/coordination, respectively.

Correlations between facet and sub-facet scores. We examined correlations among the facet and sub-facet scores to examine the orthogonality of the model. To compute facet and sub-facet scores, we first standardized (z-scores) the segment-level raw frequency counts by coder to eliminate within-coder variance. Next, we averaged the z-scores across segments for each player to generate player-level scores. Then, after multiplying negative indicators by -1 to address reverse coding, we z-scored each indicator across players so that different indicators could be combined. Finally, we averaged the indicator scores within each sub-facet to obtain the sub-facet scores, and then averaged sub-facet scores under the same facet to obtain the facet scores.

The correlations among the three main facets (Table 4) showed that negotiation/coordination was orthogonal to both the constructing shared knowledge and maintaining team function facets (*rs* = 0.29). However, constructing shared knowledge and maintaining team function were moderately correlated ($r = 0.44$), likely because the "shares understanding of problems and solutions" sub-facet of constructing shared knowledge was significantly correlated ($r = 0.38$) with the "takes initiatives to advance collaboration process" sub-facet under maintaining team function.

As expected, sub-facets significantly correlated with their corresponding main facets (*rs* between 0.40 and 0.85). Additionally, sub-facets under the same facet were orthogonal to each other for constructing shared knowledge and negotiation/coordination (*rs* < 0.09). However, the sub-facets under maintaining team function were negatively correlated ($r = -0.57$). Another surprising result was that the "responds to others' questions/ideas" (sub-facet of negotiation/coordination) significantly correlated with "fulfills individual roles on the team" ($r = 0.64$) and negatively correlated with "takes initiatives to advance collaboration process" ($r = -0.38$) (sub-facet of maintaining team function).

We further investigated those four unexpected correlations by examining their scatterplots and found linear relationships regarding the two relatively high correlations (i.e., -0.57 and 0.64). Considering relevant indicators under those sub-facets, such relationships suggested the possibility that some students were more passively involved in collaboration, illustrated by them responding to teammates questions, yet not directly taking the initiative, even though they were focused on the task (as demonstrated

Table 4
Correlation matrix (n = 33) for facets and sub-facets—*Physics Playground* data. Data cells shaded in bold correspond to correlations between facets and sub-facets. Data cells shaded in italics correspond to correlations between sub-facets. Unshaded cells correspond to correlations between facets.

	Negotiation/ Coordination	Maintaining team function	Constructing shared knowledge		Negotiation/Coordination		Maintaining team function	
			Shares understanding of problems and solutions	Establishes common ground	Responds to others' questions/ideas	Monitors execution	Takes initiatives to advance collaboration process	Fulfills individual roles on the team
Constructing shared knowledge	.29	.44**	**.81****	**.66****	**.33**	**.13**	**.13**	**.28**
Negotiation/Coordination		.29	**.17**	**.26**	**.61****	**.85****	**-.21**	**.45****
Maintaining team function			**.50****	**.11**	**.32**	**.14**	**.40***	**.53****
Shares understanding of problems and solutions				*.08*	*.27*	*.04*	*.38**	*.10*
Establishes common ground					*.22*	*.18*	*-.26*	*.34*
Responds to others' questions/ ideas						*.09*	*-.38**	*.64***
Monitors execution							*-.01*	*.13*
Takes initiatives to advance collaboration process								*-.57***

Note. $^{*}p < .05$. $^{**}p < .01$.

by not talking about irrelevant topics and being visibly focused). Whereas the small number of participants ($n = 33$) is of some concern for the correlations, it is not unusual to find some unexpected relationships when comparing empirical data to a theoretical framework.

Correlations between facet scores and external measures. We did not expect a correlation between facets scores and the physics pretest ($M = 9.61$, $SD = 3.05$) because CPS competency should be independent of incoming domain knowledge. This was confirmed by the data ($rs < 0.25$). Although we did not include a physics posttest, we did have in-game measures of the teams' performance. That is, we computed the number of gold trophies (i.e., optimal solutions) obtained throughout gameplay, which corresponds to a deeper knowledge of the underlying physics concepts (Shute et al., 2013). The number of gold trophies ($M = 15.55$, $SD = 5.96$) obtained by a team did not correlate with constructing shared knowledge ($r = 0.03$, $p = .87$) or negotiation/coordination ($r = 0.16$, $p = .38$). However, maintaining team function did marginally correlate with the number of gold trophies ($r = 0.30$, $p = .09$), providing some evidence for the external validity of the model given a small sample of 33 students.

4.3. Discussion

We established indicators that map onto theoretical facets of CPS in the *Physics Playground* study, and conducted an initial validation where middle school students collaborated in a face-to-face setting for 3 h. Students were assigned to specific roles on the team and switched their roles after solving each game level. The results from our PCA provided partial support for our CPS competency model as the indicators clustered in a way that could be explained by our theoretical model. While the alignment was imperfect, it is not surprising to find that some extracted factors contained clear mappings to indicators while others did not, especially due to our somewhat larger unit of analysis (i.e., 4-min windows) where indicators from different factors can co-occur.

The correlational analyses involving facet and sub-facet scores indicated that the three facets were somewhat orthogonal to each other, and the sub-facets significantly correlated with their corresponding main facet. Again, facets were not expected to be perfectly orthogonal due to the 4-min unit of analysis. We also examined correlations among indicators themselves and did not find any large correlations ($rs < 0.23$) for indicators from different facets. In addition, the lack of correlation between the facet scores and prior knowledge and the positive correlation between maintaining team function and in-game performance data provided further support to our CPS model. That said, there were four unexpected correlations, which might be attributable to the small sample size ($n = 33$), and need to be examined further. Accordingly, and to test the generalizability of our model, we considered a different domain—*Minecraft Hour of Code* —where undergraduate students collaborated with each other remotely for 20 min, and their roles were fixed. This study included a larger sample and both subjective and objective outcome measures.

5. Study 2—Minecraft Hour of Code

This data was collected as part of a larger project (Stewart & D'Mello, 2018). We used code.org's Minecraft-themed *Hour of Code* (https://code.org/minecraft) as the problem-solving medium (Fig. 2). *Hour of Code* is an online resource for students grades two and above to learn basic computer programming principles in an hour. It uses a visual programming language, Blockly (https://developers.google.com/blockly/), to interlock blocks of code (such as loops). Blockly eliminates syntax errors by only interlocking syntactically correct blocks, allowing students to focus on the coding logic and programming principles. This CPS task included conditions for collaboration, such as coordination and communication between peers, individual accountability, and positive interdependence between peers (Szewkis et al., 2011). Our facets and sub-facets can map to these conditions. For example, we have facets for coordination (i.e., negotiation/coordination) and communication (i.e., constructing shared knowledge).

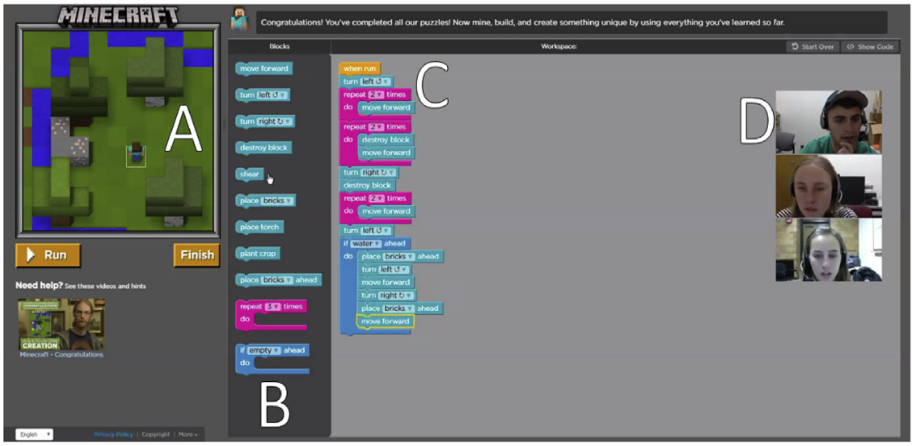

Fig. 2. Minecraft-themed *Hour of Code* from code.org. Students could watch their code run (A), choose from a code bank of possible blocks (B), generate code (C) and see their teammate's faces (D).

5.1. Method

Participants. Participants were 111 undergraduate students from a medium-sized private Midwestern university. They were compensated with course credit. Students were 63% female with a mean age of 19.4; 74.8% of the students were Caucasian, 9.9% Hispanic/Latino, 8.1% Asian, 0.9% Black, 0.9% American Indian/Native Alaskan, 2.7% "other", and 2.7% did not report ethnicity. We formed 37 groups of three based on scheduling availability. Nineteen students from ten teams indicated prior familiarity with at least one of their team members prior to participation. Students had no computer programming experience.

Procedure. Students were randomly assigned to one of three rooms, with fixed roles during the session. Each room was equipped with a computer for video conferencing and sharing via Zoom (https://zoom.us). Students could see and hear each other via a webcam and microphone. One randomly-assigned student controlled the group interaction within the environment. This student's screen was shared with the team, ensuring that everyone viewed the same content. The other two students were tasked with verbally contributing to the collaboration.

Each student individually (i.e., without screen sharing) filled out a demographic survey indicating gender, age, major, and self-reported standardized test score (ACT and/or SAT), which correlates with actual test scores (Cole & Gonyea, 2010). SAT scores were converted to ACT scores using concordance tables (ACT, 2018). Students also individually completed the validated short version of the Big Five Inventory (BFI) (Gosling, Rentfrow, & Swann, 2003) to assess extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience.

Students completed a 20-min introductory phase as a group (i.e., with screen sharing) where they solved five easy levels and viewed three accompanying videos on basic computer programming principles, such as loops and if statements. These introductory tasks familiarized students with the problem-solving environment by requiring them to build structures within the game and navigate around obstacles. Students were then asked to complete a challenging programming task, as a team, in the same *Hour of Code* environment. Students had 20 min to build a 4 × 4 brick building using at least one if statement and one repeat loop, with at least three bricks built over the water. This task had to be completed with 15 blocks of code or less. The same student who controlled the interaction with the environment during the introductory phase also controlled the interaction in this phase.

After both the introductory exercises and the challenging programming task, students individually (without screen sharing) rated their perception of the collaboration process (Barsade & Gibson, 2012; Dobao & Blum, 2013; Ku, Tseng, & Akarasriworn, 2013). Specifically, students were asked to rate the following statements on a six-point (very dissatisfied, somewhat dissatisfied, slightly dissatisfied, to slightly satisfied, somewhat satisfied, and very satisfied) scale: "I am satisfied with (1) my team's *performance* at completing the lessons, (2) how we *communicated* with each other, (3) how we mutually *cooperated* to complete the lessons, and (4) how *agreeable* my teammates are." Ratings of communication, cooperation, and agreeableness were strongly correlated (Cronbach's alpha = .89), so we averaged these three measures for each student to produce a variable reflecting subjective perceptions of the collaboration. We did this separately for the introductory exercises and challenging task.

Finally, students completed a researcher-created posttest consisting of ten multiple-choice items targeting conceptual knowledge of coding concepts (e.g., repeat loops and if statements) in the context of the Minecraft-themed Hour of Code task they just completed. The test score could range from 0% to 100%.

Data was collected over two semesters. We added a warning message 5 min before the end of the challenge task in the second semester, but everything else was identical. To account for this, we z-scored all outcomes measures by semester.

Data coding. We focused our coding efforts on the challenging programming task rather than the introductory phase. Due to technical issues, one group was excluded from the video coding. In total, we had 36, 20-min video recordings of students' interactions with the Minecraft environment, their spoken utterances, and facial expressions (see Fig. 2). We used the same indicators in Table 1 for the *Minecraft Hour of Code* data with the following adjustment. We merged "builds on others' ideas to improve solutions" into "proposes specific solutions" because we found the two to be highly similar. We also removed "asks to take action before player asks for help" because only one student had control of the interaction in Minecraft.

Through iterative processes of coding and discussion, three coders reached an agreement of 0.92 (Spearman correlation) on a random sample of three videos. Then, the three coders divided the videos among themselves and individually coded their assigned videos. In lieu of coding the entire 20-min video, we adopted a thin-slicing approach (Olsen & Finkelstein, 2017) where the coders were assigned randomly selected 90-s segments out of each 5-min time period, per video. That is, they coded four 90-s segments (360-s total) out of each 20-min video recording (30% of the data). Three groups completed the challenge task in less than 20 min, resulting in fewer segments for those particular teams. In all, we completed 417 coding sheets (note that there are three sheets per segment, one for each student).

Below is an excerpt of an actual exchange between all three participants (Players A, B, and C) during gameplay, along with the relevant indicators.

Player A: "Turn right, and then we have to move forward. Do we have to do the 'if water' again?" (Proposes specific solution + Confirms understanding by asking questions/paraphrasing)
Player C: "Yeah I think so. Cuz we'll fall in, right?" (Provides reason to support/refute a potential solution)
Player A: "Yeah that's true. Then we wanna place bedrock ahead. Oh, but don't we want to repeat that? One, two, three …" (Proposes specific solution + Asks if others have suggestions)
Player B: "And we have to move forward" (Proposes specific solution)
Player A: "Pardon?" (Confirms understanding by asking questions/paraphrasing)
Player B: "And we have to move forward to keep placing it." (Provides reason to support/refute a potential solution)

Table 5
Descriptive statistics (n = 108) for frequency counts per indicator across students—*Hour of Code*.

Indicator	Min	Max	Mean	Median	SD
Proposes specific solutions (Constructing shared knowledge)	0	22	7.66	6	4.76
Talks about givens and constraints of a specific task (Constructing shared knowledge)	0	9	2.71	2	2.10
Confirms understanding by asking questions/paraphrasing (Constructing shared knowledge)	0	15	1.93	1	2.73
Makes an attempt after discussion (Negotiation/Coordination)	0	7	1.17	0	1.86
Provides reasons to support a potential solution (Negotiation/Coordination)	0	5	1.13	1	1.28
Interrupts or talks over others as intrusion (Constructing shared knowledge)	0	8	1.07	1	1.39
Compliments or encourages others (Maintaining team function)	0	6	0.94	0	1.31
Talks about results (Negotiation/Coordination)	0	4	0.68	1	0.83
Does not respond when spoken to by others (Negotiation/Coordination)	0	2	0.31	0	0.59
Asks if others have suggestions (Maintaining team function)	0	3	0.20	0	0.53
Initiates off-topic conversation (Maintaining team function)	0	2	0.07	0	0.30
Joins off-topic conversation (Maintaining team function)	0	1	0.06	0	0.23
Brings up leaving a Challenge (Negotiation/Coordination)	0	1	0.02	0	0.14
Makes fun of, criticizes, or is rude to others (Negotiation/Coordination)	0	1	0.02	0	0.14
Visibly not focused on tasks and assigned roles (Maintaining team function)	0	0	0.00	0	0.00

5.2. Results

Descriptive Statistics. Table 5 provides descriptive statistics for each indicator across students (n = 108) during gameplay in a descending order of means. These results were computed by summing indicator counts over the thin-slice coded segments per student and then computing descriptive statistics across students. Similar to Study 1, we found that college students mostly focused on the constructing shared knowledge facet of CPS. They infrequently engaged in talking about irrelevant topics (part of the maintaining team function facet), giving up the task (part of the negotiation/coordination facet), and making fun of, or being rude to others (part of the negotiation/coordination facet). They were always visibly focused on the task (part of the maintaining team function facet), indicating high task engagement.

Principal component analysis (PCA). We removed one indicator (i.e., "visibly not focused on the gameplay") because it never occurred. We examined correlations among indicators to determine the feasibility of conducting a PCA. One indicator — "makes fun of, criticizes, or is rude to others" – did not correlate ($rs < 0.10$) with the others, so we also removed it. Only four inter-indicator correlations were greater than the recommended cut-off value of 0.30 (Tabachnick & Fidell, 2007). Nevertheless, we proceeded with an initial PCA on the remaining 13 indicators (417 × 13 matrix) using Principal Component Analysis and Varimax rotation. This yielded a KMO of 0.51, which did not quite reach the suggested value of 0.60 (Kaiser, 1974). Thus, we removed the following four indicators with communalities lower than 0.40: "talks about givens and constraints of a specific task," "provides reasons to support a solution," "does not respond when spoken by others," and "initiates off-topic conversations."

Subjecting the 417 × 9 (segments × indicators) matrix to a PCA yielded five factors with eigenvalues larger than one. The factors explained 70.7% of the variance (see Table 6). The KMO value of 0.48 was less than the recommended value of 0.6 (Kaiser, 1974), but Bartlett's test of sphericity was significant ($\chi^2 = 279.21$, $df = 36$, $p < .001$). The diagonals of the anti-image matrix were all larger than 0.40 (close to the ideal 0.50) (Yong & Pearce, 2013) and all communalities were greater than 0.50, which exceeds the cut-off value of 0.40 (Costello & Osborne, 2005).

The factor loadings shown in Table 6 indicated that Factors 1 and 4 merged indicators for negotiation/coordination and maintaining team function. Factor 1 loadings indicated a mix of engaging in off-topic conversation and suggesting giving up on the challenge, while Factor 4 reflected positive collaborations with team members complimenting each other after a successful try and encouraging each other to move on after a failure. Factor 2 reflected both constructing shared knowledge and negotiation/

Table 6
Factors extracted and loadings—*Minecraft Hour of Code* data.

% of variance explained by the factor	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
	17.74	15.63	13.71	12.42	11.18
Indicator					
Brings up giving up the challenge (Negotiation/Coordination)	0.89				
Joins off-topic conversation (Maintaining team function)	0.88				
Confirms understanding by asking questions/paraphrasing (Constructing shared knowledge)		0.83			
Makes an attempt after discussion (Negotiation/Coordination)		0.81			
Proposes specific solutions (Constructing shared knowledge)			0.78		
Interrupts or talks over others (Constructing shared knowledge)			0.76		
Compliments or encourages others (Maintaining team function)				0.74	
Talks about results (Negotiation/Coordination)				0.74	
Asks if others have suggestions (Maintaining team function)					0.97

Note. Loadings with absolute values smaller than 0.40 are not shown.

Table 7
Correlation matrix (n = 108) for facets and sub-facets—*Minecraft Hour of Code* data. Data cells shaded in bold correspond to correlations between facets and sub-facets. Data cells shaded in italics correspond to correlations between sub-facets. Unshaded cells correspond to correlations between facets.

	Negotiation/Coordination	Maintaining team function	Constructing shared knowledge		Negotiation/Coordination		Maintaining team function	
			Shares understanding of problems and solutions	Establishes common ground	Responds to others' questions/ideas	Monitors execution	Takes initiatives to advance collaboration process	Fulfills individual roles on the team
Constructing shared knowledge	.07	.08	**.72****	**.58****	**.19***	**-.03**	**.23***	**-.19***
Negotiation/Coordination		.36**	**.14**	**-.06**	**.57****	**.87****	**.33****	**.18**
Maintaining team function			**.15**	**-.06**	**.21***	**.32****	**.85****	**.60****
Shares understanding of problems and solutions				*-.14*	*.16*	*.08*	*.27***	*-.13*
Establishes common ground					.09	*-.13*	*.01*	*-.12*
Responds to others' questions/ideas						.10	*.27***	*-.02*
Monitors execution							*.24**	*.24**
Takes initiatives to advance collaboration process								*.09*

Note. $^{*}p < .05$. $^{**}p < .01$.

coordination, with indicators suggesting co-occurrence between checking their understanding and attempting a solution. Factor 3 and Factor 5 more clearly represented constructing shared knowledge and maintaining team function, respectively.

Correlations among facets and sub-facets. We computed student-level ($n = 108$) facet and sub-facet scores using the same z-score standardization and aggregation procedures used in Study 1. We additionally computed team-level scores ($n = 36$) by averaging facet and sub-facet scores across the three students in a team.

We first examined correlations (at the student level) among the facets and sub-facets (Table 7). Similar to Study 1, sub-facets were significantly correlated with their corresponding main facets (*rs* from 0.58 to 0.87). Unlike Study 1, constructing shared knowledge was orthogonal to both the maintaining team function and negotiation/coordination facets (*rs* $<$ 0.08), but the latter two facets were correlated with each other ($r = 0.36$, $p < .01$).

As we expected, the two sub-facets under each of the three main facets were uncorrelated (*rs* between −0.14 and 0.10). We found four unexpected correlations across different sub-facets, but with small magnitudes (*rs* $<$ 0.27). "Takes initiatives to advance collaboration processes" correlated with "shares understanding of problems and solutions" and "responds to others' questions/ideas." In addition, the two sub-facets of maintaining team function correlated with one sub-facet of negotiation/coordination (i.e., "monitors execution").

Study 1's unexpected correlations were not observed in Study 2—likely given the larger sample, different context (e.g., population and study design), and shorter segment-size (90-s) in Study 2. However, both studies revealed a correlation between the "takes initiative to advance collaboration processes" sub-facet (part of maintaining team function) and the "shares understanding of problem and solutions" sub-facet (part of constructing shared knowledge). The co-occurrence of these two sub-facets, even across 90-s windows, might suggest some revision to the theoretical model.

Relationships among facets and individual differences. We correlated the three main facet scores with personality traits (extraversion, agreeableness, conscientiousness, emotional stability, and openness), scholastic aptitude via ACT/SAT scores (which strongly correlates with intelligence (Frey & Detterman, 2004)), and gender, and found only one significant correlation (*rs* for the others ranged from −0.12 to 0.13) between negotiation/coordination and being male ($r = 0.25$). Thus, CPS scores as assessed by our model were independent from personality, intelligence, and to some extent gender.

Relationships between facets and external measures. We used the three main facet scores to predict students' posttest scores, subjective perceptions of team performance, and perceptions of the collaboration (i.e., average of perceptions of communication, cooperation, and agreeableness – see above).[1] We used linear mixed effects models to predict each of the outcome variables at a time, with team as a random intercept and the three facets scores as fixed effects. We included self-reported ACT/SAT scores, gender, team familiarity (i.e., whether the student reported knowing another member of the team), whether the student was the one assigned to control the interaction, and verbosity (total words spoken using the IBM Watson Speech to Text service (IBM, n.d)) as covariates.

After controlling for covariates (see Table 8), we found that students' constructing shared knowledge facet scores significantly predicted their posttest scores, but not their subjective perceptions of their teams' performance or the collaboration process. Similarly, maintaining team function was a significant and positive predictor of students' perceptions of the collaboration outcomes, whereas negotiation/coordination facet scores was a (surprisingly) significant negative predictor of this variable. Regarding the negative relationship, it may be the case that students who felt the need to negotiate with their teammates were dissatisfied with their team's progress. None of the three facets significantly predicted students' perceptions of the collaboration, but there was a non-significant ($p = .17$) positive trend for maintaining team function. With respect to the covariates, gender, knowing one's teammates, and being assigned to control the interaction were unrelated to the three outcome variables, but students with higher ACT/SAT scores scored better on the posttest and were inclined to be more critical of their team's performance.

Next we conducted team level analyses. We averaged the triads' facet scores, ACT/SAT scores, and verbosity, to obtain team-level scores. Two independent raters evaluated each team's final solution based on five solution requirements (e.g., at least one if statement, and at least three bricks over the water). The two raters reconciled any disagreements via discussion. Each criterion was worth a single point, and thus scores ranged from 0 to 5 with a mean of 2.86 ($SD = 1.06$).

We regressed (using ordinary linear regression) task score on the three main facet scores and the following covariates: ACT/SAT scores, team familiarity (whether at least one member of a team knew another), and average verbosity. We found that constructing shared knowledge was a significant positive predictor of task score ($B = 1.57$, 95% CI (0.20, 2.93), $p = .03$), but negotiation/coordination ($p = .16$) and maintaining team function ($p = .37$) were not. That said, these results should be viewed cautiously due to the small sample size for the team-level analysis ($n = 34$ after removing two teams due to missing data).

5.3. Discussion

We further validated our CPS model in Study 2 by varying the student sample, collaborative task, collaboration environment, and other pertinent factors. The PCA indicated that the indicators clustered somewhat differently compared with the results from Study 1. This suggests that the expression of CPS might be context-dependent. The correlational analysis provided some support for the orthogonality of the main facets in our CPS competency model. Importantly, we also found that the CPS facets were unrelated to individual differences in personality and intelligence, but predicted various subject-level and team-level outcomes in expected

[1] As an example, the following is the R specification for the posttest scores model: lmer (PostTest_Score ~ Constructing_Shared_Knowledge + Negotiation_Coordination + Maintaining_Team_Function + ACT/SAT + Gender [M/F] + Team_Familiarity [Y/N] + Control_Interaction [Y/N] + Total_Words_Spoken + (1| Team_ID)).

Table 8
Mixed effects models (coefficients and 95% confidence intervals in parentheses) regressing individual-level outcome variables on three main facets and covariates.

Predictors	Posttest Score	Perceptions of Outcomes	Perceptions of Collaboration
Constructing shared knowledge	0.55*	0.19	−0.04
	(0.11, 0.99)	(-0.25, 0.63)	(-0.42, 0.34)
Negotiation/Coordination	−0.28	−0.57*	−0.26
	(-.74, 0.18)	(-1.03, −.10)	(-0.67, 0.14)
Maintaining team function	0.31	0.60**	0.25
	(-0.11, 0.72)	(0.18, 1.02)	(-0.10, 0.61)
Covariates			
ACT/SAT Score	0.28**	−0.19	0.08
	(0.09, 0.47)	(-0.39, 0.00)	(-0.08, 0.25)
Gender [Male]	0.06	0.01	−0.16
	(-0.33, 0.45)	(-0.39, 0.40)	(-0.50, 0.18)
Familiarity with teammates [Yes]	−0.02	−0.02	0.01
	(-0.46, 0.50)	(-0.55, 0.52)	(-0.47, 0.50)
Controlling interaction [yes]	−0.22	−0.03	−0.11
	(-0.63, 0.18)	(-0.43, 0.36)	(-0.44, 0.23)
Verbosity	0.01	−0.08	0.11
	(-0.22, 0.20)	(-0.29, 0.13)	(-0.07, 0.29)
Intercept	−0.03	0.00	0.08
	(-0.35, 0.28)	(-0.34, 0.34)	(-0.22, 0.39)
Random Effects			
σ^2	0.71	0.68	0.48
τ_{00}	0.13_{Team}	0.23_{Team}	0.22_{Team}
ICC	0.15_{Team}	0.26_{Team}	0.31_{Team}
Marginal R^2/Conditional R^2	0.149/0.280	0.135/0.357	0.058/0.354

Note. *$p < .05$; **$p < .01$; $n = 99$ in these models after removing participants with missing data on some of these variables.

directions with the exception that the negotiation/coordination facet negatively predicted perceptions of collaborative outcomes. Overall, these results provide further support for the discriminant and predictive validity of our CPS model. And considering the differences between Study 1 and 2, our model shows promise in terms of applications across different samples and contexts.

6. Discussion

The main contribution of this paper is our proposed generalized competency model for CPS and our two validation studies. The three primary facets of our CPS model are constructing shared knowledge, negotiation/coordination, and maintaining team function. Constructing shared knowledge focuses on communicating ideas and expertise to others and establishing shared understanding among group members. Negotiation/coordination helps a team develop and execute team solutions and then revise such solutions as necessary. Maintaining team function reflects a positive group dynamic where members are conscious about being part of a team and proactively contribute to the success of the team.

In addition to the three main facets and six sub-facets of CPS concepts, our model also provides indicators for each sub-facet to aid in measurement. These indicators can be modified and re-used in different CPS contexts and across different populations. Thus, the model can be used to guide further research on the design of CPS assessments by informing CPS task design to elicit relevant evidence for the targeted skills (Scoular et al., 2017).

We also provided initial evidence for the convergence, discriminant, and predictive validity of our model. The two studies described herein differed in terms of the CPS task and interaction environment (solving puzzles in the game *Physics Playground* vs. visual programming with *Minecraft Hour of Code*), demographics (middle school students vs. college undergraduates), CPS duration (3 h versus 20 min for the main task), roles of each member (dynamic versus static), and coding scheme (coding the entire interaction versus coding thin slices). Despite these differences, both studies provide positive support for our generalized CPS model. Descriptive statistics showed that there was considerable variability in the occurrence of different indicators, with a primary focus on deriving specific solutions (part of constructing shared knowledge).

With respect to our four research questions (see Introduction), we found that (1) PCA on indicator frequencies yielded five factors that approximately mapped onto our theoretical model; (2) correlational analysis indicated the facets and sub-facets were, for the most part, orthogonal; (3) the facets showed discriminant validity in that they did not correlate with prior knowledge, intelligence, and personality; and (4) the facets showed incremental predictive validity by predicting team-level task scores and individual-level posttest scores as well as subjective perceptions of the collaboration outcome controlling for covariates.

The results are in line with current research on CPS models. Sharing knowledge, establishing shared understanding, and executing and monitoring solution plans occur frequently in CPS contexts (Andrews-Todd & Forsyth, 2018). Empirical data have shown that proactively co-constructing knowledge facilitates effective collaboration (Howard et al., 2017). Negotiation via proper argumentation to solve conflicts is the key to advancing the CPS process (Cáceres, Nussbaum, Marroquín, Gleisner, & Marquínez, 2018). Moreover, our empirical results demonstrate that our model can generalize to both face-to-face and remote/online CPS contexts, and can assess and interpret middle school and college students' CPS behaviors and skills. Our results should be generalizable to other subject areas, populations, and research designs, but that needs to be tested by further studies.

Some of the findings were less clear. For instance, several of the indicators occurred rarely, while others showed low correlation with other indicators in the same sub-facet, suggesting that minor refinements to our CPS model might be in order. In addition, the factor structures did not perfectly align with our theoretical model, with indicators associated with different facets loading onto the same factor in some cases. Similarly, correlations among facets were low (especially in Study 2), but certainly not zero. At first blush, these findings might suggest that our CPS facets are not "process pure" in that there was no perfect mapping between theoretically-specified indicators and empirically-driven factors. Regardless of the fact that theoretical models rarely are perfectly aligned with empirical data, these results suggest that multiple facets might simultaneously be at play. For example, within a 4-min (Study 1) or even 90-s (Study 2) segment, a team member might suggest a potential solution (constructing shared knowledge) (e.g., "If water, then turn right"), negotiate with team members about how to execute it (negotiation/coordination) (e.g., "And then you wanna repeat it three times, in that way we have a four by four"), and finally encourage each other if the solution fails (maintaining team function) (e.g., "Okay just reset it. We will figure this out").

Based on the empirical data, we have made some refinements to our model. For instance, the sub-facet "fulfills individual roles on the team" under the maintaining team function facet only had negative indicators in the model, and they occurred infrequently in our two studies. Therefore, we added another indicator "provides instructional support" to the sub-facet which captures instances when a student provides step-by-step instructions to direct the game interactions (e.g., if student A proposes a specific solution but student B does not fully understand, then student A directs student B to do certain actions in the game).

Second, while the CPS facets predicted the outcome measures in the expected directions, one surprising finding was that negotiation/coordination negatively predicted subjective perceptions of team performance. Negotiation is sometimes used as a means to resolve conflicts among team members, so it is plausible that those who negotiated more with their teammates had *lower* perceptions of how their team was performing. Along these lines, we did find a non-significant negative correlation between the sub-facet "responds to others' questions/ideas" with subjective perceptions of team performance ($r = -.14$, $p = .15$); but no correlation ($r = 0.006$, $p = .95$) between the sub-facet "monitors execution" and perceptions of performance. Further, the fact that none of the facet scores related to students' perceptions of the collaboration process suggest that facet scores do not tap onto the factors that students rely on when completing the survey or that the survey items were unclear. That is, in the survey, students rated their satisfaction on items such as, "*I am satisfied with how we communicated with each other*." We did not specify what cooperation or communication means and students might have their own understanding and criteria of effective collaboration.

Like all studies, ours has limitations. One limitation is the small sample sizes in both studies, with 11 teams in Study 1, and 36 teams in Study 2. Future studies should include more teams to enable better-powered team-level analyses. Another limitation pertains to our somewhat larger coding windows of 240 s and 90 s. This may prevent more stringent tests of facet orthogonality, because indicators for multiple facets can appear in those larger time windows. A more fine-grained unit of analysis, such as utterance-level coding, would alleviate this concern. Moreover, we only tested our model across two CPS tasks in laboratory settings which limits our claims of generalizability. More studies are needed to test our model across various tasks and in authentic (non-laboratory) CPS environments. For example, we need to test the same task used on multiple populations and also different CPS tasks on the same population in future studies. Additionally, we did not test our model among existing teams who have their own cultures. As such, our claims are currently limited to newly-formed teams.

There are also several directions for future work. Manual coding is time-consuming (Graesser et al., 2018), so in future studies, we plan to automate the coding of verbal indicators. Some researchers have begun investigating best practices for automated assessment using the PISA and ATC21S frameworks (e.g., Hao, Chen, Flor, Liu, & von Davier, 2017; Scoular et al., 2017; von Davier et al., 2017). Such automated coding could also be used as the basis to provide timely feedback to inform students and teachers (Awwal, Scoular, & Alom, 2017). We are currently taking a step in this direction. For instance, to accomplish utterance-level coding, we have been using IBM Watson Speech-to-Text to transcribe audio recordings. Coders were trained to code each utterance by putting the video and the transcripts side by side, so that coders could check both sources. We can then use supervised text classification techniques to learn a model that replicates the human coding from language features like we have done in related research areas (Donnelly et al., 2017; Stone et al., 2019) (note that our results are ongoing and outside the scope of this paper). We are currently pursuing work on modeling nonverbal synergy in interpersonal interaction, using machine learning classification techniques (Grafsgaard, Duran, Randall, Tao & D'Mello, 2018). We hope to extend such work to this dataset in the near future.

Future research should also address how to monitor CPS processes over time, both within a single CPS session and across multiple sessions. Do teams improve collaboration as the session progresses, and is the degree of improvement predictive of outcomes? Additionally, there is the question of how to best assess individual and group competency levels separately because individual performance depends heavily on other team members' behaviors (von Davier et al., 2017). Highly competent students, for example, tend to drive the outcome of the group performance, thus the group outcome does not necessarily reflect individual performance, especially for low-ability students (Wilczenski, Bontrager, Ventrone, & Correia, 2001).

In conclusion, to prepare students for success in school, jobs, and life in general, it is important to train them to be effective collaborative problem solvers. This requires accurately and reliably assessing their CPS skills. Our generalized model aims to serve as a guideline for CPS training and assessment given its potential to generalize across two different collaborative contexts and populations.

Declaration of interest

We wish to confirm that there are no known conflicts of interest associated with this manuscript.

Acknowledgements

This work was supported by the US National Science Foundation [award number 1660859]. We'd also like to thank Paul Nichols, Tamera McKinniss, and Aurora Dixon for their help collecting the data used in Study 1.

References

ACT (2018). *Guide to the 2018 ACT/SAT concordance.* Retrieved from https://collegereadiness.collegeboard.org/pdf/guide-2018-act-sat-concordance.pdf.

Andrews-Todd, J., & Forsyth, C. M. (2018). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in human behavior*https://doi.org/10.1016/j.chb.2018.10.025.

Andrews-Todd, J., Forsyth, C., Steinberg, J., & Rupp, A. (2018). Identifying profiles of collaborative problem solvers in an online electronics environment. *Proceedings of the 11th international conference on educational data mining* (pp. 239–245). .

Awwal, N., Scoular, C., & Alom, M. (2017). An automated system for evaluating 21st century skills using Game-based assessments. *Proceedings of 9th international conference on education and new learning technologies* (pp. 1593–1598). (Barcelona, Spain).

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *The Journal of the Learning Sciences, 9*, 403–436.

Barsade, S. G., & Gibson, D. E. (2012). Group affect: Its influence on individual and group outcomes. *Current Directions in Psychological Science, 21*, 119–123.

Bause, I. M., Brich, I. R., Wesslein, A. K., & Hesse, F. W. (2018). Using technological functions on a multi-touch table and their affordances to counteract biases and foster collaborative problem solving. *International Journal of Computer-Supported Collaborative Learning, 13*(1), 7–33. https://doi-org.proxy.lib.fsu.edu/10.1007/s11412-018-9271-4.

Cáceres, M., Nussbaum, M., Marroquín, M., Gleisner, S., & Marquínez, T. J. (2018). Building arguments: Key to collaborative scaffolding. *Interactive Learning Environments, 26*(3), 355–371.

Care, E., & Griffin, P. (2014). An approach to assessment of collaborative problem solving. *Research and Practice in Technology Enhanced Learning, 9*, 367–388.

Care, E., Scoular, C., & Griffin, P. (2016). Assessment of collaborative problem solving in education environments. *Applied Measurement in Education, 29*(4), 250–264.

Chang, C. J., Chang, M. H., Chiu, B. C., Liu, C. C., Chiang, S. H. F., Wen, C. T., & Wu, S. W. (2017). An analysis of student collaborative problem solving activities mediated by collaborative simulations. *Computers & Education, 114*, 222–235.

Chiu, M. M. (2008). Effects of argumentation on group micro-creativity: Statistical discourse analyses of algebra students' collaborative problem solving. *Contemporary Educational Psychology, 33*, 382–402.

Cole, J. S., & Gonyea, R. M. (2010). Accuracy of self-reported SAT and ACT test scores: Implications for research. *Research in Higher Education, 51*(4), 305–319.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation, 10*(7), 1–9.

Cukurova, M., Luckin, R., Millán, E., & Mavrikis, M. (2018). The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Computers & Education, 116*, 93–109. https://doi.org/10.1016/j.compedu.2017.08.007. https://search.crossref.org/?q=The+NISPI+framework%3A+Analysing+collaborative+problem-solving+from+students%27+physical+interactions.

von Davier, A. A., Hao, J., Liu, L., & Kyllonen, P. (2017). Interdisciplinary research agenda in support of assessment of collaborative problem solving: Lessons learned from developing a collaborative science assessment prototype. *Computers in Human Behavior, 76*, 631–640.

Dillenbourg, P., & Traum, D. (2006). Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *The Journal of the Learning Sciences, 15*, 121–151.

Dobao, A. F., & Blum, A. (2013). Collaborative writing in pairs and small groups: Learners' attitudes and perceptions. *System, 41*, 365–378.

Donnelly, P. J., Blanchard, N., Olney, A. M., Kelly, S., Nystrand, M., & D'Mello, S. K. (2017). Words matter: Automatic detection of questions in classroom discourse using linguistics, paralinguistics, and context. In I. Molenaar, O. X., & S. Dawson (Eds.). *Proceedings of the 7th International Learning Analytics and Knowledge Conference* (pp. 218–227). New York: ACM.

Fiore, S. M., Graesser, A., Greiff, S., Griffin, P., Gong, B., Kyllonen, P., & Soulé, H. (2017, April). *Collaborative problem solving: Considerations for the national assessment of educational progress.* Retrieved from https://nces.ed.gov/nationsreportcard/pdf/researchcenter/collaborative_problem_solving.pdf.

Fransen, J., Weinberger, A., & Kirschner, P. A. (2013). Team effectiveness and team development in CSCL. *Educational Psychologist, 48*, 9–24.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science, 15*(6), 373–378.

Funke, J., Fischer, A., & Holt, D. V. (2018). Competencies for complexity: Problem solving in the twenty-first century. In E. Care, P. Griffin, & M. Wilson (Eds.). *Assessment and teaching of 21st century skills: Research and applications* (pp. 41–53). Dordrecht: Springer.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*, 504–528.

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest, 19*(2), 59–92.

Grafsgaard, J., Duran, N., Randall, A., Tao, C., & D'Mello, S. (2018). Generative multimodal models of nonverbal synchrony ini close relationship. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition* (pp. 195–202). IEEE.

Greiff, S., Holt, D., & Funke, J. (2013). Perspectives on problem solving in cognitive research and educational assessment: Analytical, interactive, and collaborative problem solving. *Journal of Problem Solving, 5*(3), 71–91.

Griffin, P., Care, E., & McGaw, B. (2012). The changing role of education and schools. In P. Griffin, B. McGaw, & E. Care (Eds.). *Assessment and teaching of 21st century skills* (pp. 1–15). Heidelberg: Springer.

Gu, X., Chen, S., Zhu, W., & Lin, L. (2015). An intervention framework designed to develop the collaborative problem-solving skills of primary school students. *Educational Technology Research & Development, 63*, 143–159.

Hao, J., Chen, L., Flor, M., Liu, L., & von Davier, A. A. (2017). CPS-Rater: Automated sequential annotation for conversations in collaborative problem-solving activities. *ETS research report RR-17-58*. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1002/ets2.12184.

Hayashi, Y. (2018). The power of a "Maverick" in collaborative problem solving: An experimental investigation of individual perspective-taking within a group. *Cognitive Science, 42*, 69–104.

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A Framework for teachable collaborative problem solving skills. In P. Griffin, & E. Care (Eds.). *Assessment and teaching of 21st century skills: Methods and approach* (pp. 37–56). Dordrecht, the Netherlands: Springer.

Howard, C., Di Eugenio, B., Jordan, P., & Katz, S. (2017). Exploring initiative as a signal of knowledge co-construction during collaborative problem solving. *Cognitive Science, 41*(6), 1422–1449. https://doi.org/10.1111/cogs.12415.

Huang, J., Hmelo-Silver, C. E., Jordan, R., Gray, S., Frensley, T., Newman, G., et al. (2018). Scientific discourse of citizen scientists: Models as a boundary object for collaborative problem solving. *Computers in Human Behavior, 87*, 480–492. https://doi.org/10.1016/j.chb.2018.04.004.

IBM. (n.d.). IBM Watson Speech to Text service. Retrieved from https://www.ibm.com/watson/services/speech-to-text/. Accessed: 2018-05-02.

Janssen, J., & Bodemer, D. (2013). Coordinated computer-supported collaborative learning: Awareness and awareness tools. *Educational Psychologist, 48*, 40–55.

Jordan, M. E., & McDaniel, R. R., Jr. (2014). Managing uncertainty during collaborative problem solving in elementary school teams: The role of peer influence in robotics engineering activity. *The Journal of the Learning Sciences, 23*, 490–536.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31–36.

Ke, F., & Im, T. (2013). Virtual-reality-based social interaction training for children with high-functioning autism. *The Journal of Educational Research, 106*, 441–461.

Kopp, B., Hasenbein, M., & Mandl, H. (2014). Case-based learning in virtual groups–collaborative problem solving activities and learning outcomes in a virtual professional training course. *Interactive Learning Environments, 22*, 351–372.

Ku, H. Y., Tseng, H. W., & Akarasriworn, C. (2013). Collaboration factors, teamwork satisfaction, and student attitudes toward online collaborative learning. *Computers in Human Behavior, 29*, 922–929.

Lee, D., Huh, Y., & Reigeluth, C. M. (2015). Collaboration, intragroup conflict, and social skills in project-based learning. *Instructional Science, 43*, 561–590.

Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal, 25*, 95–111.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development, 50*(3), 43–59.

Moreland, R. L. (2010). Are dyads really groups? *Small Group Research, 41*, 251–267. https://doi-org.proxy.lib.fsu.edu/10.1177/1046496409358618.

National Center for Education Statistics (ED) (2017). *Collaborative problem solving skills of 15-year-olds: Results from PISA 2015. Data point. NCES 2017-249*. National Center for Education Statistics. Retrieved from https://login.proxy.lib.fsu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED577588&site=eds-live.

Nelson, L. M. (1999). Collaborative problem solving. In C. M. Reigeluth (Vol. Ed.), *Instructional design theories and models: A new paradigm of instructional theory: Vol. 2*, (pp. 241–267). New York, NY: Routledge.

Nelson, T., & Squires, V. (2017). Addressing complex challenges through adaptive leadership: A promising approach to collaborative problem solving. *Journal of Leadership Education, 16*(4), 111–123.

OECD (2017a). *PISA 2015 collaborative problem-solving framework.* Retrieved fromJuly https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf.

OECD (2017b). *PISA 2015 results (volume V): Collaborative problem solving, PISA.* Paris: OECD Publishing. https://doi.org/10.1787/9789264285521-en.

Oliveri, M. E., Lawless, R., & Molloy, H. (2017). *A literature review on collaborative problem solving for college and workforce readiness. ETS GRE® board research report. ETS GRE®-17-03. (ETS research report RR-17-06)*. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1002/ets2.12133.

Olsen, J. K., & Finkelstein, S. (2017). Through the (thin-slice) looking glass: An initial look at rapport and co-construction within peer collaboration. *Proceedings of the 12th international conference on CSCL, 511-518. Philadelphia, Pennsylvania*.

Partnership of 21st Century Learning (2016). *Framework for 21st century learning.* Retrieved from http://www.p21.org/storage/documents/docs/P21_framework_0816.pdf.

Raes, A., Schellens, T., De Wever, B., & Benoit, D. F. (2016). Promoting metacognitive regulation through collaborative problem solving on the web: When scripting does not work. *Computers in Human Behavior, 58*, 325–342.

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Ed.). *Computer supported collaborative learning* (pp. 69–97). Heidelberg: Springer.

Rosen, Y. (2015). Computer-based assessment of collaborative problem solving: Exploring the feasibility of human-to-agent approach. *International Journal of Artificial Intelligence in Education, 25*, 380–406.

Rosen, Y. (2017). Assessing students in human-to-agent settings to inform collaborative problem-solving learning. *Journal of Educational Measurement, 54*, 36–53.

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *The Journal of the Learning Sciences, 14*, 201–241.

Scoular, C., Care, E., & Hesse, F. W. (2017). Designs for operationalizing collaborative problem solving for automated assessment. *Journal of Educational Measurement, 54*, 12–35.

Shute, V. J., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games.* Cambridge, MA: MIT Press.

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in Newton's Playground. *The Journal of Educational Research, 106*, 423–430.

Siddiq, F., & Scherer, R. (2017). Revealing the processes of students' interaction with a novel collaborative problem solving task: An in-depth analysis of think-aloud protocols. *Computers in Human Behavior, 76*, 509–525. https://doi.org/10.1016/j.chb.2017.08.007. https://search.crossref.org/?q=Revealing+the+processes+of+students%27+interaction+with+a+novel+collaborative+problem+solving+task%3A+An+in-depth+analysis+of+think-aloud+protocols.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.). *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge, UK: Cambridge University Press.

Stewart, A., & D'Mello, S. K. (2018). Connecting the dots towards collaborative AIED: Linking group makeup to process to learning. *International Conference on Artificial Intelligence in Education* (pp. 545–556). Cham: Springer.

Stone, C., Quirk, A., Gardener, M., Hutt, S., Duckworth, A. L., & D'Mello, S. K. (2019). Language as thought: Using natural language processing to model noncognitive traits that predict college success. *Proceedings of the 9th International learning Analytics and Knowledge Conference* (pp. 320–329). ACM.

Strijbos, J. W., Martens, R. L., & Jochems, W. M. (2004). Designing for interaction: Six steps to designing computer-supported group-based learning. *Computers & Education, 42*, 403–424.

Szewkis, E., Nussbaum, M., Rosen, T., Abalos, J., Denardin, F., Caballero, D., et al. (2011). Collaboration within large groups in the classroom. *International Journal of Computer-Supported Collaborative Learning, 6*(4), 561–575.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon.

Teasley, S. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. B. Resnick, R. Säljö, C. Pontecorvo, & B. Burge (Eds.). *Discourse, tools and reasoning: Essays on situated cognition* (pp. 361–384). Berlin: Springer.

Teasley, S. D., Fischer, F., Weinberger, A., Stegmann, K., Dillenbourg, P., Kapur, M., et al. (2008, June). Cognitive convergence in collaborative learning. *Proceedings of the 8th international conference on ICLS: Vol. 3*, (pp. 360–367). International Society of the Learning Sciences.

Vista, A., Care, E., & Awwal, N. (2017). Visualising and examining sequential actions as behavioural paths that can be interpreted as markers of complex behaviours. *Computers in Human Behavior, 76*, 656–671.

Webb, M., & Gibson, D. (2015). Technology enhanced assessment in complex collaborative settings. *Education and Information Technologies, 20*, 675–695.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*, 71–95.

Wilczenski, F. L., Bontrager, T., Ventrone, P., & Correia, M. (2001). Observing collaborative problem-solving processes and outcomes. *Psychology in the Schools, 38*, 269–281.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*, 79–94.