Chapter 5
Challenges of Assessing Collaborative Problem Solving

Arthur C. Graesser, Peter W. Foltz, Yigal Rosen, David Williamson Shaffer, Carol Forsyth, and Mae-Lynn Germany

Abstract An assessment of Collaborative Problem Solving (CPS) proficiency was developed by an expert group for the PISA 2015 international evaluation of student skills and knowledge. The assessment framework defined CPS skills by crossing three major CPS *competencies* with four problem solving *processes* that were adopted from PISA 2012 Complex Problem Solving to form a matrix of 12 specific *skills*. The three CPS competencies are (1) establishing and maintaining shared understanding, (2) taking appropriate action, and (3) establishing and maintaining team organization. For the assessment, computer-based agents provide the means to assess students by varying group composition and discourse across multiple collaborative situations within a short period of time. Student proficiency is then measured by the extent to which students respond to requests and initiate actions or communications to advance the group goals. This chapter identifies considerations and challenges in the design of a collaborative problem solving assessment for large-scale testing.

A.C. Graesser (✉) • C. Forsyth • M.-L. Germany
University of Memphis, Memphis, TN, USA
e-mail: graesser@memphis.edu

P.W. Foltz
University of Colorado, Boulder, and Pearson, Boulder, CO, USA

Y. Rosen
Harvard University, Cambridge, MA, USA

D.W. Shaffer
University of Wisconsin, Madison, WI, USA

© Springer International Publishing AG 2018

E. Care et al. (eds.), *Assessment and Teaching of 21st Century Skills*,
Educational Assessment in an Information Age,
DOI 10.1007/978-3-319-65368-6_5

Introduction

Collaborative problem solving (CPS) was selected by the Organisation for Economic Co-operation and Development (OECD) as a new development for the Programme for International Student Assessment (PISA) for the 2015 international survey of student skills and knowledge (OECD 2013). There are excellent reasons for focusing on CPS. It is widely acknowledged that CPS is an essential skill in the home, the workforce, and the community. Indeed, much of the planning, problem solving, and decision making in the modern world is performed by teams (National Research Council 2011). The success of a team can be threatened by an uncooperative member or a counterproductive alliance, but can be strongly facilitated by a strong leader that insures that team members are all contributing. Skilled collaboration and social communication facilitate productivity in the workplace (Klein et al. 2006; Salas et al. 2008), engineering and software development (Sonnentag and Lange 2002), mission control in aviation (Fiore et al. 2014), and interdisciplinary research among scientists (Nash et al. 2003). Consequently, there is a growing discussion in national education systems for including more group-based project-based learning as well as the teaching and assessment of collaboration as part of twenty-first century skills (Brannick and Prince 1997; Griffin et al. 2012; National Research Council 2011).

One issue that repeatedly surfaces is how CPS differs from individual problem solving. Collaboration allegedly has advantages over individual problem solving because (a) there is a more effective division of labor, (b) the solutions incorporate information from multiple sources of knowledge, perspectives, and experiences, and (c) the quality of solutions is stimulated by ideas of other group members. However, the literature is mixed on whether the quality of solutions is better in a group versus a collection of individuals working independently. Problem solving solutions by a group are sometimes better than the sum of the solutions of the individual members (Aronson and Patnoe 1997; Dillenbourg 1999; Schwartz 1995; Stasser and Titus 2003; Theiner and O'Connor 2010). However, this positive emergence does not always occur when one person dominates the team or there is wasted effort in non-germane communication. Better solutions can sometimes emerge when there are differences in points of view, disagreements, conflicts, and other forms of social disequilibrium in order to minimize inferior solutions via group think (Dillenbourg 1999; Rosen and Rimor 2009). However, chronic discord may have serious negative repercussions. Success in problem solving as a group is therefore dependent on knowing how best to apply the skills at the right times. For example, the ground rules of the collaborative situation need to be understood by the group members in order to optimize group interactions and final solutions. The awareness and ability to convey to students when, why, and which aspects of collaboration are fruitful for the type of knowledge being sought may improve the quality of collaborative efforts (Mullins et al. 2011).

A core focus of CPS assessment is on the quality of the solution. In order to assess student performance in collaborative exercises, a central issue is whether the unit of assessment should be the group versus an individual within a group. That is, should the unit of statistical analyses be one individual within a particular group, the set of individuals in a group, or the solution of the group as a whole? Focus on the individual may be better for tracking individual learning and providing directed feedback. However, focus on the group can assess the more holistic emergence of the processes across the group. Researchers and practitioners undoubtedly need to consider all of these, as well as whether characteristics of the individuals in a group can predict processes and outcomes of CPS as a whole.

A major factor that contributes to the success of CPS and further differentiates it from individual problem solving is the role of communication among team members (Dillenbourg and Traum 2006; Fiore et al. 2010; Fiore and Schooler 2004). Communication is essential for organizing the team, establishing a common ground and vision, assigning tasks, tracking progress, building consensus, managing conflict, and a host of other activities in CPS. Communication further provides a window into the individual processes and team processes. Thus, communication skills are fundamental in the assessments of CPS discussed in this chapter.

Developing an assessment of CPS skills can be quite complex and multifaceted, drawing information and techniques from such fields as individual problem solving, computer-mediated collaborative work, individual and team cognition, and discourse and communication theory. This presented several challenges to the Collaborative Problem Solving Expert Group (CPEG) that developed the framework for the assessment of CPS in PISA 2015. Since this was the first time such an assessment had been developed for a large scale international test of these skills, the expert group had to incorporate facets from an emerging complex and diverse literature rather than modifying a previous assessment. This chapter provides some highlights of aspects of the CPS Framework (OECD 2013) and then focuses on particular assessment considerations and challenges.

Snapshot of Collaborative Problem Solving in PISA 2015

The following definition of CPS was articulated in the PISA 2015 framework for CPS: *Collaborative problem solving competency is 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 by pooling their knowledge, skills and efforts to reach that solution.* The unit of analysis for the competency is the individual in a group rather than the group as a whole. The competency is an assessment on how well the individual interacts with agents during the course of problem solving; this includes achieving a shared understanding of the goals and activities as well as efforts to solve the problem and pooling resources. An agent could be considered either a human or a computer agent that interacts with the student. In both cases, an agent has the capability of generating goals, performing actions, communicating messages, sensing its environment, adapting to changing environments, and learning (Franklin and Graesser 1997).

Table 5.1 Matrix of collaborative problem solving skills for PISA 2015

	(1) Establishing and maintaining shared understanding	(2) Taking appropriate action to solve the problem	(3) Establishing and maintaining team organisation
(A) Exploring and understanding	(A1) Discovering perspectives and abilities of team members	(A2) Discovering the type of collaborative interaction to solve the problem, along with goals	(A3) Understanding roles to solve the problem
(B) Representing and formulating	(B1) Building a shared representation and negotiating the meaning of the problem (common ground)	(B2) Identifying and describing tasks to be completed	(B3) Describing roles and team organisation (communication protocol/rules of engagement)
(C) Planning and executing	(C1) Communicating with team members about the actions to be/being performed	(C2) Enacting plans	(C3) Following rules of engagement (*e.g.,* prompting other team members to perform their tasks.)
(D) Monitoring and reflecting	(D1) Monitoring and repairing the shared understanding	(D2) Monitoring results of actions and evaluating success in solving the problem	(D3) Monitoring, providing feedback and adapting the team organisation and roles

Reprinted with permission from PISA Collaborative Problem Solving Framework. https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

It is beyond the scope of this chapter to articulate the details of the entire framework but we provide a background on the major competencies, skills, and levels of proficiency that define CPS within the context of PISA. Three major CPS *competencies* are crossed with the four major individual problem solving *processes* to form a matrix of 12 specific *skills*. There are three levels of *proficiency* (below, at, or above standard) for each of these 12 skills; there are associated actions, communications, cognitive and social processes, and strategies that define what it means for the student to be proficient.

Table 5.1 presents the skills of the 3 × 4 CPS Framework (see OECD 2013). The dimension of problem solving processes contains the same four components as the PISA 2012 Problem Solving Framework for individual problem solving (OECD 2010). The dimension of CPS competencies, as well as the associated skills, attempt to incorporate the CPS skills identified in other CPS frameworks, such as the CRESST teamwork processing model (O'Neil et al. 2010), the teamwork model of Salas, Fiore, and colleagues (Fiore et al. 2008, 2010; Salas et al. 2008), and ATC21S (Griffin et al. 2012).

Three CPS Competencies

The three CPS competencies are (1) establishing and maintaining shared understanding, (2) taking appropriate action, and (3) establishing and maintaining team organization.

1. *Establishing and maintaining shared understanding.* Students must have an ability to identify the mutual knowledge (what each other knows about the problem) or what is often called common ground (Clark 1996; Clark and Brennan 1991), to identify the perspectives of other agents in the collaboration, and to establish a shared vision of the problem states and activities (Dillenbourg 1999; Dillenbourg and Traum 2006; Fiore and Schooler 2004). 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. One important way to maintain a shared understanding is to have a *transactive memory*, a shared knowledge system for acquiring, storing, and retrieving information. Transactive memory facilitates group performance, learning, and transfer (Austin 2003; Lewis et al. 2005) in addition to keeping the group on task and assigning tasks to individuals with the best expertise (Littlepage et al. 2008).
2. *Taking appropriate action to solve the problem.* Students must 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. These include taking actions that solve the main substantive problem and also communication acts, such as explaining, justifying, negotiating, debating, and arguing.
3. *Establishing and maintaining group organisation.* 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 organisation, reflect on the success of the group organisation, and help handle communication breakdowns, conflicts, and obstacles. Students need to take steps to make sure that agents are completing tasks and communicating important information.

Problem Solving Processes

The problem solving processes were incorporated from the PISA 2012 Problem Solving Framework that targeted individual problem solving (Funke 2010; OECD 2010). The four cognitive processes comprise:

(A) *Exploring and understanding*. This includes interpreting the initial information about the problem and any information that is uncovered during exploration and interactions with the problem.
(B) *Representing and formulating*. Information is selected, organized, and integrated with prior knowledge. This may include the use of graphs, tables, symbols, or words. Students formulate hypotheses by identifying the relevant factors of the problem and then critically evaluate the hypotheses.
(C) *Planning and executing*. This includes identifying the goal of the problem, setting any sub-goals, developing a plan to reach the goal state, and executing the plan.
(D) *Monitoring and reflecting*. The student monitors steps in the plan to reach the goal state, marks progress, and reflects on the quality of the solutions.

Matrix of 12 CPS Skills and Proficiencies

The 12 skills in the matrix thus represent the competencies of CPS and the processes required to solve problems. A satisfactory assessment of CPS would assess the skill levels of students for each of these 12 cells. Some of these skills are reflected in actions that the student performs, such as making a decision by choosing an item on the screen, selecting values of parameters in a simulation, or preparing a requested report. Other skills require acts of communication, such as asking other group members questions, answering questions, making claims, issuing requests, giving feedback on other agents' actions, and so on. These acts of communication are needed to monitor shared understanding and team organization.

The CPEG identified various actions and communications that are associated with different levels of proficiency for each of the 12 skills. However, the main task was to specify a unidimensional scale of CPS. There needed to be a reduction in capturing the complexity of the mechanism and a focus on a robust construct that could reasonably handle a broad spectrum of problems, cultures, and students. The CPEG converged on the following three levels of proficiency for each skill: *below*, *at*, and *above* standard. However, it is important to emphasize that these levels of proficiency are being revised after item development, field testing phases, and the 2015 assessments among nations. These three categories reflect the initial theoretical framework on the CPS proficiency construct.

Below Standard of Proficiency The student is ineffective in advancing group goals and CPS. The student does not respond to requests for information and to prompts for them to take action. The student does not take actions that contribute to achieving group goals because they perform random or irrelevant actions.

At Standard of Proficiency The student is a good responsive team member, but does not assertively take the initiative and solve difficult barriers in collaboration. The student responds to requests for information and prompts for action, and selects actions that help achieve group goals. However, the student does not proactively take the initiative in requesting information from the agents, performing unprompted actions, and effectively responding to conflicts, changes in the problem situation, and new obstacles to goals.

Above Standard of Proficiency The student proactively takes the initiative and solves difficult barriers in collaboration. In addition to responding to other agents' requests and prompts, the student proactively requests information from the agents and performs unprompted actions. The student effectively responds to conflicts, changes in the problem situation, and new obstacles to goals.

Additional Components in the CPS Framework

The CPEG recognized that there are dimensions of CPS to consider other than the initiative and responsiveness of the group member. Group members are undoubtedly influenced by their psychological dispositions, cognitive skills, and traits. The complete CPS Framework acknowledged the importance of the students' prior knowledge (math, literacy, science, computer literacy) and psychological characteristics (cognitive abilities, attitudes, motivation, personality). Some of these components are assessed as part of the PISA student survey as well as part of other PISA assessments. The nature of the problem solving context (personal, social, workplace, technology, in versus out of school) and ground-rules of the problem scenario (hidden profile, consensus, negotiation, collaborative work) clearly constrain the sorts of social interactions that contribute to effective solutions. Aside from these obvious contextual influences on CPS, there are two salient dimensions of context that merit considerable attention. They are the team composition and the characteristics of the tasks.

Team composition can have a profound impact on the CPS of a group (Kreijns et al. 2003; Rosen and Rimor 2009; Wildman et al. 2012). Very little can be accomplished if no team member takes initiative or if all are fighting over control. Ideally, the members will identify who has talents or a willingness to take on a subtask, will achieve the subgoals, and will communicate progress to other members in a judicious fashion. CPS performance is compromised to the extent that the division of labour is unintelligent, subgoals are not achieved, the group goals are blocked, and there are communication breakdowns. A meaningful collaborative interaction rarely emerges spontaneously, but requires careful structuring of the collaboration to promote constructive interactions. For example, many types of collaboration have a symmetrical structure with respect to knowledge, status, and goals (Dillenbourg 1999), but the roles and tasks of the different group members may be very different. 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 status differentials, boss-subordinate relationships, and teacher-student interactions. Finally, symmetry of goals involves common group goals rather than individual goals that may conflict.

Task characteristics impose particularly specific constraints on the solution space. *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. A simple concrete example is carrying a large table across the room, a problem that cannot be accomplished by one student but rather a collection of students acting in a coordinated manner in time and space. Another example consists of jigsaw problems where a group goal requires the accomplishment of a set of tasks (A, B, and C), each of which is taken up by a particular student, and each student has limited access to the other students' knowledge (Aronson and Patnoe 1997; Stasser and Titus 2003); the puzzle can only be solved when the products of the tasks are pooled and coordinated. Tasks with high interdependency require an organization among team members that assigns tasks to team members and insures that each member is making adequate progress. Communication is essential to achieve such collaboration. The medium in which tasks take place also can have a great effect on the degree to which collaboration can be performed. A shared workspace that every team member can view provides a common ground for members to inspect progress on each other's tasks. However, in hidden profile tasks, team members have different information, so they need to explicitly ask questions and volunteer information through acts of communication in order to facilitate team members having a common understanding.

The difficulty of the problems varies in addition to challenges in collaboration. The PISA 2012 Complex Problem Solving framework and assessment defined many task characteristics that determined problem difficulty (OECD 2010) and also the extent to which assessment was domain-specific versus domain-general (Greiff et al. 2014). It is beyond the scope of this chapter to define the dimensions of problem difficulty. It suffices to say that problem difficulty increases for ill-defined problems over well-defined problems; dynamic problems (that change during the course of problem solving) over static problems; problems that are a long versus a short distance from the given state to the goal state; a large problem space over a small space; the novelty of the solution; and so on. Difficulty also should increase as a function of the number of agents involved in the collaboration and the symmetry of their roles. It is conceivable that the need for effective collaboration would increase as a function of the problem difficulty.

Challenges in Assessment of Collaborative Problem Solving

This section identifies some of the challenges in the assessment of CPS that surfaced in the discussions of the CPEG and in our coverage of the literature. The challenges focus on three issues: discourse management, group composition, and the use of computer agents in assessments.

Discourse Management

Because collaboration requires communication, an assessment of collaborative problem solving requires a suitable design of discourse management. Discourse management is the control of the communication within a task, such as turn-taking and conveying information and topics at the appropriate times. In developing items for CPS assessment, a key goal is to manage the discourse among the student and agents so that the CPS skills can be assessed in a minimal amount of time. This is a new skill for item developers who typically do not have a background in discourse processing theories (Clark 1996; Graesser et al. 2003).

Part of the challenge to students completing CPS tasks lies in creating and maintaining a shared understanding (competency 1) and team organization (competency 3), whereas another part involves taking action (competency 2) to advance a solution and handle conflict or change. Performance increases as a function of common ground of group members (Cannon-Bowers and Salas 2001; Fiore and Schooler 2004; Foltz and Martin 2008; Salas et al. 2008). When systemic group conflicts stem from fouled communication patterns, there is a need to take direct steps to intervene and to do so persistently (Dillenbourg and Traum 2006; Rice 2008). Less conflict occurs when groups have worked together previously and begin a new task (Hsu and Chou 2009). Interestingly, Asterhan and Schwarz (2009) have proposed that argumentation can be an effective means to reaching a deep level of understanding and shared vision. After the group goes through this difficult phase of conflict, which can be associated with negative affect (Barth and Funke 2010), a productive team moves forward with a better solution than if no arguments occur. That is, some amount of social-cognitive dissonance and clashes in common ground forces one to sort out facts and converge on better conclusions and solutions. This is very different from pseudo polite convergence and "group think," where the group quickly agrees with other team members instead of investing time in the task at a deep level (Stewart et al. 2007).

It is widely accepted that the core elements of collaboration (shared understanding, action, and team organization) are all related to one another and need to be coordinated rather than being exhibited in isolation. There is the standard tension between dividing the components into separate assessment modules and integrating them in ways that illuminate nontrivial mechanisms. Moreover, the particular coordination of these skills is to some extent context-dependent and could benefit from sophisticated computational models of discourse processing to assess collaborative competency (Graesser et al. 2011; Rosé et al. 2008; Shaffer and Gee 2012).

Discourse mechanisms of establishing common ground in communication have been investigated in human-human interactions and human-computer interactions (Clark 1996; Clark and Brennan 1991). Common ground is accomplished by making appropriate assumptions on what each other already knows ("old" or "given" knowledge in the common ground) and by performing acts of communication (verbal or nonverbal) about new information and business that is properly *coordinated*. People normally assume that knowledge is in the common ground if it (a) is physically co-present and salient (all parties in the communication can perceive it), (b) has been verbally expressed and understood in the discourse space, in front of the sender, recipients, and any side audience, and/or (c) is common knowledge for members of a group, culture, or target community. If a group member does not respond to a question or request for information, then the member is not taking responsibility for maintaining common ground and fails CPS competency 1. A leader manifests above average competency by taking the initiative in maintaining common ground. The student can ask questions to inquire about the status of the recipient agent with respect to levels of responsiveness and understanding: Are you there? Are you listening? Do you understand? Did you do something? Could you recap/summarize?

Turn-taking conventions facilitate the coordination of communication. For example, backchannel feedback ("uh huh," "okay," head nod) from the recipient acknowledges the sender's message and helps maintain shared understanding. There are adjacent pairs of turns between sender and recipient (Sacks et al. 1974), such as question-answer, request-acceptance, request-denial, command-promise, greeting-greeting, and so on. In dialogues with referential grounding between a leader and follower, there often are four-step handshaking sequences: Leader: "You lift the red bar." Follower: "The red bar?" Leader: "Yes, that bar." Follower: "Okay, the red bar" (Clark et al. 1986). This 4-step frame can be more economically expressed in face-to-face interaction by pointing gestures, directing a person's gaze, and other non-verbal channels (Van der Sluis and Krahmer 2007).

From the standpoint of CPS assessment, it is critical to design tasks that put students in situations that exercise collaboration through communication. It is possible to measure many of the skills related to discourse management by observing the degree to which the student responds to questions and requests for information, acknowledges other team members' actions, and initiates discourse to move the group forward. Within the PISA 2015 assessment, if a student tends to respond but does not initiate questions and requests for information, then the student could be considered as meeting minimal standards of proficiency. However, if the student initiates questions and requests, then the student would be considered as being above the standard of proficiency. Tasks need to be set up to assess both of these situations, as well as students being random or capricious. The design of every discourse episode in a task needs to be mindful of the alternative physical and discourse actions of the student. How does each alternative action map onto the skills in each of the 12 cells in Table 5.1?

Group Composition

CPS tasks with high interdependency are very sensitive to group composition. One team member who has low competency can dramatically decrease the performance of the entire team and force other team members to compensate in order to achieve team goals. An overly strong leader can prevent other team members from manifesting their talents. A team can flounder when leadership is required and no one steps up. When group-think occurs, the team politely agrees and settles for an inferior solution. When group disequilibrium occurs, the team can become dispirited and collapse. Thus, there are a large number of ways that a group can fail in CPS. These failures can be attributed to an individual or to the group process as a whole.

As mentioned early in this chapter, one question that was raised in the development of the PISA assessment was whether to measure the collaborative problem solving ability of the group as a whole or for particular individuals within the group. The CPEG decided on the latter because PISA focuses on measurement of individual skills. This decision has nontrivial consequences on the measurement logistics. In particular, it is necessary to expose each 15-year old student to multiple tasks with different team compositions, to partners with varying collaborative skills, and to multiple phases within a task that afford a broad array of situations. It is important for there to be challenges and barriers in the tasks so that an assessment can be made of the different CPS skills. With students working in teams, there is no guarantee that a particular student will be teamed up with the right combination to arrive at a sensitive measurement of any individual student's CPS skills. Consequently, there was a need to turn to technology to deliver a systematic assessment environment that is attuned to the dynamics of CPS (Griffin et al. 2012; Shaffer 2012).

Computer Agents as a Means to Assessment

The constraints of PISA required a computer-based assessment that measures CPS skills of individual students in a short assessment time window of an hour for four to five problem solving scenarios. Consequently, there was considerable discussion in the CPEG about whether to use computer agents in the assessments rather than other humans collaborating with the student through computer mediated communication (CMC). The CPEG determined that the computer agents provided more control over the interaction and could provide sufficiently valid assessments within the time constraints. Agents also provide control over logistical problems that stem from (a) assembling groups of humans (via CMC) in a timely manner within school systems that have rigid time constraints, (b) the necessity of having multiple teams per student to obtain reliable assessments, and (c) measurement error when a student is paired with humans who do not collaborate. A systematic design of computer agents could be arranged that provides control, many activities and interactions per unit time, and multiple groups of agents. These are all needed to fulfill valid measurement of relevant constructs in a short amount of time. In summary, the logistical constraints of the assessment specification lead the group to decide on the conversational agent option.

Questions have periodically been expressed on the use of computer agents in the assessments. Nevertheless, agents were used for PISA after careful consideration of various costs and benefits of human-human versus human-agent interactions. This decision is compatible with many other high stakes assessments. For example, the

Communicating and Collaborating dimension of the National Assessment of Educational Progress (NAEP 2014) will employ virtual agents, which measures skills such as, "exchange data and information with virtual peers and experts", "provide and integrate feedback from virtual peers and experts", and "debate with a virtual team member."

The amount of time available to assess CPS is very short, namely 60 min in two 30 min sessions. This requires efficiency in data collection. Agents allow a sufficient control over the interaction to get a sufficient number of assessment events that fit these time constraints and that cover the constructs that are essential to measure (Table 5.1). It often takes 5 min or more for a new group of humans to get acquainted in computer-mediated conversation before any of the actual problem solving processes begin. In contrast, agent environments can cut this time dramatically with strategic dialogue management and rapid immersion in the collaborative context. Finally, it is possible to measure a student's CPS competencies in multiple teams, with multiple tasks, and multiple phases in a controlled interaction. This would be logistically impossible with human-human interaction.

There is a broad spectrum of computer agents that have been used in tasks that involve tutoring, collaborating learning, co-construction of knowledge, and collaborative problem solving. At one extreme are fully embodied conversational agents in a virtual environment with speech recognition (*e.g.,* the *Tactical Language and Culture System*, see Johnson and Valente 2008) and tutorial learning environments that hold conversations with the student with talking heads, such as *AutoTutor* (Graesser et al. 2014; Lehman et al. 2013), *Betty's Brain* (Biswas et al. 2010), *Operation ARIES/ARA* (Halpern et al. 2012; Millis et al. 2011), and *iSTART* (Jackson and McNamara 2013). Such environments are motivating to 15-year old students, but the solution is impractical. There would be major challenges in technology, costs, and cultural variations in language and discourse. The more appropriate solution for assessments like PISA is minimalist agents that consist of printed messages in windows on the computer display, such as email messages, chat facilities, print in bubbles besides icons, and documents in various social communication media (Rouet 2006). These forms of agent-based social communication media have already been implemented in PIAAC (OECD 2009) on problem solving in electronically rich environments.

An assessment requires the human to pay attention to the agent when the agent communicates, just as a human does who takes the floor when speaking. This can also be accomplished with a minimalist agent by chat, by dynamic highlighting of messages and windows through colour or flash, and by coordination of messages with auditory signals (Mayer 2009). Humans can communicate with computer agents through a variety of channels. The simplest interface requires the student to click an alternative on a menu of optional speech acts and for there to be a limited number of options (two to seven). An advantage of this approach is that it focuses on the student's ability to know what kind of communication is required to complete the task without requiring the student to generate that communication. Other possibilities are open-ended responses that range from typing (or speaking) a single word, to articulating sentences, and composing lengthier essays.

Practical assessment of CPS can benefit from the use of agent technologies combined with communication and conventional computer technologies. Technology allows investigators to place humans in realistic orchestrated situations and observe their behavior and reactions. For example, many technological environments are based on naturalistic decision making (NDM) (Klein 2008; Lipshitz et al. 2001) in which each individual has his or her own goals, identity, and expertise that must be aligned in decisions and action in order to reach the end goal that affects both the individual and the group as a whole. Fan et al. (2010) have explored the prospects of artificial agents as collaborators during complex problem solving.

In order to validate the approach of using minimalist agents, a focused study has been conducted to investigate possible differences in student performance in human-to-agent (HA) versus human-to-human (HH) PISA-like CPS assessment tasks (Rosen 2014; Rosen and Foltz 2014). The participants were 179 14-year old students from the United States, Singapore, and Israel; there were 136 students assigned to the HA group and 43 students in the HH group. HH students were randomly assigned into pairs to work on the CPS task. The students were informed prior to their participation in the study whether they would collaborate with a computer agent or a classmate. In a case of HH setting, the students were able to see the true name of their partner. Students were exposed to identical collaborative problem solving assessment tasks and were able to collaborate and communicate by using identical methods and resources. However, while in the HA mode students collaborated with a simulated computer-driven partner, and in the HH mode students collaborated with another student to solve a problem. The findings showed that students assessed in HA mode outperformed their peers in HH mode in their collaborative skills. Collaborative problem solving with a computer agent involved significantly higher levels of shared understanding, progress monitoring, and feedback. 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 HA mode applied more attempts to solve the problem, compared to the HH mode. A process analysis of the chats and actions of the students during the CPS task showed that in the HA group the students encountered significantly more conflict situations than in the human-to-human group. The study provided initial validity evidence for HA approach in assessing CPS skills, although further research is needed to establish more comprehensive evidence. Specifically, there needs to be a systematic analysis of the extent to which the various cells in Table 5.1 are captured with reliability and validity.

These results suggest that each mode of CPS assessment can be differentially effective for different educational purposes. Non-availability of students with particular CPS skill levels in a class may limit the fulfilment of assessment needs, but technology with computer agents can fill the gaps. Thus, 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. On the other hand, HH interactions may uncover natural patterns of communication and CPS that do not occur in the HA modes.

In closing, CPS is a new field for large-scale assessment. It 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. Communication is an integral part of successful CPS in addition to the problem solving mechanisms. Group members must achieve common ground that provides a shared understanding of the task at hand, role assignments, and the contributions of team members as the problem solving evolves. In order to achieve the group goal, there needs to be adequate discourse management, group organization, and subgoal achievement. Such an assessment of CPS proficiency requires computer agents to insure that a human interacts with multiple individuals (agents) in multiple groups with different barriers and impasses. An agent-based assessment system therefore permits the development of an optimal assessment that can be administered for multiple tasks within a short time period.

References

Aronson, E., & Patnoe, S. (1997). *The jigsaw classroom: Building cooperation in the classroom*. New York: Addison Wesley Longman.

Asterhan, C. S., & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. *Cognitive Science, 33*, 374–400.

Austin, J. R. (2003). Transactive memory in organizational groups: The effects of content, consensus, specialization, and accuracy on group performance. *Journal of Applied Psychology, 88*, 866–878.

Barth, C. M., & Funke, J. (2010). Negative affective environments improve complex solving performance. *Cognition and Emotion, 24*, 1259–1268.

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

Brannick, M. T., & Prince, C. (1997). An overview of team performance measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory methods and applications* (pp. 3–16). Mahwah: Lawrence Erlbaum Associates.

Cannon-Bowers, J. A., & Salas, E. (2001). Reflections on shared cognition. *Journal of Organizational Behavior, 22*, 195–202.

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

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 221–239.

Dillenbourg, P. (1999). *Collaborative learning: Cognitive and computational approaches. Advances in learning and instruction series*. New York: 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*, 121–151.

Fan, X., McNeese, M., & Yen, J. (2010). NDM-based cognitive agents for supporting decision-making teams. *Human Computer Interaction, 25*, 195–234.

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.

Fiore, S. M., Rosen, M., Salas, E., Burke, S., & Jentsch, F. (2008). Processes in complex team problem solving: Parsing and defining the theoretical problem space. In M. Letsky, N. Warner, S. M. Fiore, & C. Smith (Eds.), *Macrocognition in teams: Theories and methodologies* (pp. 143–163). London: Ashgate Publishers.

Fiore, S. M., Rosen, M. A., Smith-Jentsch, K. A., Salas, E., Letsky, M., & Warner, N. (2010). Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. *Human Factors, 52*, 203–224.

Fiore, S. M., Wiltshire, T. J., Oglesby, J. M., O'Keefe, W. S., & Salas, E. (2014). Complex collaborative problem solving in mission control. *Aviation, Space, and Environmental Medicine, 85*, 456–461.

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 organisations and systems: Cross-disciplinary perspectives and approaches* (pp. 411–431). New York: Routledge.

Franklin, S., & Graesser, A. (1997). Is it an agent, or just a program? A taxonomy for autonomous agents. In*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., Gernsbacher, M. A., & Goldman, S. (Eds.). (2003). *Handbook of discourse processes*. Mahwah: Erlbaum.

Graesser, A. C., McNamara, D. S., & Rus, V. (2011). Computational modeling of discourse and conversation. In M. Spivey, M. Joanisse, & K. McRae (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 558–572). Cambridge: Cambridge University Press.

Graesser, A. C., Li, H., & Forsyth, C. (2014). Learning by communicating in natural language with conversational agents. *Current Directions in Psychological Science, 23*, 374–380.

Greiff, S., Würstenburg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A. C., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review, 13*, 74–83.

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.

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

Hsu, J. L., & Chou, H. W. (2009). The effects of communicative genres on intra-group conflict in virtual student teams. *International Journal of Distance Education Technologies, 7*, 1–22.

Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology, 105*, 1036–1049.

Johnson, L. W., & Valente, A. (2008). Tactical language and culture training systems: Using artificial intelligence to teach foreign languages and cultures. In M. Goker & K. Haigh (Eds.), *Proceedings of the twentieth conference on innovative applications of artificial intelligence* (pp. 1632–1639). Menlo Park: AAAI Press.

Klein, G. (2008). Naturalistic decision making. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*, 456–460.

Klein, C., DeRouin, R. E., & Salas, E. (2006). Uncovering workplace interpersonal skills: A review, framework, and research agenda. In G. P. Hodgkinson & J. K. Ford (Eds.), *International review of industrial and organisational psychology* (pp. 80–126). New York: Wiley.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior, 19*, 335–353.

Lehman, B., D'Mello, S. K., Strain, A., Mills, C., Gross, M., Dobbins, A., Wallace, P., Millis, K., & Graesser, A. C. (2013). Inducing and tracking confusion with contradictions during complex learning. *International Journal of Artificial Intelligence in Education, 22*, 85–105.

Lewis, K., Lange, D., & Gillis, L. (2005). Transactive memory systems, learning, and learning transfer. *Organization Science, 16*, 581–598.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making, 14*, 331–352.

Littlepage, G. E., Hollingshead, A. B., Drake, L. R., & Littlepage, A. M. (2008). Transactive memory and performance in work groups: Specificity, communication, ability differences, and work allocation. *Group Dynamics: Theory, Research, and Practice, 12*, 223–241.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A., & 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.

Mullins, D., Rummel, N., & Spada, H. (2011). Are two heads always better than one? Differential effects of collaboration on students' computer-supported learning in mathematics. *International Journal of Computer-Supported Collaborative Learning, 6*, 421–443.

Nash, J. M., Collins, B. N., Loughlin, S. E., Solbrig, M., Harvey, R., Krishnan-Sarin, S., Unger, J., Miner, C., Rukstalis, M., Shenassa, E., Dube, C., & Spirito, A. (2003). Training the transdisciplinary scientist: A general framework applied to tobacco use behavior. *Nicotine & Tobacco Research*, *5*, 41–53.

National Assessment of Educational Progress. (2014). *Abridged technology and engineering literacy framework for the national assessment of educational progress*. Washington, DC: National Academies Press.

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

O'Neil, H. F., 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.

OECD. (2009). *Problem solving in technology-rich environments: A conceptual framework*. OECD Education Working Papers, 36, OECD Publishing. Retrieved from http://search.oecd.org/officialdocuments/displaydocumentpdf/?doclanguage=en&cote=edu/wkp(2009)15

OECD. (2010). *PISA 2012 field trial problem solving framework*. Paris: OECD. Retrieved from http://www.oecd.org/dataoecd/8/42/46962005.pdf.

OECD (2013). PISA 2015 collaborative problem solving framework. Paris: OECD. Retrieved from http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20 Collaborative%20 Problem%20Solving%20Framework%20.pdf.

Rice, R. E. (2008). Unusual routines: Organizational (non)sensemaking. *Journal of Communication, 58*, 1–19.

Rosé, C., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*, 237–271.

Rosen, Y. (2014). Comparability of conflict opportunities in human-to-human and human-to-agent online collaborative problem solving. *Technology, Knowledge and Learning, 19*, 147–174.

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

Rosen, Y., & Rimor, R. (2009). Using a collaborative database to enhance students' knowledge construction. *Interdisciplinary Journal of E-Learning and Learning Objects, 5*, 187–195.

Rouet, J.-F. (2006). *The skills of document use*. Mahwah: Erlbaum.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696–735.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*, 540–547.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences, 4*, 321–354.

Shaffer, D. W. (2012). Models of situated action: Computer games and the problem of transfer. In C. Steinkuehler, K. Squire, & S. Barab (Eds.), *Games, learning, and society: Learning and meaning in the digital age* (pp. 403–433). Cambridge: Cambridge University Press.

Shaffer, D. W., & Gee, J. P. (2012). The right kind of GATE: Computer games and the future of assessment. In M. Mayrath, D. Robinson, & J. Clarke-Midura (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Charlotte: Information Age Publishing.

Sonnentag, S., & Lange, I. (2002). The relationship between high performance and knowledge about how to master cooperation situations. *Applied Cognitive Psychology, 16*, 491–508.

Stasser, G., & Titus, W. (2003). Hidden profiles: A brief history. *Psychological Inquiry, 14*, 304–313.

Stewart, C. O., Setlock, L. D., & Fussell, S. R. (2007). Conversational argumentation in decision making: Chinese and US participants in face-to-face and instant-messaging interactions. *Discourse Processes, 44*, 113–139.

Theiner, G., & O'Connor, T. (2010). The emergence of group cognition. In A. Corradini & T. O'Connor (Eds.), *Emergence in science and philosophy* (pp. 79–117). New York: Routledge.

Van der Sluis, I., & Krahmer, E. (2007). Generating multimodal references. *Discourse Processes, 44*, 145–174.

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*, 137–170.