Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task

Article *in* Computers in Human Behavior · October 2018

DOI: 10.1016/j.chb.2018.10.025

2 authors:

Jessica Andrews-Todd
Educational Testing Service

Carol Forsyth
Educational Testing Service

Computers in Human Behavior xxx (2018) xxx-xxx

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com

Full length article

Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task

Jessica Andrews-Todd*, Carol M. Forsyth

Educational Testing Service, Princeton, NJ, USA

ARTICLE INFO

Keywords:
Collaboration
Collaborative problem solving
Ontology
Assessment
Simulation-based assessment

ABSTRACT

Collaborative problem solving (CPS) is a complex construct comprised of skills associated with social and cognitive dimensions. The diverse set of skills within these dimensions make CPS difficult to measure. Typically, research on measuring CPS has used highly constrained environments that help narrow the problem space. In the current study, we applied the in-task assessment framework to support the exploration of CPS skills at a deep level in an open digital environment in which three students worked together to solve an electronics problem. The construct of CPS was defined in depth prior to the implementation of the environment through the development of a complex, hierarchical ontology. The features from the ontology were identified in the data and four theoretically-grounded profiles of types of collaborative problem-solvers were produced - high social/high cognitive, high social/low cognitive, low social/high cognitive, and low social/low cognitive. Results showed that students in the low social/low cognitive profile group demonstrated poorer performance than students in other profile groups. Further, having at least one high social/high cognitive member in a team facilitated performance. This study offers groundwork for future studies in measuring CPS with an approach suitable for less constrained collaborative environments.

1. Introduction

Collaborative problem solving (CPS) and related constructs are increasingly being identified as important for success in the 21st century workforce (Burrus, Jackson, Xi, & Steinberg, 2013). For example, among the top five skills considered "very important" for successful job performance are teamwork/collaboration, critical thinking/problem solving, and oral communication (Casner-Lotto & Barrington, 2006). These results are consistent with other findings identifying interpersonal and communication skills as important for current and future success (National Research Council, 2008; Trilling & Fadel, 2009). As students move through higher education and into the workforce, they will be expected to work with others to solve complex problems, make decisions, and generate novel ideas, each of which require skills associated with CPS. Educational research has shown benefits of engaging in these kinds of activities. In particular, there are positive effects of engaging in collaborative activities with outcomes associated with learning, social, emotional, and psychological well-being ((Andrews & Rapp, 2015; Dillenbourg & Traum, 2006; Gillies, 2004; Jeong & Chi, 2007; Springer, Stanne, & Donovan, 1999). Acknowledgement of the importance of CPS has motivated interest in international assessments, national assessments, and curriculum reform aimed at focusing to a greater extent on the measurement, acquisition, and development of CPS skills (Bennett & Gitomer, 2009; Fiore et al., 2017; Griffin, McGaw, & Care, 2012; National Research Council, 2011; OECD, 2013b). In the current study, we aim to demonstrate an approach for measuring CPS skills in rich and authentic contexts, which, as will be detailed in the later sections, is usually carried out by limiting the possibilities of collaboration.

2. Background

2.1. Computer environments for collaboration

A number of systems have been designed to allow students to interact with academic content in a collaborative or communicative way (Jordan, Hall, Ringenberg, Cue, & Rosé, 2007; Rowe, Shores, Mott, & Lester, 2011; Rus, D'Mello, Hu, & Graesser, 2013; Ward et al., 2013; Zapata-Rivera, Jackson, & Katz, 2015). Many of these systems have uti-

* Corresponding author. Educational Testing Service, 660 Rosedale Road, MS 16-R, Princeton, NJ 08541, USA.
Email address: jandrewstodd@ets.org (J. Andrews-Todd)

https://doi.org/10.1016/j.chb.2018.10.025
Received 12 June 2018; Received in revised form 4 October 2018; Accepted 19 October 2018
Available online xxx

lized conversational agents in which students learn by holding a conversation with artificial agents using natural language while solving a problem or answering a question. The simplest form of these environments is an interaction in which a human carries out dialogue with only one computer agent. An example of such a system includes *AutoTutor* which has an artificial agent that helps college students learn about topics such as computer literacy, physics, and scientific reasoning (Graesser et al., 2012; Nye, Graesser, & Hu, 2014). A more complex use of conversational agents incorporates two or more computer agents. Examples of these include *OperationARA* in which students discuss topics of research methodology with artificial agents (Forsyth, Graesser, Pavlik Jr, Millis, & Samei, 2014; Halpern et al., 2012; Millis et al., 2011), the Programme for International Student Assessment (PISA) 2015 CPS tasks in which test takers interact with computer agents to solve problems (OECD, 2013b), and *Betty's Brain* which utilizes a learning-by-teaching paradigm to help students learn about scientific phenomena. Specifically, students engage with Betty, a computer agent, by asking her questions and giving her quizzes provided by a mentor agent, Mr. Davis (Biswas, Jeong, Kinnebrew, Sulcer, & Roscoe, 2010). Thus, in these systems students virtually collaborate with artificial agents.

There are also environments that are quite different that utilize human to human collaboration (Dewiyanti, Brand-Gruwel, Jochems, & Broers, 2007; Hmelo-Silver, Chernobilsky, & Jordan, 2008; Liu, von Davier, Hao, Kyllonen, & Zapata-Rivera, 2015). There remain open questions regarding the extent to which human to agent and human to human interactions are comparable. Human to human collaboration provides authentic interaction that is familiar for most individuals and closely resembles the kinds of activities individuals regularly encounter in their everyday lives. These factors may lead individuals to be more engaged and motivated to communicate with teammates relative to interacting with artificial agents (Herborn, Stadler, Mustafić, & Greiff, in press). However, in some contexts (e.g., large scale assessment), human to agent collaboration may be desired to potentially increase the likelihood that individuals encounter a broad range of collaborative partners with sufficient control over confounding variables (Rosen, 2014). Given these factors, additional work is needed to better understand the potential similarities and differences between the two approaches for collaboration.

Human to human collaboration environments include systems that facilitate asynchronous or synchronous communication. For example, researchers have utilized open-source learning management systems such as Moodle that allow for asynchronous, threaded online discussion (Wise & Chiu, 2011) or massive open online courses (MOOCs) which can support collaborative and discussion-based interactions through asynchronous communication (Rosé & Ferschke, 2016). Many variations exist for synchronous communication. For example, Co-Lab is one collaborative learning environment that supports experimentation and modeling in the natural sciences. Specifically, students are able to communicate in real time via a text chat box while they experiment through simulations and remote laboratories around content in domains such as water management, greenhouse effects, and electricity (van Joolingen, de Jong, Lazonder, Savelsbergh, & Manlove, 2005). Alternatively, there are blended reality collaborative environments that allow individuals to work with and communicate with others in an immersive virtual space or a virtual world in real time. In one particular instantiation, pre-service teachers co-located in a classroom worked with other individuals participating remotely through avatars in an online, three-dimensional virtual world (Bower, Lee, & Dalgarno, 2017).

While a great deal of work has been done to develop environments in which individuals can work with others (whether human or artificial), much of the work focuses on learning, utilizing collaborative activity in a computer environment as a medium through which learning can occur. In particular, the focus is on supporting learning or evaluating learning of particular domain information as opposed to learning or evaluating collaborative skills per se. Only more recently have substantial efforts been put forth to evaluate CPS skills themselves and design computer environments to support the measurement of such skills.

2.2. Assessment and collaborative problem solving

Collaborative problem solving is defined in this paper as "the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills, and efforts to reach that solution" (OECD, 2013b, p. 6). This definition identifies a number of group processes necessary for effective collaborative problem solving such as establishing and maintaining a shared understanding, identifying and carrying out the most appropriate problem solving strategies to reach a solution, and organizing the group to ensure optimal information sharing practices. Defining the construct highlights the complexity of CPS, and this complexity makes measuring skills associated with CPS a difficult undertaking. This is because CPS involves two dimensions, a cognitive dimension associated with problem solving processes and a social dimension associated with collaboration processes. Thus, CPS involves a diverse set of skills that are researched in different disciplines such as linguistics, individual problem solving, computer-supported collaborative learning, and critical thinking (Care, Scoular, & Griffin, 2016). As such, an interdisciplinary approach is required to capture CPS.

Given the multiple interacting parts of the construct and the process-oriented nature of the construct, traditional types of assessment items such as multiple-choice questions are not suitable for measuring CPS (Davey et al., 2015). Therefore, there has been a turn to the types of computer environments already being used in learning and computer-supported collaborative learning research areas as described previously, but for assessment purposes. These environments offer promise for the measuring of constructs such as CPS because they allow individuals to demonstrate performance in complex, interactive situations. These environments further have the capability to capture all actions and discourse in the environment which provide additional sources of information about how individuals are interacting during task engagement rather than only their final product or answer choice (Honey & Hilton, 2011; Quellmalz & Pellegrino, 2009).

Despite the benefits of computer environments for measurement of complex constructs such as CPS, there are a number of challenges associated with their use. One challenge concerns operationalizing CPS skills at the level of granularity suitable for the data output from the targeted computer environment (Kerr, Andrews, & Mislevy, 2016). A second challenge concerns identifying the targeted skills in fine-grained log data generated during individuals' task engagement. There are large quantities of information in log files that are often at a small grain size (e.g., utterances, mouse clicks) and there are few guiding principles that exist for parsing, aggregating and analyzing large streams of log data from computer environments such as games and simulations (Gobert, Sao Pedro, Baker, Toto, & Montalvo, 2012). Furthermore, log data are often sparse, such that one individual may produce a large number of actions, but a given action in a particular situation may only be produced by a few individuals (Honey & Hilton, 2011; Romero, González, Ventura, del Jesús, & Herrera, 2009).

These challenges have considerably informed the design and use of assessments of CPS using computer environments, particularly for large scale assessments. There have been a number of different designs that incorporate different features such an human-agent or human-human collaboration (Care et al., 2016; Hao, Liu, von Davier, & Kyllonen, 2015; Rosen & Foltz, 2014; Rosen & Tager, 2013). Furthermore, varying projects have taken different approaches to mapping the complex interactions associated with the processes and skills of CPS (Andrews et al., 2017; Herro, Quigley, Andrews, & Delacruz, 2017; O'Neil, Chuang, & Chung, 2003; Hesse, Care, Buder, Sassenberg, & Griffin, 2015; Liu et al., 2015; OECD, 2013b). For example, PISA recently selected CPS as one of its constructs to measure in the 2015 international survey of 15-year-old students' skills and knowledge across more than three dozen countries. The PISA theoretical framework to conceptualize CPS identified three core CPS or social competencies, including establishing and maintaining shared understanding, taking appropriate action to solve the problem, and establishing and maintaining team organization. These three CPS competencies were further crossed with four problem solving or cognitive processes to form a matrix of skills. The problem solving processes included exploring and understanding, representing and formulating, planning and executing, and monitoring and reflecting. The resulting matrix included skills such as discovering perspectives and abilities of team members, identifying and describing tasks to be completed, and monitoring, providing feedback, and adapting the team organization and roles. An alternative framework developed at Educational Testing Service focused on the integrated functions of both the cognitive and social dimensions of CPS in supporting the building of knowledge. This CPS framework includes four broad CPS skills: sharing resources/ideas, assimilating and accommodating knowledge/perspective taking, regulating problem solving ideas, and maintaining positive communication (Liu et al., 2015).

In designing the PISA computer-based assessment, the team utilized a human-agent approach in which a single human interacts with one, two, or three computer agents as their team members as opposed to other humans. Human participants communicated with the agents using a predetermined set of message options throughout each problem scenario. The complexity of interactions with humans (e.g., learners may not have the opportunity to encounter a diversity of situations that may arise in collaborative settings or may be paired with others who do not collaborate well) informed this design decision, as the computer agents provide consistency and control over the social interaction (Graesser et al., 2017; OECD, 2013b). The challenge of disentangling the skills associated with CPS and identifying them in potentially large streams of log data have led other CPS projects to similarly design highly constrained environments in an effort to narrow the problem space (e.g., Chung, O'Neil, & Herl, 1999; Herborn, Mustafić, & Greiff, 2017; Hsieh & O'Neil, 2002). There has also been work in large scale assessments of CPS that has incorporated open dialogue among humans collaborating; however, many have either treated the chat messages as simply another action type and not integrated the content of the chat messages (e.g., Adams et al., 2015; Care & Griffin, 2014) or constrained the problem so that there is not much variation in the kind of actions and communications human collaborators can carry out (e.g., Hao et al., 2015; Liu et al., 2015). These design decisions do not always allow for the full scope of the CPS construct to be measured. Specifically, such design decisions do not support a detailed measurement of the actions and communication of individuals at a deep level of grain size.

3. The current study

In the current study, we have designed an open computer environment for collaboration between three humans. This design has particular advantages in providing more ecological validity than collaboration with an artificial agent, as it closely resembles the kinds of collaboration that individuals may encounter in the real world. An unconstrained or open design also allows individuals to have the opportunity to openly communicate around a complex problem as they would in many everyday contexts such as school or the workplace. The opportunity for open communication around a complex problem can further support the discovery of unexpected patterns that can inform theory and future designs of collaborative environments. The challenge however, as noted previously, is making sense of the data generated in such a complex environment to support interpretations about how individuals collaborate. To help address this challenge, we developed the in-task assessment framework (I-TAF; Kerr et al., 2016) (described in detail in section 4.3) which provides principles for operationally defining the construct according to social and cognitive dimensions of CPS and determining important features of CPS a priori through the development of an ontology.

The I-TAF approach is a departure from approaches that have often been utilized in the measurement of CPS skills. Common designs for CPS assessments have constrained the problem space to support traditional analyses. For example, as noted prior, projects have utilized human-agent collaboration for CPS measurement to provide more control over the collaborative interaction (OECD, 2013b), constrained communication to the selection of predetermined chat messages (Hsieh & O'Neil, 2002; Rosen & Foltz, 2014), or simplified the problem so that human to human communication would not contain much variation (Hao et al., 2015). With predetermined chat messages, as an example, communication essentially becomes a multiple choice problem in which psychometric approaches such as item response theory can be conducted. However, we lose out on the richness of dialogue that can occur in open collaborative environments that can facilitate important components of CPS such as negotiation around novel ideas and solutions. The advantage of I-TAF is that it provides an approach for operationalizing CPS in such a way to make it measurable in process. It further provides a detailed measurement of both actions and communication between team members in a particular CPS scenario at a deep level of grain size in ways that are theoretically grounded. As such, I-TAF facilitates theory and depth of analysis, the latter of which is often lacking in highly constrained assessments such as PISA.

The identification of defined features in this study through the I-TAF approach guided exploration into research questions that are described next. Specifically, we explored the extent to which individuals displayed the CPS skills identified a priori in an ontology. Furthermore, we investigated whether theoretically-grounded profiles developed based on the identified features relate to problem solving performance as expected based on cognitive and social psychological research or other external measures (i.e., student and teacher ratings of collaboration). Our expectation was that profiles with more of the skills associated with CPS would show a positive relationship with performance, as has been found in prior work (Andrews & Rapp, 2015; Herborn et al., 2017). Given the lack of consistency often found with self-reports (Dunlosky & Metcalfe, 2008), we did not have directional expectations for the relationships among the profiles and student and teacher ratings. Finally, we conducted an exploratory investigation into the types of team compositions displayed in the data. We were interested in exploring the constellation of profile memberships within each team and whether certain team compositions were more beneficial for performance than others. Our specific research questions are outlined next.

RQ1: To what extent do participants' discourse and actions display the skills associated with CPS identified in the ontology?
RQ2: How do profiles based on the CPS skills displayed relate to performance?
RQ3: How do the CPS skill profiles relate to teacher and student self-report ratings?
RQ4: What types of team compositions exist with respect to the CPS skill profiles? (exploratory analysis)

4. Method

4.1. Participants

Instructors teaching engineering and electronics courses in community colleges and universities and their respective classes were recruited to participate in the study. Eight classes participated with a total of 129 students randomly assembled into teams of three for each class (i.e., 43 teams). Of those who reported their race (2% were unreported), 51% were White, 6% were Asian, 7% percent were Black or African American, 2% were American Indian or Alaska Native, 10% reported being more than one race, and 2% reported Other. With respect to ethnicity, 22% of the total sample of students were Hispanic. Of the students who reported their gender (2% were unreported), 81% were males and 17% were females. The average age of the participants was 24 with a range of 16–60.

4.2. Task and measures

Prior to completing a simulation-based task on electronics concepts, students individually completed a pre-survey that asked questions about their background, including items related to their race, gender, age, highest level of education completed, parents' level of education, native language, computer use, and personality. The pre-survey also included items related to students' preference for working in groups relative to individually and the extent to which they believed collaboration was important. Instructors also completed a pre-survey in which they rated the electronics knowledge and teamwork skills for each student in their class. With students seated in a computer lab, each at their own computer, each instructor randomly assigned their students into groups of three, making sure students in the same group were not near each other in the computer lab. Though students were in the same class, they were unaware of who their partners were in the activity. Students were provided usernames (e.g., Lion, Tiger, Bear) when signing in to the activity to facilitate anonymity. Students worked together and communicated virtually through the online, computer-mediated environment described next.

The task, called the Three-Resistor Activity, deals with the relationship between current, resistance, and voltage as represented by Ohm's Law. Specifically, each student in a team worked on a separate computer that was running a fully functional simulation of portion of an electronics circuit. Each of the three circuits, each occupied by one of three teammates, were connected to form a series circuit. On each circuit board, there was a digital multimeter (DMM), two probes extending from the DMM, a resistor, a calculator, a zoom button, a chat window, and a submit button. Each of these components on the board allowed students to take measurements, view or change the resistance for their circuit, perform calculations, zoom out to view the state of (but not interact with) teammates' boards, communicate with each other, and submit answer choices. A screenshot of the task can be found in Fig. 1.

Each individual in the team had the goal of determining the appropriate value for their resistor in order to reach a specified goal voltage value. Because each of the circuits of the teammates were connected in series, any changes that one teammate makes on their board to reach their own goal voltage affects the current through the circuit and therefore the voltage drop across each circuit in the series. Thus, the students needed to coordinate their actions and work together to reach the desired goal voltage values across each circuit. There were four levels of increasing difficulty in the task. In higher levels of the task, students had additional problems to solve. Not only were they asked to reach the goal voltage values, but they were also asked to collaborate to determine the unknown resistance and voltage of an external, fourth circuit in the series that neither teammate could control. As students worked to solve these problems, all of their actions and discourse (e.g., resistor changes, measurements, calculations, chat messages, submissions) were time-stamped and logged to a database, creating the log files used for analyses.

In each of the four task levels, one variable changed in such a way that increased the difficulty of the problem. In Level 1, each student in a team was given the same goal voltage to reach and each teammate

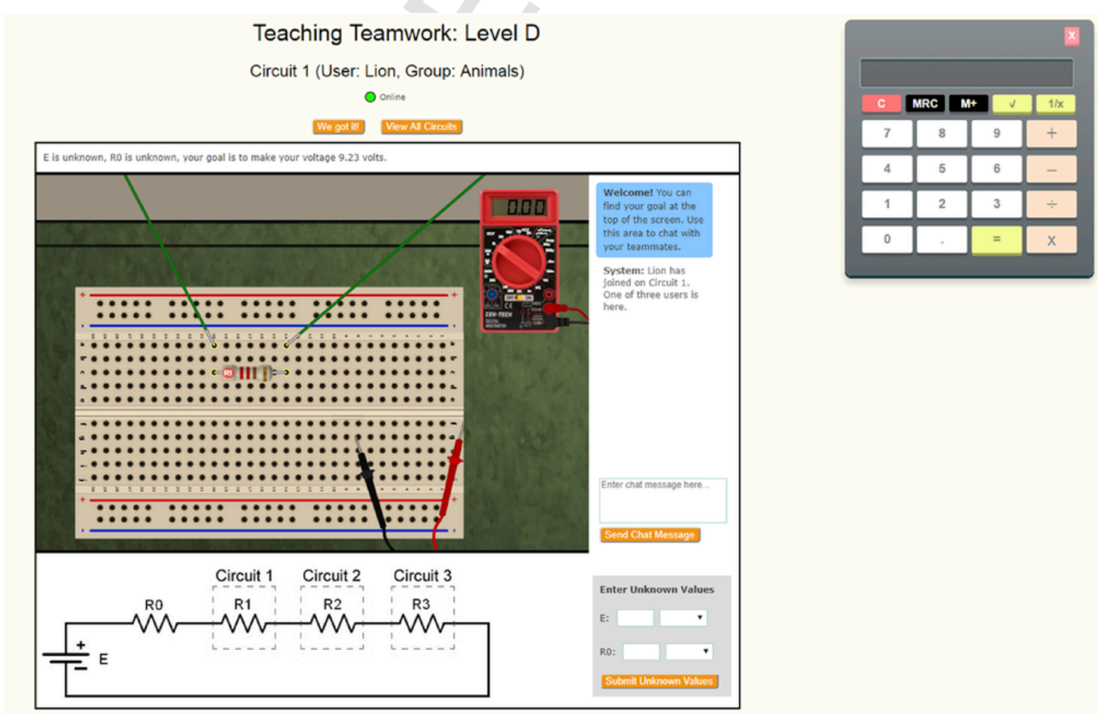

Fig. 1. Screenshot of three-resistor activity.

was also provided the values of the resistance and supply voltage of an external, fourth circuit in the series that neither teammate could control. Having the same goal values to obtain and being given all needed information about the external voltage and resistance limited the amount of information that needed to be shared among teammates. However, teammates were unaware that they had the same goal voltages so discussion was necessary to uncover this information before solving the problem. In Level 2, students were also given the values of the resistance and supply voltage of the external, fourth circuit; however, each teammate now received a different goal voltage that needed to be reached. In Level 3, students were again given different goal voltages to reach and also the resistance of the external, fourth circuit in the series, but they were asked to discover the value and unit for the supply voltage of the external circuit. In Level 4, students were asked to discover and input the value and unit for both the resistance and the supply voltage of the external, fourth circuit in the series as well as reach specified and different goal voltage values for the circuits of each teammate. Table 1 provides an overview of the characteristics of each task level.

4.3. Assessment framework

We utilized the in-task assessment framework (I-TAF) (Kerr et al., 2016) to guide the conceptualization of skills associated with the CPS construct and identification of the targeted skills in the data. I-TAF is informed by evidence-centered design (ECD) (Mislevy, Steinberg, & Almond, 2003) and specifically provides support for instantiating the student model, task model, and evidence model of ECD which are often challenging to carry out with constructs such as CPS in computer environments due to the complexity of the construct and the small grain size of the data. The student model defines the variables associated with the targeted knowledge, skills, and abilities. The task model describes task features or affordances available in the task needed to obtain evidence of targeted student model variables. The evidence model bridges the student and task models by identifying the salient features of what learners do or say and specifying how the identified evidence will be used to make claims about what learners know or can do. I-TAF provides guiding principles for how to carry out each of these models. Specifically, I-TAF reconceptualizes what it means to define the construct of interest, extract evidence of the targeted construct, and link the evidence back to the targeted construct when the evidence is fine-grained log data as opposed to standard item response data. One notable component of I-TAF we utilized was its process for developing the student model and connecting it to components of the evidence model. I-TAF advocates the use of an ontology (similar to a concept map) which lays out the targeted construct in a principled way. The ontology allows for representing the complexity of CPS by laying out all the components of CPS and their relationships at the level of granularity represented in the data (log files in our case). The ontology can help build assessment arguments by informing aspects of the evidence model. Specifically, the ontology shows the key evidence elements that need to be identified in the data to make the kinds of inferences about learners outlined in the student model.

Table 1
Overview of task levels.

Task Level	External Voltage (E)	External Resistance (R0)	Goal Voltages
1	Known by all teammates	Known by all teammates	Same for all teammates
2	Known by all teammates	Known by all teammates	Different for each teammate
3	Unknown by teammates	Known by all teammates	Different for each teammate
4	Unknown by teammates	Unknown by teammates	Different for each teammate

4.4. CPS ontology

In order to conceptualize the CPS construct and determine what skills would be included in our analyses, we created a student model in the form of an ontology. An ontology provides a theory-driven representation of the skills associated with the construct and their relationships, linking the skills to observable behaviors present in the electronics task that could provide evidence of each skill. The development of the ontology was based on an extensive literature review of frameworks associated with CPS and work from related areas such as individual problem solving, computer-supported collaborative learning, linguistics, and organizational psychology (Hesse et al., 2015; Liu et al., 2015; Meier, Spada, & Rummel, 2007; Morgan, Salas, & Glickman, 1993; OECD, 2013b; O'Neil, Chung, & Brown, 1995; Spada, Meier, Rummel, & Hauser, 2005). The top portion of the ontology provides a generalizable construct definition of CPS (e.g., sharing information as one skill of CPS) that can be utilized in other work seeking to measure CPS or other related constructs. The lower levels of the ontology provide more specificity, describing CPS within a domain (e.g., sharing status updates) and then within the specific task context being used (e.g., sharing the status of resistance in a circuit). Links between the layers in the ontology describe how behaviors at lower levels of the ontology can be combined to make inferences about high level cognitive behaviors. The structure of a portion of the ontology can be found in Fig. 2, with nodes corresponding to high-level CPS skills, subskills, features that can be extracted about each subskill given certain behaviors in the task context, and observable variables that can be inferred from the features, along with links indicating the relationships between the nodes in the ontology.

The full ontology included nine high-level CPS skills across social and cognitive dimensions and subskills that correspond to each high-level CPS skill. The social dimension included four CPS skills, namely, maintaining communication, sharing information, establishing shared understanding, and negotiating. The cognitive dimension included five CPS skills, namely, exploring and understanding, representing and formulating, planning, executing, and monitoring. Next, we describe each of the nine CPS skills in turn.

In the social dimension, maintaining communication corresponds to content-irrelevant social-oriented communications (Lipponen, 2000; Lipponen, Rahikainen, Lallimo, & Hakkarainen, 2003; Liu et al., 2015). This includes rapport building communication (e.g., using chat emoticons, praising teammates, greeting teammates, apologizing), off-topic communication (e.g., "I should have drank coffee this morning"), and inappropriate communication (e.g., curse words). Sharing information refers to content-relevant information used in the service of solving the problem. Verbalizing one's thoughts to team members is an important part of the collaborative process and has been shown to be important for team success (e.g., van Boxtel, van der Linden, & Kanselaar, 2000; Webb, 1991). This includes sharing one's own information (e.g., sharing what circuit board one is on, sharing one's goal voltage value, sharing resistance values on one's board), sharing task or resource information (e.g., sharing where the zoom button is located), or sharing the state of one's understanding (e.g., metacognitive statements such as "I don't get it"). Establishing shared understanding corresponds to communications used in the service of learning the perspective of others and establishing that what has been said is understood. This component of CPS comes from linguistics and communication theory (Clark, 1996; Clark & Brennan, 1991). This includes requesting information from teammates (e.g., "what is your resistance?" "what values do we need?"), providing responses that indicate comprehension or lack of comprehension of a contribution (e.g., "ok," "I hear you," or requests

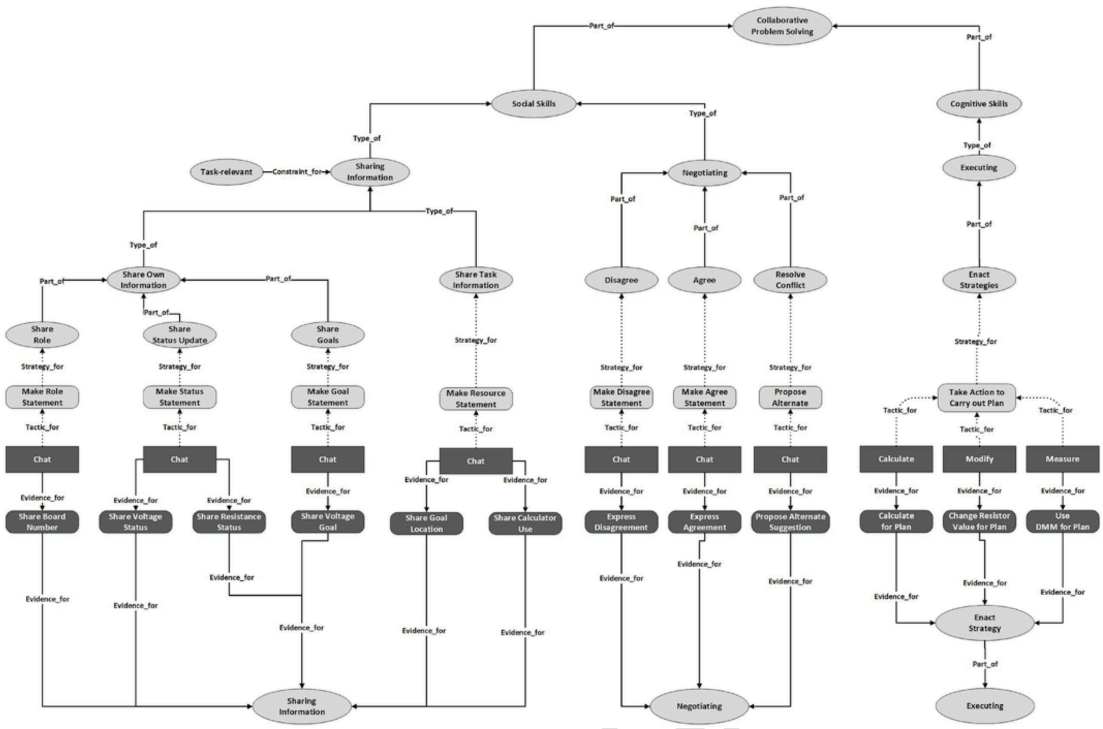

Fig. 2. Collaborative Problem Solving Ontology Fragment. The full ontology can be found in the online supplemental materials.

for clarification), and making repairs when issues in shared understanding arise. Negotiating refers to communication in the service of determining if conflicts exist and resolving the conflicts. A number of research areas such as collaborative learning and organizational psychology consider the importance of negotiation in group contexts for outcomes such as learning, decision making, and problem solving (e.g., Brodt & Thompson, 2001; Hesse et al., 2015; Kirschner, Paas, & Kirschner, 2009). Components of this skill includes expressing agreement (e.g., "you are right"), expressing disagreement (e.g., "that's not right"), and attempting to reach a compromise.

The cognitive dimensions include problem solving processes, and much of the theory from these components of CPS come from the individual problem solving literature which has a comprehensive body of work extending back decades. In the cognitive dimension, exploring and understanding corresponds to actions taken to build a mental representation of components of the problem (Frensch & Funke, 1995; OECD, 2013a). This skill includes exploring the problem space (e.g., spinning the dial on the digital multimeter) and demonstrating understanding of given information and information obtained during exploration. Representing and formulating corresponds to actions and communication used in the service of building a coherent mental representation of the whole problem space (Mayer & Wittrock, 1996; OECD, 2013a; VanLehn, 1996). This skill includes providing a representation of the problem whether it be verbal or graphical (e.g., "this is a series circuit") and formulating hypotheses. Planning, an important component for problem solving in reaching optimal results, refers to communication used to develop a plan or strategy for solving the problem (Cohen, 1989; Hesse et al., 2015; OECD, 2013a; Wirth & Klieme, 2003). This skill includes determining goals (e.g., "we need 6.69 V across our resistors), setting sub-goals or steps for carrying out a plan, and developing and revising strategies (e.g., "ok we set our values to R and find the current"). Executing corresponds to actions and communication used in the service of carrying out a plan (OECD, 2013a; Wirth & Klieme, 2003). This includes taking steps to carry out the plan (e.g., changing one's voltage value to the voltage suggested by a teammate), making suggestions for actions teammates should take to carry out the plan (e.g., "adjust yours to 300 Ω"), and communicating what steps one is taking to carry out the plan (e.g., "I am going a little lower and then readjusting"). Monitoring refers to actions and communication used to monitor progress toward the goal and monitor the team organization (OECD, 2013a, 2013b; O'Neil, 1999). This includes actions to monitor the team's progress relative the goal (e.g., clicking the submit button to receive feedback about success in solving the problem), communicating progress toward the goal to teammates (e.g., "I've got my goal voltage"), checking the progress and status of teammates (e.g., "Where is Lion?" or clicking the zoom button to view the state of a teammate's circuit board), and determining whether teammates are following the rules of engagement or any assigned roles or prompting them to do so (e.g., "Let's get a move on Lion").

5. Analyses and results

We investigated the nature of the theoretically grounded features identified in the ontology. As the reader may recall, many of these features are chat-based and required qualitative coding to make sense out of the low-level collaborative discourse. Thus, the results include the following sections: (5.1) Qualitative Coding, (5.2) CPS Skill Frequencies, (5.3) CPS Skill Profiles, (5.4) CPS Skill Profiles and Performance, (5.5) CPS Skill Profiles and Ratings, and (5.6) Exploratory Analysis: Team Composition.

5.1. Qualitative coding

The CPS ontology was used to create an extensive rubric for raters to code the log files, specifically identifying high-level CPS skills based on students' low-level discourse and actions during task engagement. The rubric outlined high-level CPS skills, subskills, their definitions, example behaviors from the log files indicative of each skill, and the action types associated with each skill (e.g., chat, measurement, calculation, submit). Evidence for two of the nine high-level CPS skills occurred in both action and chats (i.e., monitoring and executing) so in the rubric we split these into separate action and chat skills (i.e., monitoring actions, monitoring chats, executing actions, executing chats). This resulted in 11 total CPS skills for coding. These included maintaining communication, sharing information, establishing shared understanding, negotiating, exploring and understanding, representing and formulating, planning, executing actions, executing chats, monitoring actions, and monitoring chats.

Two raters coded the content of students' discourse and their actions recorded in the log data for the presence of 11 CPS skills. Coding was completed at the level of each log file event (i.e., each action submission or chat submission) and each event could only receive one code. After training, independent raters double coded a random sample of 20% of all events in the data. The interrater reliability between the raters was found to be Kappa = .84. All discrepancies among raters were discussed to reach consensus on assigning a final code. After finding that acceptable interrater reliability was reached on the subset of the data (Kappa $>.60$; Landis & Koch, 1977), the remaining data were coded by a single rater. There was a total of 20,947 log file events that were coded. After coding the data, we were then able to inspect the frequencies of each of the 11 CPS skills.

5.2. Frequencies of CPS skills (RQ1)

The over-arching CPS skills consisted of five skills corresponding to the cognitive dimension of CPS and four skills corresponding to the social dimension of CPS. Specifically, the cognitive skills included *exploring and understanding*, *representing and formulating*, and *planning*. In addition, the cognitive dimension included *executing* and *monitoring* which could both be measured via chat or actions. Therefore, when analyzing the frequencies, there were 7 measures associated with the cognitive skills as *executing* and *monitoring* were reported for chat- and action-based versions of the skill. The CPS skills corresponding to the social dimension were all measured in chat. These skills included *maintaining communication*, *sharing information*, *establishing shared understanding* and *negotiating.* Analysis of the frequencies of all 11 CPS skills revealed that participants demonstrated evidence of all identified CPS skills. The most commonly occurring CPS skills included *executing actions* (7719), *monitoring actions* (5757), and *exploring and understanding* (2425), all of which correspond to the cognitive dimension. However, the next two most commonly occurring CPS skills belonged to the social dimension and included *sharing information* (1670) and *establishing shared understanding* (1091). Thus, both the cognitive and social dimensions were represented in the most frequently displayed CPS skills within this task. A full list of frequencies along with an example for each skill can be found in Table 2.

We acknowledge that certain measures corresponding to skills discovered in the chat are sparse in nature including *representing and formulating* (84), *negotiation* (406), and *maintaining communication* (564). Sparse measures can sometimes have an unfair influence on outcome measures. However, for the purposes of the subsequent analysis, these skills are included in an aggregate measure for the social and cognitive dimensions to have a complete profile of each dimension, which reduces the impact of any single skill.

5.3. CPS skill profiles (RQ2)

After confirming students displayed all CPS skills and aggregating for distinct social and cognitive dimensions, participants were grouped based on display of social and cognitive CPS skills. Specifically, the frequencies of CPS skills displayed in the social dimension and the frequencies of CPS skills displayed in the cognitive dimension were first rank-ordered for each respective dimension. Next, the rank-ordered social dimension was split by the median into roughly two equal groups (i.e., high social vs. low social CPS skills). The same process was conducted on the cognitive dimension – the rank-ordered frequencies of

Table 2
Frequencies of CPS skills.

Dimension	CPS Skill	Measure	*Frequency*	Example
Social	Maintaining Communication	chat	564	"nice job guys"
	Sharing Information	chat	1670	"my goal voltage is 1.75"
	Establishing Shared Understanding	chat	1091	"Tell me what you got"
	Negotiation	chat	406	"no if you change the resistance you change the current"
Cognitive	Exploring and Understanding	action	2425	Changing the DMM settings several times in seconds
	Representing and Formulating	chat	84	"I think if we all got 150 Ω we'll get it"
	Planning	chat	328	"how about you both go slightly over your marks and I'll drop mine"
	Executing	action	7719	Completing a resistor change suggested by a teammate
	Executing	chat	487	"Lion, lower your resistance"
	Monitoring	action	5757	Using the Zoom button to see others' circuit boards
	Monitoring	chat	416	"Tiger, are you with us?"

cognitive CPS skills were split by the median producing roughly two equal groups (i.e., high cognitive vs. low cognitive CPS skills). In total, four separate groups were created (i.e., high social CPS skills, low social CPS skills, high cognitive CPS skills, low cognitive CPS skills). We next combined these groups to create four CPS profiles that included both social and cognitive dimensions together in a 2×2 matrix (i.e., high cognitive/high social, high cognitive/low social, low cognitive/high social, and low cognitive/low social). Students were then placed into one of four CPS profile groups based on the frequency with which they displayed social CPS skills and cognitive CPS skills during interaction in the electronics task. For example, a student who displayed a frequency of social CPS skills above the median cutoff in the social dimension and a frequency of cognitive CPS skills above the median cutoff in the cognitive dimension was placed in the high social/high cognitive CPS profile group. The frequencies of each CPS skill for each profile are shown in Table 3. The reader may notice that there is an unequal representation among skill frequencies such that some skills (e.g., executing actions) occur more frequently than others (e.g., negotiating). This is because the design of the electronics task necessitates higher frequencies of some skills over others (e.g., there are a number of actions needed to execute the task whereas negotiating is not an absolute requirement to successfully solve the problem). Thus, any attempts to standardize the frequencies of the CPS skills may artificially inflate sparse skills.

5.4. CPS Skill Profiles and Performance (RQ2)

Performance comparisons among the four CPS skill profiles corresponding to the 2×2 social and cognitive dimension matrix were next investigated. To be clear, this investigation focused on problem solving performance for the task rather than learning gains as a result of students' engagement in the task. Performance was operationalized as the number of task levels completed. The reader may recall that participants could interact with up to 4 levels of the electronics simulation-based task, with each level increasing in difficulty. Participants needed

Table 3
CPS skill frequencies for each profile type.

Dimension	CPS Skill Profile				
	CPS Skill	Low Social/Low Cognitive	Low Social/High Cognitive	High Social/Low Cognitive	High Social/High Cognitive
	N	36	27	27	39
Social	Maintaining Communication	106	57	150	251
	Sharing Information	131	177	516	846
	Establishing Shared Understanding	145	104	293	549
	Negotiating	39	29	102	236
Cognitive	Exploring and Understanding	364	967	343	751
	Representing and Formulating	5	6	17	56
	Planning	51	49	79	149
	Executing (Action)	1062	2666	1062	2929
	Executing (Chat)	35	51	120	281
	Monitoring (Action)	439	1729	250	3339
	Monitoring (Chat)	41	33	115	227

to successfully complete one level before continuing to the next level. All participants spent roughly equivalent amounts of time interacting with the task.

A Kruskal-Wallis test was conducted to investigate the relationship between the CPS skill profiles and problem solving performance (i.e., number of task levels completed). Specifically, the CPS profile membership (i.e., high cognitive/high social, high cognitive/low social, low cognitive/high social, and low cognitive/low social) was entered as the independent variable and the number of task levels completed was the dependent variable. Results from the Kruskal-Wallis test revealed a significant relationship between student CPS skill profile and performance, (X^2 (3129) = 57.28, $p < .001$, partial $\eta^2 = 0.316$), accounting for nearly 32% of the variance. Mean ranks for which higher numbers correspond to more task levels completed were as follows: low social/low cognitive (28.21), low social/high cognitive (64.28), high social/low cognitive (84.00), and high social/high cognitive (86.31). As expected, post-hoc tests revealed that students in the low social/low cognitive profile demonstrated poorer performance than students in the low social/high cognitive ($p = .001$), high social/low cognitive ($p < .001$), and high social/high cognitive ($p < .001$) groups.

5.5. *CPS Skill Profiles and Ratings (RQ3)*

5.5.1. *Student ratings*

In a pre-survey, student participants were asked two questions related to collaboration. One question asked how important students felt collaboration was in the real world and the second question asked about students' preferences for working alone versus working with others. Kruskal-Wallis tests revealed that the CPS skill profiles did not significantly correlate with students' self-report on either question: working alone vs. working with others, ($p = .404$), the importance of collaboration in the real world ($p = .285$). Specifically, on the question about working alone vs. with others, the mean ranks were as follows (higher numbers indicate stronger preference to work alone): low social/low cognitive (53.97), low social/high cognitive (56.00), high social/low cognitive (65.78), and high social/high cognitive (61.11). For the question regarding beliefs about the importance of collaboration in the real world, the mean ranks were as follows (higher number indicate higher importance for collaboration): low social/low cognitive (64.27), low social/high cognitive (71.94), high social/low cognitive (53.88), and high social/high cognitive (65.00).

5.5.2. *Teacher ratings*

Instructors also completed a pre-survey in which they answered three questions corresponding to ratings of each student's teamwork skills, their expectations for how well each student would collaborate with others, and how well each student typically performs on electronics tasks when working alone.

Results from Kruskal-Wallis tests revealed that teachers' ratings of the students' teamwork skills was significantly related to the four CPS skill profiles [$X^2(3,\ 123) = 9.638$, $p < .022$, partial $\eta^2 = 0.075$], with students in the high social/low cognitive profile having the highest mean rank (77.22) and students in the low social/high cognitive profile having the lowest mean rank (47.56). There was little difference between the ratings for the students in the low social/low cognitive profile (62.17) and students in the high social/high cognitive profile (62.50). Indeed, the only statistically significant difference on post-hoc tests was between the high social/low cognitive profile and the low social/high cognitive profile ($p = .012$). Similarly, the question regarding teachers' expectation for how well students would collaborate showed a significant relationship with the CPS skill profiles [$X^2(3123) = 10.04$, $p = .018$, partial $\eta^2 = 0.078$]. Once again students in the high social/low cognitive profile had the highest mean rank (75.48) and students in the low social/high cognitive profile had the lowest mean rank (46.10). The low social/low cognitive and high social/high cognitive profiles each had the same mean rank (63.40). Post hoc tests revealed that the only significant pairwise comparison was between the low social/high cognitive and the high social/low cognitive profiles ($p = .011$). There was no statistically significant relationship for the question concerning how well students perform on electronics tasks when working alone ($p = .117$).

5.6. *Exploratory analysis: team composition (RQ4)*

There was a total of 15 types of team compositions observed in the data with variation in the types of CPS skill profiles composed in each. The most common team composition included a homogenous team in which all three teammates were in the high social/high cognitive profile ($N = 6$). The next most frequently occurring team composition ($N = 5$) included a composition in which two teammates were in the high social/low cognitive profile and one teammate was in the high social/high cognitive profile. Because the total number of teams (43) was too low for comparisons between 15 team compositions, we used two methods to aggregate team compositions. First, we investigated performance differences between teams that had at least one high social/high cognitive team member and teams that had no team members characterized as high social/high cognitive. Second, we explored performance differences among team compositions based on heterogeneity.

5.6.1. *Teams with and without high social/high cognitive members*

Two comparison groups were created – teams that had one or more members characterized as high social/high cognitive and teams that did not have any members characterized as high social/high cognitive. In total, there were 22 teams that had at least one high social/high cognitive member, and 21 teams that did not include any high social/high cognitive members. It is important to note that among the teams with at least one high social/high cognitive member there was high variability in the number of team members characterized as high social/high cognitive. Specifically, 6 out of 22 teams had all high social/high cognitive members, 5 out of 22 teams had two members that were high social/high cognitive, and 11 out of 22 teams had only one member that was high social/high cognitive. We next explored performance differences (i.e., number of task levels completed) between teams with at least one high social/high cognitive member and teams with no high social/high cognitive member. A Kruskal-Wallis test revealed a statistically significant difference between these two types of team compositions in regards to the number of task levels completed, [$X^2(1, 41) = 10.65$, $p < .001$ partial $\eta^2 = 0.21$). Teams with at least one high social/high cognitive member out-performed teams without such a team member (mean ranks = 27.95, 15.76, respectively).

5.6.2. Team composition heterogeneity

For a second comparison, teams were grouped into three types of compositions – heterogeneous, semi-heterogeneous, and homogeneous. Heterogeneous teams included teams in which all three teammates were in different CPS skill profiles. Semi-heterogeneous teams were teams in which two teammates were a part of the same CPS skill profile and one teammate was a part of a different CPS skill profile. Homogenous teams were considered teams in which all teammates were a part of the same CPS skill profile. We found that there were 2 heterogeneous teams, 26 semi-heterogeneous teams, and 15 homogenous teams. In this particular context, the patterns suggest that there is not a relationship between the types of team compositions and performance on the task (i.e., number of task levels completed), but the sample size was not large enough within all cells for statistical analysis. A full count of the three types of team compositions and the performance level achieved within the simulation-based task can be seen in Table 4.

6. Discussion

The measurement of skills associated with CPS can be a challenging undertaking given the complexity of the construct and the difficulty in identifying skills associated with CPS in large streams of log files in computer environments. One means of reducing the complexity of measuring CPS skills has been to considerably constrain the problem, the environment, or the mode of collaboration. These design decisions can sometimes affect the scope of what can be explored with CPS, as it can hinder the examination of detailed actions and communication among individuals in complex situations and consequently the possibility for unexpected patterns or emergent roles that may help inform theory and future design for collaborative environments.

In the current study, we explored collaborative problem solving in an open online environment that used human to human collaboration. We wanted to explore the extent to which the development of a comprehensive, hierarchical ontology to lay out the construct and problem space in a principled way would help us capture complex CPS skills in an open environment in which students could use free-form chat. We found that the CPS ontology supported the identification of CPS skills and at a level of granularity that could be identified in the fine-grained log data generated by the electronics simulation-based task. In particular, participants demonstrated evidence of each of the CPS skills outlined in the ontology across the social and cognitive dimensions of CPS. Some skills occurred more frequently than others, namely, executing actions, monitoring actions, sharing information, and establishing shared understanding. These results are in line with prior work in measurement of CPS skills showing skills such as sharing information (Hao, Liu, von Davier, Kyllonen, & Kitchen, 2016), establishing shared understanding (Rosen, 2014), and executing the problem (Rosen & Foltz, 2014) displayed in CPS assessments. The high frequency of monitoring actions is likely a function of the design of the electronics task. The task provides specific affordances (i.e., a zoom button) that allow teammates to monitor the state of each other's boards. We are encouraged by the findings that the CPS skills from our ontology occur in our data because according to previous theoretical and empirical work, these components are important for CPS (e.g., Andrews-Todd, Forsyth, Steinberg, & Rupp, 2018; Hesse et al., 2015; Nouri, Åkerfeldt, Fors, & Selander, 2017; OECD, 2013b).

Using the frequencies of the CPS skills identified for each individual participant, we were able to construct profiles of types of collaborative problem solvers. The profiles corresponded to the social and cognitive dimensions of CPS outlined in the ontology and included a low social/low cognitive profile, a low social/high cognitive profile, a high social/low cognitive profile, and a high social/high cognitive profile. These profiles were used to answer research questions about the nature of the relationship between different types of collaborative problem solvers and performance on the electronics simulation-based task as well as student and teacher ratings concerning aspects of collaboration. Results revealed that the CPS skill profiles had a significant relationship with performance in expected directions. This finding provides some validity to our approach, as the demonstration of collaboration skills are often associated with positive performance outcomes (e.g., Andrews & Rapp, 2015). Similar findings, with profiles fairly consistent with the social and cognitive dimensions of CPS used in the current study, were demonstrated in a recent study. In particular, Herborn et al. (2017) developed profiles named passive low-performing (non-)collaborators which roughly correspond to a low social/low cognitive group, active high-performing collaborators which roughly correspond to a high social/high cognitive group, and compensating collaborators which roughly correspond to a low cognitive/high social group. The authors found that active high-performing collaborators showed the best performance on a number of cognitive performance indicators and passive low-performing (non-) collaborators showed the lowest performance.

With respect to the ratings, the students' ratings of their own collaborative preferences was not significantly related to the CPS skill profiles. One potential explanation for this finding is that students have notorious difficulties in rating metacognition (Dunlosky & Metcalfe, 2008), and this may apply to ratings associated with collaboration as well. Additionally, it is possible that what the students prefer do not actually align with what they are good at or students may have just selected socially acceptable responses. Another potential explanation is that the questions were just not well aligned with the type of measures used to profile participants in this manner, as we have seen significant correlations with student ratings using other types of profiles of collaborative problem solvers (Andrews-Todd et al., 2018). In contrast, the teacher ratings associated with collaboration were significantly related to CPS skill profiles. Specifically, participants in the high social/low cognitive profile were rated the highest for teamwork skills while participants in the low social/high cognitive profile were rated the lowest for teamwork skills. Teachers' expectations for how well students would collaborate showed a similar pattern in which teachers had

Table 4
Count of varying team composition success on each task level.

Team Composition Type	*N*	Number of Task Levels Completed				
		0	1	2	3	4
Homogenous	15	3	1	4	5	2
Semi-heterogeneous	26	6	6	1	7	6
Heterogeneous	2	0	1	0	0	1

higher expectation that participants in the high social/low cognitive profile would collaborate well as compared to participants in the low social/high cognitive profile. One potential explanation for this finding is that teachers may have considered the social aspects of teamwork when making judgments about their students.

In an exploratory analysis, the team compositions based on the CPS skill profiles were examined. We found a number of combinations of varying team compositions. As a result, we narrowed the number of team composition types by grouping the types of compositions based on heterogeneity (heterogeneous, semi-heterogeneous, and homogenous) and whether teams had at least one or no member characterized as high social/high cognitive. While patterns of performance (task levels completed) did not appear to differ across the heterogeneity team composition types, cell sizes were too small to conduct inferential statistics. We did, however, find performance differences between teams that had at least one high social/high cognitive member and teams with no high social/high cognitive members. Specifically, teams with at least one high social/high cognitive member performed better than teams that did not have any high social/high cognitive members on the electronics task. These results suggest that having at least one team member integrating both social and cognitive components of CPS at high levels may be important to help guide a team to successful performance. This is consistent with work in areas such as collaborative learning that show having team members who demonstrate beneficial collaborative behaviors contribute positively to team success (e.g., Barron, 2003; Chan, 2001).

6.1. Limitations and future work

There were some limitations in this study that are worth acknowledging. One limitation concerns our sample size, particularly given the number of CPS skills we were trying to explore. Additionally, our external measures were based on subjective ratings so participant interpretations of the items may have been different from that of the research team. Further, although individuals needed to reach their own goal voltage on their respective circuit boards in the electronics task, team members could contribute by providing assistance on how to do so. However, given the interdependent nature of the task, complete exclusions of group dynamics are virtually impossible. In follow-up research, we are currently collecting data on a larger scale than the current study and we have identified, revised, and created additional external measures that we believe can help build a better validation argument for profiles generated in the future. Secondary follow-up studies additionally incorporate tasks in different content domains (i.e., mathematics and physics) and with a different population of students (i.e., middle school students). These studies can allow us to investigate the generalizability of the CPS ontology in other contexts and domains and also how the design of different tasks affect the patterns of results that emerge and types of relationships we see with performance.

The time consuming nature of the human coding used in the current study can make this kind of work difficult to scale up so it is certainly a necessity to explore natural language processing techniques that can reduce time in identifying the CPS skills in complex chat data. We have begun exploring the possibility of automating the coding of the chat data using machine learning algorithms. As we get more data in our follow-up studies, we will be able to test out the algorithms and compare them to the human coding.

In the current study, the CPS profiles were based on the frequency of CPS skills displayed. Another option for future work is to develop profiles based on the variety of CPS skills displayed (e.g., whether an individual displayed three, four, or five CPS skills). This could provide additional insights into which particular CPS skills are most effective and why. Further work can also go beyond measurement of CPS skills to explore the relationships between specific CPS skills and the extent to which these skills are predictive of learning.

7. Conclusions

The approach and results of this study can have important implications for the measurement of CPS skills. Using principles of I-TAF we developed a comprehensive, hierarchical ontology that lays out the construct of CPS in a principled way. The top portion of the ontology that corresponds to theoretically-grounded, generalizable high- and low-level skills associated with CPS and behaviors that would demonstrate evidence of each skill can be used in a broader context. Specifically, the ontology can be taken out of the context of the electronics simulation-based task in this study and implemented in other task contexts and domains. For use in new contexts, additional nodes would need to be added or current nodes modified so as to lay out what affordances in the new task context would be available for individuals to demonstrate evidence of skills from the generalizable ontology layers. For even smoother transition, tasks can be designed that have similar affordances to the current task so that fewer (if any) nodes need to be modified. This approach for conceptualizing CPS and guiding the identification of components of CPS has advantages, as it can support a detailed measurement of the actions and discourse of team members in different problem contexts at a deep level of grain size in ways that are theoretically grounded. This is particularly important when dealing with unconstrained computer environments like the one used in the current study that output log files at a small grain size.

The results for the CPS skill profiles can inform research on measuring CPS skills in contributing to the conceptualization of ways to aggregate information about individuals' CPS skills to create descriptive representations of CPS. The results of the current study can further contribute to the literature in other areas such as computer-supported collaborative learning and teams in organizations by identifying collaborative behaviors that may contribute to better performance outcomes. This may provide support in determining the best team assembly rules for optimal performance. We have begun developing inventories associated with components of the CPS ontology that can potentially serve as a tool to evaluate individuals prior to collaboration and provide guidance for assembling teams according to particular components of CPS along social and cognitive dimensions.

The current study provides preliminary analyses that offer groundwork for future studies in the measurement of CPS skills, particularly work that seeks to provide fewer constraints on the collaborative context for individuals and detailed information about actions and discourse of individuals in complex situations. This area of research is still in a nascent stage, particularly for large scale assessment. Certainly additional research is needed in order to establish comprehensive validity evidence and generalizability to a number of settings.

Author note

Jessica Andrews-Todd, Cognitive and Technology Sciences Center, Educational Testing Service; Carol M. Forsyth, Cognitive and Technology Sciences Center, Educational Testing Service.

This material is based upon work supported by the National Science Foundation under Grant DUE 1535224 awarded to the first author. The opinions expressed are those of the authors and do not necessarily represent views of the National Science Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2018.10.025.

References

Adams, R., Vista, A., Scoular, C., Awwal, N., Griffin, P., Care, E., 2015. Automatic coding procedures for collaborative problem solving. In: Griffin, P., Care, E. (Eds.), Assessment and teaching of 21st century skills: Methods and approach. Springer, pp. 115–132.

Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A. C., & Halpern, D. (2011). Operation ARIES!: A serious game for teaching scientific inquiry. In M. Ma (Ed.), *Serious games and edutainment applications* (pp. 169–195). London: Springer.

Halpern, D. F., Millis, K., Graesser, A. C., Butler, H., Forsyth, C., & Cai, Z. (2012). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity*, *7*(2), 93–100.

Forsyth, C. M., Graesser, A. C., Pavlik Jr, P. I., Millis, K., & Samei, B. (2014). Discovering theoretically grounded predictors of shallow vs. deep-level learning. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014)* (pp. 229–232). London, UK: International Educational Data Mining Society.

Andrews, J. J., & Rapp, D. N. (2015). Benefits, costs, and challenges of collaboration for learning and memory. *Translational Issues in Psychological Science*, *1*, 182–191.

Kerr, D., Andrews, J. J., & Mislevy, R. J. (2016). The in-task assessment framework for behavioral data. In A. A. Rupp & J. P. Leighton (Eds.), *The Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications* (pp. 472–507). Hoboken, NJ: Wiley-Blackwell.

Andrews, J. J., Kerr, D., Mislevy, R. J., von Davier, A. A., Hao, J., & Liu, L. (2017). Modeling collaborative interaction patterns in a simulation-based task. *Journal of Educational Measurement*, *54*(1), 54–69. Herro, D., Quigley, C., Andrews, J., & Delacruz, G. (2017). Co-Measure: developing an assessment for student collaboration in STEAM activities. *International Journal of STEM Education*, *4*(1), 26. https://doi.org/10.1186/s40594-017-0094-z

Andrews-Todd, J., Forsyth, C. M., Steinberg, J., & Rupp, A. A. (2018). Identifying profiles of collaborative problem solvers in an online electronics environment. In K. E. Boyer & M. Yudelson (Eds.), *Proceedings of the 11th International Conference on Educational Data Mining* (pp. 239–245). Buffalo, NY: International Educational Data Mining Society.

Barron, B., 2003. When smart groups fail. The Journal of the Learning Sciences 12 (3), 307–359.

Bennett, R.E., Gitomer, D.H., 2009. Transforming K–12 assessment: Integrating accountability testing, formative assessment and professional support. In: Wyatt-Smith, C., Cumming, J. (Eds.), Educational assessment in the 21st century. Springer, New York, pp. 43–61.

Biswas, G., Jeong, H., Kinnebrew, J.S., Sulcer, B., Roscoe, R., 2010. Measuring self-regulated learning skills through social interactions in a teachable agent environment. Research and Practice in Technology Enhanced Learning 5, 123–152.

Bower, M., Lee, M.J., Dalgarno, B., 2017. Collaborative learning across physical and virtual worlds: Factors supporting and constraining learners in a blended reality environment. British Journal of Educational Technology 48 (2), 407–430.

van Boxtel, C., van der Linden, J., Kanselaar, G., 2000. Collaborative learning tasks and the elaboration of conceptual knowledge. Learning and Instruction 10 (4), 311–330.

Brodt, S., Thompson, L., 2001. Negotiating teams: A levels of analysis approach. Group Dynamics: Theory, Research, and Practice 5 (3), 208–219.

Burrus, J., Jackson, T., Xi, N., Steinberg, J., 2013. Identifying the most important 21st century workforce competencies: An analysis of the occupational Information network (O* NET) (No. ETS RR-13-21). Educational Testing Service, Princeton, NJ.

Care, E., Griffin, P., 2014. An approach to assessment of collaborative problem solving. Research and Practice in Technology Enhanced Learning 9 (3), 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.

Casner-Lotto, J., Barrington, L., 2006. Are they really ready for work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century U.S. Workforce. In: Washington, DC: The conference board, partnership for 21st century skills, corporate voices for working families, and society for human resource management, Retrieved from http://eric.ed.gov/?id=ED519465.

Chan, C.K., 2001. Peer collaboration and discourse patterns in learning from incompatible information. Instructional Science 29 (6), 443–479.

Chung, G.K.W.K., O'Neil, H.F., Herl, H.E., 1999. The use of computer-based collaborative knowledge mapping to measure team processes and team outcomes. Computers in Human Behavior 15 (3), 463–493.

Clark, H.H., 1996. Using language. Cambridge University Press, New York, NY.

Clark, H.H., Brennan, S.E., 1991. Grounding in communication. Perspectives on Socially Shared Cognition 13 (1991), 127–149.

Cohen, P.R., 1989. Planning and problem solving. In: Cohen, P.R., Feigenbaum, E.A. (Eds.), The handbook of artificial intelligence. Addison-Wesley, Boston, pp. 513–562.

Davey, T., Ferrara, S., Shavelson, R., Holland, P., Webb, N., Wise, L., 2015. Psychometric considerations for the next generation of performance assessment. Educational Testing Service, Princeton, NJ.

Dewiyanti, S., Brand-Gruwel, S., Jochems, W., Broers, N.J., 2007. Students' experiences with collaborative learning in asynchronous computer-supported collaborative learning environments. Computers in Human Behavior 23 (1), 496–514.

Dillenbourg, P., Traum, D., 2006. Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. The Journal of the Learning Sciences 15 (1), 121–151.

Dunlosky, J., Metcalfe, J., 2008. Metacognition. Sage Publications, Thousand Oaks, CA.

Fiore, S.M., Graesser, A., Greiff, S., Griffin, P., Gong, B., Kyllonen, P., et al., 2017. Collaborative problem solving: Considerations for the national assessment of educational progress. National Center for Education Statistics, Alexandria, VA.

Frensch, P.A., Funke, J., 1995. Complex problem solving: The European perspective. Lawrence Erlbaum, Mahwah, NJ.

Gillies, R.M., 2004. The effects of cooperative learning on junior high school students during small group learning. Learning and Instruction 14 (2), 197–213.

Gobert, J.D., Sao Pedro, M.A., Baker, R.S.J.D., Toto, E., Montalvo, O., 2012. Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds. Journal of Educational Data Mining 4, 111–143.

Graesser, A.C., Cai, Z., Hu, X., Foltz, P.W., Greiff, S., Kuo, B.-C., et al., 2017. Assessment of collaborative problem solving. In: In: Sottilare, R., Graesser, A.C., Hu, X., Goodwin, G.F. (Eds.), Design recommendations for intelligent tutoring systems: Volume 5 - assessment Vols. 275–285, U.S. Army Research Laboratory, Orlando, FL.

Graesser, A.C., D'Mello, S., Hu, X., Cai, Z., Olney, A., Morgan, B., 2012. AutoTutor. In: McCarthy, P., Boonthum-Denecke, C. (Eds.), Applied natural language processing: Identification, investigation and resolution. IGI Global, Hershey, PA, pp. 169–187.

Griffin, P., McGaw, B., Care, E. (Eds.), 2012. Assessment and teaching of 21st century skills. Springer, New York.

Hao, J., Liu, L., von Davier, A.A., Kyllonen, P., 2015. Assessing collaborative problem solving with simulation based tasks. In: Proceedings of the 11th international conference on computer-supported collaborative learning. International Society for the Learning Sciences, Gothenburg, Sweden.

Hao, J., Liu, L., von Davier, A.A., Kyllonen, P., Kitchen, C., 2016. Collaborative problem solving skills versus collaboration outcomes: Findings from statistical analysis and data mining. In: Barnes, T., Chi, M., Feng, M. (Eds.), Proceedings of the 9th international conference on educational data mining. International Educational Data Mining Society, Raleigh, NC, pp. 382–387.

Herborn, K., Mustafić, M., Greiff, S., 2017. Mapping an experiment-based assessment of collaborative behavior onto collaborative problem solving in PISA 2015: A cluster analysis approach for collaborator profiles. Journal of Educational Measurement 54 (1), 103–122.

Herborn, K., Stadler, M., Mustafić, M., Greiff, S., 2018. The assessment of collaborative problem solving in PISA 2015: Can computer agents replace humans?. Computers in Human Behavior(in press).

Hesse, F., Care, E., Buder, J., Sassenberg, K., Griffin, P., 2015. A framework for teachable collaborative problem solving skills. In: Griffin, P., Care, E. (Eds.), Assessment and teaching of 21st century skills. Springer, New York, pp. 37–56.

Hmelo-Silver, C.E., Chernobilsky, E., Jordan, R., 2008. Understanding collaborative learning processes in new learning environments. Instructional Science 36 (5–6), 409–430.

Honey, M.A., Hilton, M.L. (Eds.), 2011. Learning science through computer games and simulations. National Academies Press, Washington, DC.

Hsieh, I.-L.G., O'Neil, H.F., 2002. Types of feedback in a computer-based collaborative problem-solving group task. Computers in Human Behavior 18 (6), 699–715.

Jeong, H., Chi, M.T.H., 2007. Knowledge convergence and collaborative learning. Instructional Science 35 (4), 287–315.

van Joolingen, W.R., de Jong, T., Lazonder, A.W., Savelsbergh, E.R., Manlove, S., 2005. Co-Lab: Research and development of an online learning environment for collaborative scientific discovery learning. Computers in Human Behavior 21, 671–688.

Jordan, P.W., Hall, B., Ringenberg, M., Cue, Y., Rosé, C., 2007. Tools for authoring a dialogue agent that participates in learning studies. In: Luckin, R., Koedinger, K.R., Greer, J.E. (Eds.), Proceedings of the conference on artificial intelligence in education: Building Technology rich learning contexts that work. IOS Press, Los Angeles, CA, pp. 43–50.

Kirschner, F., Paas, F., Kirschner, P.A., 2009. A cognitive load approach to collaborative learning: United brains for complex tasks. Educational Psychology Review 21 (1), 31–42.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.

Lipponen, L., 2000. Towards knowledge building: From facts to explanations in primary students' computer mediated discourse. Learning Environments Research 3 (2), 179–199.

Lipponen, L., Rahikainen, M., Lallimo, J., Hakkarainen, K., 2003. Patterns of participation and discourse in elementary students' computer-supported collaborative learning. Learning and Instruction 13 (5), 487–509.

Liu, L., von Davier, A.A., Hao, J., Kyllonen, P., Zapata-Rivera, J.-D., 2015. A tough nut to crack: Measuring collaborative problem solving. In: Rosen, Y., Ferrara, S., Mosharraf, M. (Eds.), Handbook of research on computational tools for real-world skill development. IGI-Global, Hershey, PA, pp. 344–359.

Mayer, R.E., Wittrock, M.C., 1996. Problem-solving transfer. In: Berliner, D.C., Calfee, R.C. (Eds.), Handbook of educational psychology. Lawrence Erlbaum, Mahwah, NJ, pp. 47–62.

Meier, A., Spada, H., Rummel, N., 2007. A rating scheme for assessing the quality of computer-supported collaboration processes. International Journal of Computer-Supported Collaborative Learning 2 (1), 63–86.

Mislevy, R.J., Steinberg, L.S., Almond, R.G., 2003. On the structure of educational assessments. Measurement: Interdisciplinary Research and Perspectives 1 (1), 3–62.

Morgan, B.B., Salas, E., Glickman, A.S., 1993. An analysis of team evolution and maturation. The Journal of General Psychology 120 (3), 277–291.

National Research Council, 2008. Research on future skill demands: A workshop summary. National Academies Press, Washington, DC.

National Research Council, 2011. Assessing 21st century skills. National Academies Press, Washington, DC.

Nouri, J., Åkerfeldt, A., Fors, U., Selander, S., 2017. Assessing collaborative problem solving skills in technology-enhanced learning environments–The PISA Framework and modes of communication. International Journal of Emerging Technologies in Learning 12 (4), 163–174.

Nye, B.D., Graesser, A.C., Hu, X., 2014. AutoTutor and family: A review of 17 years of natural language tutoring. International Journal of Artificial Intelligence in Education 24 (4), 427–469.

OECD, 2013a. PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy. OECD Publishing, Paris.

OECD, 2013b. PISA 2015 collaborative problem solving framework. OECD Publishing, Paris.

O'Neil, H.F., 1999. Perspectives on computer-based performance assessment of problem solving. Computers in Human Behavior 15 (3/4), 225–268.

O'Neil, H.F., Chuang, S., Chung, G.K.W.K., 2003. Issues in the computer-based assessment of collaborative problem solving. Assessment in Education: Principles, Policy & Practice 10, 361–373.

O'Neil, H.F., Chung, G.K.W.K., Brown, R.S., 1995. Measurement of teamwork processes using computer simulation (No. CSE Tech. Rep. No. 399). National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles.

Quellmalz, E.S., Pellegrino, J.W., 2009. Technology and testing. Science 323, 75–79.

Romero, C., González, P., Ventura, S., del Jesús, M.J., Herrera, F., 2009. Evolutionary algorithms for subgroup discovery in e-learning: A practical application using Moodle data. Expert Systems with Applications 36, 1632–1644.

Rosé, C.P., Ferschke, O., 2016. Technology support for discussion based learning: From computer supported collaborative learning to the future of massive open online courses. International Journal of Artificial Intelligence in Education 26 (2), 660–678.

Rosen, Y., 2014. Comparability of conflict opportunities in human-to-human and human-to-agent online collaborative problem solving. Technology, Knowledge and Learning 19 (1–2), 147–164.

Rosen, Y., Foltz, P.W., 2014. Assessing collaborative problem solving through automated technologies. Research and Practice in Technology Enhanced Learning 9 (3), 389–410.

Rosen, Y., Tager, M., 2013. Computer-based assessment of collaborative problem solving skills: Human-to-agent versus human-to-human approach. Pearson Education, Boston, MA.

Rowe, J.P., Shores, L.R., Mott, B.W., Lester, J.C., 2011. Integrating learning, problem solving, and engagement in narrative-centered learning environments. International Journal of Artificial Intelligence in Education 21, 115–133.

Rus, V., D'Mello, S., Hu, X., Graesser, A.C., 2013. Recent advances in intelligent systems with conversational dialogue. AI Magazine 34, 42–54.

Spada, H., Meier, A., Rummel, N., Hauser, S., 2005. A new method to assess the quality of collaborative process in CSCL. In: Proceedings of the 2005 conference on computer support for collaborative learning: Learning 2005: The next 10 years!. International Society of the Learning Sciences, Taipei, Taiwan, pp. 622–631.

Springer, L., Stanne, M.E., Donovan, S.S., 1999. Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. Review of Educational Research 69 (1), 21–51.

Trilling, B., Fadel, C., 2009. 21st century skills: Learning for life in our times. Jossey-Bass, San Francisco, CA.

VanLehn, K., 1996. Cognitive skill acquisition. Annual Review of Psychology 47 (1), 513–539.

Ward, W., Cole, R., Bolaños, D., Buchenroth-Martin, C., Svirsky, E., Weston, T., 2013. My science tutor: A conversational multimedia virtual tutor. Journal of Educational Psychology 105 (4), 1115–1125.

Webb, N.M., 1991. Task-related verbal interaction and mathematics learning in small groups. Journal for Research in Mathematics Education 22 (5), 366–389.

Wirth, J., Klieme, E., 2003. Computer-based assessment of problem solving competence. Assessment in Education: Principles, Policy & Practice 10 (3), 329–345.

Wise, A.F., Chiu, M.M., 2011. Analyzing temporal patterns of knowledge construction in a role-based online discussion. International Journal of Computer-Supported Collaborative Learning 6 (3), 445–470.

Zapata-Rivera, D., Jackson, T., Katz, I.R., 2015. Authoring conversation-based assessment scenarios. In: In: Sottilare, R., Graesser, A.C., Hu, X., Brawner, K. (Eds.), Design recommendations for intelligent tutoring systems: Authoring tools Vol. 3, Army Research Laboratory, Orlando, FL, pp. 169–178.