Research and Practice in Technology Enhanced Learning
Vol. 9, No. 3 (2014) 389 − 410

ASSESSING COLLABORATIVE PROBLEM SOLVING THROUGH AUTOMATED TECHNOLOGIES

YIGAL ROSEN

Pearson Inc
Boston, MA, United States
yigal.rosen@pearson.com

PETER W. FOLTZ

Pearson Inc & University of Colorado
4940 Pearl East Circle, Suite 200, Boulder, CO, 80301, United States
Peter.Foltz@pearson.com

Today, proficiency in collaborative problem solving (CPS) is necessary for success in both college and workplace, as well as the ability to perform that collaboration in various settings. At the same time, structuring computer-based CPS assessment, specifically for large-scale programs, is challenging. In order to perform standardized assessment of CPS skills on an individual level, each student should be matched with various types of group members, should be tested on a variety of CPS skills, and must apply the skills in varied contexts. One solution to this standardization is to use computer-based agents to serve as the collaborators in the interactions. The aim of this study was to explore students' performance in human-to-agent (H-A), compared to human-to-human (H-H) CPS assessment tasks. A sample of 14-year olds across three countries participated in the study. Students in both H-H and H-A modes were able to collaborate and communicate by using identical methods and resources. However, while in the H-A mode, students collaborated with a simulated computer-driven partner and in the H-H mode students collaborated with another student to solve a problem. This study is among the first of its kind to investigate systematically students' CPS performance in H-A and H-H standardized assessment settings. Directions for future research are discussed in terms of their implications to large-scale assessment programs, teaching and learning.

Keywords: Collaborative problem solving; assessment; automated scoring.

1. Introduction

Collaborative problem solving (CPS) is a critical competency for college and career readiness. Students emerging from schools into the workforce and public life will be expected to be able to work in teams, cooperate with others, and resolve conflicts in order to solve the kinds of problems required in modern economies. They will further need to be able use these skills flexibly with various group compositions and environments (Griffin, Care, & McGaw, 2012; OECD, 2013; O'Neil & Chuang, 2008; Rosen & Rimor, 2012). Recent curriculum and instruction reforms have focused to a greater extent on teaching and learning CPS (National Research Council, 2011; US Department of Education, 2010). However, structuring standardized computer-based assessment of CPS skills, specifically for large-scale assessment programs, is challenging. Unlike some other individual student skills, CPS skills typically require using complex performance tasks, grounded in varied educational domains, with interaction among students. These factors can affect the level of control that can be applied to ensure accurate assessment of students.

Technology offers opportunities for measurement of CPS skills in domains and contexts where assessment would otherwise not be possible or would not be scalable. One of the important enhancements brought by technology to educational assessment is the capability to embed computer-based responses and behaviors into the instrument, enabling it to change its state in response to student's operations. These can be designed in such a way that the student are exposed to an expected situation and set of interactions, while the student's interactions as well as the explicit responses are captured and scored automatically.

CPS refers to problem-solving activities that involve collaboration among a group of individuals (e.g. O'Neil, Chuang, & Baker, 2010; Zhang, 1998). CPS is one of the areas that the Organisation for Economic Co-operation and Development (OECD) emphasized for major development in the Programme for International Student Assessment (PISA) in addition to scientific literacy, math and reading literacy for the 2015 assessment. This paper examines CPS skills motivated based on the OECD CPS framework and its implications for large-scale assessment. This paper addresses the challenges in CPS assessment by introducing a new methodology for scalable computer-based assessment of CPS, providing findings from an empirical pilot study conducted in three countries, as well as discussing implication of the findings on further research and development.

1.1. *Defining collaborative problem solving*

In the PISA 2015 Framework, CPS competency is defined 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, 2013). This definition treats the competency as conjoint dimension collaboration skills and the skills needed to solve a problem. For the PISA assessment, the focus is on individual capacities *within* collaborative situations. Thus, the effectiveness of collaborative problem solving depends on the ability of group members to collaborate and to prioritize the success of the group over individual successes. At the same time, this ability is still a trait in each of the individual members of the group. There are three primary competencies in CPS.

First, CPS requires students to be able to establish, monitor, and maintain the shared understanding throughout the problem-solving task by responding to requests for information, sending important information to agents about tasks completed, establishing or negotiating shared meanings, verifying what each other knows, and taking actions to repair deficits in shared knowledge. Shared understanding can be viewed as an effect, if the goal is that a group builds the common ground necessary to perform well together, or as a process by which peers perform conceptual change (Dillenbourg, 1999). CPS is a coordinated joint dynamic process that requires periodic communication between group members. Communication is a primary means of constructing a shared understanding or Common Ground (e.g. Clark, 1996). An "optimal collaborative effort" is required of all of the participants in order to achieve adequate performance in a collaborative environment (Dillenbourg & Traum, 2006).

Second, collaboration requires the capability to identify the type of activities that are needed to solve the problem and to follow the appropriate steps to achieve a solution. This process involves exploring and interacting with the problem situation. It includes understanding both the information initially presented in the problem and any information that is uncovered during interactions with the problem. The accumulated information is selected, organized, and integrated in a fashion that is relevant and helpful to solving the particular problem and that is integrated with prior knowledge. Setting sub-goals, developing a plan to reach the goal state, and executing the plan that was created are also a part of this process. Overcoming the barriers of reaching the problem solution may involve not only cognition, but motivational and affective means (Funke, 2010; Mayer, & Wittrock, 2006).

Third, students must be able to help organize the group to solve the problem; consider the talents and resources of group members; understand their own role and the roles of the other agents; follow the rules of engagement for their role; monitor the group organization; reflect on the success of the group organization, and help handle communication breakdowns, conflicts, and obstacles (Rosen & Rimor, 2012).

Among other factors that may influence student CPS are gender, race, status, perceived cognitive or collaborative abilities, motivation and attractiveness. According to Dillenbourg (1999), effective collaboration is characterized by a relatively symmetrical structure. Symmetry of knowledge occurs when all participants have roughly the same level of knowledge, although they may have different perspectives. Symmetry of status involves collaboration among peers rather than interactions involving facilitator relationships. Finally, symmetry of goals involves common group goals rather than individual goals that may conflict. The degrees of interactivity and negotiability are additional indicators of collaboration (Dillenbourg, 1999). For example, trivial, obvious, and unambiguous tasks provide few opportunities to observe negotiation because there is nothing about which to disagree. Thus, in a standardized assessment situation, it is possible that a student should be matched with various types of group members that will represent different collaboration and problem-solving skills, while controlling for other factors that may influence student performance (e.g. asymmetry of roles).

1.2. *Assessment in collaborative problem solving*

Student skills in CPS can be assessed through a number of different approaches. One key consideration in the development of CPS assessment is types of measures used to determine the quality of student performance. These measures can include the quality of the solutions and the objects generated during the collaboration (Avouris, Dimitracopoulou, & Komis, 2003); analyses of log files, intermediate results, paths to the solutions (Adejumo, Duimering, & Zhong, 2008), team processes and structure of interactions (O'Neil, Chung, & Brown, 1997); and quality and type of collaborative communication (Cooke et al., 2003; Foltz, & Martin, 2008; Graesser, Jeon, & Dufty, 2008). While there are a number of options for measurement, a key challenge is to assure that the assessment approach can accurately capture the performance as well as be able to convert it into quantifiable measure of performance.

Another challenge in CPS assessment refers to the need for synthesizing information from individuals and teams along with actions and communication dimensions (Laurillard, 2009; O'Neil, Chen, Wainess, & Shen, 2008; Rimor, Rosen, & Naser, 2010). Communication among the group members is central in the CPS assessment and it is considered a major factor that contributes to the success of CPS (Dillenbourg & Traum, 2006; Fiore & Schooler, 2004; Fiore et al., 2010). While communication can be classified as an individual collaboration skill, the output of communication provides a window into the cognitive and social processes related to all CPS skills. Thus, communication among the team members can be assessed to provide measures of these processes. One approach has been to analyze the streams of open-ended communication in collaborative situations. For example, Foltz and Martin (2008) have used semantic and syntactic analyses of team communications in order to score individual and team performance as well as classify individual statements to different collaborative skills and Erkens and Jansen (2008) have used techniques to code collaboration protocols. Therefore, analysis of the content and structure of communication streams can provide measures of grounding and precision of references among team members, mutual goal establishment, progress toward goals, negotiation, consensus, leadership, and quality of solutions generated. However, such analysis approaches require capturing the written or spoken communication stream, having robust models of human dialogs and then performing fairly intensive computational processing for scoring and classification of the language. This can be limiting, most particularly in applications in large-scale international testing which requires scoring across multiple languages.

Nevertheless, various techniques have been developed to address the challenge of providing a tractable way to communicate in CPS assessment contexts. One technique that has been tested is communication through predefined messages (Chung, O'Neil, & Herl, 1999; Hsieh & O'Neil, 2002; O'Neil et al., 1997). In these studies, participants were able to communicate using the predefined messages and to successfully complete a task (a simulated negotiation or a knowledge map), and the team processes and outcomes were measurable. Team members used the predefined messages to communicate with each other, and measures of CPS processes were computed based on the quantity and type of messages used (i.e. each message was coded a priori as representing adaptability, coordination, decision making, interpersonal skill, or leadership). The use of messages provides a manageable way of measuring CPS skills and allows real-time scoring and reporting.

As mentioned above, within the PISA 2015 assessment the focus of measurement of collaboration is on the individual rather than the group. This approach is not surprising because, in most educational systems of accountability, it is the individual who is assessed. The focus on the individual as a unit of analysis in collaborative context allows application of traditional psychometric models and direct comparisons. To ensure valid measurement on the individual level, it is possible that each student should be paired with the same number of other partners displaying the same range of CPS characteristics. This way each individual student can be situated fairly similarly to be able to show his or her proficiency.

1.3. *Computer and human agent technology in collaborative problem solving tasks*

Collaboration can take many forms, ranging from two individuals to large teams with predefined roles. For assessment purposes collaboration can also be performed using computer agents playing the role of team members, different agents or using other humans as team members. Thus, a critical distinction is whether all team members are human or some are computer agents. There are advantages and limitations for each method which are outlined below.

The Human-to-Human (H-H) approach provides an authentic human-human interaction which is a highly familiar situation for students. Students may be more engaged and motivated to collaborate with their peers. Additionally, the H-H situation is closer to the CPS situations students will encounter in their personal, educational, professional and civic activities. However, because each human will act independently, the approach can be problematic because of individual differences that can significantly affect the CPS process and its outcome. Therefore, the H-H assessment approach of CPS may not provide sufficient opportunity to cover variations in group composition, diversity of perspectives and different team member characteristics in a controlled manner, for accurate assessment of the skills on an individual level.

Simulated team members using a preprogrammed profile, actions and communication can potentially provide the coverage of the full range of collaboration skills with sufficient control. In the Human-to-Agent (H-A) approach, CPS skills are measured by pairing each individual student with a computer agent or agents that can be programmed to act as team members with varying characteristics relevant to different CPS situations. Group processes are often different depending on the task and could even be competitive. Use of computer agents provides a component of non-competitiveness to the CPS situation, as it is experienced by a student. Additionally, if the time-on-task is limited, time spent establishing common ground or discussing non-task relevant work may lower group productivity. As a result of these perceived constraints, a student collaborating in H-H mode may limit significantly the extent to which CPS dimensions, such as shared understanding, are externalized through communication with the partner. The agents in H-A communication can be developed with a full range of capabilities, such as text-to-speech, facial actions, and optionally rudimentary gestures. In its minimal level, a conventional communication media, such as text via emails, chat, or graphic organizer with lists of named agents can be used for H-A purposes. However, CPS in H-A settings deviate from natural human communication delivery. The dynamics of H-H interaction (timing, conditional branching) cannot be perfectly captured with agents, and agents cannot adjust to idiosyncratic characteristics of humans. For example, human collaborators can propose unusual, exceptional solutions; the characteristic of such a process is that it cannot be included in a system following an algorithm, such as H-A interaction.

Research shows that computer agents have been successfully used for tutoring, collaborative learning, co-construction of knowledge, and CPS (e.g. Biswas et al., 2010; Graesser et al., 2008; Millis et al., 2011). A computer agent can be capable of generating goals, performing actions, communicating messages, sensing its environment, adapting to changing environments, and learning (Franklin & Graesser, 1996). One of the examples for computer agent use in education is a teachable agent system called Betty's Brain (Biswas, Leelawong, Schwartz, & Vye, 2005; Leelawong & Biswas, 2008). In this system, students teach a computer agent using a causal map, which is a visual representation of knowledge structured as a set of concepts and their relationships. Using their agent's performance as motivation and a guide, students study the available resources so that they can remediate the agent's knowledge and, in the process, learn the domain material themselves. Operation ARIES (Cai et al., 2001; Millis et al., 2011) uses animated pedagogical agents that converse with the student in a game-based environment for helping students learn critical-thinking skills and scientific reasoning within scientific inquiry. The system dynamically adapts the tutorial conversations to the learner's prior-knowledge. These conversations, referred to as "trialogs" are between the human learner and two computer agents (student and teacher). The student learns vicariously by observing the agents, gets tutored by the teacher agent, and teaches the student agent.

In summary, CPS assessment must take into account the types of technology, tasks and assessment contexts in which it will be applied. The assessment will need to consider the kinds of constructs that can be reliably measured and also provide valid inferences about the collaborative skills being measured. Technology offers opportunities for assessment in domains and contexts where assessment would otherwise not be possible or would not be scalable. One of the important improvements brought by technology to educational assessment is the capacity to embed system responses and behaviors into the instrument, enabling it to change its state in response to student's manipulations. These can be designed in such a way that the student will be exposed to an expected scenario and set of interactions, while the student's interactions as well as the explicit responses are captured and scored automatically. Nevertheless, computer-based assessment of CPS requires advancements in educational assessment methodologies and technology. Group composition, discourse management, and the development of effective computer agents, are considered as major challenges in designing valid, reliable, and scalable assessment of CPS skills (Graesser, Foltz, Rosen, Forsyth, & Germany, in press). The aim of this study was to address part of these challenges by examining the possible differences in student performance in CPS assessment tasks comparing a human-to-agent (H-A) to a human-to-human (H-H) methodology.

2. Research Questions

The study addressed empirically the following primary question regarding students' CPS performance in H-A, compared to H-H based assessments:

> *What are the differences in student CPS performance between H-A and H-H modes of assessment, as reflected in shared understanding, problem solving, progress monitoring and providing feedback measures?*

In order to better understand the dimensionality of CPS measures and possible factors that differentiate student performance in H-A and H-H settings, the following research questions were examined:

> *What are the relationships between the CPS measures of shared understanding, problem solving, progress monitoring, and proving feedback measures in H-A and H-H settings?*
> *What are the differences in student motivation while collaborating with a computer agent or a human partner on CPS assessment tasks?*
> *What are the differences in student CPS performance between H-A and H-H modes of assessment, as reflected in time-on-task, and number of attempts to solve the problem?*

3. Method

Study participants included 179 students age 14, from the United States, Singapore and Israel. The results presented in the current article came from a larger study in which students from six countries were recruited to participate in a 21st Century Skills Assessment project investigating the innovative ways to develop computer-based assessment of critical-thinking, and CPS. The researchers collected data between November 2012 and January 2013. Recruitment of participating schools was achieved through collaboration with local educational organizations based on the following criteria: (a) the school was public, (b) the school was actively involved in various 21st Century Skills projects, (c) the population was 14 years-old students proficient in English, and (c) there was sufficient technology infrastructure (e.g. computers per student, high-speed Internet). In all, 136 students participated in the H-A group and 43 participated in the H-H group (43 additional students participated in the H-H setting, acting as "collaborators" for the major H-H group). Specifically in H-H assessment mode, students were randomly assigned into pairs to work on the CPS task. Because the H-H approach required pairs of students working together in a synchronized manner, the number of pairs was limited. This is due to the characteristics of technology infrastructures in participating schools.

Table 1. Research population by mode of assessment and country.

Group	H-A		H-H	
	Female	Male	Female	Male
United States	22	19	6	4
Singapore	12	31	8	7
Israel	30	22	13	5
Overall	*64*	*72*	*27*	*16*

Of the total students who participated, 88 were male (49.2%) and 91 were female (50.8%). Table 1 summarizes the country and gender distribution of participating students between the H-A and H-H groups. No significant differences were found in Grade Point Average (GPA), English Language Arts (ELA), and Math average scores between participants in H-A and H-H mode within the countries. This similar student background permitted comparability of student results in CPS assessment task between the two modes of collaboration.

3.1. *Collaborative problem solving assessment task*

For the CPS computer-based assessment task, the student was asked to collaborate with a partner (computer-driven agent or a classmate) to find the optimal conditions for an animal at the zoo in order to maximize an animal's life expectancy, which was continually displayed. The student was able to select different types of food, life environments, and extra features, while both partners were able to see the selections made and communicate through a phrase-chat (selections from predefined 4-5 options, which changed depending on the task context). The use of predefined messages provided a tractable way of measuring communication and collaboration skills (e.g. Hsieh & O'Neil, 2002). Team members used the predefined messages to communicate with each other, and measures of CPS processes were automatically computed based on the type of messages used in each situation. Each message was coded a priori as representing each of the CPS skills (see Rosen and Tager (2013) for further details).

At the beginning of the task, the student and the partner were prompted to discuss how to reach better conditions for an animal, and at the end of the task, the student was asked to provide written feedback on the partner's performance. It should be noted that due to the centrality of the collaboration dimension in CPS, as it was defined in this study, the difficulty level of the problem was relatively low and served primarily as a platform for the overall assessment of CPS skills. Additionally, due to the exploratory nature of the study, the students were not limited either in a number of attempts to reach optimal solution or in the time-on-task. However, the task was programmed in such a way that at least two attempts for problem solving and at least one communication act with a partner were required to be able to complete the assessment task.

Figures 1 and 2 show examples of the task screens. The major area of the screen allows the partners to view the options available for the Environment, Food, and Extras. Both partners were able to see what variables were selected. However, the selections of the variables were made by one partner only (i.e. by the student in H-A mode or by one of the students in H-H mode) as well as the ability to try out the variables selected (by clicking on "Go"). On the right side of the screen, the partners were able to communicate by using a phrase-chat. The phrases presented at the chat were based on a pre-programmed decision-tree and situated in order to allow the student to authentically communicate with a partner and to be able to cover the CPS measures defined for the task. The computer agent's (Mike) phrases were programmed to act with varying characteristics relevant to different CPS situations (e.g. agree or disagree with the student, contributing to solving the problem or proposing misleading strategies, etc.). This approach provided each individual student with similar optimal chances to show his or her CPS skills. Each student in H-A setting (regardless of student's actions) was provided with the same amount of help and misleading strategies proposed by the agent.

While in the H-H mode, the two partners were provided with exactly the same set of possible phrases for each CPS situation. There was no control over the selections made by the students. In the H-H setting, the chat started with the "leader" who asked questions and led the conversation. The other person took the role of "collaborator" and generally replied to questions asked by the leader. Certain questions or responses, however could

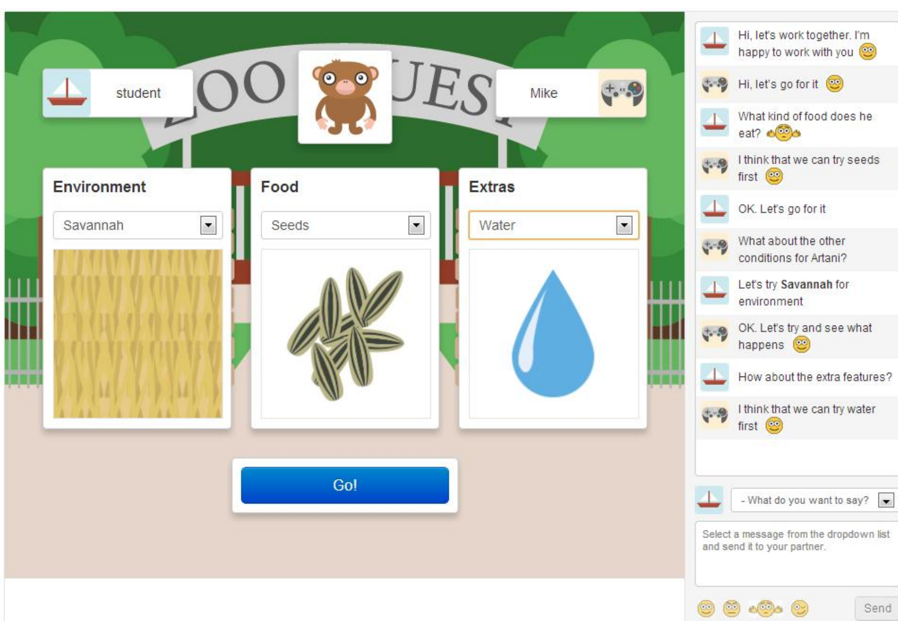

Figure 1. Collaborative problem solving task with Mike (computer agent or another student).

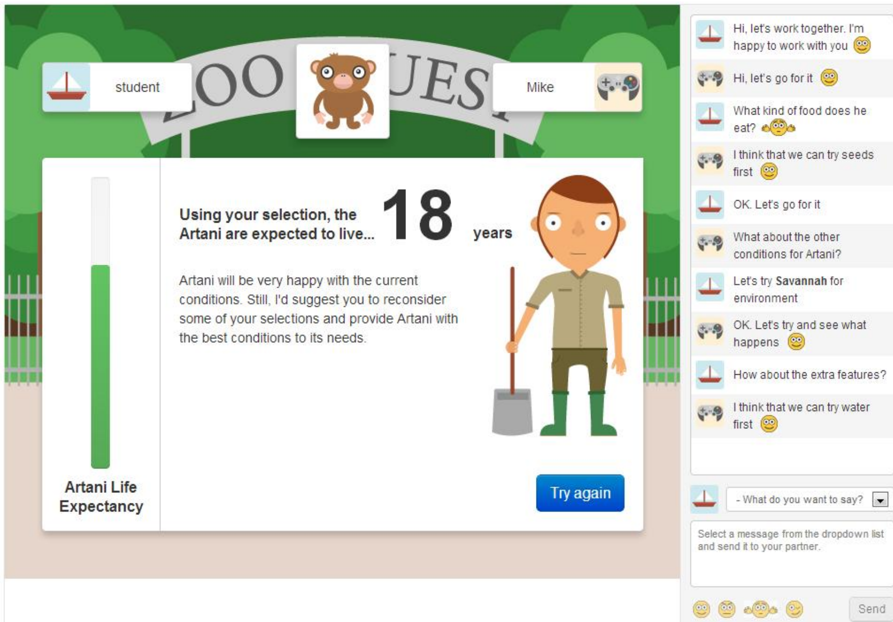

Figure 2. Viewing the results of variables selected in a CPS assessment task by student and Mike (computer agent or another student).

lead to a flipping of these roles during conversation, and so the person initially asking questions may not do so for the entire conversation. Only the leader could submit the selections for the conditions and complete the task though. If the students had submitted their attempt at solving the problem, then some additional statements could become available as part of the phrase-chat communication.

In the H-A setting, the student could choose phrases to chat and received automated responses from the computer agent to simulate a two- way chat with a collaborator. The sentences a student could use were limited to what was relevant to the context of the problem situation at each time and changed based on their progress through the task, similar to the H-H setting. Only the student could submit the selections for the conditions and complete the task though.

Clicking on “Go” provided the partners with the possibility to see the life expectancy of the animal under the variables selected (0-26 years) and to read textual information regarding the result achieved. At this stage the partners were allowed to communicate about the result and about ways to reach the optimal solution, and then decide whether to keep the selections or try again (i.e. change the variables).

3.2. *Scoring student performance in the Zoo Quest task*

The target CPS skills for this assessment consisted of shared understanding, taking appropriate actions to solve the problem, and establishing and maintaining group organization (e.g. Graesser et al., in press; OECD, 2013). In line with CPS dimensions, scores for student performance in the assessment task broken down as being: shared understanding (40 points), problem solving (26 points), and group organization represented by monitoring progress (26 points) and providing feedback (8 points). In both H-H and H-A settings, student scores in the first three CPS measures were generated automatically based on a predefined programmed sequence of possible optimal actions and communication chats that were embedded into the assessment task (see Rosen and Tager (2013) for further details). The problem-solving dimension was scored as one point per each year of the animal's life expectancy that was achieved by selecting the variables. Shared understanding score consisted of a limited number of grounding questions that were initiated by a student in pre-defined appropriate situations (e.g. questioning "What can we do to reach better conditions for the animal?") and appropriate responses to the grounding questions made by the partner. Monitoring progress score was created based on communication initiated by the student prior to the submission of the selected variables (e.g. questioning "Are you ready to go ahead with our plan?" before clicking on "Go") and the statements made by the student based on the life expectancy results that were achieved (e.g. "Should we keep this selection or try again?").

Scoring of the student feedback dimension was provided independently by two teachers from participating schools in the United States who rated the students' final written responses on a 1-4 point scale. Inter-coded agreement of feedback scoring was 92%. It should be noted that the student feedback was scored based on CPS rubrics, while spelling and grammar issues did not affect the student score.

3.3. *Questionnaire*

The questionnaire included four items to assess the extent to which students were motivated to work on the task. Participants reported the degree of their agreement with each item on a four-point Likert scale (1 = strongly disagree, 4 = strongly agree). The items were adopted from motivation questionnaires used in previous studies, and included: "I felt interested in the task"; "The task was fun"; "The task was attractive"; "I continued to work on this task out of curiosity" (Rosen, 2009; Rosen and Beck-Hill, 2012). The reliability (internal consistency) of the questionnaire was 0.85.

4. Results

All results are presented on an aggregative level across countries, since no condition by country interaction was found. First, the results of student performance in a CPS assessment are presented to determine whether there is a difference in student CPS score as a function of collaborating with a computer agent versus a classmate. Next, intercorrelations between CPS measures are presented in H-A and H-H modes as well as student motivation results. Finally, time-on-task and number of attempts to solve the problem in both modes of collaboration are presented.

4.1. *Comparing student CPS performance in H-H and H-A settings*

In order to explore possible differences in the four students' CPS scores, an analysis of variance was performed. First, MANOVA results showed significant difference between H-H and H-A groups (Wilks' Lambda=.904, F(4,174)=4.6, p<.01). Hence, we proceeded to perform individual post-hoc t-tests. The results indicated that students who collaborated with a computer agent showed significantly higher level of performance in establishing and maintaining shared understanding (ES=.4, t(177)=2.5, p<.05), monitoring progress of solving the problem (ES=.6, t(177)=4.0, p<.01), and in the quality of the feedback (ES=.5, t(177)=3.2, p<.01). The findings showed non-significant difference in the ability to solve the problem in the H-A and H-H modes of collaboration (ES=-.3, t(177)=-1.9, p=.06). Table 2 shows the results of student CPS scores in both modes.

Table 2. Student CPS scores in H-A and H-H modes of the assessment.

CPS Measure	H-A	H-H	ES	t(df)
	M(SD)	M(SD)		
Shared understanding	18.4(10.8)	14.6(8.3)	.4	2.4(177)*
Problem solving	24.6(3.2)	25.3(1.7)	-.3	-1.9(177)
Monitoring progress	5.0(7.5)	1.3(4.2)	.6	4.0(177)**
Providing feedback	4.1(2.2)	3.0(1.9)	.5	3.2(177)**

*** p <.01, * p <.05.*

4.2. *Relationships between CPS scores*

To better understand the relationship between the CPS scores, an analysis of intercorrelations among the variables was conducted. The analysis was conducted separately in H-H and H-A conditions because of the possible differences in the intercorrelations across different collaboration settings. Table 3 reports the intercorrelations between the dimensions of CPS in the H-H setting. The findings showed a significantly positive relationships between student shared understanding score and the ability to solve the problem (r=.34, p<.05), and monitor progress (r=.43, p<.01). As it shown in Table 4, the H-A setting revealed a similar pattern, mainly in relationship between shared understanding and both problem solving (r=.32, p<.01), and monitoring progress scores (r=.57, p<.01). In addition, shared understanding score in H-A settings was found in significantly positive correlation with student ability of providing constructive feedback to the partner (r=.48, p<.01). No correlation was found between student problem solving score and other CPS measures.

Table 3. Intercorrelations among dimensions of CPS in H-H setting (n=43).

	SU	PS	MP	PF
Shared understanding (SU)				
Problem solving (PS)	.34*			
Monitoring progress (MP)	.43**	.11		
Providing feedback (PF)	.21	.03	.02	

*** p <.01, * p <.05.*

Table 4. Intercorrelations between dimensions of CPS in H-A setting (n=136).

	SU	PS	MP	PF
Shared understanding (SU)				
Problem solving (PS)	.32**			
Monitoring progress (MP)	.57**	.14		
Providing feedback (PF)	.48**	.15	.26	

*** p <.01.*

4.3. *Student motivation*

In attempting to determine possible differences in student motivation of being engaged in CPS with a computer agent versus a classmate, data on student motivation was analyzed. The result demonstrated no significant difference in student's motivation whether collaborating with a computer agent or a classmate (M=3.1, SD=.7 in H-A mode, compared to M=3.1, SD=.4 in H-H mode; ES=.1, t(177)=.5, p=.64).

4.4. *Attempts to solve a problem and time-on-task*

In order to examine possible differences in the number of attempts for problem-solving as well as time-on-task, a comparison of these measures was conducted between H-A and H-H modes of collaboration. The average number of attempts for problem solving in H-A mode was 8.4 (SD=7.3), compared to 6.1 (SD=5.7) in a H-H mode (ES=.3, t(177)=2.1, p<.05). No significant difference was found in time-on-task (t(177)=-1.6, p=.11). On average, time-on-task in H-A mode was 7.9 minutes (SD=3.6), while student in the H-H mode spent 1.1 more minutes on a task (M=9.0, SD=4.5).

5. Discussion

Policymakers, researchers, and educators are engaged in vigorous debate about assessing CPS skills on an individual level in valid, reliable and scalable ways. The challenges facing implementing CPS in large-scale assessment programs suggest that both H-H and H-A approaches in CPS assessment should be explored. The goal of this study was to investigate differences in student CPS performance in H-A and H-H modes. Students in each of these modes were exposed to identical assessment tasks and were able to collaborate and communicate by using identical methods and resources. However, while in the H-A mode students collaborated with a simulated computer-driven partner, and in the H-H mode students collaborated with another student to solve a problem.

The findings showed that students assessed in H-A mode outperformed their peers in H-H mode in their collaborative skills. CPS with a computer agent involved significantly higher levels of shared understanding, progress monitoring, and feedback. The results suggest that the space of collaboration in H-A settings can be extremely large even when there are a limited number of fixed actions or discourse moves at each point in a conversation. The design of agent-based assessment was flexibly adaptive to the point where no two conversations were ever the same, just as is the case of collaborative interactions among humans. Although students in both H-H and H-A modes were able to collaborate and communicate by using identical methods and resources, full comparability was not expected. This is due to the fact that each student in H-H mode represented a specific set of CPS skills, while in the H-A mode each individual student collaborated with a computer agent with a predetermined large spectrum of CPS skills. Differences across H-H groups could be affected by a given performance of the collaborator. Additionally, because of the relatively low difficulty of the problem that was represented by the CPS task, and much larger emphasis on collaboration, students in H-A were faced with more opportunities to show their collaboration skills. Research shows that in H-H CPS settings there is a tendency to avoid disagreements in order to achieve a rapid consensus on how to solve a problem (e.g. Rosen & Rimor, 2012). It is possible that some students that acted as collaborators in H-H settings did not involve themselves in disagreements, questioning, alternative interpretations of results and other possible resources for sharing understanding, monitoring progress, and providing feedback that can be performed by the leader student. This was not the case with a computer agent. The agent was programmed to partially disagree with the student, occasionally misinterpret the results, or propose misleading strategies.

One major possible implication of CPS score difference in collaboration measures between the H-A and H-H modes is that assessments delivered in multiple modes may differ in score meaning and impact. Each mode of CPS assessment can be differently effective for different educational purposes. For example, a formative assessment program which has adopted rich training on the communication and collaboration construct for its teachers may consider the H-H approach for CPS assessment as a more powerful tool to inform teaching and learning, while H-A may be implemented as a formative scalable tool across a large district or in standardized summative settings. Non-availability of students with certain CPS skill levels in a class may limit the fulfilment of assessment needs, but technology with computer agents can fill the gaps. In many cases, using simulated computer agents instead of relying on peers is not merely a replacement with limitations, but an enhancement of the capabilities that makes independent assessment possible. Furthermore, the phrase-chat used in this study could be replaced by an open-chat in cases where automated scoring of student responses is not needed.

The results further showed no significant differences in other student performance measures to solve the problem with a computer agent or a human partner, although on average students in H-A mode applied more attempts to solve the problem, compared to the H-H mode. It should be noted that the student performance studied here was in the context of well-structured problem-solving, while primarily targeting collaborative dimensions of CPS. The problem-solving performance in this task was strongly influenced by the ability of the students to apply a vary-one-thing-at-a-time strategy (e.g. Vollmeyer, & Rheinberg, 1999), which is also known as control of variables strategy (Chen, & Klahr, 1999). This strategy is suggested by the agent in the logfile example in Table 2 ("Let's try to change one condition per trial") and was part of a phrase-chat menu that was available for a student "collaborator" in H-H settings. While the computer agent was programmed to suggest this strategy to each participant in a standardized way (before the second submission), there was no control over the suggestions made by the human partner. However, participants in H-A mode did not outperform participants in H-H mode in their problem-solving score, while the major difference between the students' performance in H-H and H-A settings were the collaboration-related skills.

Interdependency is a central property of tasks that are desired for assessing collaborative problem solving, as opposed to a collection of independent individual problem solvers. A task has higher interdependency to the extent that student A cannot solve a problem without actions of student B. Although, interdependency between the group members was required and observable in the studied CPS task, the collaboration in both settings was characterized by asymmetry of roles. A "leader" student in the H-H setting and the student in the H-A setting were in charge of selecting the variables and submitting the solutions in addition to the ability to communicate with the partner. According to Dillenbourg (1999), asymmetry of roles in collaborative tasks could affect each team member's performance. Thus, a possible explanation for these results is the asymmetry in roles between the "leader" student and the "collaborator" in the H-H setting and the student and the computer agent in the H-A setting. In a more controlled setting (i.e.- H-A) the asymmetrical nature of collaboration was associated with no relationship to the quality of collaborative skills that were observed during the task. While in the H-H setting, in which the human "collaborator" was functioning with no system control over the suggestions that he or she made, the asymmetry in roles was associated with no relationship to the quality of collaborative skills that were observed during the task.

Intercorrelations between the collaboration dimensions in both settings suggest that student performance was assessed in a consistent manner in both H-H and H-A mode, while highlighting different facets of CPS competency. The ability to create and maintain shared understanding during the collaborative task is the major strand in CPS that organized the other facets of student performance (Graesser et al., in press). Indeed, shared understanding skills were positively correlated with the ability to solve the problem and monitor progress in both H-H and H-A settings. The shared understanding score consisted of grounding questions that were initiated by a student in appropriate situations, while problem solving was associated with the ability to take the optimal actions to solve the problem. Additionally, monitoring progress and the ability to provide constructive feedback were associated with communication initiated by the student based on grounding achieved during the task as well as the actions that were taken to solve the problem.

In examining the level of motivation and time-on-task in collaborating with a computer agent or a human partner on CPS assessment task, we found no evidence for differences between the two modes. In other words, students felt motivated and efficient in collaborative work with computer agents just at the same level as when collaborating with their peers. Previous research found that examinee motivation tended to predict test performance among students in situations in which the tests had low or no stakes for the examinees (Sundre, 1999; Sundre & Kitsantas, 2004; Wise & DeMars, 2005). To the degree to which students do not give full effort to an assessment test, the resulting test scores will tend to underestimate their levels of proficiency (Eklöf, 2006; Wise & DeMars, 2006). We believe that two major factors in computer agent implementation contributed to student motivation in CPS assessment tasks. On the one hand, the student and the agent shared the responsibility to collaborate in order to solve the problem. A computer agent was capable to generate suggestions to solve the problem (e.g. "Let's change one condition per trial") and communicate with the student in a contextual and realistic manner. On the other hand, a shared representation of the problem-solving space was implemented to provide a concrete representation of the problem state (i.e. life expectancy) and the selections made (e.g. selection of the conditions).

5.1. *Research limitations and directions for future studies*

Problem difficulty can vary in cases such as ill-defined problems versus well-defined problems, dynamic problems versus static problems, problems that are a long distance versus a short distance to the goal state, a large problem space versus a small space, the novelty of the solution, and so on. It is conceivable that the need for effective collaboration would increase as a function of the problem difficulty. More cognitively challenging problem-solving space in CPS tasks can possibly lead to differential results in H-A and H-H settings. Thus an important question to raise is: "How well would the results found here generalize to ill-structured problem-solving assessment context?" Specifically, do the similarities and differences between H-A and H-H group performance found in this study overestimate what would be found in other assessment contexts? Future studies could consider exploring differences in student performance in a wide range of problem-solving complexity and ill-structured tasks that cannot be solved by a single, competent group member. Such tasks require knowledge, information, skills, and strategies that no single individual is likely to possess. When ill-structured tasks are used, all group members are more likely to participate actively, even in groups featuring a range of student ability (Webb, Nemer, Chizhik, & Sugrue, 1998).

The current study had several limitations. First, it is based on a relatively small and non-representative sample of 14-years-old students in three countries. However, due to lack of empirical research in the field of computer-based assessment of CPS skills, it is necessary to conduct small-scale pilot studies in order to inform a more comprehensive approach of CPS assessment. Further studies could consider including a representative sample of students in a wider range of ages and backgrounds. Second, the study operationalized the communication between the partners in CPS through a phrase-chat to ensure standardization and automatic scoring, while other approaches could be considered, including verbal conversations and open-chat. Third, it is possible that the comparability findings between H-A and H-H performance in other problem-solving and collaboration contexts will be different. Future studies could consider exploring differences in student performance in a wide range of problems and collaboration methods, including tasks that require adaptivity and high interdependency between the team members. Finally, the study implemented a certain set of measures and techniques to assess CPS. Various research methodologies and measures developed in previous studies of CPS, collaborative learning, and teamwork processes potentially can be adapted to CPS assessment (e.g. Biswas et al., 2005; Hsieh & O'Neil, 2002; O'Neil & Chuang, 2008; Rosen & Rimor, 2012; Weinberger & Fischer, 2006).

6. Conclusions

The CPS assessment methods described in this article offer one of the few examples today of an approach to direct, large-scale assessment targeting social and collaboration competencies. CPS brings new challenges and considerations for the design of effective assessment approaches because it moves the field beyond standard item design tasks. The assessment must incorporate concepts of how humans solve problems in situations where information must be shared and considerations of how to control the collaborative environment in ways sufficient for valid measurement of individual and team skills. The quality and practical feasibility of these measures are not yet fully documented. However, these measures can rely on the abilities of technology to engage students in interaction, to simulate others with whom students can interact, to track students' ongoing responses, and to draw inferences from those responses. Group composition is one of the important issues in large-scale assessments of collaborative skills (Webb, 1995; Wildman et al., 2012). Overcoming possible bias of differences across groups by using computer agents or other methods becomes even more important within international large-scale assessments where cultural boundaries are crossed. The results of this study suggest that by using computer agents in a CPS task the students were able to show their collaborative skills at least at the level of that of their peers who collaborated with human partners. However, each approach to assessment of collaboration still involves limitations and challenges that must be considered in the design of the assessments. Further research can continue to establish comprehensive validity evidence and generalization of findings both in H-A and H-H CPS settings.

References

Adejumo, G., Duimering, R. P., & Zhong, Z. (2008). A balance theory approach to group problem solving. *Social Networks, 30*(1), 83–99.

Avouris, N., Dimitracopoulou, A., & Komis, V. (2003). On analysis of collaborative problem solving: An object-oriented approach. *Computers in Human Behaviour, 19*(2), 147–167.

Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, A. R. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology-Enhanced Learning, 5*(2), 123–152.

Biswas, G., Leelawong, K., Schwartz, D., & Vye, N. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence, 19*, 363−392.

Cai, Z., Graesser, A. C., Forsyth, C., Burkett, C., Millis, K., Wallace, P., Halpern, D., & Butler, H. (2011). Trialog in ARIES: User input assessment in an intelligent tutoring system. In W. Chen & S. Li (Eds.), *Proceedings of the 3rd IEEE International Conference on Intelligent Computing and Intelligent Systems* (pp. 429–433). Guangzhou: IEEE Press.

Chen, Z., & Klahr, D., (1999). All other things being equal: Children's acquisition of the control of variables strategy. *Child Development, 70*(5), 1098–1120.

Chung, G. K. W. K., O'Neil, H. F., Jr., & Herl, H. E. (1999). The use of computer-based collaborative knowledge mapping to measure team processes and team outcomes. *Computers in Human Behavior, 15*, 463–494.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Cooke, N. J., Kiekel, P. A., Salas, E., Stout, R., Bowers, C., & Cannon-Bowers, J. (2003). Measuring team knowledge: A window to the cognitive underpinnings of team performance. *Group Dynamics: Theory, Research and Practice, 7*(3), 179–219.

Dillenbourg, P. (Ed.). (1999). *Collaborative learning: Cognitive and computational approaches*. Amsterdam, NL: Pergamon, Elsevier Science.

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

Eklöf, H. (2006). Development and validation of scores from an instrument measuring student test-taking motivation. *Educational and Psychological Measurement, 66*(4), 643–656.

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

Fiore, S., Rosen, M., Smith-Jentsch, K., Salas, E., Letsky, M., & Warner, N. (2010). Toward an understanding of macrocognition in teams: Predicting process in complex collaborative contexts. *The Journal of the Human Factors and Ergonomics Society, 53*(2), 203–224.

Fiore, S., & Schooler, J. W. (2004). Process mapping and shared cognition: Teamwork and the development of shared problem models. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 133–152). Washington, DC: American Psychological Association.

Foltz, P. W., & Martin, M. J. (2008). Automated communication analysis of teams. In E. Salas, G. F. Goodwin & S. Burke (Eds.), *Team effectiveness in complex organizations and systems: Cross-disciplinary perspectives and approaches* (pp. 411–431). New York: Routledge.

Franklin, S., & Graesser, A. C. (1996). Is it an agent or just a program? A taxonomy for autonomous agents. *Proceedings of the Agent Theories, Architectures, and Languages Workshop* (pp. 21–35). Berlin: Springer-Verlag.

Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing, 11*, 133–142.

Graesser, A. C., Foltz, P., Rosen, Y., Forsyth, C., & Germany, M. (in press). Challenges of assessing collaborative problem solving. In B. Csapo, J. Funke & A. Schleicher (Eds.), *The nature of problem solving*. OECD Series.

Graesser, A. C., Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes, 45*(4), 298–322.

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 21st century skills* (pp. 1–15). Heidelberg: Springer.

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

Laurillard, D. (2009). The pedagogical challenges to collaborative technologies. *International Journal of Computer-Supported Collaborative Learning, 4*(1), 5–20.

Leelawong, K., & Biswas, G. (2008). Designing learning by teaching systems: The Betty's Brain System. *International Journal of Artificial Intelligence in Education, 18*(3), 181–208.

Mayer, R. E., & Wittrock, M. C. (2006) Problem solving. In P. A. Alexander & P. Winne (Eds.), *Handbook of educational psychology (2nd ed.)* (pp. 287–304). Mahwah, NJ: Lawrence Erlbaum Associates.

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, A. Oikonomou & J. Lakhmi (Eds.), *Serious games and edutainment applications* (pp. 169–195). London: Springer-Verlag.

National Research Council. (2011). *Assessing 21st century skills*. Washington, DC: National Academies Press.

OECD. (2013). *PISA 2015 Collaborative Problem Solving Framework*. OECD Publishing.

O'Neil, H. F., Jr., & Chuang, S. H. (2008). Measuring collaborative problem solving in low-stakes tests. In E. L. Baker, J. Dickieson, W. Wulfeck & H. F. O'Neil (Eds.), *Assessment of problem solving using simulations* (pp. 177–199). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr., Chuang, S. H., & Baker, E. L. (2010), Computer-based feedback for computer-based collaborative problem solving. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 261–279). New York: Springer-Verlag.

O'Neil, H. F., Chung, G., & Brown, R. (1997). Use of networked simulations as a context to measure team competencies. In H. F. O'Neil, Jr. (Ed.), *Workforce readiness: Competencies and assessment* (pp. 411–452). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr., Chen, H. H., Wainess, R., & Shen, C. Y. (2008). Assessing problem solving in simulation games. In E. L. Baker, J. Dickieson, W. Wulfeck & H. F. O'Neil (Eds.), *Assessment of problem solving using simulations* (pp. 157–176). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr., Chung, G. K. W. K., & Brown, R. (1997). Use of networked simulations as a context to measure team competencies. In H. F. O'Neil, Jr., (Ed.), *Workforce readiness: Competencies and assessment* (pp. 411–452). Mahwah, NJ: Erlbaum.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2004). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Rimor, R., Rosen, Y., & Naser, K. (2010). Complexity of social interactions in collaborative learning: The case of online database environment. *Interdisciplinary Journal of E-Learning and Learning Objects, 6*, 355–365.

Rosen, Y. (2009). Effects of animation learning environment on knowledge transfer and learning motivation. *Journal of Educational Computing Research, 40*(4), 439–455.

Rosen, Y., & Beck-Hill, D. (2012). Intertwining digital content and one-to-one laptop learning environment. *Journal of Research on Technology in Education, 44*(3), 223–239.

Rosen, Y., & Rimor, R. (2012). Teaching and assessing problem solving in online collaborative environment. In R. Hartshorne, T. Heafner & T. Petty (Eds.), *Teacher education programs and online learning tools: Innovations in teacher preparation* (pp. 82–97). Hershey, PA: Information Science Reference, IGI Global.

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

Sundre, D. L. (1999, April). *Does examinee motivation moderate the relationship between test consequences and test performance?* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Sundre, D. L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and nonconsequential test performance? *Contemporary Educational Psychology, 29*(1), 6–26.

U.S. Department of Education (2010). *Transforming American education – Learning powered by technology: National Education Technology Plan 2010*. Washington, DC: Office of Educational Technology, U.S. Department of Education.

Vollmeyer, R., & Rheinberg, F. (1999). Motivation and metacognition when learning a complex system. *European Journal of Psychology of Education, 14*, 541–554.

Webb, N. M. (1995). Group collaboration in assessment: Multiple objectives, processes, and outcomes. *Educational Evaluation and Policy Analysis, 17*(2), 239–261.

Webb, N. M., Nemer, K. M., Chizhik, A. W., & Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal, 35*(4), 607–651.

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

Wildman, J. L., Shuffler, M. L., Lazzara, E. H., Fiore, S. M., Burke, C. S., Salas, E., & Garven, S. (2012). Trust development in swift starting action teams: A multilevel framework. *Group & Organization Management, 37*(2), 138–170.

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*(1), 1–17.

Wise, S. L., & DeMars, C. E. (2006). An application of item response time: The effort-moderated IRT model. *Journal of Educational Measurement, 43*(1), 19–38.

Zhang, J. (1998). A distributed representation approach to group problem solving. *Journal of the American Society for Information Science, 49*, 801–809.

Appendix A. Simplified Sample Log File for Student Performance on the Collaborative Problem Solving Task

The "Student" was in charge of the selections of the conditions and could communicate with an "Agent" through a phrase-chat, while the Agent could see the selections made by the Student and could communicate through a phrase chat. It should be noted that an Agent could be another student (H-H setting) or a computer agent (H-A setting).

ATTEMPT	**1**
Student	*Hi, let's work together. I'm happy to work with you. (Emoticon = smile)*
Agent	Hi, let's go for it! (Emoticon=smile)
Student	*What kind of food does he eat? (Emoticon = question)*
Agent	I think that we can try seeds first. (Emoticon=smile)
Student	*How about the extra features?*
Agent	I think that we can try water first.
Student	*OK. Let's go for it.*
Agent	What about the other conditions for Artani?
Student	*Let's try [Aquatic] for environment. (Emoticon=smile)*
Agent	OK. Let's try and see what happens (Emoticon=smile)
Student	*Are you ready to go ahead with our plan?*
Agent	Click Go if you are ready to tryout the conditions
CHOICES	**Aquatic, Seeds, Water**
RESULT	**10 [animal life expectancy in years]**
PROMPT	Artani will be unhappy with the current conditions. I'd suggest you to reconsider some of your selections and provide Artani with conditions more suitable to its needs.
ATTEMPT	**2**
Student	*Let's try [Rainforest] for environment*
Agent	I don't think that [Rainforest] is the best choice for environment
Student	*What can we do to reach better conditions for the animal?*
Agent	Let's try to change one condition per trial.
CHOICES	**Rainforest, Seeds, Water**
RESULT	**18 [animal life expectancy in years]**
PROMPT	Artani will be very happy with the current conditions. Still, I would suggest you to reconsider some of your selections and provide Artani with the best conditions to its needs.
ATTEMPT	**3**
Student	*Should we keep this selection or try again?*
Agent	The target is 26 years, let's get as close as we can to that!

Student	*Let's try [Plants] for food.*
CHOICES	**Rainforest, Plants, Water**
RESULT	**20 [animal life expectancy in years]**
PROMPT	Artani will be very happy with the current conditions. Still, I would suggest you to reconsider some of your selections and provide Artani with the best conditions to its needs.
ATTEMPT	**4**
Student	*Let's try [Tire swing] for extra features. (Emoticon=question)*
Agent	I don't think that [Tire swing] is the best choice for extra feature
CHOICES	**Rainforest, Plants, Tire swing**
RESULT	**26 [animal life expectancy in years]**
PROMPT	A job well done! You've chosen the best possible conditions for Artani.

In ATTEMPT 1, the Student starts the collaboration with a short introduction ("Hi, let's work together. I'm happy to work with you"), and then asks the Agent to propose a solution ("What kind of food does he eat?"). The Student confirms the solution proposed by the Agent ("OK. Let's go for it."), makes the selections according to the proposed solution and submits the solution. However, the result is not optimal (10 years).

During ATTEMPT 2, the Student proposes a solution ("Let's try [Rainforest] for environment"). However, the Agent disagrees with the solution ("I don't think that [Rainforest] is the best choice for environment"). This situation puts the Student in a challenging conflict-type situation that is essential in CPS assessment. The Student can ignore the fact that the Agent disagrees with the proposed solution and make the selections without reaching a consensus, or try to better understand Agent's point of view or propose a different solution. In this case, the Student decides to better understand Agent's perspective by asking for advice ("What can we do to reach better conditions for the animal?"). This request is followed by an advice from the Agent ("Let's try to change one condition per trial."). Based on the advice, the Student submits the selections. Although, the result shows a progress it is not optimal (18 years).

In ATTEMPT 3, the Student continues with providing the Agent with the possibility to propose a new solution ("Should we keep this selection or try again?"). Then the Student makes a decision on the selections and executes the solution that leads to almost an optimal result (20 years).

ATTEMPT 4, the Student proposes a solution in order to optimize the result ("Let's try [Tire swing] for extra features"). However, the Agent does not agree with the proposed solution ("I don't think that [Tire swing] is the best choice for extra feature"). Then the Student makes the choice without reaching an integrative consensus with the Agent. Despite this conflict-type situation the optimal solution is reached (26 years).

END of TASK: The Student is prompted ("A job well done! You've chosen the best possible conditions for Artani.").