JEM JOURNAL of EDUCATIONAL MEASUREMENT NCME national council on measurement in education

Journal of Educational Measurement
Spring 2017, Vol. 54, No. 1, pp. 12–35

Designs for Operationalizing Collaborative Problem Solving for Automated Assessment

Claire Scoular and Esther Care
University of Melbourne
Friedrich W. Hesse
University of Tuebingen and Leibniz-Institut für Wissensmedien in Tuebingen

Collaborative problem solving is a complex skill set that draws on social and cognitive factors. The construct remains in its infancy due to lack of empirical evidence that can be drawn upon for validation. The differences and similarities between two large-scale initiatives that reflect this state of the art, in terms of underlying assumptions about the construct and approach to task development, are outlined. The goal is to clarify how definitions of the nature of the construct impact the approach to design of assessment tasks. Illustrations of two different approaches to the development of a task designed to elicit behaviors that manifest the construct are presented. The method highlights the degree to which these approaches might constrain a comprehensive assessment of the construct.

Collaborative problem solving (CPS) is one of several "21st century skills" (Binkley et al., 2012) appearing in national educational statements as essential for students to develop. Others include critical thinking, decision making, creativity, and digital literacies. Notwithstanding the enthusiasm for exploring these skills for the purposes of inclusion in curriculum, teaching, and assessment, the field remains in its early days. Efforts around definition of the skills, their descriptions, how they might be included in curricula, taught, and assessed are myriad and diverse. The majority of research effort has been dedicated to expert-driven hypotheses concerning the nature of the skills, design of tasks to elicit behaviors hypothesized to demonstrate their existence, and data collection from students' responses to tasks to determine if demonstrated behaviors align with the hypothesized construct. In this article, the implications are explored in the definition of CPS for task development. Two initiatives that have focused on this skill are the Assessment and Teaching of 21st Century Skills project (ATC21S; Griffin, McGaw, & Care, 2012), and the Organization for Economic Co-operation and Development's Programme for International Student Achievement (PISA; OECD 2013). The two initiatives base their task development on their respective, and different, definitions. The consequences of these differences for task design are explored in order to understand how nuances in definition impact on task development. Interest in the assessment of CPS is relatively recent. Its nature is not well understood, and its development from basic to sophisticated levels is not well known. Accordingly, definitional differences and their implications for the design of assessment tasks may be significant in informing the assessment and educational communities about the nature of the construct.

ATC21S was an international project (2009–2012) that was designed to draw some consensus from researchers, educators, and policymakers concerning the nature of 21st century skills, and to explore ways of assessing and teaching them. It involved leading technology corporations and rested on the efforts of six countries (Singapore, Finland, Netherlands, the United States, Costa Rica, and Australia) that participated in a research effort to explore innovative assessment approaches to measurement of the skills among 11–15-year-olds. PISA evaluates the quality, equity, and efficiency of school systems. Its program is implemented in over 70 countries, and constitutes an effort to monitor the outcomes of education systems within an internationally agreed framework. In terms of evidence, it relies on the collection of 15-year-old students' achievement data derived from assessments of capabilities such as mathematics, reading, and problem solving. ATC21S in particular focused on CPS, a choice guided by the perception that this skill set is complex and contributed to by many subskills of interest that are identified as 21st century skills. PISA included CPS in its 2015 global assessment, building on previous experience in assessing individual problem solving.

Skills

Complex skills have been defined across many contexts such as physical education (Galligan et al., 2000), literacy (Mehta, Foorman, Branum-Martin, & Taylor, 2005), and communication (Verderber, Verderber, & Berryman-Fink, 2010). Some definitions base the label of complexity on mere multiplicity of demand; some base the label on the higher cognitive or physical demand that requires interactions between subskills; and others base it on bundling of multiple and differentiable subskills.

In the literature around problem solving—the area most strongly aligned with the construct of CPS—there has been a growing interest in the "complexity" of the skill (Quesada, Kintsch, & Gomez, 2005). For example, Greiff et al. (2013) argue that general intelligence (Spearman's *g*; Spearman, 1927) is insufficient as an explanatory variable for achievement, and that strategic and processing aspects of mental functioning should be brought more strongly into the limelight. Funke, Fischer, and Holt (in press) identify a shift away from well-defined analytical conceptions of problem solving toward focus on more complex problems that involve dynamic interaction with the task, echoing the perception of problem solving as a complex skill set rather than a sequential set of processes—a "bundle of skills, knowledge, and abilities that are required to deal effectively with complex and dynamic nonroutine situations in different domains." This work identifies the challenges in identifying manifested behaviors in a less defined problem space. Indicators of performance in complex problem solving extend well beyond the standard solution criteria with indicators measuring breadth and depth of exploration, decision making, and the coordination of different action types (Fischer, Greiff, & Funke, 2012). One of the most difficult concepts to convey to those interested in the application of problem solving skills is how these constructs and their manifestations through assessment tasks actually relate to the demands placed on individuals in real life settings. There is widespread acceptance, for example, of the notion that one needs to learn specific and isolated skills that contribute to reading in the early years (including phonological awareness and phonics, fluency, and vocabulary), so that in later years one makes sense of the written word. The same understanding is not typically extended to some of the skills that are increasingly valued in the 21st century; this is presumably due to our more limited understanding of these skills and their developmental trajectories. However, the knowledge that a complex skill like reading comprehension can be taught and learned through its component skills needs to be applied equally to a complex skill such as CPS. And so, to map and measure the early component skills of CPS, simple tasks can be designed. These simple tasks stimulate and draw on skills in the early developmental stage. These will not, particularly in isolation, look like the complex skills that we will see applied when dealing with a wicked problem; but neither does sounding out the alphabet look like reading and understanding Shakespeare. In this article, we provide examples of tasks that draw on these early skills. The tasks are designed to elicit particular subskills that are part of the complex skill set that we call CPS.

The structures of CPS as proposed by Hesse, Care, Buder, Sassenberg, and Griffin (2014) and OECD (2013) imply a complex skill set by virtue of the hypothesized interactions across its components. These interactions consist of those between the student and the problem space as well as among students working together to solve a problem. The construct is hypothesized to consist of a number of subskills, from both cognitive and social domains, that can be enacted together to form a recognizable and differentiable skill set.

Nature of Collaborative Problem Solving

The nature of CPS as seen through the lens of assessment and its conceptualizations over the past 10–25 years is best traced back to the work of O'Neil and colleagues who completed several studies in the 1990s to develop a theoretical framework and methodology to assess workforce readiness skills (e.g., O'Neil, Allred & Baker, 1992a, 1992b). The National Center for Research on Evaluation, Standards, and Student Testing (CRESST) model of problem solving included not only content and cognitive strategies, but also the components of self-regulation, metacognition, and motivation. The model was examined specifically in the context of computer-based assessment. Akin to the ATC21S and PISA initiatives, O'Neil et al. identified five competencies—use of resources, interpersonal skills, information, systems, and technology. In their computer-based assessment approach, they focused on cognitive (including metacognitive), negotiation, and affective skills and processes. Their simulation tasks modeled a human to agent approach (O'Neil, Allred, & Dennis, 1993) with different motivations and resources available to the two partners (one an agent) in order to stimulate different levels of expertise. The work of O'Neil and colleagues moved to identify CPS explicitly (Chung, O'Neil, & Herl, 1999; O'Neil, Chuang, & Chung, 2003) with identification of teamwork skills. O'Neil et al. (2004) studied interactions between team members through use of predefined messages, which were categorized across adaptability, communication, coordination, decision making, interpersonal skills, and leadership, the focus primarily on the degree to which networked computer facilities could capture these processes. O'Neil et al. (2004) described the problem-solving construct in terms of understanding, strategies, and self-regulation, and CPS as "problem-solving activities that involve interactions among a group of individuals" (p. 364). These interactions were elucidated as adaptability, coordination, decision making, interpersonal skills, leadership, and communication. O'Neil et al. (2003) summarized CPS as divided into two components (collaboration and problem solving). O'Neil, Chuang, and Baker (2010) covered many of the components that are now included explicitly in definitions of CPS from a teamwork approach. O'Neil's work essentially laid the theoretical basis for the Hesse et al. (2014) and PISA CPS frameworks (OECD, 2013), although it should be noted that the teamwork model adopted by CRESST is not reflected in ATC21S. The delineation between how an individual collaborates within a shared problem space, and teamwork as defined by O'Neil, Chung, and Brown (1997), consisting of the six skills of adaptability, coordination, decision making, interpersonal skills, leadership, and communication, shares common ground, but is approached from radically different directions.

CPS is a process that relies on both cognitive and social skills contributions by those involved in the joint activity. The underlying justification for activating the process is that the nature of a problem requires input of diverse resources for resolution. These resources may be the different knowledge and skills that each individual brings, or it may be physical resources. Documentation of the nature of CPS typically draws attention to these two primary contributions. For example, in the ATC21S project, Care, Griffin, Scoular, Awwal, and Zoanetti (2015) propose the cognitive and social as third-order components. At the second order, the major strands within each of these are proposed; and at the first order, several elements are proposed within the strands. As with any such theoretical structures, it is presumed that there is shared variance within each order. The PISA CPS framework defines the construct 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" (OECD, 2013, p. 6).

Of interest in this article is identification of the nature and components of CPS in the context of their implications for design of assessment tasks. The theoretical contributions to the construct are covered in depth in Hesse et al. (2014) for ATC21S, and in OECD's draft framework (2013) for PISA and are discussed in detail in the subsequent sections. As acknowledged by both these sources, the literature on the components that are integral to the construct is diverse, and drawn from the fields of psychology, education, human resources, to name a few.

Collaborative problem solving in PISA. PISA conceptualizes CPS as a set of three collaborative competencies (establishing and maintaining shared understanding; taking appropriate action to solve the problem; establishing and maintaining team organization) and four problem-solving processes (exploring and understanding; representing and formulating; planning and executing; and monitoring and reflecting) that form a matrix of 12 skills cells. Sources for the conceptualization are identified as including O'Neil et al. (2004), O'Neil et al. (2010), Salas,

Table 1
PISA 2015 Collaborative Problem-Solving Skills Matrix (Adapted from OECD, 2013)

	1. Establishing and Maintaining Shared Understanding	2. Taking Appropriate Action to Solve the Problem	3. Establishing and Maintaining Team Organization
A. Exploring and understanding	A1. Discovering perspectives and abilities of team members	A2. Discovering the type of collaborative interaction to solve the problem	A3. Understanding roles to solve problem
B. Representing and formulating	B1. Building a shared representation and negotiating the meaning of the problem	B2. Identifying and describing tasks to be completed	B3. Describe roles and team organization
C. Planning and executing	C1. Communicating with team members about the actions to be performed	C2. Enacting plans	C3. Following rules of engagement
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 organization and roles

Dickenson, Converse, and Tannenbaum (1992), Salas, Cooke, and Rosen (2008), and ATC21S (Griffin, Care, & McGaw, 2012). The development of the construct from a theoretical perspective rests on literature and initiatives from "computer-supported cooperative work, team discourse analysis, knowledge sharing, individual problem solving, organizational psychology, and assessment in work contexts (e.g., military teams, corporate leadership)" (OECD, 2013, p. 5). In defining the domain for assessment, PISA relies on the problem-solving framework developed for 2003–2012 assessments, with the addition of the collaborative component. The focus is on assessment of the individual's capacities in collaborative contexts (Table 1).

The context for development of the PISA assessment was "collaborative skills in project-based learning in schools" and in the "workplace and civic settings" (OECD, 2013, p. 4). In addition, the assessment had to be operationalized within a computer-based environment. Rosen and Foltz (2014) report on initial validation work of the PISA approach. Another study that draws from both the PISA approach and ATC21S is reported by Liu, Hao, von Davier, Kyllonen, and Zapata-Rivera (2015).

Collaborative problem solving in ATC21S. Hesse et al. (2014) define CPS as "a joint activity where dyads or small groups execute a number of steps in order to transform a current state into a desired goal state" (p. 39). The authors make the point that the activity is a coordinated set of processes, and that it may not follow a linear sequence. Whereas individual problem solving might follow such a process, the complexity added by the collaborative dimension challenges this possibility. In addition, Hesse et al. make clear that particular skills will be applied across multiple processes such as problem identification, problem representation, planning, executing, and monitoring. Hesse et al. propose a framework to describe the skill set that includes the social and cognitive components. Within social, there is specification of three strands: participation, perspective taking, and social regulation. Within cognitive, there are two strands: task regulation and knowledge building. Within each of these five strands, there are subskills, or elements identified (Table 2). The theories drawn upon for formulation of the structure make clear the dependencies across elements and strands and the two components.

The framework described by Hesse et al. (2014) pays major attention to the skills each individual brings to the problem-solving space, rather than to team work or group work. It sees the social components of CPS as skills the individual brings to bear in the same way that they bring the cognitive to the arena. For the development of ATC21S tasks, the cognitive side of CPS was conceptualized from an information processing approach (Newell & Simon, 1972) that facilitated development of online assessments. It included purloining from the knowledge building estate, with a metacognitive component as well—although the latter is located in the Hesse framework as a "social" component. Hence, this framework leads to assessment tasks that rely on cognitive and social skills. The ATC21S focus on the construct is developmental. Its framing is to guide development of the skill set through teaching and learning in a way that will equip students to draw on and demonstrate skills in problem solving in ambiguous contexts and where neither most efficient paths nor final solutions are known. The focus therefore is not solution-focused, but skills-focused.

Implications of Differences in Framing the Construct

There are clear differences between the two initiatives in their conceptualization of the construct. The following task development descriptions highlight the implications of these differences for task design.

Problem definition. Both ATC21S and PISA initiatives accept an information processing approach in which the tension between the current and desired goal constitutes the problem space. A difference emerges across the two frameworks in perspectives on the nature of a problem. ATC21S tends toward the position in which there is no clarity about what the solution to a problem may be, mirroring real life problems. In contrast, PISA tends toward the position that the primary unknown is how to reach the solution, reflecting the perspective of Mayer (1990). Mayer stated "problem solving is the cognitive processing directed at transforming a given situation into a goal situation when no obvious method of solution is available" (p. 284).

Another way of describing these perspectives is in terms of task definition. The PISA framework articulates the concept of a "well defined" task, describing

Table 2
Social and Cognitive Elements With Summary Descriptions

	Social Elements	Summary Description
Participation	Action	Activity within the working environment varying from minimal to across both familiar and unfamiliar contexts
	Interaction	Interacting with others varying from mere acknowledgment, to substantive responses through initiating and promoting interaction
	Task completion/ perseverance	Is aware of the task through attempting and then persevering through multiple attempts or strategies
Perspective taking	Adaptive responsiveness	Ignoring input from others through accepting, and then adapting contributions of others to suggest possible solution paths
	Audience awareness (Mutual modeling)	Acting without regard to the receptive communication needs of others through intuitive modification of one's own contributions
Social regulation	Negotiation	Awareness of differences through achieving common understanding and resolution of differences
	Self-evaluation (Metamemory)	Aware of one's own performance through recognition of adequacy in the context of the immediate and broader contexts
	Transactive memory	Builds awareness of other's strengths and weaknesses based on performance in the specific context
	Responsibility initiative	Assumes responsibility for one's own contributions through being aware of joint responsibility within the group
	Cognitive Elements	**Summary Description**
Task regulation	Organizes (Problem analysis)	Accepts problem at face level through analysis and identification of subtasks
	Sets goals	Moves from goal of task completion to setting of subtasks that recognize links between them
	Resource management	Individually manages resources through consensual resource allocation
	Flexibility and ambiguity	Inaction in ambiguous situations through acknowledgement and exploration
	Collects elements information	Identifies need for information and explores to map other possible needs
	Systematicity	Uses trial and error through intentional sequences of actions to exhaust possible solutions

(*Continued*)

Table 2
Continued

	Cognitive Elements	Summary Description
Knowledge building	Relationships	Identifies isolated elements of information through linking these and identifying patterns among them
	Rules “If … then”	Takes action without consideration of consequence through identification of course and effect and its use
	Hypothesis “What if…”	Takes a single approach through adaptation based on additional information and reconstruction of solution pathway

problem situations “for which there are clearly specified goals, given states, and legal actions” (OECD, 2013, p. 10). Further, PISA differentiates between static and interactive tasks. Static tasks include all information required for solution from the outset. Interactive tasks require the individual to explore the space in order to access required information. In contrast, ATC21S does not specify the nature of the task (Care & Griffin, 2014) beyond the essential requirement that it demands multiple and different resources. The hypothesized cognitive skills of task regulation and knowledge building help to operationalize this through analysis of processes, goal setting, resource management, tolerance for ambiguity, information collection and systematicity, identification of relationships, cause and effect, and testing of hypotheses as competencies of interest. This identification strongly implies that lack of information about the problem space—both the artifacts within it, and the processes that can be brought to bear on it—is a primary requisite of a task that demands CPS efforts. The ATC21S initiative assumed that problems, even as designed for student assessment and teaching, should represent the real-world characteristic of lack of definition (Care, Scoular, & Griffin, 2016).

Focus on teaching and learning or problem solution. Both initiatives focus on individual performance in a collaborative event, as opposed to group performance. Notwithstanding, an interesting aspect of the PISA description is its identification of the success of the group (OECD, 2013, p. 7) over the individual, reflected somewhat by Liu et al. (2015). The focus on success, and problem solution to a slightly lesser extent, differentiates this approach from ATC21S. The natural consequence of the problem definition difference is that progress through problem-solving processes will be less predictable in the ATC21S approach than in the PISA approach. This aligns with the ATC21S focus on teachability and learnability, and can be seen in the construct definition’s explication of tolerance for ambiguity, and for flexibility in problem-solving strategies.

Symmetry and interdependence. PISA maintains Dillenbourg’s (1999) stance that effective collaboration relies on symmetry with respect to the knowledge, status, and goals that individuals bring to bear. This means that individuals should have similar amounts of knowledge, be equal in the collaborative space, and hold common goals.

ATC21S reflects the same values in terms of status and goals, but not in terms of knowledge, or “resources.” Of essence to the ATC21S CPS construct is that it is called upon precisely due to the need for resources (skills, knowledge, or artifacts) that do not reside equally in the collaborating partners. Reflecting that perspective, assessment tasks would vary in degree of symmetry in order to scaffold students through a learning experience.

The concept of interdependency endorsed by PISA nullifies its symmetry perspective. PISA identifies two types of symmetry—one concerning the individuals and one concerning the task. ATC21S rests on asymmetry, both concerning the individuals and the task. The task demands different resources, and different individuals bring these to bear variably.

Communication and its centrality to the social strand. PISA is explicit about collaboration, requiring individuals to reflect on how their skills and knowledge contribute to problem solution, identify different points of view that may lead to conflict, and be explicit about opinions and interpretations. The collaborative competencies of “establishing and maintaining shared understanding . . . and . . . establishing and maintaining team organization” (OECD, 2013, p. 9) rely on communication.

ATC21S similarly values “participation” by the individual, as well as “interaction” as demonstrated by communicating with a partner and “responsiveness” by responding to communications or actions. Communication is core to the ATC21S social strand, with participation representing both engagement with the task and engagement with the partner; with interaction clearly residing in interplay between partners; and with social regulation dependent on the capacity of the partners to negotiate and delineate their contributions and responsibilities. The two initiatives are reasonably consistent on the central role of communication. A major design requirement for assessment tasks is that this core competency is reflected.

Context and Rationale of the Two Initiatives

The rationale underlying development of the ATC21S tasks is to demonstrate the single steps and subskills that are involved in tasks requiring CPS input. This makes these processes visible to the teacher and negotiable by the student. Exemplar assessment tasks can demonstrate the thinking processes of students, and student level of sophistication across the subskills can be determined. Examples should provide the potential for teachers to identify what resources and strategies to bring to bear for instruction of the student in such a way that the skills can be enhanced. However, this does not necessarily contribute to the individual’s capacity to generalize the skill. In order for problem-solving skills to be generalized, or in Mayer’s (2002) terminology, transferred, they need to be taught within a meaningful environment and across multiple contexts. Memorizing specific algorithms may be sufficient for replicating the same process in a similar context, but is insufficient for the implementation of the problem-solving processes more generally. Students require explicit instruction to transfer problem-solving skills from one context to another until such time that they can recognize that the solution to one type of problem may be useful in solving another problem. Design of assessment tasks should ideally include the elements identified within the CPS framework in such a way that students can recognize the commonality of particular processes and strategies for problems that occur in a wide range of contexts.

The context for PISA is the need for a summative assessment system designed to inform systems of education, not specific teachers and their students. This means that the data derived from assessment should be interpretable at general and aggregate levels as the priority. In addition, the history of problem-solving assessment in PISA means that the construct development for PISA 2015 CPS needed to link to the construct description used to develop previous assessments of individual problem solving.

A question of interest is whether the different conceptualizations of and rationales for these two initiatives necessarily lead to different approaches to assessment. Comparison of the two design approaches will highlight the extent to which they are congruent with their frameworks, and whether differences may imply measurement of different skill sets.

Method

The information presented here is targeted to illustrate alignment between concept and assessment design. The goal is to identify the degree to which designs for assessment task development have the capacity to respond to the characteristics and imperatives of the ATC21S and PISA initiatives.

Two designs are described. The human-to-human (H-H) design requires two students to partner virtually in an online assessment environment to work through a task together. The human-to-agent (H-A) design requires one student to work online through a task with a presumed partner who is a “virtual other” to represent a real collaborator. The two designs are applied to one task with slightly different methods, in order to highlight the capacity of each design to represent the CPS constructs as presented by ATC21S and PISA. The actual problems used to illustrate the two design approaches are derived from the ATC21S initiative.

The H-H task design approach is based broadly on the ATC21S design, while the H-A approach is based on the PISA design. The H-H assessment tasks are designed for two students to work together online and potentially collaborate on the assessment irrespective of location. The tasks are designed to elicit specific behaviors in alignment with the skills outlined in the construct. These behaviors can be identified through log files that are generated during task play recording students’ actions and communications in time sequence. An embedded chat box facilitates communications. Students cannot progress through the assessment independent of one another. The H-H task design is deliberately ill defined so that students are presented with an ambiguous problem space they need to navigate.

The H-A task design is somewhat more defined, presenting all required information to the student from the outset. In this scenario, the student works alone, with a simulated agent built in to the task to present the notion of collaboration. The student is presented with a series of constructed multiple-choice responses that they can use to navigate through the problem space, somewhat aligned with a traditional

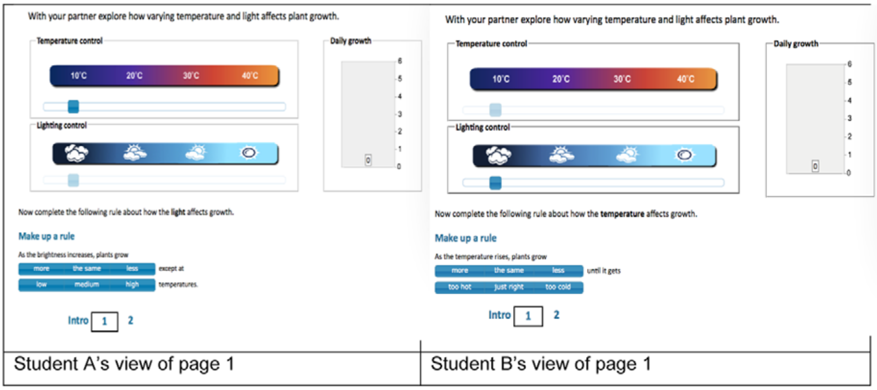

Figure 1. Screenshots from Plant Growth H-H scenario.

assessment format. The H-A eliminates the need for a free-form chat box, because H-H communication cannot take place.

Comparison of H-H and H-A Scenarios

The Plant Growth task. Both H-H and H-A designs are applied to the ATC21S task called Plant Growth. The task requires students to find the best conditions for plant growth by varying temperature and light density. This involves students exploring patterns, sharing information, generating rules, and investigating cause and effect. The first page requires students to trial combinations of variables to identify the optimal conditions. On the second page, students identify the problem principle and apply the conditions.

The H-H design presents students with an asymmetric task in which they can only manipulate their own resources, one controlling the temperature and the other the brightness. This design is consistent with the ATC21S view of task types that both merit and require CPS activities. Asymmetry of resources is an essential design feature, situating students in a real-world replica. This imposes the need for collaboration and requires students to pool their respective knowledge, information, and resources.

In the H-H design, students are initially faced with the social aspect of the task in which they build mutual understanding with their partners through the use of the chat box (Figure 1).

Students can explore the social space through acknowledgment of their partners and asking them questions. Students can explore the problem space sharing their own perspectives and enquiring as to the information and resources held by their partners. The task is ill defined, providing minimal instructions to the students and an ambiguous task space, which requires exploration both socially and cognitively. Students may proceed through iterations of trial and error, not just in regard to the problem's resources, but also in regard to their attempts to collaborate constructively. Identifying the problem principle can only be achieved through the pooling and analysis of both their resources. Students need to regulate the social space to manage

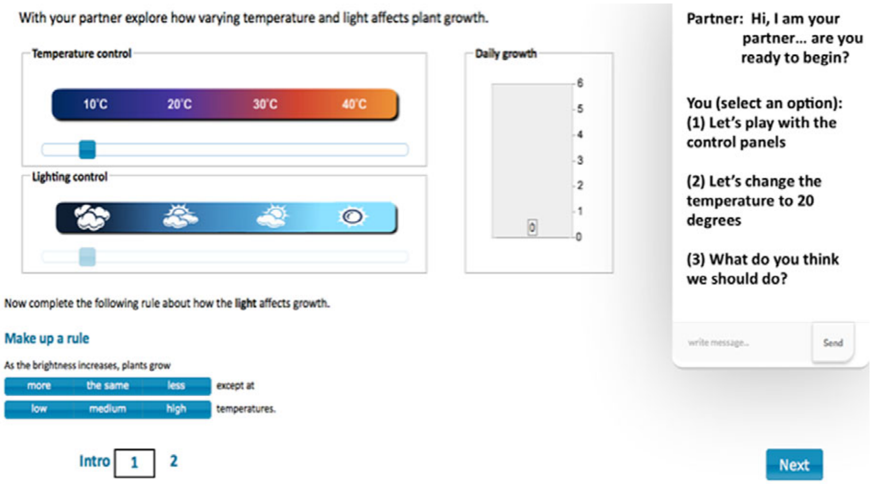

Figure 2. Example of Plant Growth displayed in the H-A design.

different perspectives. Although the student goal may well be to solve the problem, it is the process of working within the social and cognitive hurdles of the task that will provide the learning opportunity.

In the alternative H-A design, the student is presented with the problem space and navigates through the task as prescribed by multiple-choice pathway options from the computerized agent. Students are required to select the answer they deem most appropriate that will lead the agent to provide a response with further multiple-choice options. The constructed multiple-choice responses identify a particular option as more accurate than the others. This format means that students are constrained by a strict turn and turn-about structure. The agent is nonpersonalized, or neutral. There are no social cues or opportunities that might prompt different social factor responses from the student. Therefore, in this design, the student only deals with the problem-solving component. The problem space is well defined, presenting all the information to the student from the outset, with each possible solution path embedded within the multiple-choice options. This scaffolds the student and presents options that may not otherwise have been considered.

Figure 2 presents an example of Plant Growth displayed in the H-A design. In this example, the student and agent are depicted as having chosen to share their view of the controls; the agent choice is programmed as an autoresponse to the student choice. Now the student is presented with three multiple-choice constructed responses: (1) Let's play with the control panels. (2) Let's change the temperature to 20°. (3) What do you think we should do? If the student selects (1), the agent will ask for more detail regarding the plan, which will lead to (2). If the student selects (3), the agent will prompt the student to trial the controls also leading to (2). If the student selects the optimal response, (2), the agent will agree and students will be scored not just on their ability to select the correct response but on whether their preceding actions are aligned with their response (in this case if they changed the temperature to 20°). If the student deviates from this plan, they will again be prompted by the agent to remain on task. The responses are constructed in this manner until several combinations of temperature and light have been trialed. The student is then prompted to be systematic in trialing all possible combinations of temperature and light with the provided options (1) What do you think the answer is? (2) Let's play with the controls again. (3) We should now try all of the temperature and light combinations. If the student selects (1) or (2), they are prompted to proceed with (3). If the student selects (3), they will achieve the correct answer but can also be assessed on whether they trial all combinations were trialed before proceeding to the next page in the task.

The agent is one of several artifacts in the problem space, and limits the learning through the task process to a defined set of options, with learning stemming from the constructed responses rather than the iterations and explorations of the social relationship and problem space. Hence, the agent is merely another artifact of the task itself rather than an active participant. The dynamic design of H-H requires students to transfer numeric data into verbal relationships in order to solve the problem collaboratively. In the H-A design, this opportunity to rerepresent the problem is not available. This design restricts opportunity for individuals to generalize problem solving and its sequences in different ways.

Capturing data and identifying behaviors. The difference in the physical appearance of the two approaches to task design are perhaps more obvious than the capacity for data generation in the background. The H-H scenario allows for every action and chat during task play to be captured in a time-sequenced log file. Behaviors can be identified in the format of sequences of action and chat events and those behaviors can be mapped back onto the framework with the justification that the behaviors identified are indicative of a student's level of proficiency in a particular skill. Alternatively, the H-A scenario only allows for the multiple-choice option selected, time spent on task, and the correctness of selection to be captured. In this sense, the H-A design is similar to a traditional multiple-choice test, where a student achieves a score for selecting the correct answer.

To further demonstrate the method, Table 3 presents an excerpt of chat events in a log file that has been generated by two students working through the Plant Growth task. The first column represents the role allocated to each student (A or B). All of the events in this example occur on page 1 and are chat events (the event is shown in column 2 and the chat itself is shown in column 3). Column 4 indicates the sequence of events by the use of a time stamp for each event. In the interest of linking the behaviors observed in the log file to the framework, column 5 has been added to indicate the skill that could be observed. As can be observed in the log file, the students immediately interact by sharing information in order to build a mutual understanding of the problem space. They become aware that they do not have access to all the resources, and collect additional information by trialing the controls and asking their partners questions. They proceed in trialing the resources and persevere in their attempt to complete the task successfully. They take turns guiding their progress through the task and each takes responsibility by reporting their hypothesis about the problem to their partner. By communicating their progress to one another, they are able to develop an understanding of the relationships between the two controls as well as

Table 3
An Excerpt of Chat Events From a Log File Generated I the H-H Scenario During the Plant Growth Task

Student Role	Event	Data	Time Stamp	Observable Subskill
A	chat	I have some controls and a growth chart	1/08/2015 14:33	Interaction
A	chat	you	1/08/2015 14:33	
B	chat	yeah same	1/08/2015 14:33	
A	chat	oh the brightness control just moved	1/08/2015 14:33	
B	chat	that was me	1/08/2015 14:34	
A	chat	I must control temperature then	1/08/2015 14:34	
B	chat	yeh	1/08/2015 14:34	
A	chat	You put where ?	1/08/2015 14:34	Collecting information
B	chat	i put on very bright	1/08/2015 14:34	
B	chat	growth now 2	1/08/2015 14:34	
A	chat	ok ill try	1/08/2015 14:34	
A	chat	I have moved to 20 degrees, growth = 4	1/08/2015 14:34	
B	chat	k so growth is better with more light	1/08/2015 14:35	Task completion
A	chat	yeh but we should try other temperature	1/08/2015 14:35	Perseverance
B	chat	for me quite dark = growth 2	1/08/2015 14:35	
A	chat	I kept on 20 degrees	1/08/2015 14:35	
A	chat	Ill move to 20 degrees now	1/08/2015 14:36	
B	chat	ok so 30 deg & quite dark = 3	1/08/2015 14:36	Systematicity
B	chat	yes!	1/08/2015 14:36	
A	chat	Brighter and warmer seems best	1/08/2015 14:37	Relationships
B	chat	ok what now	1/08/2015 14:37	Tolerance for ambiguity
A	chat	Lets complete the rule!	1/08/2015 14:37	Responsibility
B	chat	Yes but should we retest to make sure?	1/08/2015 14:37	Goal setting
A	chat	Ok start from low to high	1/08/2015 14:37	Resource management
A	chat	Ive moved to 10 degrees	1/08/2015 14:37	
B	chat	ok so lets do same again	1/08/2015 14:37	Systematicity

noting the variations in the growth chart. The students demonstrate a high level of tolerance for the ambiguity of the task in which there is a lack of definition surrounding the most efficient solution path. By navigating the ambiguous space together, they set joint goals about how to proceed. One of the main challenges for this task is trialing both controls systematically and noting the perceived outcomes in the growth chart. The team presented in the example manages the resources well by planning to retry each variable from lowest to highest. In the full log (not shown here), the students trial each variable sequentially before making a joint decision to go back and retest to check whether their shared hypotheses were correct.

The empirical data from the H-H generated log files such as the one presented Table 3 can be used in the development of H-A constructed responses. This ensures that the H-A constructed responses imitate an H-H scenario, as far as possible. Figure 3 presents an example of constructed responses developed for the Plant Growth task. The agent asks the student what they think they should do and the student is presented with four options to respond. If any of the first three options are selected, the agent provides additional information and prompts the student to try their controls. The fourth option is the desired response because the student is stating that they have some idea of how to proceed. The agent then seeks further information from the student about how they are going to proceed and presents four subsequent responses. If the student selects either of the bottom two responses, they are prompted by the agent again to trial their controls. If either of the top two responses are selected, the student is progressing along the desired solution path, and therefore is trialing the controls systematically. The student is then prompted by the agent to report the information presented in the growth chart. The student progresses through the task in this manner until prompted to make a statement about how the controls work. Due to the nature of the "branching" approach for the H-A, there is a desired solution path presented. This is different from the H-H design in which there is no expected or desired solution path and students can progress through the task successfully in many different ways.

Mapping behaviors to the frameworks. Example behaviors from the Plant Growth task are presented in Table 4 along with the elements in the Hesse et al. (2014) framework to which these are mapped. These behaviors have also been mapped to the skills in the PISA framework. The first example presented in Table 4 is a student's response to chat from their partner. This can be mapped to a high level of interaction in the Hesse framework, and to skill A2 in the PISA framework, which involves the discovery of the type of collaborative interaction to solve the problem. This behavior can be captured in the H-H design by identifying the sequences of chat between students in the log file. In the H-A design, response alone can be identified when the student responds to the agent by selecting a multiple-choice option. Given that the selection of a multiple-choice option is essential for task progress, it is not possible to ascertain in this scenario whether this could be attributed to regulation of the social space or specific to the problem space. The second example indicator is the identification that the student's answer is different from their partner's; they chat and then the student changes their answer to match their partner's. This can be mapped to a high level of negotiation in the Hesse framework and skill B1 in the PISA

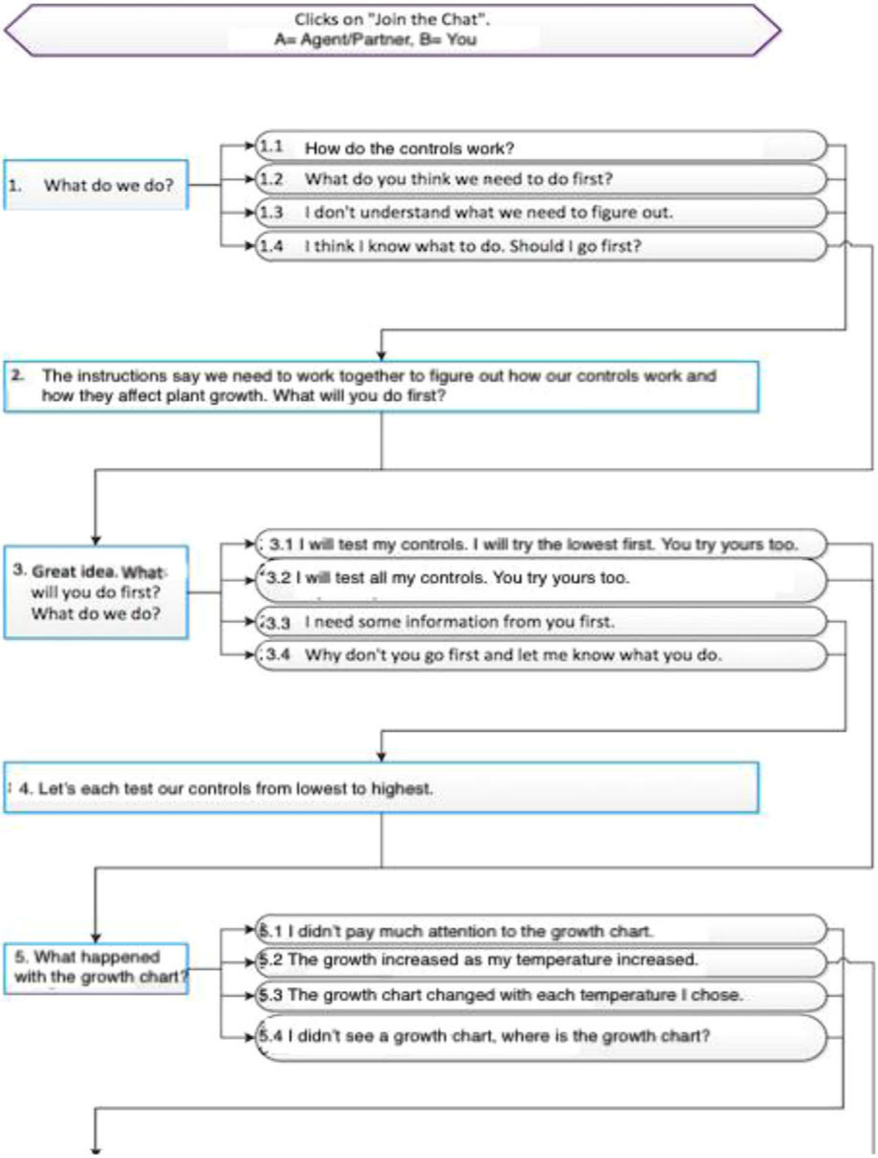

Figure 3. Example of H-A branching constructed responses in the Plant Growth task.

framework, which is building a shared representation and negotiating the meaning of the problem. In the H-H design, student inputs can be observed in the log file and compared readily across students; if there is a change in one student's answer to match the others and there are instances of chat in between these events then students can be seen to be negotiating the answer. There is less opportunity to assess such a skill in the H-A design since the agent is not designed to disagree or formulate an

Table 4
Plant Growth Task Indicators Mapped to Skills in the Hesse and PISA Frameworks

Task Behavioral Indicator	Hesse et al. (2014) Subskill	PISA Skill
1. Responds to chat from partner	Interaction—Responds to cues in interaction	A2—Discovering the type of collaborative interaction to solve the problem along with goals
2. Student's answer is different from partner, chat, and change answer	Negotiation—Achieves resolution of difficulties	B1—Building a shared representation and negotiating the meaning of the problem (common ground)
3. Reports variable outcomes to partner	Responsibility initiative—Reports to others on progress of activities	C1—Communicating with team members about the actions to be/being performed

alternative option. The third example indicator is a student reporting variable outcomes to their partner, for example, stating that the temperature has been increased to 20° and as a result, the growth has increased. In the H-H design, key variables can be searched for within-chat communication (e.g. "20°," "30°") from one student to another to identify whether they are reporting such information to one another. In the H-A design, the responses from student to agent are constrained. There is less opportunity in the H-A design than in the H-H design to capture the wide range behaviors implicit in the CPS construct due to constraints on choice and sequence of actions and on chat.

Discussion

The descriptions of approaches to presenting a task to elicit CPS behaviors make clear the limitations inherent in online capture capacity. This discussion is specific, however, to the capacity of each approach to target the particular characteristics of the definitions adopted by the two assessment initiatives, rather than the broader discussion about online capture of complex behaviors. The usefulness of the analysis lies in the degree to which each assessment approach is aligned with its definition and meets the goal of measuring CPS.

The main rationale behind the ATC21S approach is that the assessment should inform both teaching and learning. Therefore, the tasks need to elicit documentable behaviors that can demonstrate the processes to be taught and learnt. The main rationale for the PISA approach is the identification of student CPS capacities for provision of large-scale information to countries about their student populations. In both cases, elicited behaviors need to be consistent with hypotheses about the nature of problems to be solved and the nature of the required cognitive actions and social actions and interactions.

Problem definition. The H-A design prioritizes ambiguity and lack of information as characteristic of tasks that require CPS resources. The H-A design reflects a constrained environment in which preselected paths only are available. Consistent with the differences in definition of the problem is the likely capture of information. Where more diverse behaviors can be captured, there is more scope for interpretation of skills. The flexibility of movement within the H-H tasks leads to greater variability in how students proceed through the task and how they interpret the problem. Students presented with the H-A scenario are guided in their understanding of the problem by the constrained nature of their choices, with all students ultimately reaching a solution point and the variance will lie in the correctness of that solution. This aspect of the design in both scenarios accurately matches that which is intended in the respective frameworks. The ATC21S approach is not designed to present clarity surrounding the problem; the goal is to present the problem as ill defined, and this is achieved through presentation of minimal instructions and an environment to explore at will. The PISA approach aims to have well defined problems that lack clarity only insofar as correctness of responses is not known. This is achieved in the H-A more accurately in the scenarios since the students are scaffolded through the task by the agent; it is only the final answer and the efficiency with which they reached that point that can be assessed for each student.

Focus on teaching and learning or problem solution. The H-A use of multiple-choice options that reflect the postulated responses of an agent can be argued to represent the perspective that there is a correct path toward a solution and, by extension, a correct solution. To the degree that this is a true reflection of the PISA perspective, the task design is well aligned with the PISA position. The H-A approach provides a highly scaffolded environment for the student to proceed through, perhaps similar to a traditional assessment and an approach that operates in the majority of classrooms globally. The H-H task design also has constraints around what steps a student can take in the online problem space, but is significantly less scripted. While the H-H approach can capture whether a student's progress results in a correct answer, it is the process by which the students progress through the task that is of real interest. This permits the collaborating students to take more or less efficient paths, to experiment, and to explore. These processes are strongly aligned with the construct description, in which the focus is on the capacities of students to implement the processes, but not in a linear fashion. Being able to assess *how*, and not just *if*, a student could solve the problem collaboratively will provide much more information to the teacher and student, which is the goal of the ATC21S initiative. In this sense, solution becomes just one of many indicators of CPS ability. Being able to break down the elements of CPS and feedback-specific information to teachers and students should provide informative insights into this complex skill set.

In the H-A approach, only an individual outcome will be possible. Since there is only one student, a group outcome is not applicable. The PISA initiative is interested in both an individual outcome and a group outcome but the H-A approach does not allow for its capture. Notwithstanding, the exploratory work reported by Liu et al. (2015) provides some interesting variations on this. In contrast, the ATC21S initiative is interested primarily on the individual outcome. The extent to which this is possible is yet to be determined. While individual outcomes can be estimated, the extent to which one student's ability impacts on the other is still uncertain. The dependence between students in the H-H scenario raises a very interesting and worthy topic for future research in this field.

Symmetry and interdependence. In the assessment event, both designs require the individual to operate within the constraints that characterize the online environment. In the case of H-H, the capacity of an individual to move in the problem-solving space is to varying degrees determined by the CPS capacities of the partner in the event. If the partner does not provide resources or communicate information that is required for the individual to progress, this necessarily determines the parameters of what is possible for the individual. In the case of H-A, the capacity of the individual to demonstrate skills is constrained totally by what the system architecture permits through the multiple-choice options provided. The issue of symmetry does not arise for H-A, given the individual problem-solving nature of the environment. For H-H, the example task presented in this article, Plant Growth, is asymmetric. Beyond this, the interaction itself promotes further asymmetry, as well as dependence due to the changing conditions that emanate from any one partner's acts or chats. The H-H environment is thus less predictable through the course of the problem solving. Considering this, the H-A does not really account for symmetry of roles, goals, or knowledge, as identified in the PISA initiative, since there is only one student. The ATC21S initiative rests on the assumption of asymmetry for both students and tasks and this is played out in the H-H scenarios to varying degrees. The extent to which variations in asymmetry impact the measures of CPS remains to be explored.

Communication and its centrality to the social strand. Interpersonal communication has been described as having five major characteristics (Verderber et al., 2010) that are useful to draw on for the purposes of analysis. Note that the descriptor of communication as "interpersonal"—a necessary component of CPS—draws attention to the need for responsiveness between collaborating partners. It is this responsiveness that builds the understanding of each partner about the resources that are, and that become, available through the partnership. Where assessment tasks are designed to capture a particular construct, it is essential that the task can capture valid indications of its components. It is clear from the examples provided that there are questions concerning the degree to which these tasks can accomplish this, and even clearer that the H-A context cannot capture some of the subskills within this complex skill set.

Communication is an essential component of perspective taking. The five characteristics of interpersonal communication (Verderber et al., 2010) raise some interesting issues in the consideration of the degree to which both H-H and H-A activities in an online space can be perceived to reflect communication. In addition, the H-A condition limits further the characteristics of social media technologies. It could be argued that awareness of online presence of another is not necessarily *continuous* in the H-H condition; but it is without question that such consciousness cannot exist in the H-A condition. That communication is *transactional* in the H-H condition is reasonable, and examination of the log files in these examples indicates clearly the learning (or knowledge building) that occurs through the interactions. The H-A condition cannot provide learning to each participant, given the nonentity of one; however, it may be argued that the human participant can learn from the automated messaging (multiple-choice option) in the multiple-choice options. The *irreversible* nature of communication would stand for both conditions. The *interpersonal* nature of communication could not be argued to exist, however, since it is a nonhuman interaction. The degree to which the messaging could be regarded as communication is limited by this fact. And finally, any *emotion* that exists as a result of the messaging for the human must be an artifact of response to text alone, not to what may be regarded as an intentional communication from another that is personalized to the receiver. Accordingly, with respect to the five characteristics, there are major limitations in the degree to which the H-A condition can be regarded as capable of capturing communication competencies, and by extension capable of capturing the other social skills required for CPS that rely on this, such as perspective taking and social regulation. Given that communication is identified as having a central role in CPS by both the ATC21S and PISA initiatives, it is unlikely an H-A scenario would be able to elicit sufficient behaviors for good representation of the construct. The role of the agent is merely an artifact of the problem space. Communication is not possible with an agent and there are no social cues to interpret or differing of opinions to negotiate.

Limitations

While there are advantages and disadvantages to both types of scenario, it is apparent that the H-H scenario allows for the generation of more and richer data. Time-structured log files present much more detailed and readily interpretable information about the processes and skills elicited by the students than do constrained multiple-choice options. The constraints in actions and lack of communication make it difficult to measure many of the complex skills outlined in both the Hesse and PISA frameworks—such as negotiation, goal setting, and perspective taking. Both frameworks outline CPS as a complex skill set and it seems only logical that complex behaviors would need to be identified and captured to draw inferences about ability. The four example behaviors outlined in Table 4 are mostly difficult to capture in the H-A scenarios. An additional factor that requires exploration concerns group composition in the H-H approach. Both approaches are interested in individual student ability. However, it is logical that the capacity of a student to collaborate will be constrained by the performance of the partner; the latter will provide either more or less opportunity for the student to demonstrate the skill level of which they are capable. This variability does not apply to the H-A approach. An ability estimate is not applicable for the agent, and allowances for dependence are less relevant when the path of the agent is a known one. Rigorous trials that record across tasks how different student partnering is associated with student performance estimates are required. The developmental learning approach adopted by ATC21S requires assessment tasks that elicit early learning indicators of the complex skill set as well as more sophisticated research components. At the task development level, this provides challenges in being able to demonstrate face validity of the tasks, since the early development components rest on simple skills (in the same way as early development of what will finally be reading comprehension, rests on recognition of letters of the alphabet) that might not be recognizable as informing the complex skill set. Empirical data derived from such tasks at large scale is required to confirm the hypothesized progression.

Rigorous psychometric analyses will need to be conducted to identify if both approaches are measuring the same construct. A comparative analysis of data collected through both approaches is required. Additional validation work will need to be conducted to support the claims that each approach is measuring CPS. The ATC21S project's focus is on the development of a teaching and learning tool, while the OECD's focus is on large-scale assessment. The ATC21S approach requires measures that provide performance interpretation for students, while the PISA approach is norm-referenced and summative. From this perspective, it is understandable why the H-A approach might be favored given that the large-scale nature of PISA would require stable Internet connectivity and suitable technology access within and across sample countries, which is perhaps just not possible at this stage. In this article, the rationale for task design and development has been explicated—it remains to establish the implications of these designs through large-scale data. In terms of task design, it would appear that there could be a tiered approach, H-H being the optimal scenario and H-A an optional approach to at least capture parts of the construct where the optimal approach is not possible. It is difficult to develop H-A scenarios that appear genuine and natural for the students without first trialing such tasks in the H-H scenario. The data collected from H-H scenarios, however, have the potential to inform graduated approaches to H-H tasks. One example could be player selection of prescripted chat events rather than free communications. Such an option could be brought into play when it triggers events demonstrating that the players are not moving forward in the task. The prescripted chat would act as a scaffold for the dyad, in much the way that the current H-A tasks operate, although without the presumption of correct and incorrect responses.

Conclusion

The ATC21S and PISA CPS initiatives are based on strong similarities in terms of construct definition. Their assessment designs include some notable similarities, but respond to the imperatives of the purpose of the assessments to produce different assessment environments for students. How the task design has responded to the initiatives' imperatives and state of the art in online assessment represents a state of tension, and the degree to which either approach can comprehensively capture the construct of CPS remains to be established empirically. The advantage of the design developed by the PISA initiative is ease of programming, data capture, and scoring. The advantage of the design developed by the ATC21S initiative is rich data collection. The richness of the CPS skill set is difficult to capture, and these two initiatives are indicative of these difficulties. Responses to multiple-choice items in the tasks described are difficult to attribute accurately to particular subskills described in the conceptual frameworks due to the multiple factors that might cause the responses. The actions and chat of students to ATC21S design tasks, although more readily associated with particular subskills, present a similar challenge for inferring of cognitive and social processes.

An interesting aspect of the analysis of construct to task design across the two initiatives is the alignment of design with construct. In particular, the emphasis on problem definition with PISA is echoed in the PISA task design; the emphasis on ambiguity in Hesse et al. (2014) is echoed the ATC21S task design. The social aspects of CPS are less well managed in both initiatives, although it is clear that the communication aspects of the ATC21S approach are more ambitious. This highlights the challenge for future work in the field—how to assess social skills when the opportunity to present interpersonal skills is significantly reduced due to technology constraints.

References

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Dordrecht, The Netherlands: Springer.

Care, E., & Griffin, P. (2014). An approach to assessment of collaborative problem solving special issue: Assessment in computer supported collaborative learning. *Research and Practice in Technology Enhanced Learning*, *9*, 367–388.

Care, E., Griffin, P., Scoular, C., Awwal, N., & Zoanetti, N. (2015). Collaborative problem solving tasks. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 85–104). Dordrecht, The Netherlands: Springer.

Care, E., Scoular, C., & Griffin, P. (2016). Assessment of collaborative problem solving in education environments. *Applied Measurement in Education*, *29*, 250–264.

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

Dillenbourg, P. (1999), What do you mean by "collaborative learning"? In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 1–19). Oxford, UK: Elsevier.

Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *Journal of Problem Solving*, *4*(1), 19–42.

Funke, J., Fischer, A., & Holt, D. V. (in press). Competencies for complexity: Problem solving in the 21st century. In E. Care, P. Griffin, & M. Wilson (Eds.). *Assessment and teaching of 21st century skills: Research and applications*. Dordrecht, The Netherlands: Springer.

Galligan, F., Barry, T., Crawford, D., Howe, D., Maskery, C., Ruston, A., & Spence, J. (2000). *Advanced PE for edexcel*. Portsmouth, NH: Heinemann.

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts—Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, *105*, 364–379.

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 of 21st century skills* (pp. 1–15). Dordrecht, The Netherlands: Springer.

Griffin, P., McGaw, B., & Care, E. (Eds.) (2012). *Assessment and teaching of 21st century skills*. Dordrecht, The Netherlands: Springer.

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2014). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 37–56). Dordrecht, The Netherlands: Springer.

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

Mayer, R. E. (1990). Problem solving. In M.W. Eysenck (Ed.), *The Blackwell dictionary of cognitive psychology* (pp. 284–288). Oxford, UK: Blackwell.

Mayer, R. E. (2002). A taxonomy for computer-based assessment of problem solving. *Computers in Human Behavior*, *18*, 623–632.

Mehta, P. D., Foorman, B. R., Branum-Martin, L., & Taylor, W. P. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1 to 4. *Scientific Studies in Reading*, *9*(2), 85–116.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

OECD. (2013). *Draft PISA collaborative problem solving framework*. Paris, France: OECD Publishing.

O'Neil, H. F., Jr., Allred, K., & Baker, E. L. (1992a). *Measurement of workforce readiness: Review of theoretical frameworks*. (CSE Technical Report 343). Pittsburgh, PA: Carnegie Mellon University.

O'Neil, H. F., Jr., Allred, K., & Baker, E. L. (1992b). *Measurement of workforce readiness: Design of prototype measures*. (CSE Technical Report 344). Pittsburgh, PA: Carnegie Mellon University.

O'Neil, H. F., Jr., Allred, K., & Dennis, R. (1993). *Designs for assessing individual and group problem solving*. Los Angeles CRESST, University of Southern California.

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, NY: Springer.

O'Neil, H. F., Chuang, S., & Chung, G. K. W. K. (2004). Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education: Principles, Policy and Practice*, *10*, 361–374.

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

Quesada, J., Kintsch, W., & Gomez, E. (2005). Complex problem-solving: A field in search of definition. *Theoretical Issues in Ergonomics Science*, *6*(1), 5–33.

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

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, *50*, 540–548.

Salas, E., Dickenson, T. L., Converse, S., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3–29). Norwood, NJ: Ablex.

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York, NY: Macmillan.

Verderber, K. S., Verderber, R. F., & Berryman-Fink, C. (2010). *Inter-act: Interpersonal communication concepts, skills, and contexts*. New York, NY: Oxford University Press.

Authors

CLAIRE SCOULAR is Research Fellow at the Assessment Research Centre, University of Melbourne, Parkville, 3010, Australia; c.scoular@unimelb.edu.au. Her research interests include the measurement of 21st century skills, online learning, and computerized assessment.

ESTHER CARE is Professorial Fellow at the University of Melbourne, Parkville, 3010, Australia, and Senior Fellow at the Brookings Institution, 1755 Massachusetts Avenue, Washington DC 20036, USA; e.care@unimelb.edu.au. Her research interests include psychoeducational assessment, education reform, and 21st century skills.

FRIEDRICH W. HESSE has a Chair in Applied Cognitive Psychology at the University of Tuebingen and is Director of the Leibniz-Institut für Wissensmedien in Tuebingen, Schleichstrasse 6, 72076 Tuebingen, Germany; fh@iwm-tuebingen.de. His primary research interests include technology-based learning and knowledge processes.