Psychological Science in the
Public Interest
2018, Vol. 19(2) 59–92

DOI: 10.1177/1529100618808244
www.psychologicalscience.org/PSPI
SAGE

Advancing the Science of Collaborative Problem Solving

Arthur C. Graesser[1], Stephen M. Fiore[2], Samuel Greiff[3], Jessica Andrews-Todd[4], Peter W. Foltz[5,6], and Friedrich W. Hesse[7,8]

[1]Department of Psychology and Institute for Intelligent Systems, University of Memphis; [2]Department of Philosophy and Institute for Training & Simulation, University of Central Florida; [3]Cognitive Science and Assessment, University of Luxembourg; [4]Educational Testing Service, Princeton, New Jersey; [5]Pearson, Boulder, Colorado; [6]Institute of Cognitive Science, University of Colorado; [7]Leibniz-Institut fur Wissensmedien, Tübingen, Germany; and [8]Department for Applied Cognitive and Media Psychology, University of Tübingen

Abstract

Collaborative problem solving (CPS) has been receiving increasing international attention because much of the complex work in the modern world is performed by teams. However, systematic education and training on CPS is lacking for those entering and participating in the workforce. In 2015, the Programme for International Student Assessment (PISA), a global test of educational progress, documented the low levels of proficiency in CPS. This result not only underscores a significant societal need but also presents an important opportunity for psychological scientists to develop, adopt, and implement theory and empirical research on CPS and to work with educators and policy experts to improve training in CPS. This article offers some directions for psychological science to participate in the growing attention to CPS throughout the world. First, it identifies the existing theoretical frameworks and empirical research that focus on CPS. Second, it provides examples of how recent technologies can automate analyses of CPS processes and assessments so that substantially larger data sets can be analyzed and so students can receive immediate feedback on their CPS performance. Third, it identifies some challenges, debates, and uncertainties in creating an infrastructure for research, education, and training in CPS. CPS education and assessment are expected to improve when supported by larger data sets and theoretical frameworks that are informed by psychological science. This will require interdisciplinary efforts that include expertise in psychological science, education, assessment, intelligent digital technologies, and policy.

Keywords

assessment, collaboration, collaborative problem solving, education, groups, PISA, teams

Collaborative problem solving (CPS) is an important 21st century skill that is increasingly recognized as being critical to efficiency, effectiveness, and innovation in the modern global economy (Fiore, Graesser, & Greiff, 2018; Organisation for Economic Co-operation and Development [OECD], 2017a). CPS has attracted interest in international assessments; national assessments of middle and high school students; and training in colleges, industry, and the military (Care, Griffin, & Wilson, 2018; Fiore et al., 2017; Graesser, Foltz, et al., 2018; Hesse, Care, Buder, Sassenberg, & Griffin, 2015; National Research Council, 2011; OECD, 2017a, 2017b; Sottilare, Burke, et al., 2018). CPS is an essential skill in the home, the workforce, and the community because many of the problems faced in the modern world require teams to integrate group achievements with team members' idiosyncratic knowledge. CPS requires both cognitive and social skills. From the cognitive standpoint, team members must be able to define the problem, understand who knows what on the team, identify gaps in what is known and what is required, integrate these to generate candidate solutions, and monitor progress in achieving

Corresponding Author:
Arthur C. Graesser, Department of Psychology & Institute for Intelligent Systems, 365 Innovation Dr., 403B, University of Memphis, Memphis, TN 38152-3230
E-mail: graesser@memphis.edu

the group goals. From the social perspective, the success of a team requires that members establish shared understanding, pursue joint and complementary actions, and coordinate their behavior in service of generating and evaluating solutions. Successful collaboration can be threatened by a social loafer, an uncooperative unskilled member, or a counterproductive alliance, whereas it can be facilitated by a strong team member who draws out different perspectives, helps negotiate conflicts, assigns roles, promotes team communication, and guides the team to overcome troublesome obstacles (Fiore, Rosen, et al., 2010; Letsky, Warner, Fiore, & Smith, 2008; Salas, Cooke, & Rosen, 2008).

For the first time ever, an international assessment of CPS was conducted in 2015, with the results reported in December of 2017 by the OECD (2017a, 2017b). CPS was selected by the OECD as a new assessment for the Program for International Student Assessment (PISA) in the 2015 international survey of student skills and knowledge. More than 500,000 15-year-old students from 52 countries completed the PISA CPS 2015 assessment (hereafter called PISA assessment) in addition to assessments of mathematics, science, literacy, and other proficiencies. The 2017 report on PISA assessment of CPS (OECD, 2017b) had a large number of important results that have far-reaching implications for the public. Only 8% of students throughout the globe performed at the highest level of proficiency (as defined later in this article), whereas 29% of students scored at the lowest levels. Female students had substantially higher CPS proficiencies than male students (in all countries), whereas male students showed a modest advantage in individual problem solving in a previous PISA assessment in 2012 (OECD, 2017). The cultural and ethnic diversity of team members in schools was found to predict CPS proficiencies positively rather than negatively after implementing statistical controls for socioeconomic status. Results suggest that participation in group school activities, such as band, plays, sports, newspapers, and volunteer service activities, were training grounds for developing CPS skills. These are just a few of the many intriguing results in the OECD report. The momentum of interest in CPS has recently had an impact on the United States. In particular, it stimulated a report to the National Assessment of Educational Progress on testing CPS in U.S. schools (Fiore et al., 2017). Assessments like these have the potential to stimulate research, new curricula in K–16 (i.e., kindergarten through 12th grade plus postsecondary education), and new standards in industry, the government, and the public at large.

It is important for psychological science to be part of national and international discussions. This requires fundamental research on CPS to develop and refine socio-cognitive theory as well as more applied research for implementation and evaluation in real-world settings. As researchers in a field of psychological science, we need to identify what we can contribute to help shape answers to key questions and provide solutions to overcome challenging obstacles. Following are some representative questions that cut across cognitive, social, and psychometric research: What proficiencies are included in CPS and how should these proficiencies be assessed? What are the psychological mechanisms that explain and improve CPS skills? What cognitive and social psychological theories can inform and improve our understanding of CPS processes and outcomes? What sort of CPS training should be in school curricula and the workforce? This article was written in part to encourage psychological scientists to be more active partners in this interdisciplinary and international movement to understand and improve CPS proficiencies.

It is important to take stock of where psychological science fits into this evolving CPS landscape. Psychological science has provided a significant body of empirical research on group processes, teams, collaboration, and communication. There have been attempts to differentiate team member contributions from the performance of the team as a whole. However, psychological science research on teams has typically focused on learning in groups (Slavin, 2017), group decision making (Gigone & Hastie, 1997; Hastie & Kameda, 2005; Mesmer-Magnus & DeChurch, 2009), or team training (Mathieu, Hollenbeck, van Knippenberg, & Ilgen, 2017), rather than group problem solving per se. There also is a rich history of investigating the psychology of individual problem solving (Funke, 2010; Greiff, Wüstenberg, et al., 2014; Mayer & Wittrock, 2006; Sternberg, 1995), but there has been only sporadic progress in investigating problem solving in groups. Although social psychology has occasionally investigated problem solving in groups, that work has focused on the generation of ideas with little or no accountability of the solutions, integration of knowledge, and interdependencies among team members (Dennis & Williams, 2005; Mullen, Johnson, & Salas, 1991). The signature features of CPS are the existence of a group goal involving a novel problem to be solved (i.e., as opposed to completing a routine task), objective accountability (i.e., the quality of the solution is visible to team members), differentiation of roles (i.e., team members complete different tasks), and interdependency (i.e., a single person cannot solve the problem alone). This constellation of features is not necessarily the same as that in in collaborative learning, decision making, memory, and work. These latter types of collaboration become more complex and more integrated when considering CPS in real-world settings (Letsky et al., 2008).

We begin by defining CPS and justifying why it is important to study. We then cover the major theoretical frameworks that have articulated CPS mechanisms, measurement, and assessment. Unfortunately, research on CPS is sparse, and the sample sizes in the empirical studies of groups are small, so psychological scientists need to fortify or modify their claims with systematic empirical research on larger samples of group data.

Psychological scientists will need to partner with experts in other fields in interdisciplinary research efforts in order to collect data that will have a maximal impact. Toward this end, this article points out innovative technologies, data science, and quantitative-assessment models. In particular, technological advances have outpaced theoretical and empirical work in CPS. Over the past 2 decades, there has been an evolution of digital technologies that include computer-supported collaboration among humans in chat facilities; automated analyses of natural language; and computer agents that simulate team members, tutors, or mentors. Developments in intelligent digital technologies and quantitative modeling have increased the ability to track CPS processes in rich detail and transform logged data into meaningful psychological measures (Dowell, Nixon, & Graesser, 2018; Foltz & Martin, 2008; Gilbert et al., 2018; Graesser, Cai, Morgan, & Wang, 2017; Shaffer, 2017; von Davier, Zhu, & Kyllonen, 2017). The measures not only evaluate CPS performance but also generate feedback to problem solvers and instructors in summative and formative assessments. These advances are available to psychologists to explore mechanisms of CPS.

Finally, this article considers issues associated with education and workforce policy-level decisions that would draw from research on CPS. For example, we identify national guidance on methods to train individuals on CPS competencies (i.e., relevant knowledge, skills, and strategies). Many of these suggestions are speculative at this point because the fields of psychology and education are only beginning to tackle CPS training, now that the importance of CPS proficiency is acknowledged nationally and internationally. Making CPS skills understandable and teachable represents a policy-level goal that requires close cooperation between stakeholders in research, education, and the government. CPS skills acquired in school activities are expected to improve students' college and career readiness and thereby benefit society at large (Fiore et al., 2017, 2018; Hesse et al., 2015; OECD, 2017a, 2017b; Quellmalz, Timms, Silberglitt, & Buckley, 2012). There are no widely accepted CPS curricula or standards in school systems at this point (Scoular & Care, 2018). These need to be developed to prepare students to be productive and healthy citizens in the 21st century. CPS curricula will hopefully be influenced by psychological science rather than being entirely dominated by anecdotes, intuition, and folklore.

What Is Collaborative Problem Solving?

A number of features differentiate CPS from other forms of collaboration. This article will offer some nuanced descriptions and types of CPS, but we start with its typical distinctive characteristics.

First, a group has a goal of solving a novel problem by formulating a plan to move from a starting state to a goal state when no routine plan or script is available (Mayer & Wittrock, 2006; Newell & Simon, 1972). It is a *group* goal because it is impossible or unlikely that an individual will be able to solve the problem alone (for those problems that are suitable for CPS). The *novelty* of the problem-solving solution differentiates CPS from collaborative work, which normally focuses on coordinating groups of people to perform actions that implement well-established solutions to routine tasks.

Second, the *quality of the solution* can be evaluated during problem solving and is visible to team members. That is, the team members can detect whether the group goal is achieved and the extent to which the problem is solved during the course of problem solving. This objective accountability of the quality of a solution differentiates CPS from collaborative learning, which focuses on helping individual students learn a subject matter or set of skills through collaboration, without the requirement that any problem be solved. The extent to which the individual team members learn is typically not visible to team members during the process of collaboration; this learning becomes manifest later on when team members receive test scores on what they know about the subject matter.

Third, there is a *differentiation of roles* among the team members who take on different tasks to solve different aspects of the problem. This is different from collaborative-decision-making tasks in which all team members are presented the same set of questions to answer, items to judge, or decisions to make, often with access to the same information.

Fourth, CPS requires *interdependency* among team members, each bringing different resources to the table, rather than a single team member being able to solve the problem alone. Interdependency is essential for many problems in the 21st century, the complexity of which requires multiple domains of expertise and diverse perspectives that a single individual cannot provide. Individual problem solving is efficient and pragmatically wise to recruit for some problems, but not for a widening array of problems to solve in the 21st century.

Although CPS has distinctive features that distinguish it from collaborative work, learning, judgment, and decision making, there are also some features common to all forms of collaboration. Collaboration has potential advantages and disadvantages compared with completing tasks alone. Some advantages of collaboration are as follows: (a) A division of labor can enhance quality; (b) there are multiple sources of knowledge, perspectives, and experiences; (c) team members are stimulated by ideas of other team members so emergent ideas might evolve from the interaction; and (d) multiple members enhance evaluation of the products of collaboration. In contrast, there are potential disadvantages to the extent that (a) communication is inefficient (e.g., team members waste time with irrelevant discussion); (b) social loafing occurs (e.g., a team member does not deliver); (c) diffusion of responsibility occurs (e.g., team members assume that other team members will complete tasks); and (d) conflict, disagreements, and false information disrupt productive discussion, idea generation, and evaluation.

Research on the various forms of collaboration (i.e., work, learning, judgment, decision making, or problem solving) have identified a number of factors that influence the quality of the collaboration. For example, simply assigning individuals into groups is not always sufficient to establish effective collaboration. Instruction, guidance, or an activity with clear structure is often needed to prepare individuals for group work. Research in collaborative learning has shown that students in structured groups show better learning outcomes than those in unstructured groups (Gillies, 2008; Vogel, Wecker, Kollar, & Fischer, 2017). Social loafing, task complexity, and group composition are additional factors that influence collaborative learning, as discussed later in this article. We expect that these same factors would predict performance in CPS.

Research in collaborative learning and decision making has often compared individual and team measures of performance. In collaborative learning, the question is whether learning is better for individuals when they study alone versus in groups (Slavin, 2017). Measures of learning for particular subject matters are collected later, after the learning process is completed, as opposed to during the learning process. Researchers investigate whether students overall or in various student subcategories show better learning when they were part of groups versus learning alone. In collaborative decision making, the question is whether the accuracy of a judgment or decision is higher when an individual provides his or her data alone versus after a group deliberation (Hastie & Kameda, 2005). There is an objective answer in these tasks, so researchers can measure the discrepancy between the judgments or decisions and an objective standard. The researchers can measure the added value of a group deliberation over and above individuals in different categories (best, second best, average, etc.) after quantitative and statistical adjustments that control for various measurement biases (Gigone & Hastie, 1997). As one might expect, available empirical research does not converge on a simple conclusion regarding whether learning, judgment, and decision making are better in groups versus alone.

Research on collaborative memory is another paradigm that investigates how the social interactions associated with collaboration influence how individuals encode, store, and retrieve information, sometimes showing memory costs associated with collaboration (Andrews & Rapp, 2015; Rajaram, 2011). For example, although collaborators tend to recall more information than one person does alone, they recall less information than do nominal groups, which are groups formed by pooling together the nonredundant recalls of noninteracting individuals. The superior memory performance of nominal groups relative to collaborating groups has been termed *collaborative inhibition* (Andersson, Hitch, & Meudell, 2006; Weldon & Bellinger, 1997). When recalling information collaboratively, group contributions can not only disrupt the retrieval of individual memory through collaborative inhibition but also induce the forgetting of information that was previously known but never mentioned during group discussion (Cuc, Koppel, & Hirst, 2007). These effects of retrieval-induced forgetting are not only observed for the individual recalling a target item but also can occur for conversational partners present during recall (Stone, Barnier, Sutton, & Hirst, 2010). Moreover, during collaborative discussion, group members may introduce information that is inaccurate. A number of studies have demonstrated that such inaccurate contributions often go undetected and become encoded by group members and used on subsequent tasks, resulting in an effect called the *social contagion of memory* (Andrews & Rapp, 2014; Davis & Meade, 2013; Roediger, Meade, & Bergman, 2001; Thorley & Christiansen, 2018).

These comparisons of findings from collaborative versus individual performance in learning, memory, judgment, and decision-making tasks would presumably generalize to the CPS tasks that do *not* have differential roles and interdependency among team members. At this point in team-science research, the body of CPS research is insufficient to assess such generalizations. The impact of a single team member on a group is expected to be different for CPS and collaborative work than for collaborative learning and decision making. An underperforming team member can slide by in group learning and decision making because an individual's achievements are either invisible or inconsequential (e.g., a single vote is unimportant unless the team vote is close), and the group can readily move on unless there is a stalemate. In contrast, an underperforming member of a CPS team (or a work team) can create a major impasse for the entire team and require replanning and reassignment of team members' responsibilities. Therefore, there are reasons to be skeptical of whether research on collaborative learning and decision making will generalize to CPS.

In our view, the most interesting CPS tasks are sufficiently challenging that they require differential roles and interdependency of team members. Consequently, the comparisons between groups and individuals in these other forms of collaboration arguably have minimal or secondary relevance to CPS. In essence, we argue that the complexity of problems in the 21st century are sufficiently difficult that an individual could never solve the problems alone! Team members depend on the contributions of others in different fields, with different knowledge, skills, abilities, and expertise, to make significant progress in solving the challenging problems of today. Whenever it is obvious that a problem cannot be solved by a single individual, then the important questions for science and practice address how team members can advance team goals in CPS. This leads to the question of why it is important to investigate CPS mechanisms and improve CPS proficiency in schools and the public.

Why Does Society Need to Improve CPS Proficiency?

The importance of CPS is increasing with the complexity of human social systems and the problems to be solved. As the 21st century progresses, the complexity of socio-technological systems across industry, the military, and academia is ever-increasing (Autor, Levy, & Murnane, 2003; Letsky et al., 2008). Concomitant with this, collaborative cognition is becoming increasingly prevalent as societies involve multiple stakeholders and become more dependent on deep knowledge for solutions to difficult problems (Fiore, 2008; Hall et al., 2018). Inquiry into CPS mechanisms continues to evolve, given the need to better understand how to improve collaborative processes and solutions to complicated problems. In these environments, teams are required to solve complex problems the solutions to which require integration of knowledge across any number of interconnected systems that are distributed across people and machines (Fiore & Wiltshire, 2016; A. Fischer, Greiff, & Funke, 2012). We are not suggesting that individual work will disappear, nor are we denying the inherent creativity that comes from solo endeavors. Rather, we are arguing that, by necessity, many of the pressing problems of the 21st century require collaboration. Because of this, research on CPS must similarly evolve to understand and address the mutual needs of learning and performance by individuals and teams.

The PISA assessment unveiled problematic deficiencies when it comes to student competencies in collaboration (OECD, 2017b). Approximately 8% of students across OECD member countries scored at the highest level (Level 4, as defined in Fig. 1). This level demands that students complete tasks requiring complex forms of collaboration on very challenging problems to be solved. The collaboration requires them to overcome social obstacles in team behavior, to resolve or circumvent team conflicts, and to take the initiative to lead the team to handle the most difficult challenges. The results showed that 28% of the students scored at Level 3 and 35% at Level 2 (Fig. 1). The remaining 29% of the students were limited to solving items requiring the lowest form of complexity in collaboration and problem solving (Level 0 or Level 1; Fig. 1). This extremely low success rate is very troubling because these are precisely the skills that are needed in the workforce. The experiences of students in and out of the classroom are not preparing them to have the skills that are needed as adults.

Speculations about this deficit have indeed been circulating for a number of years. More than a decade ago, reports explicitly acknowledged the lack of a sufficient workforce that is capable of contributing to the modern knowledge-intensive economy (National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, 2007, 2010). At the same time, there were reports that recognized the lack of a workforce capable of collaborating across disciplines (National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, 2005; National Research Council, 2015). The results discussed in the PISA report provide converging evidence that deficiencies in CPS are a global problem.

In studies of workforce preparation, reports have identified gaps between the knowledge, skills, and abilities that employers are seeking and those held by college graduates (Hart Research Associates, 2015; National Academy of Sciences, Engineering, & Medicine, 2016). These knowledge, skills, and abilities predominantly include teamwork, interpersonal skills, and communication across professions, but unfortunately the existing linkages with education and curricula development do not line up with workforce preparation (National Research Council, 2011, 2012). Federal policy recommendations point to the evidence showing that studies of learning in school predict process and

Figure V.3.5 ■ **Summary descriptions of the four levels of proficiency in collaborative problem solving**

Level	Score range	What students can typically do
4	Equal to or higher than 640 score points	At Level 4, students can successfully carry out complicated problem-solving tasks with high collaboration complexity. They can solve complex problems with multiple constraints, keeping relevant background information in mind. These students maintain an awareness of group dynamics and take actions to ensure that team members act in accordance with their agreed-upon roles. At the same time, they can monitor progress towards a solution and identify obstacles to overcome or gaps to be bridged. Level 4 students take initiative and perform actions or make requests to overcome obstacles and to resolve disagreements and conflicts. They can balance the collaboration and problem-solving aspects of a presented task, identify efficient pathways to a solution, and take actions to solve the given problem.
3	540 to less than 640 score points	At Level 3, students can complete tasks with either complex problem-solving requirements or complex collaboration demands. These students can perform multi-step tasks that require integrating multiple pieces of information, often in complex and dynamic problems. They orchestrate roles within the team and identify information needed by particular team members to solve the problem. Level 3 students can recognise the information needed to solve a problem, request it from the appropriate team member, and identify when the provided information is incorrect. When conflicts arise, they can help team members negotiate a solution.
2	440 to less than 540 score points	At Level 2, students can contribute to a collaborative effort to solve a problem of medium difficulty. They can help solve a problem by communicating with team members about the actions to be performed. They can volunteer information not specifically requested by another team member. Level 2 students understand that not all team members have the same information and can consider differing perspectives in their interactions. They can help the team establish a shared understanding of the steps required to solve a problem. These students can request additional information required to solve a problem and solicit agreement or confirmation from team members about the approach to be taken. Students near the top of Level 2 can take the initiative to suggest a logical next step, or propose a new approach, to solve a problem.
1	340 to less than 440 score points	At Level 1, students can complete tasks with low problem complexity and limited collaboration complexity. They can provide requested information and take actions to enact plans when prompted. Level 1 students can confirm actions or proposals made by others. They tend to focus on their individual role within the group. With support from team members, and when working on a simple problem, these students can help find a solution to the given problem.

Fig. 1. Summary descriptions of the four levels of proficiency in Programme for International Student Assessment collaborative problem solving. Reprinted from Organisation for Economic Co-operation and Development (2017b). Used with permission.

performance in later contexts (National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, 2005; National Research Council, 2015).

The results of a survey by the American Management Association (2012) showed that high-level managers believed that students coming from college do not possess the skills needed for collaboration. They cite an overemphasis on what we would call "task knowledge": course content focusing on high-tech skills such as math and science, without concomitant emphasis on communication and collaboration. According to a report on career preparation recently commissioned by the Association of American Colleges and Universities, college graduates' self-perceptions of their own knowledge, skills, and abilities diverge from employers' perceptions (Hart Research Associates, 2015). For example, nearly two thirds of college graduates believe they can effectively work in a team, whereas only approximately a third of managers stated that college graduates demonstrated this competency. Likewise, more than half of college graduates felt they were able to work with others with different backgrounds but less than a fifth of managers saw this to be the case.

These results support the recommendation to develop pedagogical approaches that incorporate CPS into academic curricula. In schools and colleges, CPS has had either a secondary status or has been relegated to extracurricular activities (e.g., band, sports, student newspapers) rather than being part of the core curriculum. We argue, therefore, that CPS training needs to be *curricular* rather than *extracurricular*. There is a need to better understand how to develop and adopt methods for learning CPS processes (Fiore et al., 2017, 2018; Gilbert et al., 2018; Graesser, Dowell, et al., 2018; Graesser, Foltz, et al., 2018; Hesse et al., 2015; Scoular, Care, & Hesse, 2017). The PISA report points to the kinds of educational initiatives that help develop collaborative competencies for all students. The projects span from the intellectual work of collaborative writing and science projects to the procedural and psychomotor activities of marching bands and sports teams (Kniffin, Wansink, & Shimizu, 2015).

Unfortunately, instructors have rarely had the professional development that informs student training on collaboration competencies and that would allow them to provide useful feedback to the students. There presumably is value in systematic training of students to collaborate in varying contexts, disciplines, and projects, over and above simply helping students master course-content knowledge and occasionally assigning group projects with minimal feedback on CPS processes and outcomes. Given the current state of research on CPS training, it is premature to prescribe a specific CPS curriculum. However, there is enough suggestive evidence to point the way forward for initial implementation of alternative forms of CPS training in the classroom as well as informal and extracurricular activities. At present, students rarely receive meaningful instruction, modeling, and feedback on collaboration (Fiore et al., 2018). They are typically graded on the outputs of the task-relevant content of their projects rather than the quality of the process to complete the tasks. Moreover, students may sometimes receive feedback on their teamwork, but the training on CPS is rarely informed by psychological science. Psychological science can play a role in gaining a better understanding of effective training of CPS proficiencies in addition to investigating CPS mechanisms.

CPS Theoretical Frameworks

This section provides an overview of theoretical frameworks of CPS. These have been developed out of an integration of findings on team-based research from the social, cognitive, and learning sciences, as well as national or international assessment frameworks. One fundamental goal has been to reconceptualize the process of problem solving from individual and isolated work to one that involves multiple people with different roles working interdependently toward a common goal. Collaborators attempt to construct a shared understanding of the problem and team goals from a complex set of inputs and subsequently to develop a plan of action that considers the roles, responsibilities, constraints, and tasks of individual team members. This typically is a dynamic, emergent process. That is, individuals within the team, and the team itself, come to comprehend the elements of the problem situation by interacting and by interpreting the information provided by team members as well as by the dynamically evolving situation.

Most CPS theoretical frameworks have two overarching components: (a) the collaborative, communicative, or social aspects that are coupled with (b) the cognitive problem-solving aspects. However, the theories sometimes differ in the details of how the coupling is accomplished between the teamwork and task work. Teamwork is fundamentally social, comprising communication, the exchange of ideas, and a shared identification of the problem and its elements. There are negotiated agreements on connections between the parts of the problem, tasks to accomplish, and potential solutions. There is the need to manage relationships between people, their actions, and the effects that those actions produce. As an obvious contrast to individual problem solving, CPS requires that these elements are transparent to most or all members of the team and that team members are aware of the important elements. A transparent, visible, shared vision and series of updates is critical to the success of groups. In comparison with individual problem solving, in which these steps are internally managed as one works through the problem in a more private manner, CPS introduces added layers of psychological processing associated with social cognition.

CPS has challenges at multiple levels that can hinder the achievement of group problem-solving goals. Regarding task work, there are challenges in accessing, combining, and synthesizing multiple forms of data and information in the service of knowledge integration. There are challenges in formulating plans, tracking progress toward goals, and revising plans when unexpected obstacles occur. Regarding teamwork, there are challenges associated with the collaboration. These include unwise assignment of team-member roles, interpersonal problems that create conflict, communication problems, attitudinal problems, low group cohesion or trust, and coordination problems.

In the remainder of this section, we first provide a summary of social factors and how they might influence CPS. The earliest work on collaboration was conducted in social psychology, so that is where we will start. We then summarize some of the foundational theoretical work developed to study collaborative cognition and problem solving. Next come the two broad frameworks used in international CPS assessments, namely PISA and Assessment and Teaching of 21st Century Skills (ATC21S). This theory-review section sets the stage for the second half of our article, in which we review some quantitative and technological approaches to investigating CPS processes, followed by instructional methods for training CPS skills. As we cover these sections, we identify some of the debates and disagreements that have surfaced in the short history of CPS research, assessment, and training.

Social factors

Social factors influence the extent to which the CPS processes succeed, but the factors that are particularly relevant to CPS have yet to be documented empirically. A first step is to identify the social factors that are known to be important in group collaborations more generally. Important moderator variables are associated with team composition, such as personality, diversity of team-member perspectives, their knowledge, and other background characteristics. Perceptions of group members can influence team dynamics and how members interact. For example, beliefs about a member's competence or knowledge in a particular domain can influence whether individuals accept and rely on the information from that person (Andrews & Rapp, 2014). People use group membership to make inferences about people's credibility (Horry, Palmer, Sexton, & Brewer, 2012). That is, people tend to perceive members of their in-group as more credible than members of an out-group (R. D. Clark & Maass, 1988; Doosje, Branscombe, Spears, & Manstead, 2006). These judgments about competence in turn influence whether group members rely on information that is shared in the group (Andrews & Rapp, 2014).

Personality undoubtedly influences how individuals behave in group settings. Diversity in group members' personality has been shown to influence performance outcomes. Specifically, variability in agreeableness and neuroticism among group members has a negative impact on performance (Mohammed & Angell, 2003), whereas higher average agreeableness in a team can positively affect performance outcomes (Barrick, Stewart, Neubert, & Mount, 1998; S. T. Bell, 2007). Agreeable group members are presumably more likely to engage in the positive interpersonal behaviors known to support performance outcomes (LePine, Piccolo, Jackson, Mathieu, & Saul, 2008). However, if all of the members are too agreeable, there is a risk of "group think," in which group members agree with each other to be polite or to minimize the time and effort required to arrive at a good solution (Dillenbourg, 1999; Janis, 1982; Stewart, Stelock, & Fussell, 2007). Prior work has also shown that diversity in extraversion among group members can positively affect performance outcomes (Mohammed & Angell, 2003; Neuman, Wagner, & Christiansen, 1999). A recent dissertation (Herborn, 2018) examined the correlations between students' PISA scores and the Big Five personality constructs (Costa & MacCrae, 1992); the author found a positive correlation between CPS scores and both openness and agreeableness, but no correlation with neuroticism, extroversion, and conscientiousness. Therefore, available research has not established a clear picture on the relationship between personality and different types of collaboration.

These personality results, as well as those shown for other factors associated with team composition, are important to consider when developing assessments of and training for CPS. However, available empirical findings have not converged on adequate guidelines on how to assemble groups with particular team-member characteristics and roles in different CPS tasks. There is a frequent intuition that particular ways of grouping individuals can have implications for the ways in which group members interact and perform. However, there are potential disagreements on the particular principles that predict successful grouping. For example, a student who identifies with a teammate based on some social characteristic (e.g., gender or ethnicity) might be more willing to accept and rely on information shared by that teammate compared with information shared by a student considered to be in an out-group. This would suggest that there are advantages to having group members be similar in CPS tasks. However, the PISA 2015 results (OECD, 2017b) support the conclusion that experience with diversity in schools positively predicts CPS performance, presumably because diversity spawns multiple perspectives and planning strategies. These clearly are two incompatible predictions. And there are many other potentially conflicting predictions. For example, high-ability team members perform differently, presumably better, when working with a team whose members are relatively high in ability. On the other hand, assigning high-ability members to the same team may have liabilities, such as power struggles. The psychological sciences are needed to explore the validity of these incompatible predictions, some of which may be sensitive to different phases of CPS.

Social cohesion is another important factor to consider for CPS. Research on cohesion in teams is extensive, and the consensus is that team cohesion is positively related to team effectiveness (Beal, Cohen, Burke, & McLendon, 2003; Evans & Dion, 1991; Gully, Devine, & Whitney, 1995; Mathieu, Kukenberger, D'innocenzo, & Reilly, 2015; Mullen & Copper, 1994; Oliver, Harman, Hoover, Hayes, & Pandhi, 1999). The relationship between cohesion and team effectiveness has been found to be moderated by a number of factors, such as task interdependence and group size. Specifically, the positive relationship between cohesion and team effectiveness is stronger when team members are more interdependent (Gully et al., 1995) and in smaller groups (Mullen & Copper, 1994). Cohesion is also associated with leadership in teams (Mathieu et al., 2015) and can mediate diversity differences in teams (Liang, Shih, & Chiang, 2015). These results have implications for the kinds of tasks most suitable for supporting the cohesion–performance relationship and the potential consequences of different group characteristics (e.g., size, composition).

In some group situations, loss of motivation, such as social loafing, can occur as individuals withhold or reduce their efforts during engagement in a task. In these situations, individuals do not exert as much effort as when they are working alone. A number of factors can contribute to individuals behaving in this way. One factor is the identifiability of one's individual effort in a group setting. That is, people may neither receive credit for their inputs nor receive blame for their lack of inputs in group settings (Latané, Williams, & Harkins, 1979). As group size increases, individual anonymity increases and so does social loafing (Liden, Wayne, Jaworski, & Bennett, 2004). Dispensability of effort can be another factor that contributes to individuals engaging in social loafing. That is, individuals have little motivation to exert much effort on a task if they believe their effort will have little impact on whether the team succeeds (Karau & Williams, 1993; N. L. Kerr, 1983).

The potential for social loafing and associated underlying social factors has important implications for the design of CPS assessments and training. Designing tasks and instructions so that people are aware that their individual contributions are identifiable and will be evaluated can decrease the likelihood of social loafing (Williams, Harkins, & Latané, 1981). Choosing to use smaller group sizes in CPS can increase the identifiability of the efforts of particular individuals during task engagement and thus decrease the likelihood of social loafing. Perceptions of the dispensability of group members' efforts can be reduced by design decisions that make group members believe their input is important for success of the group. Such decisions include creating tasks that are more challenging or providing group members with unique roles or aspects of the task (Liden et al., 2004). Some pedagogical methods for accomplishing these goals have been developed in the collaborative learning literature, but whether these also apply to CPS remains an open question.

Here is one noteworthy example: *Jigsaw* is a method in the collaborative learning literature in which each group member is responsible for particular subject-matter topics in the group task, and all of the topics are essential for a complete understanding of the subject matter (Aronson & Patnoe, 1995). None of the members at the beginning of the task has all of the relevant knowledge, so the individual contributions of each group member are important for the success of the group; this can reduce the likelihood of group members' losing interest or failing to contribute to the group product. However, the potential liability of the jigsaw method is that there is less common knowledge among team members at the beginning, and that might present an obstacle in CPS as well as other forms of collaboration.

In summary, psychological scientists have conducted research on group processes as well as the personality, knowledge, motivation, attitudes, and other characteristics of group members. The research has concentrated on group learning and decision making and associated tasks, so how well this research generalizes to CPS remains an open question. We identified a number of trade-offs among factors, so there are uncertainties in generating predictions about their impact on CPS performance.

Another important limitation of empirical investigations of group processes and products is that they have had small to modest sample sizes and, for some research paradigms, a narrow landscape of populations and contexts. As discussed later, we believe that some digital technologies and quantitative methods can help overcome these hurdles. Automated analyses of CPS interactions with computer-mediated communication are expected to accommodate larger samples of groups, broaden the diversity of populations, provide more detailed observations of CPS processes, and substantially speed up the process of data analyses.

Foundational frameworks for CPS

Researchers have drawn from a variety of disciplines during the course of developing a comprehensive approach for understanding and improving CPS. A number of frameworks on CPS were reviewed in the PISA 2015 report (OECD, 2017a). All of these frameworks assume that CPS includes a suite of subskills that are needed for effective teamwork in service of problem solving. Contributions in the following disciplines are noteworthy and influenced the CPS assessments described later in this section.

Education research. O'Neil, Chuang, and Chung (2003) drew from the literature on teamwork and the study of processes driving team effectiveness (Morgan, Salas, & Glickman, 1993) to identify the features needed to optimize CPS performance outcomes. The first is *adaptability*, which involves monitoring both the team and the task, as well as responding appropriately when problems emerge. Next is *coordination*, which includes the synchronization and integration of group activities to accomplish the task in a timely fashion. Third, *interpersonal* skills include cooperating with and accommodating other members of the group in service of the task. *Leadership* is also required to offer direction for the group, and, finally, *communication* skills are needed so that members exchange information clearly and accurately. These subskills are executed in combination with what O'Neil et al. (2003) refer to as problem-solving strategies. In addition to the traditional cognitive strategies in individual problem solving (Mayer & Wittrock, 2006; Newell & Simon, 1972), these cognitive strategies include a *self-regulatory* component whereby problem solvers autonomously monitor and modify behavior as needed and regulate motivation to ensure effort. Last is *metacognition*, which in CPS consists of planning for problem solving and periodic checking to ensure strategies are being executed appropriately.

Cognitive science. Researchers have pursued quantitative modeling of groups as social-cognitive systems (Theiner, Allen, & Goldstone, 2010). A framework of *multiple, interacting levels of cognitive systems* (MILCS) attempts to account for the cognitive capacities of individuals and the groups to which they belong (Goldstone & Theiner, 2017). Group cognition is assumed to dynamically emerge from the individuals in the group during the course of their interactions. The dynamic mechanisms are modeled by lateral inhibition within a network, diffusion processes for accumulating evidence, and other computational models that have a long history in cognitive science in modeling individual cognition. These dynamic models support the development of collective intelligence whereby a group can show performance over and above what would be possible by any individual (Goldstone & Theiner, 2017). In a study of one kind of collective problem solving, Goldstone, Roberts, and Gureckis (2008) studied path finding in a spatial environment. They showed that a kind of collective intelligence emerges when individuals implicitly solve the problem by moving between paths for themselves, which leads to a more efficient path for everyone when footpaths are imprinted in the landscape. The MILCS framework is relevant to the information search part of CPS in which the group needs to find an optimal solution among a large space of alternatives. CPS requires identification and sharing of information relevant to a particular problem faced by the group. Sometimes it is optimal for different team members to search locally in depth in a limited region of the space or have differentiated roles, rather than having a wide net of team members vote on a particular decision or direction. Computational models are developed to simulate and develop a mechanistic explanation of various phenomena of relevance to CPS.

Macrocognition-in-teams model. In a field that integrates the cognitive and organizational sciences, the *macrocognition-in-teams model* (MITM) was developed to help improve understanding of complex CPS. The MITM captures the iterative processes that unfold during collaborative cognition (Fiore, Rosen, et al., 2010; Fiore & Wiltshire, 2016). It is a multilevel framework that incorporates individual and team level factors. It considers cognition that is both internalized (e.g., knowledge) and externalized (e.g., artifacts). It considers temporal characteristics to examine phases of collaboration and how these alter process and performance. The MITM has five overarching components:

1. *Internalized team knowledge* is the knowledge held by each individual team member that forms the bases for shared cognition. This might include knowledge about a problem domain or about members of a team.
2. *Individual knowledge building* consists of the processes involved in increasing understanding of the problem area.
3. *Team knowledge building* comprises the processes engaged to develop actionable knowledge related to the problem on which the team is working.
4. *Externalized team knowledge* is the knowledge made explicit and shared by the team. This includes knowledge communicated verbally or through artifacts to understand what happens when something internally held by an individual is distributed across members of the team.
5. *Team-problem-solving outcomes* are the products of prior actions, solutions, and objectives. These are evaluated by the team to determine how well they meet objectives, and revised as needed.

These major macrocognitive components include a number of subprocesses engaged individually and collaboratively as the team works through the problem (Fiore, Elias, Salas, Warner, & Letsky, 2010; Fiore, Rosen, Salas, Burke, & Jentsch, 2008; Fiore, Rosen, et al., 2010; Fiore, Smith-Jentsch, Salas, Warner, & Letsky, 2010).

Researchers from a variety of domains have applied the MITM to examine the processes that teams use when engaged in complex work settings. For example, the MITM has been used to study expert teams experienced in air-operations centers to better understand their collaborative processes (S. G. Hutchins & Kendall, 2010); these teams primarily engaged in individual information gathering and team information exchange. Others have used the MITM in organizational studies examining how teams collaborate to develop complex information systems (Seeber, Maier, & Weber, 2013). More recently, the MITM was used to examine CPS in teams at NASA's Johnson Space Center (Fiore, Wiltshire, Oglesby, Okeefe, & Salas, 2014). This research showed how knowledge building was driven by internalized knowledge of team members and how they externalized that knowledge collaboratively with situation updates.

Recent research in the cognitive sciences has used the MITM to study dynamic and emergent collaborative processes. Transitions between phases of problem solving were identified using the processes detailed in the MITM (Wiltshire, Butner, & Fiore, 2017). Team communications were used to identify phase transitions in CPS to illustrate meaningful change in coordination behaviors. A measure of entropy showed how communication patterns change when teams transition from one problem-solving phase to another. When communication had high levels of entropy, this signaled a shift in coordination for that team. Other measures of entropy were indicative of the team's stability, flexibility, and problem-solving performance.

In summary, a number of theoretical frameworks underlie CPS and associated social and collaborative processes. There clearly is no lack of theory, and the frameworks have many points of commonality. For example, they agree that it is useful to somehow contrast the collaborative social aspect (sometimes called teamwork) from the cognitive problem-solving aspect (task work). However, the frameworks differ in the details on how these two aspects are integrated and which components fall under the umbrella of each aspect. A resolution of these differences will undoubtedly require substantially more empirical analyses of CPS on different tasks in different contexts for different populations.

We now turn to two CPS frameworks that were developed based on much of the aforementioned literatures. These frameworks were used to assess students around the world in competencies related to CPS.

PISA CPS 2015 framework

As discussed earlier, PISA had the first international assessment of CPS (OECD, 2017a, 2017b). The PISA expert group was instructed by OECD's Program Governing Board to reconstruct the scientific literature on CPS and to formulate an assessment of CPS proficiency that could be collected from several dozen countries with citizens who speak several dozen different languages. As with all prior PISA assessments, the test takers were 15-year-old students and had a minimum of 1,500 students per country. The expert group was instructed to consider the cognitive aspects of CPS but not the personality and emotions of the test takers. A student taking the CPS assessment was limited to two 30-min sessions.

The PISA expert group was tasked with coming up with a succinct definition of CPS competency so the following definition was developed (OECD, 2017a):

> 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, and pooling their knowledge, skills and efforts to reach that solution. (p. 7)

This definition served the purpose of conceptually describing the construct, whereas an assessment framework was also needed to offer more concrete guidance in operationalizing measures. According to the definition, an "agent" could be either a human team member or a computer agent that interacts with the student. The PISA assessment ended up having individual test takers interact with one or more computer agents.

The PISA assessment framework has both a cognitive dimension (task work) and a collaborative dimension (teamwork), which is compatible with most theoretical articulations of CPS. The problem-solving dimension in the PISA framework incorporated the four PISA 2012 competencies that targeted individual problem solving (Funke, 2010; Greiff, Kretzschmar, Müller, Spinath, & Martin, 2014; OECD, 2010).

- *Exploring and understanding*. Interpreting the initial information about the problem and any information that is uncovered during the course of exploring and interacting with the problem.
- *Representing and formulating*. Identifying global approaches to solving the problem, relevant strategies and procedures, and relevant artefacts (e.g., graphs, tables, formulae, symbolic representations) to assist in solving the problem.
- *Planning and executing*. Constructing and enacting goal structures, plans, steps, and actions to solve the problem. The actions can be physical, social, or verbal.
- *Monitoring and reflecting*. Tracking the steps in the plan to reach the goal states, marking progress, and reflecting on the quality of the progress or solutions.

There were three processes on the collaborative dimension in PISA CPS 2015 (OECD, 2017a, 2017b):

1. *Establishing and maintaining shared understanding*. Keeping track of what each other team member knows about the problem (i.e., shared knowledge, common ground, H. H. Clark, 1996), the perspectives of team members, and a shared vision of the problem states and activities (Cannon-Bowers & Salas, 2001; Dillenbourg & Traum, 2006; Wegner, 1986).

2. *Taking appropriate actions to solve the problem.* Performing actions that follow the appropriate steps to achieve a solution. This includes physical actions and communication acts that advance the solution to the problem.
3. *Establishing and maintaining group organization.* Helping to organize or reorganize the group by considering the knowledge, skills, abilities, and resources of particular group members during the assignment of roles. This also includes following the rules of engagement for particular roles and the group, as well as handling obstacles to tasks assigned to other team members.

Crossing the four problem-solving processes with the three collaboration processes results in 12 skills in the *CPS assessment matrix*, as shown in Figure 2. A satisfactory assessment of CPS would assess the skill levels of students for each of these 12 cells, and these would contribute to a student's overall *CPS proficiency measure*. Within each cell, one can assess the extent to which the student takes the initiative or leadership in overcoming obstacles rather than being merely responsive to the requests of others or being unresponsive or unhelpful to solving the problem.

The PISA theoretical framework report (OECD, 2017a) was much broader in scope than this assessment framework. A full description of all of the relevant factors discussed in the OECD report is beyond the scope of this article. We note simply that, in addition to covering the available theory and empirical research on CPS, the report identified the following factors that should be considered in a CPS *theoretical* framework but were not directly incorporated in the *assessment* framework summarized in Figure 2:

Task characteristics. A number of distinctions were made, such as between interdependent or independent solutions, well-defined or ill-defined problems, static or dynamically changing problems, and team members having the same or different goals.

Problem scenarios. The scenarios were classified into different problem categories (e.g., Jigsaw, consensus, or negotiation) and content distinctions, such as private or public problems, technical or nontechnical subject matters, and school or nonschool contexts.

Team composition. The team members could have symmetrical or asymmetrical status, same or different roles, and be part of groups that varied in size.

Team-member characteristics. The individual team members varied in knowledge (math, reading and writing, science, or everyday knowledge) and psychological attributes (dispositions and attitudes, motivation, cognitive abilities).

An example problem scenario and task that illustrates how CPS was assessed are presented later in this article. The assessment covers the skills in Figure 2 and produced the CPS proficiency scale in Figure 1. The sample of test-takers consisted of approximately 540,000 15-year-old students in 52 OECD countries. This sample size is both large and diverse across different languages and cultures.

Three major limitations of the PISA assessment framework have been expressed since the inception of this first international test of CPS proficiencies. The first limitation is that students interacted with computer agents rather than other students through face-to-face interactions or computer-mediated communication. The use of computer agents satisfied various logistical challenges, as will be discussed later, but it still raised the concern that the assessment environment deviated from naturalistic, ecologically valid CPS activities. The second limitation is that the constraints of the agent-based assessment cut off a number of the discourse patterns that are central to CPS. In particular, negotiation is a very important conversational pattern that is part of establishing shared knowledge, making a decision, or agreeing on a course of action. It often takes a multiturn exchange between team members to negotiate, but the PISA assessment allowed only one exchange rather than multiple exchanges. Third, as requested by OECD, the PISA assessment did not consider the personality of the team members and their emotions. This logistical decision of OECD is understandable in order to accommodate many countries and cultures, but a mature theory of CPS would need to consider the personalities and emotions of team members.

Assessment and Teaching of 21st Century Skills

The ATC21S (Griffin & Care, 2015) project was designed to develop a concept of both assessing and teaching CPS skills. The result was a framework for teachable CPS skills (Hesse et al., 2015). Unlike PISA, ATC21S was not developed to focus on a subset of skills that could be measured in a large-scale assessment with many countries. ATC21S did, however, attempt to formulate standardized measures of various CPS skills among individuals in human-to-human interactions in order to permit comparisons in subsequent research.

The ATC21S theoretical framework includes skills in participation, perspective taking, and social regulation as relevant components.

Figure V.2.1 ■ **Skills evaluated in the PISA 2015 collaborative problem-solving assessment**

		Collaborative problem-solving competencies		
		(1) Establishing and maintaining shared understanding	**(2) Taking appropriate action to solve the problem**	**(3) Establishing and maintaining team organisation**
Problem-solving processes	(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

Fig. 2. Matrix of collaborative problem-solving skills for Programme for International Student Assessment 2015. Reprinted from Organisation for Economic Co-operation and Development (2017b). Used with permission.

Participation potentially includes a long-term process of becoming part of community of practice, as described by Lave and Wenger (1991), along with action, interaction, and task completion. *Action* refers to the general level of participation (E. Hutchins, 1995; Nardi, 1996), whereas *interaction* describes behavior with and in response to others and is a minimum requirement for successful coordination (Crowston, Rubleske, & Howison, 2006). *Task-completion* skills include motivational aspects of participation that are needed to stay engaged in the problem-solving activity.

Perspective taking involves understanding and responding to others. From a discourse perspective, it refers to the ability to contextualize utterances of the others in a CPS situation, to integrate their contributions, and to tailor one's utterances to others, or what is called *audience design* (H. H. Clark, 1996; Schober & Clark, 1989). Audience-awareness skills are needed to integrate one's own contributions with those of others (Dehler, Bodemer, Buder, & Hesse, 2011) and to avoid too much of an egocentric bias.

Social regulation starts from the assumption that groups can profit from the diversity of their members (De Wit & Greer, 2008; Doise & Mugny, 1984; Roschelle, 1992; Salomon, 1993; van Knippenberg & Schippers, 2007). However, diversity per se is not enough, because team members need to work with the diversity productively (Thompson, Wang, & Gunia, 2010; van Knippenberg, De Dreu, & Homan, 2004). The framework distinguishes four aspects that can be related to social regulation: metamemory, transactive memory, negotiation, and initiative. The first two refer to the ability to recognize group diversity, which breaks down into knowledge about oneself (*metamemory*; Flavell, 1976) and knowledge about the knowledge, strength and weakness of the collaborators (*transactive memory*; Wegner, 1986). The presence or absence of *negotiation* skills can be observed when conflicts arise between the group members. Problem solvers negotiate differences between individual approaches by formulating compromises or by determining rank orders among alternative solutions. The *initiative* skills refer to the responsibility of a group member for the progress of the problem-solving process in the group. If the collective responsibility is too low, lurking behavior or disengagement from the task might show up, and tasks may become unsolvable.

The APC21S was the first effort to identify the important components of CPS assessment and teaching and the first to collect data from a large sample of students in countries that speak English. The theoretical framework was broad and covered some detailed discourse mechanisms that are prominent in CPS activities. The fact that human-human interaction data were collected was an important advance so that hypotheses could be tested on effective CPS processes, and the data could be mined for discovering new promising collaboration patterns that predict CPS performance (Care, Scoular, & Griffin, 2016; Scoular, Care, & Hesse, 2017).Unfortunately, it has taken many years to annotate and analyze the data. Consequently, it was not practical to provide summative assessments of CPS in schools and timely formative assessments to students and instructors on improving CPS skills in individuals or groups. The effort also did not evolve into a curriculum for CPS education that has been widely accepted (Scoular & Care, 2018).

Once again, automated analyses of CPS processes can play a role in providing summative and formative assessment. There are promising opportunities to incorporate recent advances in digital technologies so that the detection of CPS processes and the assessment of various CPS competencies can be automated. If this is successful, then extremely large data sets across a wide landscape of tasks, contexts, and populations can be collected and analyzed. That would provide a notable advance in discriminating theories, testing hypotheses, and guiding the development of an educational curriculum for CPS education and training.

Comparisons of PISA and ATC21S frameworks

In comparing the two frameworks, it is important to distinguish between the theoretical framework and the operationalization of the assessment framework. There are only small differences between PISA and ATC21S with respect to their theoretical frameworks (Care & Griffin, 2017; Care et al., 2016; Harding, Griffin, Awwal, & Alom, 2017). The PISA theoretical framework (OECD, 2017a) integrated the ideas from ATC21S, which had started years earlier, so it is not surprising that the differences were small. Both frameworks postulate two main dimensions: a cognitive dimension (problem solving, task work) and a social dimension (collaboration, teamwork). The main differences between the two frameworks lie in the operationalization and assessment of the construct by applying human-to-agent (H-A) versus human-to-human (H-H) interactions. Accordingly, each framework goes along with specific advantages and disadvantages that are primarily related to internal and external validity aspects.

Aside from the dimension of teaching and learning CPS skills, the ATC21S framework appears to be more ecologically valid for several reasons: (a) Interactions between real humans are realized through computer-mediated communication; (b) the differential status of team members could be seen as more authentic and related to real life; (c) the problem-solving and interaction processes are less scripted, and the set of predefined messages and actions is not limited; (d) a large range of individual variables and social processes can be studied, such as emotions and microinteractions; and (e) group-level outcomes are measureable (Care & Griffin, 2017; Care & Kim, 2018; Pöysä-Tarhonen, Care, Awwal, & Häkkinen, 2018; Scoular et al., 2017) as opposed to being limited to the responses of a single test taker with a constellation of computer agents.

On the other hand, the more standardized process within the PISA assessment framework can be seen as having higher internal validity because (a) tasks and problem spaces are better defined and (b) assessment items have a number of characteristics (e.g., the use of "minimalist" agents and predefined response sets) that are described below. These characteristics minimize several biases, ensure control over confounding variables, and allow a systematic and high-granularity variation of specific variables of interest to test specific hypotheses in future studies (Care & Griffin, 2017; Graesser, Forsyth, & Foltz, 2017). More pragmatically, the data-collection and scoring processes in the PISA assessment is less complex and time-consuming than the processes within the ATC21S assessment framework, which require expert annotation of the H-H interactions (Care & Griffin, 2017; Scoular et al., 2017).

A descriptive, empirical, comparative analysis of the PISA and ATC21S assessments would be worthwhile to conduct. A major step in that direction was pursued by Herborn, Stadler, Mustafić, and Greiff (2018), who compared an H-A condition with an H-H condition. In that study, the PISA framework design and PISA CPS tasks were used, but in the H-H condition, the computer-simulated agents were replaced with humans in the same classroom (i.e., an actual peer), and students were told that they were working with one of their classmates. The interaction interface remained unchanged (i.e., there was a limited set of predefined choices and chat messages). Herborn et al. (2018) found no CPS performance differences between the H-A and H-H conditions; that is, both the H-A and the H-H conditions loaded on one overarching CPS factor with very little method variance. However, compared with the H-A condition, students interacted more in the H-H condition. The authors interpreted this as the first evidence of the validity of the H-A approach adopted in PISA.

A similar study also applied the PISA framework design and compared audio and text chats with regard to H-H interactions (Nouri, Åkerfeldt, Fors, & Selander, 2017). More utterances were found in the audio chat condition, and some subskills from the PISA system were more prevalent with audio chat. Only two subskills showed significant correlations (surprisingly, both positive and negative) with the overall task performance, and only in the audio condition. Hence, the predictive validity of the PISA assessment framework was questioned by the authors of this study (Nouri et al., 2017).

At this point the debate continues whether the PISA assessment framework with computer agents provides a valid assessment of CPS proficiency. There is a growing literature of empirical studies that compare H-A and H-H interactions (Herborn, 2018; Liu, von Davier, Hao, Kyllonen, & Zapata-Rivera, 2015; Rosen, 2014; Rosen & Foltz, 2014). Some additional details about the use of agents and the conversations in the PISA implementation can help clarify some issues in the debate.

Implementation of PISA assessment with agents

The PISA proficiency measure assessed how well a single human interacted with computer agents during the course of problem solving. The test taker was one team member along with the team of computer agents who solved a problem in a CPS scenario. Therefore, the assessment was of an individual in a CPS group rather than of the group as a whole, and there was no possibility of assessing emergence of group CPS competency from the competency of individual group members. The test taker performed actions in an activity window or selected what to say next from a set of chat-message alternatives in a chat window. There was always a limited set of response options (two to five, typically) from which the test taker could select at each assessment episode, much like a multiple-choice test that is interwoven in a CPS scenario. This limitation in fixed options at each assessment observation permitted standard psychometric methods to be applied in the assessment analyses. However, this constraint did not allow open-ended test-taker responses or multiturn exchanges on conversation threads that the computer system did not anticipate. There were no multiturn exchanges in negotiations or co-constructions that built on other team members' ideas unless the computer system designed particular threads ahead of time. Moreover, the computer agents were static agents in a chat facility, without spoken messages, animation, or visual depictions. The use of minimalist agents was necessary to eliminate biases of culture, personality, and emotions, which were beyond the scope of the PISA assessment.

For example, Figure 3 shows a screenshot of an assessment unit in one of the scenarios of PISA. The CPS scenario is about students in a class hosting several out-of-town visitors from another country for a day. They have options of visiting one or more of three major city highlights (museum of local history, community market, electric car factory). However, there are constraints of time, transportation (train or bus), important locations (the three city highlights, school, food court, Internet café, stations), and distances between locations. The

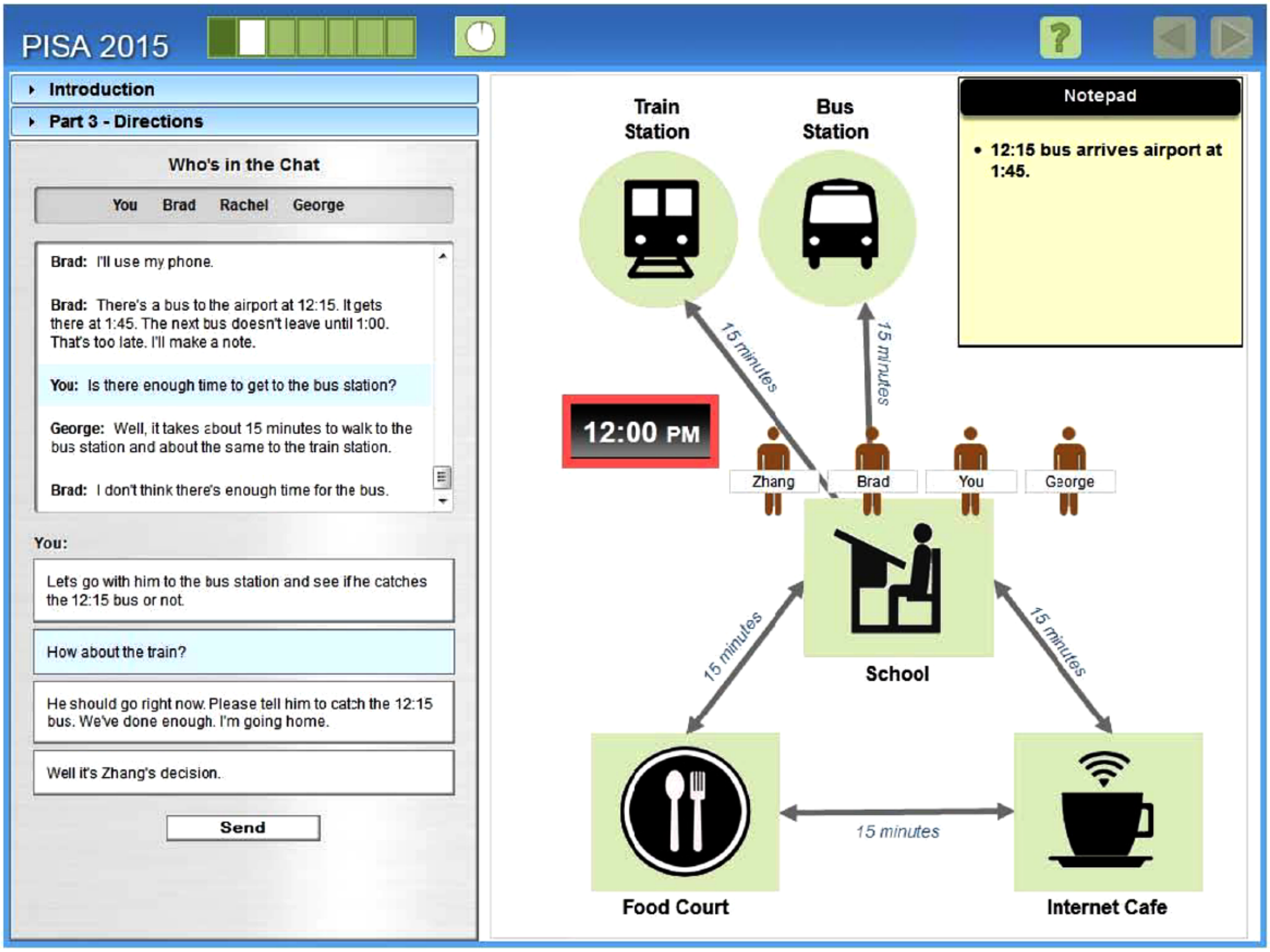

Fig. 3. Screenshot of an assessment unit in the Programme for International Student Assessment 2015. The left window of the screenshot is dedicated to the conversation among the team members. The right window sometimes allows the test taker to perform actions by clicking on an option in a turn. Image used with permission from the Organisation for Economic Co-operation and Development.

right window of the Figure 3 screenshot has images that show some of these constraints, along with a note pad. The right window also indicates that there are four team members, namely the test taker ("you") and three agent peers (George, Rachel, Brad). The right window sometimes allows the test taker to perform actions by clicking on an option in a turn. The left window of the screenshot is dedicated to the conversation among the team members. There is a chat history among the team members in the upper half of the left widow. The most recent conversational turn is by Brad, who expresses "I don't think there's enough time for the bus." At that point, it is time for the test taker to decide what to say in the chat. The test taker selects one of four options:

Let's go with him to the bus station and see if he catches the 12:15 bus or not.

How about the train?

He should go right now. Please tell him to catch the 12:15 bus. We've done enough. I'm going home.

Well it's Zhang's decision. [Zhang is a visitor]

The best option at this point in the problem is the second chat option because time is limited, the train is faster, and Zhang needs to abruptly return to his country to attend to an unexpected emergency. The test taker gets credit in one of the 12 cells in Figure 2 if the correct option is selected.

There is a diverse set of situations in the PISA assessment, which makes sure that important conditions are covered. A student who responds randomly to the response options will obviously receive low values on CPS proficiency as well as on the collaboration and problem-solving dimensions. A student might be a good team player and might be responsive but might not take the initiative when there are difficulties (e.g., an unresponsive agent or the occurrence of an unexpected new obstacle in the problem) or might take some initiative when there are breakdowns but might not be able to handle very complex cognitive problems. High scores in CPS proficiency indicate that a student takes the initiative in moving the team to achieve group goals during difficult times (conflicts, incorrect actions, unresponsive team members) and can also handle complex problems with many cognitive components that burden working memory and require reasoning. These conditions were needed to have an adequate CPS assessment. In contrast, many of these situations might not ever arise when a student interacts with other humans in many if not most teams. The absence of such conditions would create missing scores in some of the 12 cells of Figure 2 and would thereby threaten the validity of the assessment. As shown earlier, in Figure 1, there are four levels of CPS proficiency in the PISA assessment. Some students did not have the ability to reach the first level and would be assigned a 0 level, or not proficient.

Results of the PISA assessment revealed that the CPS proficiency scores statistically support the view that the PISA assessment was uncovering some variance that was unique to the CPS theoretical construct. The CPS scores were significantly correlated with literacy, numeracy, and science (between $r = .70$ and $r = .77$), whereas literacy, numeracy, and science had higher intercorrelations (between $r = .80$ and $r = .88$). These correlations used the student as a unit of analysis within countries. In addition, there was a high correlation ($r = .85$) between individual problem solving in the PISA 2012 assessment and the CPS of PISA 2015 when using country as a unit of analysis. Therefore, some unique variance taps problem solving as opposed to other cognitive constructs.

To provide a richer understanding of how social and demographic factors were related to CPS performance, questionnaires were administered to the students, parents, and teachers in PISA. This allowed researchers to correlate CPS proficiencies with demographic variables, classroom settings, and psychological characteristics (including the personality and attitudes of the students) in addition to assessments of CPS, numeracy, literacy, and science (OECD, 2017b). For example, CPS performance was higher for female students than for male students, higher for students exposed to diverse classrooms, and higher for students who participated in group activities in school. These correlational relationships were confirmed after statistically adjusting for obvious variables, such as socioeconomic status. The OECD report also assessed the extent to which CPS performance could be predicted by the attitudes of students, parents, and teachers toward CPS activities. However, it is beyond the scope of this article to review the major findings from of the PISA CPS 2015. The important take-home messages, from the present standpoint, are that (a) there is a theoretically grounded framework for assessing CPS proficiency; (b) the assessment has been administered to approximately 540,000 15-year old students in 52 countries; (c) there are dozens of intriguing findings on predictors of CPS proficiency from characteristics of students, parents, teachers, classroom settings, and cultures; and (d) systematic training of CPS skills is insufficient in educational settings throughout the globe (Fiore et al., 2018).

Once again, however, there was the worry that the students' individual interaction with agents would not be sufficiently similar to students' interaction with other students in a computer-mediated conversation. It was logistically necessary to implement human-agent interactions in PISA 2015 because the assessments needed to be scored automatically by computer within two 30-min sessions and to accommodate several dozen languages and cultures. Networking two to four students in teams via computer also was not possible in many countries and school systems. However, the use of agents instead of groups of humans raised questions about the ecological validity of the PISA assessment. There were no opportunities for the lengthy conversation threads to handle different types of negotiation and team members building on each other's ideas. The personality and emotions of the team members were out of scope. Team members are known to experience a wide array of emotions in CPS, such as confusion, frustration, or even anger, when a team member disagrees with the perspectives of others, when a person's favorite solution is dropped, or when there is conflict.

The alternative to the PISA assessment with agents is to analyze the naturalistic conversations and interactions among human team members. In the past, this has been extremely time-consuming and expensive, as we know from the barriers in scaling up ATC21S. It requires researchers to collect data from teams, record the process of their interactions, sometimes transcribe audio or visual communication, segment the interactions into units, annotate each observational unit into one or more theoretical categories, discover or confirm expected patterns of interaction, perform statistical analyses, systematically assess the generality of the findings, develop assessment measures that can be standardized, and so on. Researchers have conducted projects that implement these steps systematically, but these projects involve small or modest sample sizes and a narrow landscape of tasks, contexts, and populations. Because of these methodological and logistical challenges, we believe that technology is needed to step in and fill this gap. There are promising opportunities to incorporate recent advances in digital technologies to automate the detection of CPS processes and the assessment of various CPS competencies. If this is successful, researchers can collect and analyze extremely large data sets across a wide landscape of tasks, contexts, and populations. That would provide a notable advance in discriminating theories, testing hypotheses, and guiding in the development of an educational curriculum for CPS education and training. We hope that psychological scientists will be essential players in the use and development of these technologies.

Role of Technology in CPS Assessment

The world has changed with advances in artificial intelligence (Elliot, 2017), computational linguistics (Jurafsky & Martin, 2008), semantic representations from large corpora (Landauer, McNamara, Dennis, & Kintsch, 2007), speech-to-text transcription, educational data mining (R. S. J. D. Baker & Yacef, 2009), conversational agents (W. J. Johnson & Lester, 2016), and other digital technologies that are distinctively relevant to CPS assessment. A review of these developments is beyond the scope of this article. Instead, we summarize specific efforts that have distinctive relevance to modeling and analyzing the CPS processes and assessments. Our objective with this section is to illustrate how psychological scientists can partner with experts in these technologies to advance CPS research and applications.

Use of computer agents in CPS assessment

We have discussed how the PISA assessment adopted conversational agents. A single human test taker interacted with one to three agents throughout the assessment. This orchestrated interaction was needed to cover the 12 cells in Figure 2 and provide a valid CPS assessment depicted in Figure 1. A central advantage of assessments with computer agents is the degree of control over the conversation. The discourse contributions of the two agents (a1, a2) and any associated digital media (m) can be coordinated so that each [a1, a2, m] sequential display is functionally a single assessment unit to which the human responds through language, action, or silence in a particular human turn. Thus, there is an orchestrated transition network that alternates between assessment units and human turns—essentially a dialogue. The test takers' CPS proficiency scores are computed from the correctness of the responses to these assessment units. This is different from a collaboration in which many people can speak in an unconstrained manner, often simultaneously (H. H. Clark, 1996).

With the use of agents, there can be some complexity with optional conversation threads in the discourse, but not too much complexity. In conditional branching, the computer's generation of the assessment unit U_{n+1} at turn $n + 1$ is contingent on the state of the human turn HT_n at turn n. The degree of conditional branching is limited to a small number of states associated with each HT_n in the PISA assessment; specifically, there were two to five options at each turn (i.e., either chat options or alternative actions to be performed in the activity window). Consequently, the conditional branching was not complex. In addition, the turn taking frequently converged at points of assessment rather than diverging in many directions. Only one score was associated with each assessment unit, and each unit was aligned with only one of the 12 cells in the CPS assessment matrix in Figure 2.

The design of the PISA assessment was compatible with the normal psychometric modelling in the world of assessment, in which multiple-choice tests are ubiquitous. PISA had fixed sequences of assessment units (U_1, U_2, . . . U_m) that occurred at specific points as the problem was solved and the responses of the human were automatically recorded (as clicks on action options or chat options). The conversations were designed so that the conversations would naturally close shortly after the human responded to an assessment unit and the subsequent assessment unit was launched (e.g., "Thanks for your input, let's go on"). Assessment scores were collected for each test taker for the assessment units that collectively covered each of the 12 cells in the CPS assessment matrix of Figure 2 and contributed to overall CPS proficiency measures in Figure 1.

The conversational branching, albeit limited, is a promising capability but does not go the full distance in accommodating some of the interactivity of authentic CPS among humans. Negotiation is an excellent example of its limitation. It may take many cycles of back and forth between two parties in a negotiation, whereas there was only a one-step response in the PISA assessment (i.e., a single response regarding whether the test taker agreed with an agent's proposal). It is important to acknowledge that negotiation is not always a competitive exchange between two parties but is ubiquitous in many levels of language and communication. A common example relevant to CPS among disparate entities is the negotiation of common ground among team participants on the meaning or referent of a word (H. H. Clark, 1996). It can be challenging to successfully negotiate a common ground. The challenge is obvious when a stranger asks for directions to a location in a city. The stranger asks "Where is City Hall?" followed by a multiturn conversation on where City Hall is, including various landmarks, until the stranger hopefully pursues the correct path (but often fails!). Grounding of meanings and referents is very difficult to accomplish with computer agents, so multiturn grounding was not attempted in PISA.

The central assumption is that the agents are designed in such a way as to elicit the relevant behaviors from test takers in sufficiently realistic situations. For example, cell C2 in Figure 2 describes the relevant skill and behavior of "enacting plans." Once this has been identified as a relevant skill within the assessment framework, the agents' behavior can be orchestrated in a way that endangers the group goal, and whether the test taker adaptively responds to make progress on the group goal can be observed. The assessment determines whether students do or do not exhibit the relevant response and, per statistical inference, whether they will show it in similar situations in real-world situations. This example shows the importance of a theoretical underpinning of the assessment and of a straightforward conceptualization of CPS assessment tasks.

Quite clearly, the agents cannot perfectly emulate all psychological mechanisms of humans. Unlike humans, agents do not have a sophisticated theory of other people's minds and do not respond to the emotions of conversation participants and to subtle social pragmatics. That being said, researchers have been developing some agents that do detect and respond to human emotions in intelligent tutoring systems (D'Mello & Graesser, 2012; Graesser, 2016; W. J. Johnson & Lester, 2016), and advances in artificial intelligence are moving forward in simulating more sophisticated agents that will have a major impact on the workforce and communities (Elliot, 2017). There is no simple answer to the question of whether agents do or do not capture the psychological mechanisms of human communication. In some ways they do, and in some ways they do not. As the agent technologies mature, answers to these questions will change.

Evidence-centered design in tracking of actions and conversation in assessment

Evidence-centered design (ECD; Mislevy, Steinberg, & Almond, 2003) provides guidance for assessment design, deployment, and data analysis. Its conceptual-assessment framework provides a blueprint for the design structures of an assessment. The conceptual-assessment framework is broken up into a number of components, called models, that correspond to *what* is being measured (student model), *where* it is being measured (task model), and *how* it is being measured (evidence model). The *student model* defines the variables corresponding to the knowledge, skills, and abilities being measured. The *task model* describes task and environment features (sometimes called affordances) needed to obtain evidence to support claims about student model variables. The *evidence model* bridges the student model and task model through two components: evidence rules and a statistical model. The evidence rules identify the salient features of what learners say and do and provide rules for evaluating those work products. The statistical model specifies how the evidence collected about learners will be used to make claims about what learners know or can do. These three models correspond to the primary components of an assessment argument.

CPS is a complex construct, so it is often assessed in digital environments that capture (in a log file) all of the actions and discourse in the environment during individuals' performances. The complexity of the construct and the environments used to assess it present challenges for building the student, evidence, and task models of ECD. Specifically, there are challenges with respect to conceptualizing what skills make up CPS and making sense of the large streams of log data. D. Kerr, Andrews, and Mislevy (2016) have developed an ECD-based framework that provides technological support for instantiating the student, evidence, and task models when seeking to assess proficiency in complex skills derived from high-granularity log data. The *in-task assessment framework* (I-TAF) reconceptualizes what it means to define the construct of interest, extract evidence from learner performances, and link the evidence to the construct of interest when the evidence is high-granularity log data of individuals' actions and discourse rather than standard item response data. I-TAF further outlines a process for extracting evidence of the ontology concepts from low-level log data and connecting behaviors and discourse from a task environment to components of the ontology.

A concrete example should clarify the application of ECD. I-TAF has been applied in a project designed to assess engineering and electronics students' CPS skills using log data from a simulation-based task on electronics concepts. The simulation-based task, called the three-resistor activity (Fig. 4), targets the relationship between resistance, voltage, and current as represented by Ohm's law. The task is a simulation of a series circuit. Students work in a team of three, each on a separate computer; however, each computer runs a fully functional simulation of only a portion of the electronics circuit. Task levels involve reaching a given goal voltage across each resistor in the circuit across three computers. Because the resistors on each student's circuit board are connected in series, the actions that a student takes on his or her board affects the state of the other students' boards. Thus, the students must work together and coordinate their actions to reach the goal voltages for each circuit board. To complete the task levels, students can measure the voltage, view and change their resistances, perform calculations on an in-task calculator, use a zoom feature to view the state of teammates' boards, communicate with teammates through a chat window, and submit their answer choices. Each of these actions and communications are time stamped and logged to a database. These behaviors, at the level of utterances and mouse clicks, were used as evidence to make inferences about students' CPS skills.

The three-resistor activity is an open environment: There are few constraints on what students can do, and there are open conversations among students. These features add a layer of complexity for assessment. I-TAF is suitable for addressing this complexity by better

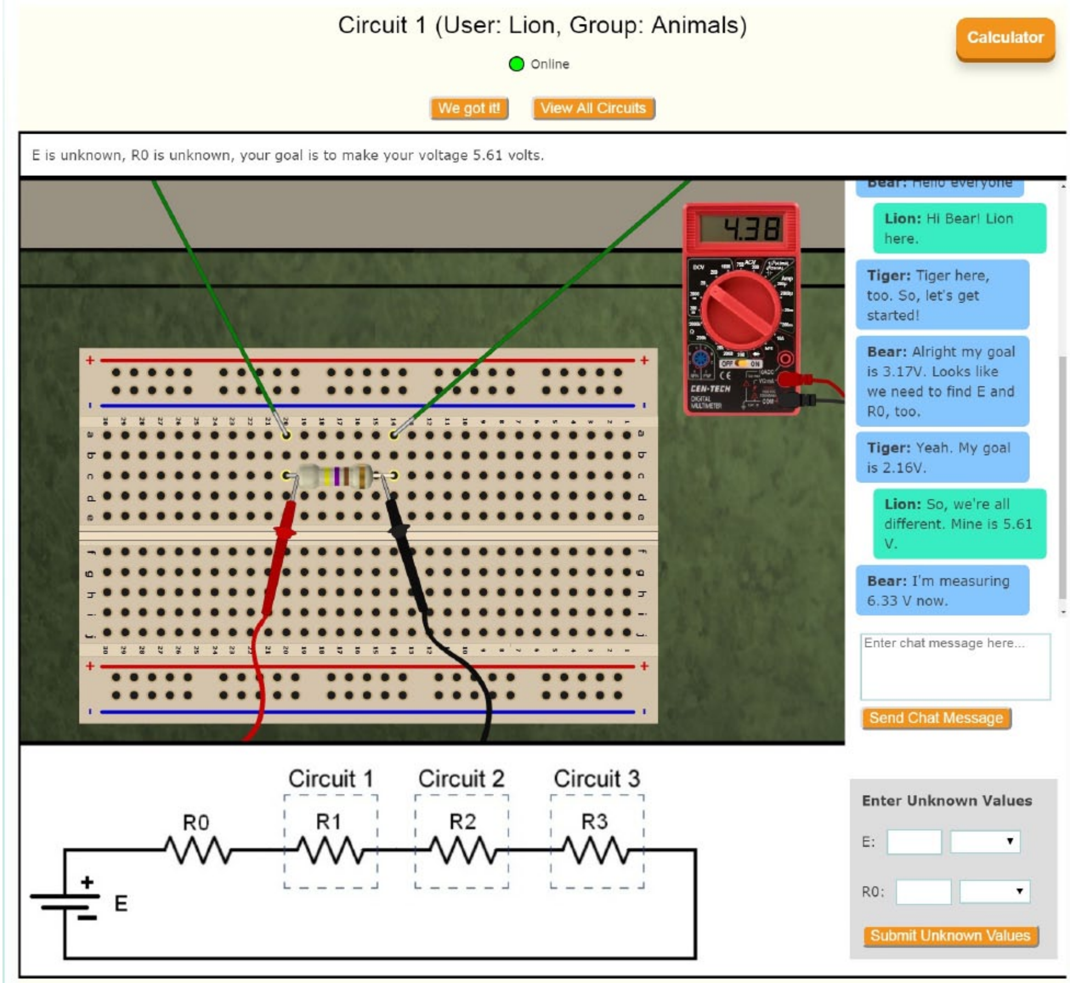

Fig. 4. Screenshot of the three-resistor activity. The main window of the screen includes a simulation of an electrical board with a resistor and digital multimeter for taking measurements. The bottom portion of the screen includes a schematic of a series circuit and an input box to submit answer choices. The right side of the screen has a text chat box to facilitate communication among team members. Image used with permission from the Concord Consortium (https://concord.org).

defining the construct space at multiple levels of granularity and providing principles for connecting behavioral evidence to high-granularity constructs. The I-TAF team created a CPS ontology that laid out concepts or skills associated with CPS and their relationships. Determining the components of the top layer of the ontology was based on an extensive literature review of existing CPS frameworks and literature from related areas, such as linguistics, individual problem solving, organizational psychology, and computer-supported collaborative learning (Grand, Braun, Kuljanin, Kozlowski, & Chao, 2016; Hesse et al., 2015; Liu et al., 2015; Meier, Spada, & Rummel, 2007; Morgan et al., 1993; OECD, 2017a; O'Neil et al., 2003; Zhuang, MacCann, Wang, Liu, & Roberts, 2008). This layer of the ontology provides a generalizable construct definition of CPS that can be used across other tasks and content domains.

The project involved a digital environment that captured data at the level of mouse clicks. Therefore, it was important to conceptualize the construct at multiple layers of granularity to induce patterns from the low-level actions and discourse in the log files to high-level proficiencies associated with CPS. The ontology includes (a) additional layers to operationalize the construct in terms of concepts, behaviors, or strategies that can be used to demonstrate mastery of concepts in the ontology and (b) specificity with respect to how the behaviors would look in the log files in the environment being used for assessment. A fragment of the CPS ontology can be found in Figure 5. The upper portion of the ontology in Figure 5 specifies that social skills and cognitive skills are part of CPS. Below that are the various layers and components that are organized hierarchically. In each branch of the ontology, further layers

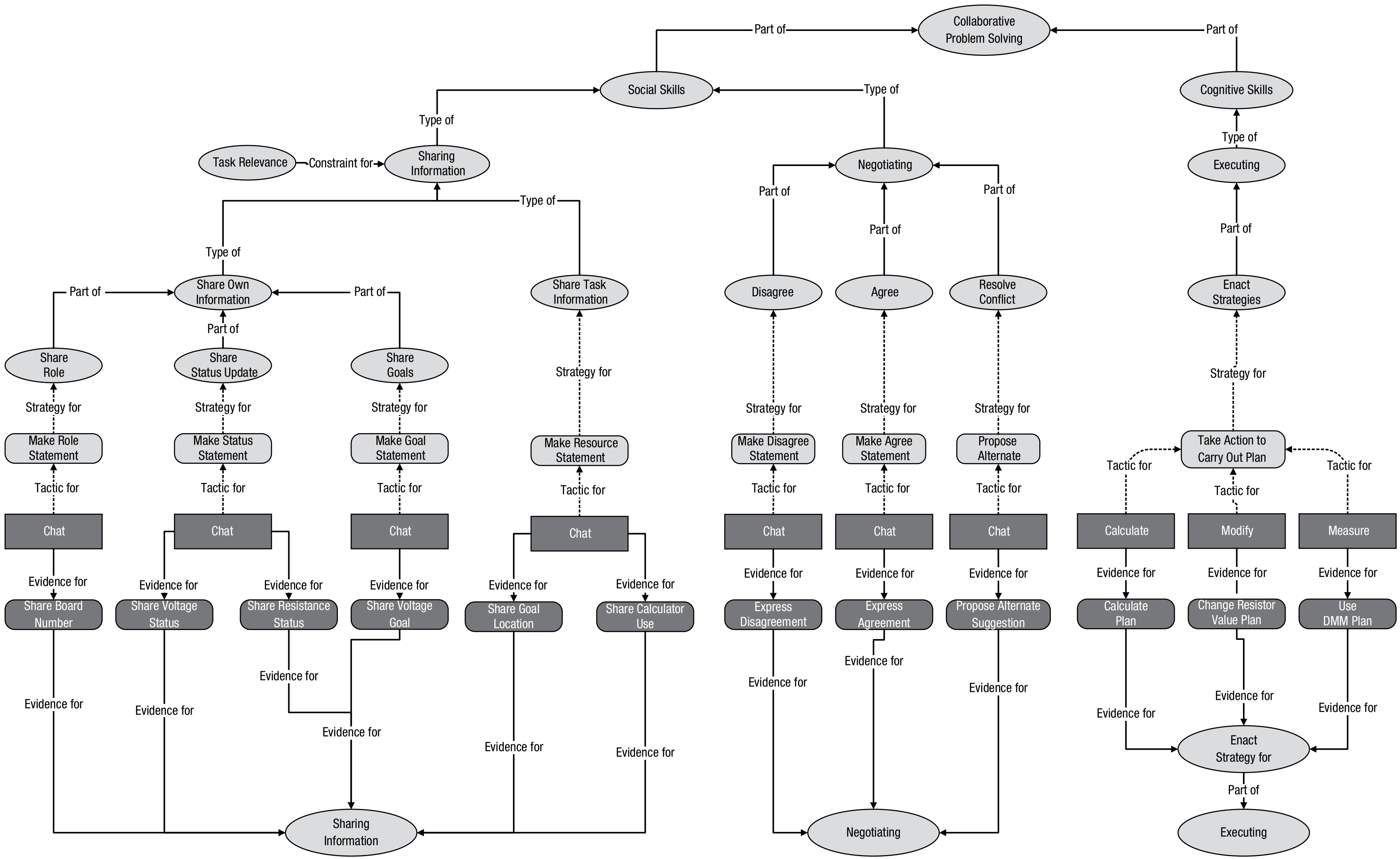

Fig. 5. Fragment of collaborative problem solving (CPS) ontology for the three-resistor activity. The upper portion of the ontology specifies that social skills and cognitive skills are part of CPS. Below that, *sharing information* is a type of social skill and *task relevance* is a constraint for *sharing information*, whereas *share own information* and *share task information* are types of *sharing information*. Further, *share role, share status update*, and *share goals* are part of *share own information*. Moving to the right in the ontology, *negotiating* is a type of social skill, whereas *disagree, agree*, and *resolve conflict* are parts of *negotiating*. In the last branch, the ontology denotes that *executing* is a type of cognitive skill whereas *enact strategies* is a part of *executing*. In each branch of the ontology, further layers denote what strategies can be used to provide evidence of the higher-level nodes, what affordances are available in the task for students to carry out a given strategy, and what inferences can be drawn about students if the behaviors are displayed. Reprinted from Andrews-Todd and Forsyth (2018) with permission from Elsevier.

denote what strategies can be used to provide evidence of the higher-level nodes, what affordances are available in the task for students to carry out a given strategy, and what inferences can be drawn about students if the behaviors are displayed. For example, for the executing branch of the ontology, a strategy for enact strategies is *take action to carry out the plan*. As another example, there are three affordances available for *take action to carry out the plan*, and they include *performing calculations, modifying the resistor*, and *taking measurements*. When these behaviors are used in a specific way, they can provide evidence for inferences associated other nodes in the ontology (*calculate for plan, change resistor value for plan, use digital multimeter for plan*), each of which can ultimately provide evidence for students' ability in executing.

The ontology provides explicit information about the evidence that needs to be identified in the data in order to make inferences about students. The evidence can be identified in the data using human annotation and machine-learning approaches (R. S. J. D. Baker & Yacef, 2009). Regarding human annotation, the ontology is used to develop rubrics for raters to qualitatively code log files for the presence of ontological categories. This method of developing rubrics for qualitative coding has been successful because interrater reliability between raters' coding actions and discourse of study participants has been high (κ = .84). Further work has explored computational methods for aggregating information from the coded data to develop profiles that categorize types of collaborative problem solvers for reporting about CPS competency. One method was an exploratory cluster analysis of the frequencies of CPS skills displayed on an individual level. Specifically, clustering with Ward's method identified four unique profiles of collaborative problem solvers: chatty doers, social loafers, active collaborators, and group organizers, all based on social and cognitive theory that aligned with the CPS behaviors in each profile. The chatty doers displayed high levels of content-irrelevant social communication and high levels of executing actions in the service of solving the problem. The social loafers were characterized by low levels of CPS skills in general, whereas active collaborators were characterized by high levels of all CPS skills except content-irrelevant social communication. Group organizers were characterized by CPS skills associated with establishing and maintaining organization for the problem and the group (Andrews-Todd, Forsyth, Steinberg, & Rupp, 2018).

A second method was to group students on the basis of aggregate frequencies of skills corresponding to the higher level social and cognitive dimensions of CPS. The frequencies for each dimension (i.e., social skills vs. cognitive skills) were ranked and split into two roughly equal groups of high and low displays for each dimension. Calculating these independently for each dimension led to the creation of four groups, corresponding to a 2 × 2 social-cognitive dimension matrix (i.e., a group with high cognitive skills and high social skills, a group with a high cognitive skills and low social skills, a group with a low cognitive skills and high social skills, and a group with low cognitive skills and low social skills; Andrews-Todd & Forsyth, in press).

Work to validate all of the profiles is ongoing, but preliminary evidence has shown statistically significant relationships between CPS skill profiles and performance in expected directions. For example, students in the group with low cognitive skills and low social skills have demonstrated poorer performance in the three-resistor activity than students in any of the other groups. Furthermore, for the clustering approach, the active collaborators showed the highest levels of performance, whereas the social loafers showed the lowest levels of performance. The performance of chatty doers and group organizers fell in between these groups (Andrews-Todd & Forsyth, in press; Andrews-Todd et al., 2018).

Although conceptually complex, this example provides an important illustration of how interdisciplinary research can help to develop such approaches that are automatically applied to large samples of teams. Psychological scientists are needed to make sure these judgments are made with reliability, validity, and theoretical integrity. Regarding machine learning, computational algorithms are formulated to discover sequences of actions that predict the ontological categories as well as successes on CPS at various levels of granularity. Once again, psychological scientists are needed in such collaborations to interpret the psychological plausibility of the outputs from machine learning. This is a critical nexus in the interdisciplinary collaboration between psychological scientists and those in the arena of machine learning.

The advantage of this ECD approach is that it provides a detailed assessment of the actions and conversations of the team members in a particular CPS scenario at a deep level of granularity and in ways that are theory-motivated. This approach provides the depth of interaction that was lacking in the highly constrained, theory-heavy PISA assessment. Once the ECD model is created, it can be launched with participants in an unlimited number of teams.

There are challenges in achieving these goals, however. It takes substantial expertise, effort, and sample sizes to produce reliable and valid model parameters. Actions need to be categorized by the designer, and the natural-language contributions need to be automated and compared with expected answers. The good news is that there has been substantial progress in automated natural-language text extraction in recent years (Rus, Lintean, Banjade, Niraula, & Stefanescu, 2013), so these efforts are within the horizon. Another limitation of this approach is the lack of data available to help determine whether the ECD model of one CPS scenario can transfer to another scenario. Nevertheless, this is the model to pursue for those who want theory and depth of analysis.

Measuring open-ended communication in collaboration

Collaboration requires communication to establish common ground, enact plans, monitor progress, and reflect on the state of collaboration. Communication among team members is often through spoken or written language, but it also may be implicit through observations of the actions taken by one or more team members. Although the successful solution of a CPS task may be the final output from the team, the communication stream generated during the problem-solving process is arguably the richest source of information about the team's knowledge, skills, and abilities that are applied during the task. Thus, collaboration can be characterized by the behaviors generated during the process rather than just the product (Lai, DiCerbo, & Foltz, 2017).

Communication streams generated during CPS tasks provide information about the structure of the social network of the collaborators in addition to the content and quality of information flowing through that network (Shaffer, 2017). This information reflects team-member roles, their connectedness, and how each individual is performing his or her tasks. The content and manner of the communication provides information about the team's cognitive and emotional states, knowledge, errors, information sharing, coordination, leadership, stress, workload, intent, and situational status. From an ECD perspective (Mislevy et al., 2003), the communication behaviors generated during a group problem-solving task provide the evidence of the CPS skills. Thus, communication serves as both the primary mechanism for allowing collaboration to occur and one of the key windows into monitoring and assessing CPS skills.

Studies of communication in collaborative situations have measured various aspects of communication to characterize how the behaviors correspond to collaborative skills. For example, research has shown that communication behaviors that correspond to better CPS performance include (a) inquiring about others' goals and interests and soliciting input from everyone (Stevens & Campion, 1994), (b) providing planning and coordinating statements (Ellis, Bell, Ployhart, Hollenbeck, & Ilgen, 2005), (c) showing openness in arguing a particular position or modifying a position to recognize other teammates' arguments (Chen, Donahue, & Klimoski, 2004; U. Fischer, McDonnell, & Orasanu, 2007), (d) acknowledging teammates (Achille, Schulze, & Schmidt-Nielsen, 1995), (e) asking for or giving help (Baghaei, Mitrovic, & Irwin, 2007), and (f) engaging in small talk (Stevens & Campion, 1994).

Although the evidence that CPS skills can be measured through communication is strong, the analysis of the behaviors in CPS tasks is complex. These analyses require observing or recording communication streams, coding communication behaviors, and rating the performance of individuals as well as the team. This can be quite time consuming and resource intensive, limiting instructors' ability to easily monitor and provide feedback to teams. In contrast, techniques have been developed to assess performance by automatically analyzing language of teams in collaborative situations. The conversations have been analyzed by a variety of automated text-analysis tools, which is beyond the scope of this article (Dowell, Graesser, & Cai, 2016; Dowell et al., 2018; Foltz & Martin, 2008; Liu et al., 2015; Mu, Stegmann, Mayfield, Rosé, & Fischer, 2012; Shaffer, 2017; Tausczik & Pennebaker, 2013).

One challenge for assessing collaboration has been interpreting the content of what team members communicate. One notable technique for assessing content has been latent semantic analysis (LSA; Landauer, Foltz, & Laham, 1998; Landauer et al., 2007), which has been frequently used to analyze and interpret what is being said during communication. The technique analyzes meaning rather than merely the words explicitly used, so it can account for variability in how different people express similar events or situations in collaborative situations. The approach has been applied to analyze communication across a range of complex collaborative tasks and has been shown to be able to classify interaction types (Gorman, Martin, Dunbar, Stevens, & Galloway, 2013), predict overall scores of individuals and teams (Martin & Foltz, 2004), and alert instructors when students are drifting from effective collaboration patterns (Foltz & Martin, 2008). Dowell et al. (2018) successfully used LSA to analyze teammates on the basis of their interaction profiles, which had six measures derived from the flow of conversation in natural language: participation, social impact on others, responsivity to others, internal cohesion within a speaker, the newness of information to the conversation, and the density of content. These indices of semantic analysis with LSA provided an automated foundation for classifying team members into different roles (e.g., task leader, social driver, follower, lurker, overrider, and socially detached).

Automated analyses of communication provide a means to monitor the collaborative process and assess aspects of performance. Although automated approaches currently are not as accurate as having humans interpret the discourse contributions of a team, they significantly reduce the workload of humans and have promise for handling big data. However, psychological scientists need to be directly in the loop to assess the reliability, validity, and integrity of these automated analyses. This is where psychological scientists need to partner with computational linguists. These automated analyses of communication can also be integrated into digital learning and work environments that allow teams to receive faster feedback and training based on the quality of their performance. There can be a snapshot of the performance profile of team members and the extent to which team members interact with other particular team members.

Situational-judgment tests

Situational-judgment tests (SJTs) have been extensively studied in the organizational sciences during a long history of job-related testing. Self-report observations can be collected electronically and serve as a fingerprint of the team members' traits and experiences during collaboration. At the more general level, SJTs provide a vignette describing a situation that requires some behavior or course of action for successful resolution. In the organizational sciences, these vignettes are based on critical incidents necessary for job performance, such as confrontations with an upset customer in a call-center scenario. SJTs are argued to have more face validity than standard testing (Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993).

One element of SJTs that makes them ideal for testing is the development and use of stems with multiple options that are selected by test takers. The options are often designed to indicate incorrect, partially correct, and best answers. In this way, one is better able to determine the level of knowledge acquired by a learner. SJTs can be created for paper-and-pencil forms of assessment as well as more complex video-based assessment, and they allow for multiple forms of constructs to be assessed (McDaniel & Nguyen, 2001). Meta-analytic support has been shown for SJTs in criterion-related validity, such as criteria used to evaluate job performance. Given that SJTs are fairly robust predictors of job performance, researchers have occasionally adopted this approach in educational high-stakes testing. In a longitudinal study of medical-school admissions that assessed more than 5,000 applications, SJTs were shown to add value to traditional cognitive assessment when examining interpersonal skills (Lievens, 2013).

These advances in technology show promise in analyzing large samples of CPS interactions so that systematic comparisons can be made in different populations and contexts. The validity of the claims obviously depends on the accuracy of the automated analyses, but the fidelity of the data can be verified by annotation by trained experts, including psychological scientists, on randomly sampled excerpts. This will provide a statistically reliable empirical foundation for the theoretical claims and assumptions.

These advances in technology for automated CPS assessment are vulnerable to a large array of criticisms, however. There are a number of important concerns aside from the issue with computer agents that we addressed in detail in the last section. There are liabilities when theory-driven components of the assessments bias the lens of observation and miss important phenomena relevant to CPS processes. The ontologies of the ECD and the situations generated in the SJTs enjoy both the advantages and liabilities of theoretical perspectives. However, there are also liabilities to data-mining and discovery-oriented approaches to exploring CPS processes. The results may be difficult to interpret, limited in generalizability to other contexts, and difficult to explain to critics who ask questions about validity, reliability, and replication. Criticisms will always be raised that the automated technologies miss a large number of important features and dimensions of context—some being too simple to explain and others being too complex.

Education and Training in CPS

In this section, we turn to the prospects of educating students and the public on CPS competencies. As discussed at the beginning of this article, the level of CPS proficiency is low around the globe (OECD, 2017b) because CPS curricula are conspicuously rare, students receive indirect training on CPS skills, and the quality of the training is unimpressive. In essence, we are nearly at ground zero in terms of identifying pedagogical approaches to improving CPS skills. Moreover, empirical research on CPS processes and products in educational settings is sparse compared with that for collaborative learning, decision making, and other types of group activities. To help address this gap, we provide representative examples of the kinds of training that can be used for developing varied forms of CPS competencies. Although these may not have been designed specifically for CPS, their approach aligns well with the core features of associated competencies. Furthermore, even though not all the approaches have been tested in the context of CPS, they are generalizable to numerous contexts. We therefore offer these as promising interventions that can be introduced for educational or professional development in support of training a workforce to be proficient in CPS. Therefore, suggestions in this section are speculative, with an invitation to psychological scientists to conduct relevant research in the future.

Research on groups in educational settings

There is a rich literature on group learning in educational settings (Slavin, 2017), but unfortunately very little of this research has focused on CPS per se. Nevertheless, it is important to identify ways that group learning is likely to be relevant to education and training in CPS. This subsection describes a set of methods that have been empirically examined in educational settings. All of these approaches attempt to facilitate *active learning* because they require groups of students to engage in and take control of the learning process. These approaches build on *constructivist* approaches to cognition that underscore the importance of individuals constructing their mental, physical, and social worlds rather than passively observing the worlds around them (National Research Council, 2000). Moreover, students are not likely to learn collaboration skills just because they are assigned more group work (Lai et al., 2017). Students must practice their collaboration skills with the mind-set of noticing what they are doing wrong and formulating strategies to do better (Rotherham & Willingham, 2010). In absence of this mind-set, they need training and feedback on the CPS process.

Activities involving collaborative learning require groups of students to work toward shared goals. In some but not most cases, the target of the learning directly addresses problem solving and critical thinking. Collaborative learning shows advantages over lecture-based methods (Slavin, 2017). However, the variety of pedagogical methods and the diversity of instructional strategies applied make it difficult to discern where problem solving fits into the mix. In a recent review of active learning approaches involving collaboration, Gabelica and Fiore (2013) examined four methods that have been implemented in the classroom and have vestiges of problem solving: problem-based learning, case-based learning, team-based learning, and studio-based learning.

Problem-based learning. In this method, facilitators or tutors are used to guide small-group learning. Problems are selected or derived from actual real-world problems, called *authentic problems*. Students are first encouraged to produce their naive understanding of the problem, identify similarities across the group, and generate potential hypotheses and solutions (Gijselaers, 1996). A key part of this process is that students discuss any lack of understanding they have and what knowledge needs to be acquired to solve the problem. From this, learning goals are identified, and students work in class and outside of class to gather and integrate the knowledge necessary to produce a solution. Finally, a reflective component is built into the process during which students are debriefed on what they have learned.

Case-based learning. This is very similar to problem-based learning. The primary difference is the degree of intervention provided by a facilitator or teacher. Case-based learning is popular in medical education, because the teacher needs to take a more active role. For example, during impasses in problem solving in case-based learning, more directive guidance is provided, such as probes meant to align student thinking with an appropriate path to consider. Although some debate exists as to the specifics of the strategies used within problem-based learning (Hmelo-Silver, 2004), meta-analyses suggest that small-group learning was related to academic achievement on some measures of learning compared with more conventional learning environments (Norman & Schmidt, 2000) and that group debate was a feature that improved the development of shared knowledge and problem solving (Hmelo-Silver, 2004). These reviews also suggest that there should be a flexible amount of self-direction relative to the student's place along the learning trajectory (Hmelo-Silver, 2004; Vermunt & Verloop, 1999). Perhaps most important were the outcomes that contrasted knowledge acquisition and knowledge application. Compared with problem-based learning, traditional classroom-based instruction showed some benefits on factual knowledge and standardized tests (Vernon & Blake, 1993). However, problem-based learning showed some benefit for knowledge application, retention, and better study habits (Dochy, Segers, Van den Bossche, & Gijbels, 2003).

Team-based learning. This method incorporates a research component as team members work on projects inside and outside the classroom. Team-based learning is more structured in its application because there is a specific sequence of activities inside and outside of class. The use of complex tasks produces a great deal of discussion within the group along with constructive conflict (D. W. Johnson, Johnson, & Smith, 2000). This requires formal individual pre-class preparation, followed by small-group problem solving in which concepts are applied, and then by some in-class reflection involving all the students (Michaelsen, Parmelee, McMahon, & Levine, 2008). Team-based learning is challenging to implement, and there is ongoing debate about its core features and impact on performance.

Studio-based learning. This is the most recent addition to investigations of collaborative-learning approaches that embrace some semblance of problem solving, even though it is the one with the earliest origins. Gabelica and Fiore (2013) traced it back to Germany in the early 1900s and thence to its use in design-related curricula (Bayer, 1975), such as architecture, interior design, and industrial design. It also emphasizes application of knowledge through practical experience and learning that is situated in a relevant context. However, it is different in that the problems addressed are much more ill-structured, and many problem elements are unknown or poorly defined. There is an iterative process of inquiry and critique within an interactive studio space, where students work with experts who use prompts and discussion to guide the process. This approach has been adopted in settings as varied as mathematics, chemistry, engineering, and computer programming (Hundhausen, Agrawal, Fairbrother, & Trevisan, 2010). However, there is much less research on the effectiveness of studio-based learning; most studies have reviewed the methods only qualitatively.

These different forms of collaborative learning no doubt have some relevance to training in CPS. The most ambitious attempts to directly train students in CPS skills recently took place in Australia (Scoular & Care, 2018). The training had alignments with the ATC21 framework (Hesse et al., 2015). The empirical studies were modest in scale but represented different approaches to training the skills. The first approach was experimental in a two-condition study. Approximately half of 44 students received explicit training in various CPS skills, whereas the other half did not. The second empirical study embedded CPS skills in courses across different subject matters in the curriculum (science, English, social studies) in two schools (N = 262 students). The third embedded CPS in interdisciplinary training. There were no quantitative analyses of the results with respect to problem-solving performance and other outcome measures associated with CPS. However, there were some qualitative observations that merit some attention. Perhaps the most obvious conclusion from this study is that there is very little empirical research on training and testing CPS skills in systematic ways across large samples. That opens the door for a new generation of psychological scientists to explore methods of improving CPS training so that a broad set of cognitive and noncognitive skills can be mastered.

In sum, research must explore the adaptation of empirical findings for CPS to determine their relevance to the acquisition of competence in CPS. The goals of collaborative learning are somewhat different from those of CPS, so the similarities and differences need to be identified. The collaborative-learning approaches sometimes improve and sometimes reduce learning of individuals compared with traditional classroom environments.

Research on team training

Researchers in team science have not yet identified the specific pedagogical methods that have promise in facilitating CPS competencies. Yet it is reasonable to draw from research on team training to illustrate that a strong foundation of evidence exists on which to make recommendations for research on learning competencies associated with CPS. In particular, meta-analyses on studies of training research support the claim that interventions can improve the knowledge, skills, and attitudes of teams (Delise, Gorman, Brooks, Rensch, & Steele-Johnson, 2010; Klein et al., 2009; Salas et al., 2008). Salas et al. (2008) examined the impact of specific training contents on various outcome measures (e.g., affective, cognitive, process, and performance) and found that team training has generally had a moderate, positive impact on team functioning and depended on the type of training administered and types of outcomes examined. The specific team-training strategies that can be related to CPS (including team knowledge training, critical thinking, and coordination training) showed some of the largest effect sizes.

Another meta-analysis examined how team-building interventions influenced process and affective outcomes (Klein et al., 2009). Relevant to competencies fitting with CPS, goal-setting and role clarification had the most effect on overall team performance; interpersonal relations and problem-solving skills demonstrated moderate effects on performance. Given the importance of improvement of problem-solving competencies in teams, and the fact that affective outcomes (e.g., trust and team potency) and process outcomes (e.g., coordination and cooperation) are helpful to team performance, the coupling of team building with CPS is worthwhile to explore in the future research.

Another team-training meta-analysis more directly related to CPS found that, in general, team training had positive effects on performance outcomes (Delise et al., 2010). Team training had the greatest positive impact on team-cognition outcomes and persisted on both training and transfer environments. It is noteworthy that the effect of training on cognition was larger in the transfer environments than in the training environments. This shows that training may be more effective in changing cognition when individuals have the opportunity to use these skills in the transfer environment.

Use of computer agents in CPS training

In line with our discussion of agents for assessments, another promising area for research is the use of conversational agents in the training of individuals to improve their CPS proficiencies. In such interventions, these agents step in and offer recommendations under specific conditions. This vision is compatible with recent trends in the development of intelligent tutoring systems that are being applied to CPS training (Gilbert et al., 2018; Sottilare, Graesser, et al., 2018). The intelligent tutoring system would need to track the contributions of team members and the group as a whole automatically to provide timely feedback and recommendations to team members for improvement. An agent could step in as a tutor or mentor and express recommendations in natural language, following rules such as those articulated below (Graesser, Cai, Hu, et al., 2017; Graesser, Cai, Morgan, & Wang, 2017):

1. If the team is stuck and not producing contributions on the relevant topic, then the agent says "What's the goal here?" or "Let's get back on track."
2. If the team meanders from topic to topic without much coherence, then the agent says "I'm lost!" or "What are we doing now?"
3. If the team is saying pretty much the same thing over and over, then the agent says "So what's new?" or "Can we move on?"
4. If a particular team member (e.g., Harry) is loafing, the agent says "What do you think, Harry?"
5. If a particular team member is dominating the conversation excessively, the agent says "I wonder what other people think about this?"
6. If one or more team members use unprofessional language, the agent says "Let's get serious now. I don't have all day."

These production rules would be easy to implement if there was adequate formative assessment and interpretation of natural language, as discussed in the previous section. As automated assessments become more sophisticated, the production rules would be more nuanced and follow the theoretical CPS frameworks and learning principles of psychological science. However, simulated agents like these have not yet been developed and tested, so this is a vision for the future.

It should be noted that the use of computer-based assessments and agents for CPS is more than just a practical convenience. Collaboration is inherently complex; highly interactive environments involve multiple team members who display a wide range of cognitive and social processes. A large amount of data can be collected and mined in these efforts. Computer-based environments provide an infrastructure that can sufficiently constrain complexity for experimentation to test empirical hypotheses and evaluate theoretical constructs.

In light of the research covered in this section, CPS learning and training is destined to involve a mix of instructional components, which include either explicit instruction or implicitly learned behaviors within the context of collaborative situations. Technological scaffolds, whether through the inclusion of intelligent agents or simply prompts for particular forms of process (M. Baker & Lund, 1997), are available to use in any CPS research program. The evidence on record supports the claim that it is important to include practice in conjunction with feedback from instructors or peers in addition to didactic instruction on principles of effective CPS. Multiple interacting components contribute to the learning process, so research should examine the impact of particular scaffolding structures that provide support and guidance through the process. More specifically, some promising components of training would be (a) immediate and regular feedback along with formative assessment that allows for adjustment, (b) regular communication of ongoing work, (c) complex and real-world problems to solve, (d) prompting for reflection and meta-level discussions, and (e) modeling and scaffolding from an instructor (Fiore et al., 2018; Lai et al., 2017). However, because considerable uncertainty remains about the amount of transfer and the conditions under which CPS will improve, more systematic study of these interventions across contexts is warranted as well as studies with a substantive longitudinal component.

Skeptics may pose the argument that it is too early to develop a curriculum to train students in CPS because of the paucity of research on CPS compared with other types of collaboration. Perhaps it would be preferable to spend a decade or more conducting research on CPS mechanisms and training before spreading CPS into school curricula and assessment. There are two rebuttals to this argument. One is that numerous national and international reports have expressed an urgent need to improve CPS competencies in the workforce and the public at large. A second is that most knowledge and skills (e.g., math, reading, writing, science, individual problem solving) were introduced in school systems decades or centuries before there was a solid body of research that measured competencies on these knowledge and skills. Despite these arguments, scientific evidence should underlie practice when relevant.

Conclusions and Recommendations

The previous sections have covered findings from a broad and disparate literature on the CPS theory, assessment, and training. In this section, we succinctly identify a number of conclusions and associated calls for action of psychological scientists. CPS is receiving considerable attention both internationally and nationally, so it is important to identify how psychological scientists fit into the landscape and to point a path forward for research in this area.

We also believe that findings on CPS, and the researchers involved in studying CPS, can inform policy-level decision making in a number of arenas. First, in the context of education, the federal government often convenes committees composed of experts who help scope out a domain of research and application. Such committees help policymakers consider issues around assessment design based on theory and suitable measures and metrics. A recent report by the National Center for Education Statistics (NCES) was developed to help that organization consider the issues associated with the inclusion of CPS for the National Assessment of Educational Progress (Fiore et al., 2017). This was an important development because the implementation of such assessment can potentially influence policy at multiple levels of government, new standards for education, and the curriculum in schools.

Second, in consideration of workforce development, the National Academies of Sciences, Engineering, and Medicine similarly draw from experts in the research community. Because CPS is part of the 21st-century skills identified by the National Research Council, findings from research can potentially lead to recommendations for policy related to professional education and training. This is a separate path of leverage than researchers informing members of Congress, the White House, and other branches of the federal government in support of education. Policymakers are expected to understand the kinds of skills students must demonstrate, and those in the workforce rely on such findings for hiring and recruiting. Research in CPS can potentially affect a broad range of stakeholders, including policymakers, educators, and business leaders.

We now turn to enumerating the specific points about CPS to foster an explicit path forward for research on this important topic:

1. CPS has been identified as an important skill in the international community and workforce, but recent assessments have revealed that students and adults have low CPS proficiency. This calls for an analysis of CPS mechanisms, frequent problems, and methods of solving these problems. Psychological scientists could play a major role in this broad effort by partnering with stakeholders.
2. CPS is rarely trained in schools and the workforce, and the existing training is not informed by psychological science. This opens the door to the value of psychological scientists' being part of national and international efforts drawing from their expertise in science, learning and training.
3. Psychological scientists have developed a body of empirical research and theory of team science over the years, but much of this work has focused on group learning, work, memory, and decision making rather than CPS per se. We need to sort out how much of the existing research in team science applies to CPS. Psychological scientists are encouraged to direct their focus on CPS per se in the team science research landscape.
4. Intelligent digital technologies have the potential to automatically analyze large samples of group interactions at multiple levels of language, discourse, and interactivity. This is landmark progress because existing research on teams has had small samples and time-consuming annotation of the interactions. There is a need for psychological scientists to partner with the developers of these technologies to recommend psychological characteristics to track and to scrutinize the validity of automated measures.
5. A curriculum for training CPS competencies has not been developed and adequately tested. There is a need to develop a program of research on CPS curriculum design for both students and instructors. Psychological scientists are an important asset to generate potential curricula and to test their efficacies.

To conclude, we have discussed a set of factors associated with CPS that set the stage for contributions by psychological science. There are a number of open questions, debates, and challenges that psychological scientists can help answer, resolve, or solve. We have reviewed extant CPS frameworks and assessments that have been implemented around the world. These converge on a global deficit in CPS, something that motivates our call to action to address this pressing international need (Fiore et al., 2018). Toward this end, we discussed how technological advances can be used to foster improvements in assessment, learning, and training. We have identified a set of recommendations that point the way toward improving collaborative competencies. Simply put, cooperation is needed between a variety of stakeholders to study CPS at fundamental levels as well as to more closely examine the systematic implementation of instruction, practice, and feedback for CPS. We have drawn from the extant evidentiary base, but it is clear that there is much work to be done. When psychological scientists collaborate with educational researchers, computer scientists, psychometricians, and educational experts, we hope to move forward in addressing this global deficit in CPS.

Declaration of Conflicting Interests

A. C. Graesser. S. M. Fiore, S. Greiff, and P. W. Foltz served as paid expert group members or advisory board members of the Organisation for Economic Development's Programme for International Student Assessment (PISA) of collaborative problem solving (CPS). J. Andrews-Todd works with Educational Testing Service, which was the nonprofit organization that collected data and analyzed PISA data, but she was not involved with the collection or analysis of the PISA data. P. W. Foltz worked with Pearson, which was involved with the PISA CPS framework paper, but not the collection and analysis of the PISA data. F. W. Hesse served as an advisor on the Assessment and Teaching of 21st Century Skills assessment of CPS. None of the authors have received financial gain from any commercial application of collaborative problem solving.

References

Achille, L. B., Schulze, K. G., & Schmidt-Nielsen, A. (1995). An analysis of communication and the use of military terms in Navy team training. *Military Psychology*, *7*, 95–107.

American Management Association. (2012). *Executive summary: AMA critical skills survey. Workers need higher level skills to succeed in the 21st century*. Retrieved from http://www.amanet.org/uploaded/2012-Critical-Skills-Survey.pdf

Andersson, J., Hitch, G., & Meudell, P. (2006). Effects of the timing and identity of retrieval cues in individual recall: An attempt to mimic cross-cueing in collaborative recall. *Memory*, *14*, 94–103.

Andrews, J. J., & Rapp, D. N. (2014). Partner characteristics and social contagion: Does group composition matter? *Applied Cognitive Psychology*, *28*, 505–517.

Andrews, J. J., & Rapp, D. N. (2015). Benefits, costs, and challenges of collaboration for learning and memory. *Translational Issues in Psychological Science*, *1*, 182–191.

Andrews-Todd, J., & Forsyth, C. M. (2018). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in Human Behavior*. Advance online publication. doi:10.1016/j.chb.2018.10.025

Andrews-Todd, J., Forsyth, C. M., Steinberg, J., & Rupp, A. A. (2018). Identifying profiles of collaborative problem solvers in an online electronics environment. In K. E. Boyer & M. Yudelson (Eds.), *Proceedings of the 11th International Conference on Educational Data Mining* (pp. 239–245). Retrieved from http://educationaldatamining.org/files/conferences/EDM2018/EDM2018_Preface_TOC_Proceedings.pdf.

Aronson, E., & Patnoe, S. (1995). *The jigsaw classroom: Building cooperation in the classroom*. Reading, MA: Addison-Wesley.

Autor, D., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, *118*, 1279–1334.

Baghaei, N., Mitrovic, A., & Irwin, W. (2007). Supporting collaborative learning and problem-solving in a constraint-based CSCL environment for UML class diagrams. *International Journal of Computer-Supported Collaborative Learning*, *2*, 159–190.

Baker, M., & Lund, K. (1997). Promoting reflective interactions in a CSCL environment. *Journal of Computer Assisted Learning*, *13*, 175–193.

Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*, 3–16.

Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, *83*, 377–391.

Bayer, H. (1975). *Bauhaus 1919-1928*. New York, NY: Museum of Modern Art.

Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology*, *88*, 989–1004.

Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology*, *92*, 595–615.

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

Care, E., & Griffin, P. (2017). Assessment of collaborative problem-solving processes. In B. Csapó & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 227–243). doi:10.1787/9789264273955-en

Care, E., Griffin, P., & Wilson, M. (Eds.). (2018). *Assessment and teaching of the 21st century skills: Research and applications*. Cham, Switzerland: Springer.

Care, E., & Kim, H. (2018). Assessment of twenty-first century skills: The issue of authenticity. In E. Care, P. Griffin, & M. Wilson (Eds.), *Assessment and teaching of 21st century skills: Research and applications* (pp. 21–39). Cham, Switzerland: Springer.

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

Chen, G., Donahue, L. M., & Klimoski, R. J. (2004). Training undergraduates to work in organizational teams. *Academy of Management Learning and Education*, *3*, 27–40.

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

Clark, R. D., & Maass, A. (1988). The role of social categorization and perceived source credibility in minority influence. *European Journal of Social Psychology*, *18*, 381–394.

Costa, P. T., & MacCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources.

Crowston, K., Rubleske, J., & Howison, J. (2006). Coordination theory: A ten-year retrospective. In P. Zhang & D. Galletta (Eds.), *Human-computer interaction in management information systems* (pp. 120–138). Armonk, NY: M.E. Sharpe.

Cuc, A., Koppel, J., & Hirst, W. (2007). Silence is not golden: A case for socially shared retrieval-induced forgetting. *Psychological Science*, *18*, 727–733.

Davis, S. D., & Meade, M. L. (2013). Both young and older adults discount suggestions from older adults on a social memory test. *Psychonomic Bulletin & Review*, *20*, 760–765.

Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Guiding knowledge communication in CSCL via group knowledge awareness. *Computers in Human Behavior*, *27*, 1068–1078.

Delise, L. A., Gorman, A. C., Brooks, A. M., Rensch, J. R., & Steele-Johnson, D. (2010). The effects of team training on team outcomes: A meta-analysis. *Performance Improvement Quarterly*, *22*, 53–80.

Dennis, A. R., & Williams, M. L. (2005). A meta-analysis of group side effects in electronic brainstorming: More heads are better than one. *International Journal of E-Collaboration*, *1*, 24–42.

De Wit, F. R. C., & Greer, L. L. (2008). The black-box deciphered: A meta-analysis of team diversity, conflict, and team performance. *Academy of Management Proceedings*, *2008*. doi:10.5465/ambpp.2008.33716526

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

D'Mello, S. K., & Graesser, A. C. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems*, *2*, 231–238.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction*, *13*, 533–568.

Doise, W., & Mugny, G. (1984). *The social development of the intellect*. Oxford, England: Pergamon Press.

Doosje, B. E., Branscombe, N. R., Spears, R., & Manstead, A. S. (2006). Antecedents and consequences of group-based guilt: The effects of in-group identification. *Group Processes & Inter Group Relations*, *9*, 325–338.

Dowell, N. M., Graesser, A. C., & Cai, Z. (2016). Language and discourse analysis with Coh Metrix: Applications from educational material to learning environments at scale. *Journal of Learning Analytics*, *3*, 72–95.

Dowell, N. M., Nixon, T., & Graesser, A. C. (2018). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multi-party interactions. *Behavior Research Methods*. Advance online publication. doi:10.3758/s13428-018-1102-z

Elliot, S. (2017). *Computers and the future of skill demand*. Retrieved from the OECD Publishing. doi:10.1787/9789264284395-en

Ellis, A. P., Bell, B. S., Ployhart, R. E., Hollenbeck, J. R., & Ilgen, D. R. (2005). An evaluation of generic teamwork skills training with action teams: Effects on cognitive and skill-based outcomes. *Personnel Psychology*, *58*, 641–672.

Evans, C. R., & Dion, K. L. (1991). Group cohesion and performance: A meta-analysis. *Small Group Research*, *22*, 175–186.

Fiore, S. M. (2008). Interdisciplinarity as teamwork: How the science of teams can inform team science. *Small Group Research*, *39*, 251–277.

Fiore, S. M., Elias, J., Salas, E., Warner, N., & Letsky, M. (2010). From data, to information, to knowledge: Measuring knowledge building in the context of collaborative cognition. In E. Patterson & C. Miller (Eds.), *Macrocognition metrics and scenarios: Design and evaluation for real-world teams* (pp. 179–200). London, England: Ashgate.

Fiore, S. M., Graesser, A., Greiff, S., Griffin, P., Gong, B., Kyllonen, P., . . . von Davier, A. A. (2017). *Collaborative problem solving: Considerations for the National Assessment of Educational Progress*. Alexandria, VA: National Center for Education Statistics.

Fiore, S. M., Graesser, A. C., & Greiff, S. (2018). Collaborative problem solving education for the 21st century workforce. *Nature Human Behavior*, *2*, 367–369.

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, England: Ashgate.

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., Smith-Jentsch, K. A., Salas, E., Warner, N., & Letsky, M. (2010). Toward an understanding of macrocognition in teams: Developing and defining complex collaborative processes and products. *Theoretical Issues in Ergonomic Science*, *11*, 4250–4271.

Fiore, S. M., & Wiltshire, T. J. (2016). Technology as teammate: Examining the role of external cognition in support of team cognitive processes. *Frontiers in Psychology*, *7*, Article 1531. doi:10.3389/fpsyg.2016.01531

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

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

Fischer, U., McDonnell, L., & Orasanu, J. (2007). Linguistic correlates of team performance: Toward a tool for monitoring team functioning during space missions. *Aviation, Space, and Environmental Medicine*, *78*(5 Suppl.), B86–B95.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231–236). Hillsdale, NJ: Erlbaum.

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

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

Gabelica, C., & Fiore, S. M. (2013). What can training researchers gain from examination of methods for active learning (PBL, TBL, and SBL). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*, 462–466. doi:10.1177/1541931213571100

Gigone, D., & Hastie, R. (1997). Proper analysis of the accuracy of group judgments. *Psychological Bulletin*, *121*, 149–167.

Gijselaers, W. H. (1996). Connecting problem-based practices with educational theory. *New Directions for Teaching and Learning*, *68*, 13–21. doi:10.1002/tl.37219966805

Gilbert, S., Slavina, A., Sinatra, A. M., Bonner, D., Johnston, J., Holub, J., . . . Winer, E. (2018). Creating a team tutor using GIFT. *International Journal for Artificial Intelligence in Education*, *28*, 286–313.

Gillies, R. M. (2008). The effects of cooperative learning on junior high school students' behaviours, discourse and learning during a science-based learning activity. *School Psychology International*, *29*, 328–347.

Goldstone, R. L., Roberts, M. E., & Gureckis, T. M. (2008). Emergent processes in group behavior. *Current Directions in Psychological Science*, *17*, 10–15.

Goldstone, R. L., & Theiner, G. (2017). The multiple, interacting levels of cognitive systems (MILCS) perspective on group cognition. *Philosophical Psychology*, *30*, 338–372.

Gorman, J. C., Martin, M. J., Dunbar, T. A., Stevens, R. H., & Galloway, T. (2013). Analysis of semantic content and its relation to team neurophysiology during submarine crew training. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Foundations of augmented cognition. AC 2013. Lecture notes in computer science* (Vol. 8027, pp. 143–152). Berlin: Springer.

Graesser, A. C. (2016). Conversations with AutoTutor help students learn. *International Journal of Artificial Intelligence in Education*, *26*, 124–132.

Graesser, A. C., Cai, Z., Hu, X., Foltz, P. W., Greiff, S., Kuo, B.-C., . . . Shaffer, D. W. (2017). Assessment of collaborative problem solving. In R. Sottilare, A. Graesser, X. Hu, & G. Goodwin (Eds.), *Design recommendations for intelligent tutoring systems. Volume 5: Assessment methods* (pp. 275–286). Orlando, FL: United States Army Research Laboratory.

Graesser, A. C., Cai, Z., Morgan, B., & Wang, L. (2017). Assessment with computer agents that engage in conversational dialogues and trialogues with learners. *Computers in Human Behavior*, *76*, 607–616.

Graesser, A. C., Dowell, N., Hampton, A. J., Lippert, A. M., Li, H., & Shaffer, D. W. (2018). Building intelligent conversational tutors and mentors for team collaborative problem solving: Guidance from the 2015 Program for International Student Assessment. In J. J. Johnston, R. Sottilare, A. Sinatra, & C. S. Burke (Eds.), *Building intelligent tutoring systems for teams: What matters* (pp. 173–214). West Yorkshire, England: Emerald.

Graesser, A. C., Foltz, P. W., Rosen, Y., Shaffer, D. W., Forsyth, C., & Germany, M. (2018). Challenges of assessing collaborative problem solving. In E. Care, P. Griffin, & M. Wilson (Eds.), *Assessment and teaching of 21st century skills: Research and applications* (pp. 75–91). Cham, Switzerland: Springer.

Graesser, A. C., Forsyth, C. M., & Foltz, P. (2017). Assessing conversation quality, reasoning, and problem solving with computer agents. In B. Csapó & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 245–261). Paris, France: OECD Publishing. doi:10.1787/9789264273955-17-en

Grand, J. A., Braun, M. T., Kuljanin, G., Kozlowski, S. W. J., & Chao, G. T. (2016). The dynamics of team cognition: A process-oriented theory of knowledge emergence in teams. *Journal of Applied Psychology*, *101*, 1353–1385.

Greiff, S., Kretzschmar, A., Müller, J. C., Spinath, B., & Martin, R. (2014). The computer-based assessment of complex problem solving and how it is influenced by students' information and communication technology literacy. *Journal of Educational Psychology*, *106*, 666–680.

Greiff, S., Wüstenberg, 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. (2015). ATC21S method. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 3–33). Dordrecht, The Netherlands: Springer.

Gully, S. M., Devine, D. J., & Whitney, D. J. (1995). A meta-analysis of cohesion and performance: Effects of level of analysis and task interdependence. *Small Group Research*, *26*, 497–520.

Hall, K. L., Vogel, A. L., Huang, G. C., Serrano, K. J., Rice, E. L., Tsakraklides, S. P., & Fiore, S. M. (2018). The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *American Psychologist*, *73*, 532–548.

Harding, S.-M. E., Griffin, P. E., Awwal, N., & Alom, B. M. M. (2017). Measuring collaborative problem solving using mathematics-based tasks. *AERA Open*, *3*(3). doi:10.1177/2332858417728046

Hart Research Associates. (2015). *Falling short? College learning and career success*. Washington, DC: Association of American Colleges and Universities.

Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review*, *112*, 494–508.

Herborn, K. (2018). *The assessment of collaborative problem solving in PISA 2015: An investigation of the validity* (Doctoral thesis, University of Luxembourg, Luxembourg). Retrieved from http://hdl.handle.net/10993/35890

Herborn, K., Stadler, M., Mustafić, M., & Greiff, S. (2018). The assessment of collaborative problem solving in PISA 2015: Can computer agents replace humans? *Computers in Human Behavior*. Advance online publication. doi:10.1016/j.chb.2018.07.035

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). 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.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, *16*, 235–266.

Horry, R., Palmer, M. A., Sexton, M. L., & Brewer, N. (2012). Memory conformity for confidently recognized items: The power of social influence on memory reports. *Journal of Experimental Social Psychology*, *48*, 783–786.

Hundhausen, C., Agrawal, A., Fairbrother, D., & Trevisan, M. (2010). Does studio-based instruction work in CS1? An empirical comparison with a traditional approach. In G. Lewandowski, S. Wolfman, T. J. Cortina, E. L. Walker,

& D. R. Musicant (Eds.), *SIGCSE'10: Proceedings of the 41st ACM Technical Symposium on Computer Science Education* (pp. 500–504). New York, NY: Association for Computing Machinery. doi:10.1145/1734263.1734432

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Hutchins, S. G., & Kendall, T. (2010). The role of cognition in team collaboration during complex problem solving. In K. L. Mosier & U. M. Fischer (Eds.), *Informed by knowledge: Expert performance in complex situations* (pp. 69–89). Abington, England: Taylor & Francis.

Janis, I. L. (1982). *Groupthink*. Boston, MA: Houghton Mifflin.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (2000). Constructive controversy: The educative power of intellectual conflict. *Change: The Magazine of Higher Learning*, *32*(1), 28–37. doi:10.1080/00091380009602706

Johnson, W. J., & Lester, J. (2016). Face-to-face interaction with pedagogical agents, twenty years later. *International Journal of Artificial Intelligence in Education*, *26*, 25–36.

Jurafsky, D., & Martin, J. (2008). *Speech and language processing* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, *65*, 681–706.

Kerr, D., Andrews, J. J., & Mislevy, R. J. (2016). The in-task assessment framework for behavioral data. In A. A. Rupp & J. P. Leighton (Eds.), *Handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 472–507). Hoboken, NJ: Wiley-Blackwell.

Kerr, N. L. (1983). Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology*, *45*, 819–828.

Klein, C., DiazGranados, D., Salas, E., Le, H., Burke, C. S., Lyons, R., & Goodwin, G. F. (2009). Does team building work? *Small Group Research*, *40*, 181–222.

Kniffin, K. M., Wansink, B., & Shimizu, M. (2015). Sports at work: Anticipated and persistent correlates of participation in high school athletics. *Journal of Leadership & Organizational Studies*, *22*, 217–230.

Lai, E. R., DiCerbo, K. E., & Foltz, P. (2017). *Skills for today: What we know about teaching and assessing collaboration*. Retrieved from http://www.p21.org/storage/documents/Skills_For_Today_Series-Pearson/Collaboration_White_Paper_FINAL.pdf

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Taylor & Francis.

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, *37*, 822–832.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York, NY: Cambridge University Press.

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, *61*, 273–307.

Letsky, M. P., Warner, N. W., Fiore, S., & Smith, C. A. P. (2008). *Macrocognition in teams: Theories and methodologies*. London, England: Ashgate.

Liang, H. Y., Shih, H. A., & Chiang, Y. H. (2015). Team diversity and team helping behavior: The mediating roles of team cooperation and team cohesion. *European Management Journal*, *33*, 48–59.

Liden, R. C., Wayne, S. J., Jaworski, R. A., & Bennett, N. (2004). Social loafing: A field investigation. *Journal of Management*, *30*, 285–304.

Lievens, F. (2013). Adjusting medical school admission: Assessing interpersonal skills using situational judgement tests. *Medical Education*, *47*, 182–189.

Liu, L., von Davier, A. A., Hao, J., 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* (pp. 344–359). Hershey, PA: IGI-Global.

Martin, M. J., & Foltz, P. W. (2004, May). *Automated team discourse annotation and performance prediction using LSA*. Paper presented at the annual meeting of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Boston, MA. Retrieved from https://aclanthology.info/papers/N04-4025/n04-4025.

Mathieu, J. E., Hollenbeck, J. R., van Knippenberg, D., & Ilgen, D. R. (2017). A century of work teams in the Journal of Applied Psychology. *Journal of Applied Psychology*, *102*, 452–467. doi:10.1037/apl0000128

Mathieu, J. E., Kukenberger, M. R., D'innocenzo, L., & Reilly, G. (2015). Modeling reciprocal team cohesion–performance relationships, as impacted by shared leadership and members' competence. *Journal of Applied Psychology*, *100*, 713–734.

Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 287–303). Mahwah, NJ: Erlbaum.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, *9*, 103–113.

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, *2*, 63–86.

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, *94*, 535–546.

Michaelsen, L. K., Parmelee, D. X., McMahon, K. K., & Levine, R. E. (Eds.). (2008). *Team-based learning for health professions education: A guide to using small groups for improving learning*. Sterling, VA: Stylus.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*, 3–62.

Mohammed, S., & Angell, L. C. (2003). Personality heterogeneity in teams which differences make a difference for team performance? *Small Group Research*, *34*, 651–677.

Morgan, B. B., Salas, E., & Glickman, A. S. (1993). An analysis of team evolution and maturation. *Journal of General Psychology*, *120*, 277–291.

Mu, J., Stegmann, K., Mayfield, E., Rosé, C., & Fischer, F. (2012). The ACODEA framework: Developing segmentation and classification schemes for fully automatic analysis of online discussions. *International Journal of Computer-Supported Collaborative Learning*, *7*, 285–305.

Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin*, *115*, 210–227.

Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and Applied Social Psychology*, *12*, 3–23.

Nardi, B. A. (1996). *Context and consciousness: Activity theory and human-computer interaction*. Cambridge, MA: MIT Press.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2005). *Facilitating interdisciplinary research*. Washington, DC: The National Academies Press. doi:10.17226/11153

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: The National Academies Press. doi:10.17226/11463

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2010). *Rising above the gathering storm, revisited: Rapidly approaching category 5*. Washington, DC: The National Academies Press. doi:10.17226/12999

National Academies of Sciences, Engineering, and Medicine. (2016). *Developing a national STEM workforce strategy: A workshop summary*. Washington, DC: The National Academies Press. doi:10.17226/21900

National Research Council. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press. doi:10.17226/9853

National Research Council. (2011). *Assessing 21st century skills: Summary of a workshop*. Washington, DC: The National Academies Press. doi:10.17226/13215

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academies Press. doi:10.17226/13398

National Research Council. (2015). *Enhancing the effectiveness of team science*. Washington, DC: The National Academies Press. doi:10.17226/19007

Neuman, G. A., Wagner, S. H., & Christiansen, N. D. (1999). The relationship between work team personality composition and the job performance of teams. *Group & Organization Management*, *24*, 28–45.

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

Norman, G. R., & Schmidt, H. G. (2000). Effectiveness of problem-based learning curricula: theory, practice and paper darts. *Medical Education*, *34*, 721–728.

Nouri, J., Åkerfeldt, A., Fors, U., & Selander, S. (2017). Assessing collaborative problem solving skills in technology-enhanced learning environments—The PISA framework and modes of communication. *International Journal of Emerging Technologies in Learning*, *12*, 163–174.

Oliver, L. W., Harman, J., Hoover, E., Hayes, S. M., & Pandhi, N. A. (1999). A quantitative integration of the military cohesion literature. *Military Psychology*, *11*, 57–83.

O'Neil, H. F., Chuang, S., & Chung, G. (2003). Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education: Principles, Policy & Practice*, *10*, 361–373.

Organisation for Economic Co-operation and Development. (2010). *PISA 2012 field trial problem solving framework*. Retrieved from the OECD website: http://www.oecd.org/pisa/pisaproducts/46962005.pdf

Organisation for Economic Co-operation and Development. (2017a). *PISA 2015 assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving* (Rev. ed.), Retrieved from the OECD website: http://dx.doi.org/10.1787/9789264281820-en

Organisation for Economic Co-operation and Development. (2017b). *PISA 2015 results (volume V): Collaborative problem solving*. doi:10.1787/9789264285521-en

Pöysä-Tarhonen, J., Care, E., Awwal, N., & Häkkinen, P. (2018). Pair interactions in online assessments of collaborative problem solving: Case-based portraits. *Research and Practice in Technology Enhanced Learning*, *13*, Article 12. doi:10.1186/s41039-018-0079-7

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching*, *49*, 363–393.

Rajaram, S. (2011). Collaboration both hurts and helps memory: A cognitive perspective. *Current Directions in Psychological Science*, *20*, 76–81.

Roediger, H. L., Meade, M. L., & Bergman, E. T. (2001). Social contagion of memory. *Psychonomic Bulletin & Review*, *8*, 365–371.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *The Journal of the Learning Sciences*, *2*, 235–276.

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. W. (2014). Assessing collaborative problem solving through automated technologies. *Journal of Research and Practice in Technology Enhanced Learning*, *9*, 389–410.

Rotherham, A. J., & Willingham, D. T. (2010). 21st-century skills. *American Educator*, *34*, 17–20.

Rus, V., Lintean, M. C., Banjade, R., Niraula, N. B., & Stefanescu, D. (2013). SEMILAR: The semantic similarity toolkit. In M. Butt & S. Hussain (Eds.), *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 163–168). Retrieved from http://www.aclweb.org/anthology/P13-4028.

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

Salomon, G. (Ed.). (1993). *Distributed cognitions*. Cambridge, England: Cambridge University Press.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*, 211–232.

Scoular, C., & Care, E. (2018). Teaching 21st century skills: Implications at system levels in Australia. In E. Care, P. P. Griffin, P., & M. Wilson (Eds.), *Assessment and teaching of the 21st century skills* (pp. 145–162). Cham, Switzerland: Springer.

Scoular, C., Care, E., & Hesse, F. W. (2017). Designs for operationalizing collaborative problem solving for automated assessment. *Journal of Educational Measurement*, *54*, 12–35.

Seeber, I., Maier, R., & Weber, B. (2013). Macrocognition in collaboration: Analyzing processes of team knowledge building with CoPrA. *Group Decision and Negotiation*, *22*, 915-942.

Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.

Slavin, R. E. (2017). Instruction based on cooperative learning. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 388–404). New York, NY: Routledge Press.

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, *46*, 49–76.

Sottilare, R., Graesser, A. C., Hu, X., & Sinatra, A. (Eds.). (2018). *Design recommendations for intelligent tutoring systems: Team science* (Vol. 6). Orlando, FL: United States Army Research Laboratory.

Sottilare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2018). Designing adaptive instruction for teams: A meta-analysis. *International Journal of Artificial Intelligence in Education*, *28*, 225–264.

Sternberg, R. J. (1995). Expertise in complex problem solving: A comparison of alternative concepts. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 295–321). Hillsdale, NJ: Erlbaum.

Stevens, M. J., & Campion, M. A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management*, *20*, 503–530.

Stewart, C. O., Setlock, L. D., & Fussell, S. R. (2007). Conversational argumentation in decision-making: Differences across cultures and communication media. *Discourse Processes*, *44*, 113–139.

Stone, C. B., Barnier, A. J., Sutton, J., & Hirst, W. (2010). Building consensus about the past: Schema consistency and convergence in socially shared retrieval-induced forgetting. *Memory*, *18*, 170–184.

Tausczik, Y. R., & Pennebaker, J. W. (2013). Improving teamwork using real-time language feedback. In W. E. Mackay, P. Baudisch, & M. Beaudouin-Lafon (Eds.), *CHI'13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 459–468). doi:10.1145/2470654.2470720

Theiner, G., Allen, C., & Goldstone, R. L. (2010). Recognizing group cognition. *Cognitive Systems Research*, *11*, 378–395.

Thompson, L. L., Wang, J., & Gunia, B. C. (2010). Negotiation. *Annual Review of Psychology*, *61*, 491–515.

Thorley, C., & Christiansen, P. (2018). The impact of own and others' alcohol consumption on social contagion following a collaborative memory task. *Memory*, *26*, 727–740.

Van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology*, *89*, 1008–1022.

Van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology*, *58*, 515–541.

Vermunt, J. D., & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction*, *9*, 257–280.

Vernon, D., & Blake, R. L. (1993). Does problem-based learning work? A meta-analysis of evaluative research. *Academic Medicine*, *68*, 550–563.

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, *29*, 477–511.

von Davier, A. A., Zhu, M., & Kyllonen, P. C. (Eds.). (2017). *Innovative assessment of collaboration*. Cham, Switzerland: Springer.

Wegner, D. M. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–205). New York, NY: Springer.

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1160–1175.

Williams, K., Harkins, S. G., & Latané, B. (1981). Identifiability as a deterrent to social loafing: Two cheering experiments. *Journal of Personality and Social Psychology*, *40*, 303–311.

Wiltshire, T. J., Butner, J. E., & Fiore, S. M. (2017). Problem solving phase transitions during team collaboration. *Cognitive Science*, *42*, 129–167. doi:10.1111/cogs.12482

Zhuang, X., MacCann, C., Wang, L., Liu, L., & Roberts, R. D. (2008). *Development and validity evidence supporting a teamwork and collaboration assessment for high school students* (Educational Testing Service Research Report No. RR-08-50). Retrieved from https://files.eric.ed.gov/fulltext/EJ1111347.pdf.