

New Methods in Online Assessment of Collaborative Problem Solving and Global Competency

Yigal Rosen (yigal.rosen@pearson.com)

Pearson Research & Innovation Network, USA

and

Maryam Mosharraf (maryam.tager@pearson.com)

Pearson Research & Innovation Network, USA

Abstract

Today, proficiency in collaborative problem solving and global competency is requisite for success in college and workplace. Collaborative problem solving and global competency are the two major areas that the Organisation for Economic Co-operation and Development (OECD) nominated in 2015 and 2018 for major development in Programme for International Student Assessment (PISA) in addition to scientific literacy, math and reading literacy. In PISA 2015, the competency is assessed by evaluating how well the individual collaborates with agents during the problem-solving process. At the same time, structuring computer-based of collaborative problem solving and global competency assessment, specifically for large-scale programs, is challenging. In a standardized assessment situation, a student should be matched with various types of group members that will represent different skills, cultures and contexts. In addition, the discourse between the group members should be manageable and predictable. This paper addresses these challenges by introducing new methodologies for scalable human and computer-agent based online assessment of collaborative problem solving and global competency, and discussing findings from an empirical pilot study conducted in the United States, Singapore, and Israel. Directions for future research are discussed in terms of their implications for large-scale computer-based assessment programs, teaching, and learning.

Keywords: collaborative problem solving, performance assessment, global competency

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 two major questions thus are: 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, providing findings from an empirical pilot study conducted in three countries, as well as discussing implication of the findings on global competency assessment.

Defining Collaborative Problem Solving

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 Programme 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 agent or a computer agent that interacts with the student. The competency is assessed by evaluating 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.

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.” 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, as modeled in Common Ground Theory (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).

Assessing Collaborative Problem Solving Skills

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 a preprogrammed profile, actions and communication would 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 an essential component of non-competitiveness to the CPS situation, as it is 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, 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 CPS purposes. However, CPS in H-A settings deviate from natural human communication delivery and can cause distraction and sometimes irritation. 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. If educators rely on CPS teaching of students in H-A interactions exclusively, there may be the risk that these students will build up expectations that do exactly follow such an algorithm.

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. 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 as the major challenges in designing valid, reliable, and scalable assessment of CPS skills (Graesser, et al., in press). The paper addresses these challenges by studying student CPS performance in two modes of CPS assessment.

Research Questions

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

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

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

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?

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 is public, (b) the school is actively involved in various 21st Century Skills projects, (c) the population is 14 years-old students proficient in English, and (c) there is 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. Of the total students who participated, 88 were boys (49.2%) and 91 were girls (50.8%). 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 allowed comparability of student results in CPS assessment task between the two modes of collaboration.

In this 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. 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). An animal’s life expectancy under the given conditions was 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 (1-3 stars) and 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. The task was checked with ten 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 eight students representing the target population to validate various CPS actions and communication programmed for the computer agent and to establish automatic scoring of student responses.

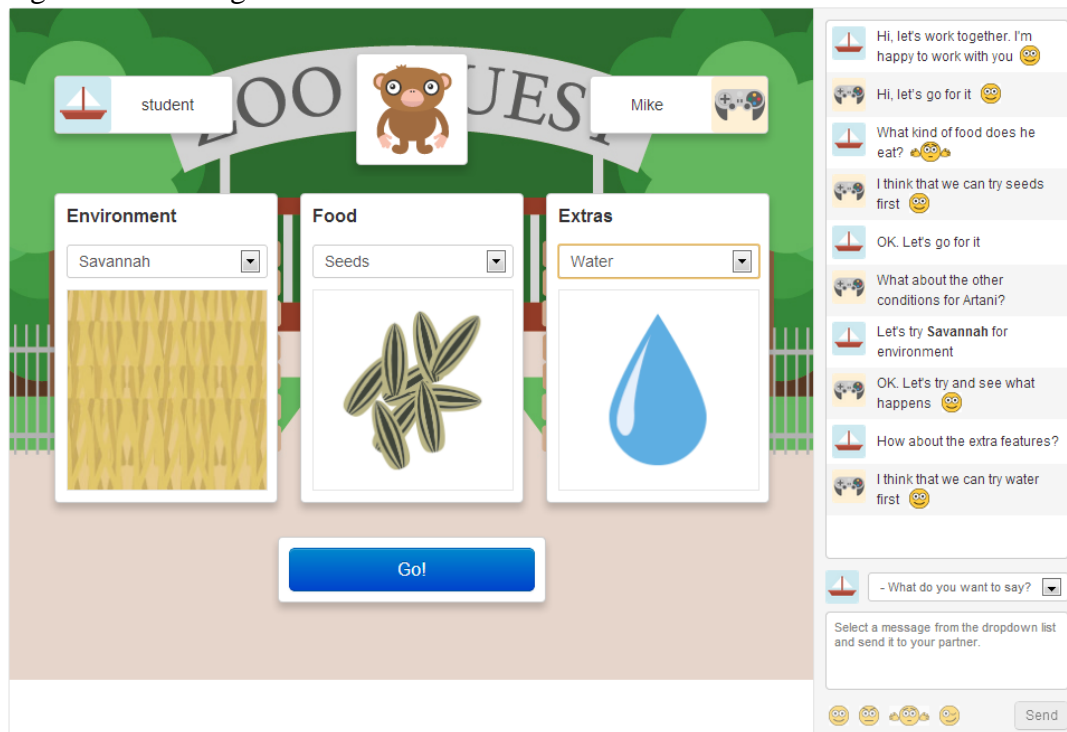
CPS scores for the assessment task consisted of shared understanding (40 points), problem solving (26 points), monitoring progress (26 points), and providing feedback (8 points). Both in H-H and H-A settings, student scores in the first three CPS dimensions were generated automatically based on a predefined programmed sequence of possible optimal actions and communication that was embedded into the assessment task. 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 number of grounding questions that were initiated by a student in appropriate situations (e.g., explaining the reason for a variable selection, 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 student feedback was provided independently by two teachers from participating schools in the United States. The teachers were trained through a one-day workshop to consistently evaluate whether student’s feedback indicated both successful and challenging aspects of working with the partner on the task, and acknowledged the contributions the partner made toward reaching a solution. Spelling and grammar issues did not affect student score. Overall, the scoring strategy was discussed with a group of ten teachers from participating countries in order to achieve consensus on CPS scoring strategy and reduce cultural biases as much as possible. Inter-coded agreement of feedback scoring was 92%.

It should be noted that the ‘collaborator’ student’s performance in H-H setting was not scored because of the non-comparability of this performance to the full CPS actions performed by the ‘leader’ student. Figure 1 shows the sample screenshot from the task.

Figure 1. Selecting variables in a CPS assessment task



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, Beck-Hill, 2012). The reliability (internal consistency) of the questionnaire was .85.

Students were also asked to indicate the background information, including gender, GPA, and Math and ELA average score, as measured by school assessments.

Results

All results are presented on an aggregative level beyond the countries, since no interaction with student-related 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 a classmate. Next, student motivation results are presented to indicate possible differences in H-A and H-H modes. Last, time-on-task and number of attempts to solve the problem in both modes of collaboration are demonstrated.

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 indicated that students who collaborated with a computer agent showed significantly higher level of performance in establishing and maintaining shared understanding ($ES=.4$, $t(df=177)=2.5$, $p<.05$), 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 non-significant difference in the ability to solve the problem in the H-A and H-H mode of collaboration ($ES=-.3$, $t(df=177)=-1.9$, $p=.06$).

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 that it is a matter of indifference 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; $ES=.1$, $t(df=177)=.5$, $p=.64$).

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. In practice, 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 ($ES=.3$, $t(df=177)=2.1$, $p<.05$). No significant difference was found in time-on-task ($t(df=177)=-1.6$, $p=.11$). 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 ($M=9.0$, $SD=4.5$).

Discussion

Policymakers, researchers, and educators are engaged in vigorous debate about 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 both H-H and H-A approaches in CPS assessment should be further explored. The goal of this study was to explore 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 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 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 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.

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.

Conclusions

Assessment methods described in this article offer one of the few examples today of a direct, large-scale assessment targeting social and collaboration competencies. As international assessment programs design assessments for 21st century skills such as collaborative problem solving and global competencies (e.g., PISA 2015 and 2018) new technological and psychometric challenges require moving beyond standard assessment items. The assessment must incorporate concepts of how humans from different backgrounds solve problems in situations where information must be shared and considerations of how to control the assessment environment in ways sufficient for valid measurement of individual and team skills. For instance, similarly to CPS, global competency assessment tasks may require working with computer or human agents to solve problems that require intercultural understanding, empathy and perspective taking. 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. 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 settings.

References

- 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.
- 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.
- 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.
- 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.
- 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. Wulfbeck, & H. F. O'Neil (Eds.), *Assessment of*

- problem solving using simulations (pp. 177-199). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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.