Computer-based Assessment of Collaborative Problem-Solving Skills: Human-to-Agent versus Human-to-Human Approach

Research Report

Yigal Rosen, Ph.D.
Maryam Tager

March 2013
About Pearson
Pearson, the global leader in education and education technology, provides innovative print and
digital education materials for pre-K through college; student information systems and learning
management systems; teacher licensure testing; teacher professional development; career
certification programs; and testing and assessment products that set the standard for the
industry. Pearson’s other primary businesses include the Financial Times Group and the
Penguin Group. For more information about the Assessment & Information group of Pearson,
visit http://www.pearsonassessments.com/.

About Pearson’s Research Reports
Pearson’s research report series provides preliminary dissemination of reports and articles
prepared by Test, Measurement, and Research Services (TMRS) staff, usually prior to formal
publication. Pearson’s publications in pdf format may be obtained at
Abstract

How can environments and activities in which collaborative skills of an individual are measured be standardized? In order to understand how students perform on collaborative problem-solving (CPS) computer-based assessment, it is necessary to examine empirically the multifaceted performance that may be distributed across environments, collaboration methods, and problem-solving spaces. The aim of this study was to explore patterns in student CPS performance and motivation in human-to-agent (H-A) settings compared to human-to-human (H-H) settings. One hundred seventy-nine 14-year-old students from the United States, Singapore, and Israel participated in the study. Students in both H-A mode and H-H mode 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; in the H-H mode, students collaborated with other students to solve a problem. Overall, the findings showed that in standardized CPS settings, students assessed in H-A mode outperformed their peers in H-H mode in collaborative skills. CPS with a computer agent involved significantly higher levels of shared understanding, progress monitoring, and feedback. However, no significant difference was found in a student’s ability to solve the problem and in student motivation with a computer agent or a human partner. This study is among the first of its kind to investigate systematically the effect of collaborative problem solving in H-A and H-H standardized assessment settings. Directions for future research are discussed in terms of their implications for large-scale assessment programs, teaching, and learning.

Keywords: collaborative problem solving, computer-based assessment, assessment in groups
Collaborative Problem-Solving Assessment

Computer-based Assessment of Collaborative Problem-Solving Skills:

Human-to-Agent Approach versus Human-to-Human Approach

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 have CPS skills as well as the ability to perform that collaboration in 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. In a standardized assessment situation, a student should be matched with various types of group members that will represent different CPS skills and contexts. In addition, the discourse between the group members should be manageable and predictable. The three major questions thus are: How can environments and activities in which collaborative skills of an individual are measured be standardized? Can partners for CPS be simulated but still maintain authentic human aspects of collaboration? And, how can manageable and predictable group discourse spaces be created within the assessment? This paper addresses these challenges by introducing a new methodology for scalable computer-based assessment of CPS, proving findings from an empirical pilot study conducted in three counties, as well as discussing limitations and implications of the findings on further research and development.
Collaborative Problem-Solving Assessment

Defining Collaborative Problem Solving

Currently, the terms collaborative problem solving, cooperative work, and group work are used interchangeably in the education research literature to mean similar constructs. Collaborative problem solving refers to problem-solving activities that involve collaboration among a group of individuals (O’Neil, Chuang, & Baker, 2010; Zhang, 1998). CPS is a conjoint construct that consists of collaboration: “coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem” (Roschelle, & Teasley, 1995, p. 70); and problem solving: “cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver” (Mayer, & Wittrock, 1996). According to Griffin, Care, and McGaw (2012), CPS refers to the abilities to recognize the points of view of other persons in a group; contribute knowledge, experience, and expertise in a constructive way; identify the need for contributions and how to manage them; recognize structure and procedure involved in resolving a problem; and as a member of the group, build and develop group knowledge and understanding. CPS is one of the two major areas that the Organisation for Economic Co-operation and Development (OECD) selected in 2015 for primary development in the Program for International Student Assessment (PISA). In PISA 2015, 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). An agent could be considered either a human or a computer agent that interacts with the student. The competency is an assessment on how well the individual collaborates with agents during the problem-solving process. This includes establishing and
maintaining shared understanding, taking appropriate actions to solve the problem, and establishing and maintaining group organization. CPS-proficient students must 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. Students should be able to identify the type of CPS activities that are needed to solve the problem and to follow the appropriate steps to achieve a solution. 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.

In our research, an operational definition of CPS refers to the capacity of an individual to effectively engage in a group process whereby two or more agents attempt to solve a problem by sharing knowledge and understanding, organizing the group work and monitoring the progress, taking actions to solve the problem, and providing constructive feedback to group members. CPS skills identified by this research attempt to incorporate skills identified in other CPS frameworks, such as the PISA (OECD, 2013), ATC21S (Griffin, Care, & McGaw, 2012), and teamwork processing model (O’Neil, Chuang, & Baker, 2010).

**Assessing Collaborative Problem-Solving Skills**

Student performance in CPS can be assessed through a number of different methods. These include measures of the quality of the solutions and the objects generated during the
collaboration, analyses of log files, intermediate results, paths to the solutions, team processes and structure of interactions, and quality and type of collaborative communication. There are distinct tradeoffs between the large amount of information that can be collected in a collaborative activity and what can be measured. 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 will be situated similarly to show his or her proficiency in CPS. Educators are urged to carefully consider group composition when creating collaborative groups or teams (Fall, Webb, & Chudowsky, 1997; Rosen, & Rimor, 2009; Webb, 1995; Wildman et al., 2012). Additionally, students should act in different roles (e.g., team leader) and be able to work collaboratively in various types of environments. Gender, race, status, perceived cognitive or collaborative abilities, motivation, and attractiveness are other factors that may influence student CPS. 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 degree of interactivity and negotiability is an additional indicator 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; students 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).

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 (Fiore, & Schooler, 2004; Dillenbourg, & Traum, 2006; Fiore et al., 2010). Various techniques were developed to address the challenge of providing a tractable way of communication in the CPS assessment context. One interesting technique that has been tested is communication through predefined messages (Chung, O’Neil, & Herl, 1999; Hsieh, & O’Neil, 2002; O’Neil, Chung, & Brown, 1997). In these studies, participants were able to communicate using the predefined messages 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). Using messages provides a manageable way of measuring CPS skills as well as allows real-time scoring and reporting.

**Computer Agent Approach in CPS Assessment**

Collaboration can take many forms, ranging from two individuals to large teams with predefined roles. Thus, there are a number of dimensions that can affect the type of collaboration and the processes used in problem solving. For example, there can be different-sized teams (two
equal team members vs. three or more team members working together), different types of social hierarchies within the collaboration (all team members equal vs. team members with different levels of authority), and, for assessment purposes, different agents – whether all team members are human or some are computer agents. There are advantages and limitations for each method. 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, pairing 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 enough opportunity to cover variations in group composition, diversity of perspectives and different team member characteristics in controlled manners, which are all essential for assessment on an individual level. Simulated team members for collaboration with preprogrammed profiles, actions, and communication would potentially ensure that the assessment could cover 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(s) that can be programmed to act as a team member(s) with varying characteristics relevant to different CPS situations. Research shows that computer agents can be successfully used for tutoring, collaborative learning, co-construction of knowledge, and CPS (e.g., Biswas et al., 2010; Graesser, Jeon, & Dufty, 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 of a computer agent at use in education is a teachable agent system called
Betty’s Brain (Biswas et al., 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 the agent’s performance as motivation and a guide, a student studies the available resources so that he or she can remediate the agent’s knowledge, and, in this process, learn the domain material for him or herself. Operation ARIES (Cai et al., 2001; Millis et al., 2011) uses animated pedagogical agents that converse with the student in a game-based environment to help the student 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,” take place between the human learner and two computer agents (student and teacher). The student learns vicariously by observing the agents, receiving tutoring from the teacher agent, and teaching the student agent.

Group processes are often different depending on the task and could even be competitive. Using computer agents provides an essential component of non-competitiveness to the CPS situation, as experienced by a student. Additionally, if the time-on-task is limited, taking the time to explain to each other may lower group productivity. As a result of these perceived constraints, 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 e-mails or chat) or graphic organizers with lists of named agents can be used for H-A CPS purposes.

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 a 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. Computer-based assessment of CPS involves the need for advancements in educational assessment methodologies and technology. Group composition, discourse management, and the use of computer agents are considered the major challenges in designing valid, reliable, and scalable assessment of CPS skills (Graesser, et al., in press). This paper addresses these challenges by studying student CPS performance in two modes of CPS assessment.

**Research Questions**

The study addressed empirically the following questions regarding student CPS performance and motivation in H-A and H-H CPS settings:

1. 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 proving feedback measures?

2. What are the differences in CPS performance within H-A and H-H modes of assessment between female and male students as reflected in shared understanding, problem solving, progress monitoring, and proving feedback measures?
3. How are a student’s GPA, ELA and Math school achievement, as reported by the students, related to student performance in CPS in H-A and H-H modes of assessment?

4. What are the differences in student motivation while collaborating with a computer agent or a human partner on a CPS assessment task?

5. 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?

**Method**

The study participants included 179 students, all 14 years old and 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 studying innovative ways of developing computer-based assessment in critical thinking, creativity, and collaborative problem solving. The researchers collected data through November 2012 to January 2013. Recruitment of participating schools was achieved through collaboration with local educational organizations based on the following criteria: (a) the school is actively involved in various 21st Century Skills projects, (b) population of 14-year-old students proficient in English, and (c) sufficient technology infrastructure (e.g., computers per student, high-speed Internet). Specifically in H-H assessment mode, students were randomly divided into pairs to work on the CPS task. In all, 136 students participated in H-A mode, and 43 participated in H-H mode.

Of the total students who participated, 88 were boys (49.2%) and 91 were girls (50.8%). Table 1 summarizes the country and gender distribution of participating students between the H-A and H-H groups.
Table 1

*Research population by mode of assessment and country*

<table>
<thead>
<tr>
<th>Group</th>
<th>H-A</th>
<th>H-H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>United States</td>
<td>22</td>
<td>19</td>
</tr>
<tr>
<td>Singapore</td>
<td>12</td>
<td>31</td>
</tr>
<tr>
<td>Israel</td>
<td>30</td>
<td>22</td>
</tr>
<tr>
<td>Overall</td>
<td>64</td>
<td>72</td>
</tr>
</tbody>
</table>

No significant differences were found in GPA, ELA, and Math average scores between participants in H-A and H-H modes within the countries. This similarity in student background allowed comparability of student results in CPS assessment tasks between the two modes of collaboration.

**Collaborative Problem-Solving Assessment**

In this CPS computer-based assessment task, the student was asked to collaborate with a partner (a computer-driven agent or a classmate) to find the optimal conditions for an animal at a zoo. 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 4 to 5 predefined options). The animal’s life expectancy under the given conditions is presented after each trial of the conditions. The student and the partner were prompted to discuss how to reach better conditions for an animal at the beginning of the task. By the end of the task, the student was asked to rate the partner (1 to 3 stars) and provide written
feedback on his or her 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 the number of attempts to reach an optimal solution or 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 in order to complete the assessment task.

The task was checked by teachers from the three participating countries to ensure that students would be able to work on the task, that the task could differentiate between high and low levels of CPS ability, and that the task was free of cultural biases. Interviews were conducted with students representing the target population to validate various CPS actions, and communication was programmed for the computer agent as well as automatic scoring of student responses.

The following information was presented interactively to the student during the task:

Episode #1: “It was a normal zoo... But suddenly disaster struck! A rare animal called an Artani was found dead! You and a friend have been tasked with saving the other Artani by finding the most suitable conditions for them within the Zoo.”

Episode #2: “Collaborate with your partner, help the Artani: In this task you will be in control of the selections made by you and your friend. Work with your partner to determine the best living conditions for Artani. You must change the three elements of Environment, Food and Extras to find the best living conditions. Your friend can help you plan a strategy to improve the conditions for Artani, before you make your selection.”
Episode #3: 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 can 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 cover the CPS measures defined for the task. The computer agent’s 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. While in the H-H mode, the partners were provided with exactly the same set of possible phrases for each CPS situation; there was no control over the phrases or the responses selected by the students.
Figure 1. Selecting variables in a CPS assessment task

Figure 2. Communication through a phrase-chat in a CPS assessment task
Collaborative Problem-Solving Assessment

Figure 3. Viewing the results of variables selected in a CPS assessment task

Clicking on “Go” provided the partners with the possibility to see the life expectancy of the animal under the variables selected (0 to 26 years) and read textual information regarding the result achieved (see Figure 3). At this stage, the partners were allowed to communicate about the result, ways to reach the optimal solution, and whether to keep the selections or try again (i.e., change the variables). Figure 4 presents a simplified sample log file for student performance on the task.
Collaborative Problem-Solving Assessment

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 it is the best choice.  

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  
OK. Let's try and see what happens.  

Student  
Let's go ahead with our plan?  

Agent  
Click Go if you are ready to tryout the conditions.  

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.  

Figure 3. Communication through a phrase-chat in a CPS assessment task
Episode #4: “Give feedback to your partner: Having saved the Artani, you need to provide feedback on your partner. Give you partner a star rating (1 to 3) and add written feedback below.”

Episode #5: “Partner's feedback: Thanks for your feedback. You have done a great job on your part! Hope to work with you again sometime.”

Students could achieve an overall CPS score of up to 100-points for the assessment task. The overall score consisted of shared understanding (40 points), problem solving (26 points), monitoring progress (26 points), and providing feedback (8 points). Student performance in the CPS assessment task was scored automatically based on predefined programming that was embedded into the assessment. Scoring of the student feedback dimension was provided independently by two teachers from participating schools in the United States. 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.

Collaborative Problem-Solving Questionnaire

The questionnaire included 4 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 4-point Likert scale (1 = strongly disagree, 4 = strongly agree). The items were adopted from motivation questionnaires used in previous studies, included (Rosen, 2009; Rosen, Beck-Hill, 2012): I felt interested in the task; The task was fun; The task was attractive; I continued to work on this task out of curiosity. The reliability (internal consistency) of the questionnaire was .85.
Students were also asked to indicate the background information, including gender, and Grade Point Average (GPA), Math and English Language Arts (ELA) average score. It should be noted that this informant was provided directly by the students and was not validated with the teachers. The background information was collected because of potential interaction with study variables.

**Results**

All results are presented on an aggregative level beyond the countries because no interaction with country 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 classmate. Next, gender-related results are presented to indicate possible differences in student performance in H-A and H-H modes, as well as the relationship with a student’s school achievement. Last, student motivation, time-on-task, and number of attempts to solve the problem in both modes of collaboration are demonstrated.

**Student CPS Performance**

In order to explore possible differences in students’ CPS scores analysis of variance was performed. First, MANOVA results showed significant difference between H-H and H-A groups (Wilks’ Lambda=.904, \( F(df=4,174)=4.6, \ p<.01 \)). Hence, we proceed to perform t-tests. The results of the CPS scores indicated that students who collaborated with a computer agent on an assessment task significantly outperformed the students who worked with a classmate in overall score achieved (M=52.1, SD=18.2 in H-A mode, compared to M=44.2, SD=10.9 in H-H mode; ES=.5, \( t(df=177)=3.5, \ p<.01 \)). Students who collaborated with a computer agent showed significantly higher levels of performance in establishing and maintaining shared understanding.
Collaborative Problem-Solving Assessment

(ES=.4, \( t(df=177)=2.5, \ p<.05 \)), monitoring progress of solving the problem (ES=.6, \( t(df=177)=4.0, \ p<.01 \)), and in the quality of the feedback (ES=.5, \( t(df=177)=3.2, \ p<.01 \)). The findings showed a non-significant difference in the ability to solve the problem in the H-A mode versus the H-H mode (ES=-.3, \( t(df=177)=-1.9, \ p=.06 \)). Table 2 shows the results of student CPS scores in both modes.

Table 2

<table>
<thead>
<tr>
<th>CPS Measure</th>
<th>H-A</th>
<th>H-H</th>
<th>ES</th>
<th>t(df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared understanding</td>
<td>18.4(10.8)</td>
<td>14.6(8.3)</td>
<td>.4</td>
<td>2.4(177)*</td>
</tr>
<tr>
<td>Problem solving</td>
<td>24.6(3.2)</td>
<td>25.3(1.7)</td>
<td>-.3</td>
<td>-1.9(177)</td>
</tr>
<tr>
<td>Monitoring progress</td>
<td>5.0(7.5)</td>
<td>1.3(4.2)</td>
<td>.6</td>
<td>4.0(177)**</td>
</tr>
<tr>
<td>Providing feedback</td>
<td>4.1(2.2)</td>
<td>3.0(1.9)</td>
<td>.5</td>
<td>3.2(177)**</td>
</tr>
</tbody>
</table>

** \( p <.01 \), * \( p <.05 \).

Gender and CPS Performance

The findings indicated that female students outperformed males in the overall CPS score both in H-A and H-H modes of collaboration, although only the difference in the H-A mode was statistically significant (M=55.9, SD=18.5, in female students, compared to M=48.7, SD=17.4 in males; ES=.4, \( t(df=134)=2.3, \ p<.05 \)). As shown in Table 3, a significant difference was found in favor of female students working in the H-A mode in their ability to establish and maintain shared understanding (ES=.4, \( t(df=134)=2.1, \ p<.05 \)), monitor progress during the process of problem solving (ES=.3, \( t=2.0, \ p<.05 \)), and provide feedback to a computer agent (ES=.5, \( t(df=134)=3.0, \ p<.01 \)).
Table 3

*Student CPS scores in H-A mode of the assessment by gender*

<table>
<thead>
<tr>
<th>CPS Measure</th>
<th>Female M(SD)</th>
<th>Male M(SD)</th>
<th>ES</th>
<th>t(df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared understanding</td>
<td>20.5(10.9)</td>
<td>16.6(10.5)</td>
<td>.4</td>
<td>2.1(134)*</td>
</tr>
<tr>
<td>Problem solving</td>
<td>24.4(3.3)</td>
<td>24.8(3.1)</td>
<td>-.1</td>
<td>.8(134)</td>
</tr>
<tr>
<td>Monitoring progress</td>
<td>6.4(8.5)</td>
<td>3.7(6.4)</td>
<td>.3</td>
<td>2.0(134)*</td>
</tr>
<tr>
<td>Providing feedback</td>
<td>4.7(1.9)</td>
<td>3.6(2.4)</td>
<td>.5</td>
<td>3.0(134)**</td>
</tr>
</tbody>
</table>

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

*Student School Achievement and CPS Performance*

Correlations between the variables were conducted in order to determine potential relationships between student-reported GPA, ELA achievement, and Math achievement as measured by traditional school assessments and student performance in CPS in H-A and H-H modes of assessment. The findings showed a significantly positive relationship between ELA student score and CPS score on providing feedback both in H-A and H-H mode (r=.42, p < .01 and r=33, p < .05, respectively). No significant correlations were found between student achievement and other dimensions of CPS performance.

*Student Motivation, Attempts to Solve a Problem, and Time-on-Task*

Data was analyzed to determine possible differences in student motivation of being engaged in CPS with a computer agent versus a classmate. The results demonstrated that it did not matter for student’s motivation whether he or she collaborated 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,
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), was greater than the average of 6.1 (SD=5.7) attempts for the H-H mode (ES=.3, $t(df=177)=2.1$, $p<.05$). No significant difference was found in time-on-task (ES=-1.9, $t(df=177)=-1.6$, $p=.11$). On average, time-on-task in H-A mode was 7.9 minutes (SD=3.6), while students in the H-H mode each spent an additional 1.1 minutes on the task (M=9.0, SD=4.5).

**Discussion**

Policymakers, researchers, and educators are engaged in vigorous debate over assessing CPS skills on an individual level in valid, reliable, and scalable ways. Analyses of the list of challenges facing CPS in large-scale assessment programs suggests that the computer-agent approach in CPS assessment is one of the most promising approaches that should be further explored. The goal of this study was to explore patterns in student CPS performance in H-A and H-H modes. Students in each of these modes were exposed to an 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; 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 is flexibly adaptive to the point where no two conversations are 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 because each student in H-H mode represented a specific set of CPS skills, while in 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). Further analyses have shown 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 (Rosen, & Tager, 2013). 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 the 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 uniquely 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 more powerful at informing 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 a certain CPS level 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, a phrase-chat used in this study can be replaced by an open-chat in cases where automated scoring of student responses is not needed.

In contrast, we found that it is a matter of indifference in student ability 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. Student performance studied here was in the context of well-structured problem solving, while primarily targeting collaborative dimensions of CPS. More challenging problem-solving space of the CPS task, can lead to differential results in H-A and H-H settings. So an important question is how well the results found here would generalize to an ill-structured problem-solving assessment context. Specifically, does the similarity 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 et al., 1998).
With respect to possible differences in CPS performance within H-A and H-H modes of assessment between female and male students, we found that female students outperformed males in the overall CPS score both in H-A and H-H modes of collaboration, but the difference was statistically significant in only the H-A mode. Female students who worked in H-A mode were able to better establish and maintain shared understanding in the group, monitor progress during the process of problem solving, and provide feedback to a computer agent. These findings are consistent with previous studies that have shown female and male students’ to have different collaboration and communication styles (Lay, 1992; Li, 2002; Prinsen, Volman, & Terwel, 2007; Underwood, Underwood, & Wood, 2001). For example, male students tend to give their ideas more directly than female students, whereas females are more likely to initiate conversation by asking questions and establishing a common ground among participants. Research results suggest that gender can impact both patterns of interaction and the effects of collaboration; educators should attempt to create groups of students that can work productively while considering H-H assessment of CPS.

Further analysis examined the extent to which student GPA, ELA achievement, and Math achievement, as reported by the students, are related to student performance in H-A and H-H modes of assessment. We found evidence for a positive relationship between ELA and CPS student scores on providing feedback both in H-A and H-H modes, but no significant correlations were found between student achievement and other dimensions of CPS performance. These results suggest that the current CPS approaches for assessment both in H-A and H-H modes are distinctive from measurement of the conventional domains in schools. The problem space and communication component operationalized by phrase-chat allowed each student to communicate with a group member regardless his or her reading, writing, or math skills.
However, based on the results of the study, the ability to provide constrictive feedback to a partner for collaboration may require a higher level of writing skills. Again, it should be noted that this informant was provided directly by the students and was not validated with the teachers.

Concerning 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 their collaborative work with computer agents at the same level as 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 the CPS assessment task. 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).

The current study had several limitations. First, it is based on a relatively small and non-representative sample of 14-year-old students in three countries and self-reported school achievement. 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
more comprehensive approaches of CPS assessment. Further studies could consider including a representative sample of students with a wider range of ages and backgrounds and collecting school achievement data in a more controlled manner.

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 performances 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. 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 can be potentially adapted to CPS assessment (e.g., Biswas, et al., 2005; Hsieh, & O’Neil, 2002; O’Neil, & Chuang, 2008; Rosen, & Rimor, 2012; Weinberger, & Fischer, 2006).

Conclusions

CPS assessment methods described in this article offer one of the few examples today of a 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 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, as discussed in this article, each mode of collaboration involves limitations and challenges. Further research is needed in order to establish comprehensive validity evidence and generalization of findings both in H-A and H-H CPS settings.
References


