Review of Psychology,
2012, Vol. 19, No. 2, 111-121 UDC 159.9

From interactive to collaborative problem solving: Current issues in the Programme for International Student Assessment

SAMUEL GREIFF

Problem solving is a central topic in modern educational contexts and has been receiving increasing interest in educational large-scale assessment studies like the Programme for International Student Assessment (PISA). However, conceptualizing and assessing problem solving skills is not a trivial task and up to now research has provided little assistance in developing sound measures. In this article the rationale behind PISA and behind two domain-unspecific strands in problem solving research, the fields of interactive problem solving (IPS) and collaborative problem solving (ColPS), which will be assessed as cross-curricular abilities in the PISA 2012 and the PISA 2015 survey, respectively, are introduced. After giving basic information on PISA, both constructs, IPS and ColPS, are reviewed with regard to (a) their conceptual understanding and theoretical background, (b) specific items used to translate the theoretical background into empirical scales, (c) existing research on predictive and construct validity, and (d) general obstacles as well as obstacles specific to large-scales assessments, which are to be met in an assessment of IPS and ColPS. The internationally high impact of the PISA survey renders further research on both constructs indispensable. The subsequent implications for problem solving research and educational systems are discussed.

Key words: interactive problem solving, collaborative problem solving, PISA, large-scale assessment, complex problem solving, dynamic problem solving

Throughout history, seminal inventions such as the use of electricity or travelling by plane have strongly altered mankind's way of life. However, hardly any change has been so rapid, so dramatic, and so widespread as the one ignited by the invention of personal computers. In fact, due to this invention literally all skills needed for a successful career have undergone great changes in the last 20 years and many studies show that tasks at school, at university, and at work are continuously less routine, but increasingly involve non-routine, dynamic, and complex skills mediated by computer interaction such as general problem solving or collaborating in a group.

These tasks are not limited to educational contexts or to the job arena, but also involve the capability to naturally use computers and other technical devices in private life. For instance, the ability to handle an MP3 Player is taken for granted in the 21st century and today's generation seldom struggles with these kinds of requirements. On the other hand, people born a few generations ago would have been utterly lost confronted with a device which was supposed to digitally play music. Imagine you came across an MP3 Player for the first time having only a slight idea what it might be used for (Figure 1; Funke, 2001; Greiff, 2012). What would be your natural approach to master this device?

Figure 1. MP3 Player as an item example for interactive problem solving. This item was published by the OECD (2011) as an example of the PISA 2012 interactive problem solving assessment.

Samuel Greiff, Department of Psychology, Heidelberg University, Hauptstrasse 47-51, 69117 Heidelberg, Germany.
E-mail: samuel.greiff@psychologie.uni-heidelberg.de (the address for correspondence).

First, you would (a) press buttons randomly (i.e., give inputs) in order to receive a reaction from the device (i.e., generate output; Kröner, Plass, & Leutner, 2005). From the observed connections between in- and outputs, you would (b) acquire knowledge and a mental representation of the underlying structure would emerge (Markman, 1999). You could (c) then apply your mental representation to reach desired goal states, for instance playing a particular song repeatedly (Funke, 2001; Novick & Bassok, 2005). While doing so, you would (d) have to use your metacognitive knowledge and strategies to monitor your progress and to reflect on your problem solving behavior (Wirth & Leutner, 2008).

On a general level, the MP3 Player, which is encountered for the first time, and the four processes necessary to master it describe a problem situation and the subsequent process of problem solving. According to Frensch and Funke (1995), a problem situation consists of an initial state, a goal state, and barriers between initial and goal state that cannot be removed by routine operations (cf. Funke & Frensch, 2007). Mayer and Wittrock (2006) define the process of problem solving as transforming a given state into a goal state when no obvious method of solution is available. This process can be divided into four subprocesses, which are illustrated in the MP3 Player-example: (a) *exploring and understanding* by generating information about the problem situation, (b) *representing and formulating* by reducing and integrating the information gathered into a mental model of the problem, (c) *planning and executing* by carrying out actions to reach a desired goal state, and (d) *evaluating and reflecting* by using metacognitive skills to monitor the entire process of problem solving (The Organization for Economic Co-operation and Development [OECD], 2010).

The objective of this paper is to introduce the assessment of problem solving in international large-scale assessments (LSA) and to discuss unsolved issues and implications for educational research from a scientific perspective. More specifically, after introducing the concept of problem solving, I will give some background information on the organization and rationale behind the Programme for International Student Assessment (PISA), arguably the most influential and comprehensive LSA worldwide. I will then present and discuss two different concepts of problem solving and their assessment, *interactive problem solving* (IPS), which will be part of the PISA 2012 survey, and *collaborative problem solving* (ColPS), which will be part of the PISA 2015 survey. The public attention PISA has received in the past and the psychometric issues associated with assessing any kind of problem solving (Greiff, 2012; Wüstenberg, Greiff, & Funke, 2012) warrant a scientific examination of the concept's realization in LSA. In fact, the very limited assessment experience particularly with regard to ColPS renders such a review even more necessary.

The understanding of a problem situation and of the four problem solving processes mentioned above is neither scientifically nor in LSA limited to technical devices such as the MP3 Player but is also applied to non-technical problems in varying contexts such as personal, social, or global (Funke & Frensch, 2007; OECD, 2010). An individual's level of problem solving is considered a crucial determinant for one's success in private life and in the job arena. Even more general, all non-routine behavior can be seen as problem solving rendering the concept potentially relevant for educational contexts. As the eminent Hungarian mathematician George Pólya put it already 40 years ago (1971, p. 4): "What is good education? Giving systematically opportunity to the student to discover things by himself". Discovering things within an unknown system with the aim to build and apply this knowledge is at the core of the understanding of problem solving mentioned above. Thus, to allocate people according to their individual problem solving level to a fitting environment and to foster their abilities is one of society's most crucial and essential tasks, which needs to be based on an accurate assessment of problem solving. As international LSA such as the PISA survey aim at measuring cross-curricular skills of relevance not only in school but also in real life (OECD, 2009a), and problem solving is considered such a skill (OECD, 2010), obviously such an assessment should be included. The underlying motivation is to gain information on a skill that goes beyond domain-specific and content-bound knowledge and provides information on students' capability when faced with ecologically valid real world problems.

However, problem solving research is no consistent field of study. Very much to the contrary, few areas of research are as deeply divided as problem solving and it is widely acknowledged that there are two main research lines apart from each other (Frensch & Funke, 1995; Sternberg, 1995). The first line conducts research on domain-specific problem solving in different content areas (cf. Sugrue, 1995) such as mathematical (e.g., Daniel & Embretson, 2010), scientific (e.g., Dunbar & Fugelsang, 2005), or technical (e.g., Baumert, Evans, & Geiser, 1998) problem solving often comparing experts and novices whereas the other line is concerned with general mental processes associated with problem solving performance in microworlds (Dörner, 1986; Greiff, 2012). Both lines are – at least indirectly – represented in PISA as the assessment of mathematics or science skills, which is mandatory in each cycle, represents domain-specific problem solving going beyond specific school curricula and content knowledge (OECD, 2009a). Additionally and for the first time in PISA 2012, IPS is conceptualized as a cross-curricular skill representing domain unspecific problem solving. Arguably, both domain-specific and domain-general processes are involved in problem solving and there are some recently published studies supporting this point of view (Abele et al., 2012; Wüstenberg et al., 2012) suggesting the importance of a specific and a general understanding of problem solving. Additional support for the integrated understanding of problem solving as a combination of domain specific and domain unspecific processes is found in Sternberg (1995). He criticizes that some research on problem solving focuses too strongly on the comparison of experts and novices and specific differences between the two thereby overstating domain specific processes and neglecting domain general processes. Further, he certifies educational psychology a general neglect of domain general aspects of problem solving by stating that they have "not captured their [researchers' and practitioners'] imagination, at least not in the United States" (Sternberg, 1995, p. 300). He asks for comprehensive research on complex problems, which (a) are more about real life (e.g., represent what happens in the classroom and is required in educational contexts) and (b) can be solved by anybody (i.e., also by students being considered novices).

According to Novick, Hurley, and Francis (1999), domain general processes are important when solving problems because abstract representation schemas are more useful than specific example problems for understanding the structure of novel problems. This is because these general representations are not contaminated by content knowledge (Holyoak, 1985). This view is empirically supported by Chen and Klahr (1999) and Klahr, Triona, and Williams (2007) who showed that training students in how to conduct experiments that allow for causal inferences by teaching them the principle of isolated variation led to an increase in the knowledge acquired even though it was gathered in a particular context (i.e., science education). In fact, knowledge was successfully transferred to different contexts. That is, students in the trained group performed better in tasks comparable to the original one but also in generalizing the knowledge to various tasks with differing characteristics and in other contexts (Chen & Klahr, 1999).

Interestingly, this kind of domain-general understanding of problem solving was the very idea when the decision to include IPS and ColPS in the PISA survey was made: to provide the means and measures necessary to specifically target those general problem solving and collaborative skills needed for a successful participation in today's society that are not captured within domain specific problem solving. Before I present these two concepts in detail, I will give some background information on the PISA survey and its impact in research and policy.

THE PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT (PISA)

During the 1950ies great pioneering work was accomplished by implementing LSA, mostly on a national level, to gain a broader understanding of educational systems for the first time ever (e.g., Nagy, 2000). A few decades later, these efforts were carried further and LSA advanced on international grounds with the first international LSA receiving noteworthy public attention being the Trends in International Mathematics and Science Study (TIMSS). TIMSS is still carried out in a 4-year cycle and provides important insights on mathematics and science achievement across almost 50 countries (e.g., Gonzales et al., 2009; Köller & Baumert, 2001). Whereas TIMSS seeks to measure what students know in relation to the intended curriculum, the PISA survey, which has been carried out by the Organization for Economic Co-operation and Development (OECD) for over a decade starting in 2000, is aimed at establishing an understanding of what students can do with their acquired knowledge at the end of compulsory schooling and is less concerned with curricular domain-specific contents. More specifically, "rather than focusing on the extent to which […] students have mastered a specific school curriculum, it [the assessment] looks at their ability to use their knowledge and skills to meet real-life challenges" (OECD, 2009a, p. 1). Even though scientific experts vividly discuss whether PISA sufficiently fulfils this intention (e.g., Hopmann, Brinek, & Retzl, 2007; Sjoberg, 2007), the targeted constructs are obviously skill- and not knowledge-based. That is, the assessment aims at testing skills that students meet when negotiating their daily lives (OECD, 2009a).

Taking place in three year cycles, classical skills in mathematics, science, and reading and their application in contents that are not bound to specific curricula are assessed in PISA with one of them being the major domain in every cycle. That is, in the 2009 assessment, reading was for the second time after 2000 a major domain allowing one hour of assessment to reliably and validly reflect students' level of reading achievement (OECD, 2009b). Besides the three classical domains, additional options to assess cross-curricular competencies are included on an irregular basis. For instance, participating countries could opt for an assessment of digital reading skills in 2009. Both IPS in PISA 2012 and ColPS in PISA 2015 are an optional assessment of cross-curricular skills in the main survey and the increasing number of countries choosing it displays the internationally growing interest in domain-unspecific measures of achievement.

In 2009 a representative sample of approximately half a million students completed the main survey in almost 70 participating countries representing about 26 million 15-year olds (OECD, 2009a) and politicians and practitioners use the PISA results as a way to evaluate their policy and their national educational systems yielding large practical, political, and public implications. For example, the German school system is probably unique in placing students already after Grade 4 in different school tracks according to their individual ability (low, medium, or high track) with transmission rates between tracks at a later stage being extremely low (Jürges & Schneider, 2006). The underlying theoretical assumption is that fostering students in a way tailored at their intellectual ability and their individual needs, enhances overall achievement. However, this theoretical assumption, which has been at the core of the German educational system for several decades, was proven plainly wrong by the PISA studies (e.g., OECD, 2009b; PISA-Konsortium Deutschland, 2004). That is, variances between German schools were among the highest worldwide and the level of achievement was average or even below average depending on the specific cycle and the domain considered, whereas other countries like Finland with a considerably later or no differentiation into different school tracks exhibited substantially lower variances between schools and high average performance. It does not surprise that – in light of these results – experts frequently discuss necessary and fundamental interventions into the educational system and even wonder whether different school tracks are not an overdue relict of old times. These considerations gain their empirical right to exist directly from the PISA studies.

Clearly, such exemplary results demonstrate the potentially major implications for educational systems and the challenge PISA currently faces in moving beyond purely descriptive measures of achievement but to provide sound measures to answer questions about the correlates and causes of school performance. Efforts into this direction involve using student, parent, and school questionnaires as well as linking contextual information such as economic indicators or educational markers to levels of achievement (OECD, 2009a). Further, the repeated PISA cycles now allow moving from a cross-sectional to a longitudinal design linking interventions into an educational system to outcomes a few years later within and across nations. This serves two overarching purposes: to describe the level and the development of important skills in the 21st century at different educational levels over time on the one hand, and to explain the process of skill acquisition theoretically and empirically thereby deriving means to understand and enhance these skills on the other hand. That is, LSA launched by the OECD become increasingly explanatory and longitudinal and widen their view to later stages of life. For instance, the Programme for the International Assessment of Adult Competencies, which tests adults between the age of 16 and 65, took place the first time in 2011 with results to be published in 2013.

Looking forward to the upcoming PISA cycles with a steadily increasing number of participating countries (OECD, 2009a), it is in 2012 and 2015 for the first time ever that cross-curricular competencies are explicitly included in the assessment. After allowing for the assessment of analytical problem solving in 2003 with a strong affiliation to mathematics and of digital reading in the 2009 with a strong affiliation to reading, the conception of both IPS and ColPS are profoundly domain-general. That is, even though the assessment is contextually embedded into certain domains solving the tasks is explicitly not necessitated by specific knowledge. However, considering the existing body of knowledge about IPS and ColPS, including them into a high impact assessment as PISA may be challenging from a conceptual, empirical, and technical point of view. I will now review both concepts and their role in the PISA survey.

INTERACTIVE PROBLEM SOLVING IN PISA 2012

Some problem situations – for instance those used in the PISA 2003 assessment of analytical problem solving – contain all information necessary to solve the problem at the outset. This differs from the PISA 2012 understanding, in which problems “that require direct interaction by the solver to uncover and discover relevant information” (OECD, 2010, p. 7) are considered a central feature of the assessment. That is, without some active and systematic exploration and interaction between problem solver and task, the problem cannot be solved. This unique feature is reflected in the term IPS and there is some evidence that skills additional to those involved in traditional reasoning-based problem solving and general mental ability, which do not necessarily require active exploration, are required in IPS (Klieme, 2004; Raven, 2000; Wüstenberg et al., 2012). Formally, OECD (2010) understands problem solving skills as “an individual’s capacity to engage in cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious. It includes the willingness to engage with such situations in order to achieve one’s potential as a constructive and reflective citizen” (p. 10). This rather unspecific definition largely reflects the self-concept of the PISA survey with its focus on external validity and its direct reference to participation in society, whereas little information about the concept itself is contained.

To gain a deeper understanding about the theoretical background of IPS, one needs to search existing literature. However, when doing so the term IPS cannot be found, but there is a large body of literature dealing with interactive problems under different labels: Funke (2001, 2010) emphasized the dynamics inherent in each problem as the problem situation may change by itself over time using the term *dynamic problem solving* whereas Dörner (1986) introduced the original term *complex problem solving* referring to the complexity of the underlying system. That is, changing one variable in a task may lead to manifold changes in other variables. This European line of research (cf. Frensch & Funke, 1995) has worked on interactive problems for several decades and produced a variety of interesting findings even though not primarily from an assessment perspective and not under the flag of interactivity.

Within this line, the process of problem solving is defined as “the successful interaction with task environments that are dynamic (i.e., change as a function of user’s intervention and/or as a function of time) and in which some, if not all, of the environment’s regularities can only be revealed by successful exploration and integration of the information gained in that process“(Buchner, 1995, p. 14). This definition highlights the aspect of interactivity and also acknowledges that prior knowledge or previous experience as well as the context may influence how the problem is tackled but these aspects are not of elementary concern and problems are designed to be solvable without prior knowledge and independent of the specific context they are embedded into (OECD, 2010). However, not only this European understanding of complex or IPS is resembled in the PISA 2012 understanding. It also includes a theoretical notion of problem solving in particular content domains as based on research largely carried out in the United States, which emphasizes two main demands placed on a problem solver: (a) to represent a problem (*representation*) and (b) to carry out a solution to solve the problem (*solution*; Novick & Bassok, 2005). Mayer and Wittrock (2006) further subdivide these two overarching processes into four more narrow dimensions, which are closely resembled in the four PISA processes mentioned above. Different authors conceive and label processes involved in problem solving differently, but there is a considerable overlap in their views (e.g., Baxter & Glaser, 1998; Blech & Funke, 2005), which is largely represented in the PISA 2012 understanding of IPS (OECD, 2010). Interestingly, despite acknowledging that problem solving consists of several distinct processes, PISA 2012 will report only one dimension paying tribute to the limited availability of testing time and to easiness of public perception.

Besides coming to an understanding of the processes IPS is composed of, translating the theoretical understanding into items poses a great challenge. Clearly, interactive problems changing dynamically as reaction to user's intervention have to be assessed computer-based and it is for the first time in 2012 that the PISA survey provides a comprehensive computer-delivery platform - one of the reasons why interactive problems were not included in previous assessment cycles. Further, unsolved measurement issues adjourned a meaningful assessment of IPS until a few years ago. More specifically, in the early years of research on interactive problems, complex microworlds (e.g., Blech & Funke, 2005; Dörner, 1986; Frensch & Funke, 1995) were used to test problem solving behavior often requiring several hours of testing time and being associated with severe psychometric issues (Greiff, 2012). One major advance in assessing problem solving skills was made when Funke (2001) introduced (a) finite state automata (FSA) and (b) linear structural equation (LSE) models paving the way for recent psychometric improvements (cf. Greiff, 2012; Kröner et al., 2005; Wüstenberg et al., 2012). These formalisms allowed to formally describe the underlying structure of different interactive problems and to develop measures from the data produced. Interactive problems based on the first formalism, FSA, are composed of qualitative connections between variables. That is, changing one variable may transfer parts of the system or the entire system into a different state. Problems based on the second formalism, LSE, on the other hand, are composed of quantitative connections between variables. Decreasing or increasing an input variable may in turn lead to a decrease or increase in one or several output variables

The MP3 Player in Figure 1, one of the items published by the OECD to demonstrate the principle of IPS, is an example of a (a) FSA and displays the kind of interaction mentioned above: In its original state, the MP3 Player contains only a fraction of the information necessary to fully represent the device or to solve a specific problem. Formally, the underlying structure is composed of qualitatively different states as is mandatory in FSA and each of the four problem solving processes is targeted by specific items within the PISA 2012 assessment. For instance, some statements about the player's functionality have to be evaluated due to their correctness assessing representing and formulating (e.g., You need to use the lower button to change the type of music. Is this statement correct?), whereas others require a specific goal to be reached assessing planning and executing (e.g., Set the MP3 Player to Pop, Volume 6, and Bass 1).

In a (b) LSE example, which is not directly taken from the upcoming PISA items but conceptually similar to a considerable amount of them is displayed in Figure 2, input variables (i.e., different chemical substances) are related to output variables (i.e., different attributes of these substances) and again different processes of the problem solving process are tested. That is, generating information by systematically entering values into the input variables and by moving only one slider at a time allowing to directly relate input to output variables indicates a high level of exploring and understanding, whereas unsystematic intervention patterns indicate low performance levels. Further, one potentially sound way to assess the metacognitive process of evaluating and reflecting, which is largely taking place implicitly during the entire problem solving process (Wirth, 2004; Wirth & Leutner, 2008), is to ask for a general rule how to explore the kind of system presented in Figure 2 (e.g., What is the best way to move the sliders in order to gain as much information about the system as possible?).

Research on complex microworlds in general and on FSA and LSE systems specifically looks back on an extensive body of studies and interesting findings (cf. Frensch & Funke, 1995). FSA and LSE are eligible to provide both a theoretical conception (Fischer, Greiff, & Funke, 2012) and an overarching item framework for IPS in PISA 2012. Despite this research (Frensch & Funke, 1995), the orientation towards skill assessment and diagnostics of individual problem solving levels has just begun. In fact, only recently severe psychometric problems within formal frameworks were resolved in a second major measurement advance (Greiff, 2012; Wüstenberg et al., 2012) and now allow for a proper assessment perspective on IPS serving the emerging public interest in this cross-curricular skill. But what do we currently know empirically about the assessment of IPS?

Recent results suggest that representation and solution as overarching processes can be empirically separated in different student populations (e.g., Kröner et al., 2005; Wüstenberg et al., 2012) and, depending on the specific sample under study and the item layout, the first three problem solving processes defined in PISA (OECD, 2010; evaluating and reflecting is usually not assessed directly) can be dis-

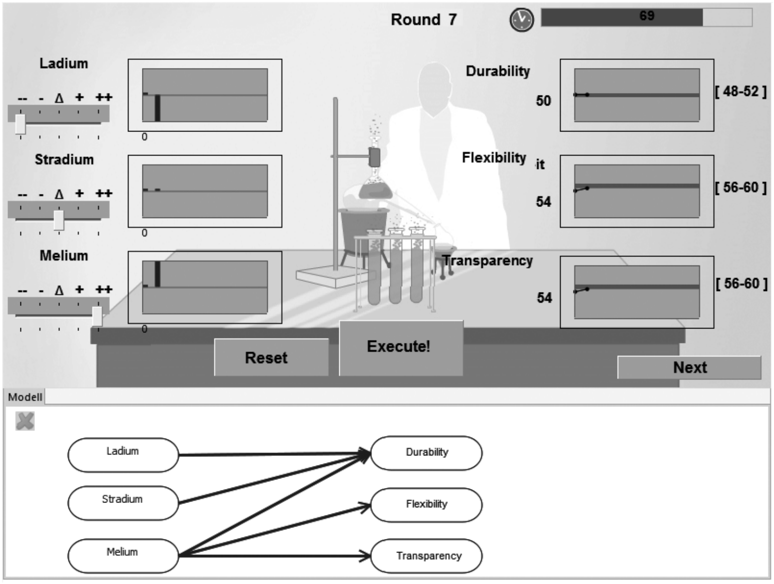

Figure 2. Laboratory item example within the framework of linear structural equation models (cf. Greiff, 2012).

tinguished as well (Greiff, 2012). The correlation between the processes is moderate to high, but considerably too low to suggest identity. IPS as measured via LSE and FSA is correlated with and yet distinct from general intelligence (Kröner et al. 2005; Sonnleitner et al., 2012; Wüstenberg et al., 2012) and from working memory (Bühner, Kröner, & Ziegler, 2008). When incremental validity in assessing external criteria is checked, IPS substantially explains variance beyond measures of intelligence. Wüstenberg et al. (2012) show that 6% of variance in grade point average is additionally predicted when an LSE-based test is used complementing a test of reasoning. Further, Abele et al. (2012) report that IPS is predictive of building up domain-specific knowledge, which, in turn, leads to domain-specific problem solving skills. That is, domain-specific problem solving is largely determined by knowledge in the respective area, which was also shown by the PISA 2003 results on analytical problem solving (Leutner, Klieme, Meyer, & Wirth, 2004). In order to gain this domain-specific knowledge IPS is a significant prerequisite beyond intelligence rendering IPS an essential while indirect factor in solving domain specific problems. In general, IPS exhibits good psychometric characteristics when measured via formal frameworks (Funke, 2001) and general measurement desiderata are taken into account (Greiff, 2012). However, empirical results so far largely originate from two research institutions as the repeated mentioning of some authors above show and the general importance of IPS in a large-scale context has yet to be shown. No doubt, results obtained so far are promising and measuring IPS in PISA 2012 may lead to important insights on how educational systems foster cross-curricular problem solving skills in largely domain-bound classes, but considering that theoretical concepts – even though existing – are now for the first time applied to an assessment perspective and given the scarce empirical evidence compared to other cognitive constructs (e.g., intelligence), it is quite courageous to include IPS at this stage in the PISA survey. At the same time, much less is known about ColPS, which will be assessed in PISA 2015 – an even more challenging enterprise. I will now turn to ColPS and introduce the concept along with potential operationalisations and challenges awaiting test developers.

COLLABORATIVE PROBLEM SOLVING IN PISA 2015

A substantial and rapidly increasing amount of tasks around the world is carried out collaboratively in teams (Brannick & Prince, 1997). That is, not only that analytical non-routine skills performed by individuals are becoming more important but the considerable changes of today's work requirements involve team activities to an extent unthinkable 50 years ago. Surprisingly, research on ColPS – the application of individual problem solving skills in a group context – is scarce at best, which is also reflected in the original description for ColPS issued by the OECD for PISA 2015. There, the importance of ColPS is underlined without further specifying what kind of collaboration and which facets may be of importance. One advantage of the continuity within PISA is that the assessment can draw on previous cycles, for instance by linking items or only slightly altering frameworks. Assessing IPS in PISA 2012 is already pioneering work without links to past surveys, which at least can rely on a decent body of research, whereas this is not the case for ColPS. More specifically, there is research on problem solving as mentioned above and some research on collaboration and group behavior coming from cognitive and social psychology, but both lines have not been merged up to this point in time (cf. Frensch & Funke, 1995; O'Neil, Chuang, & Chung, 2003; OECD, 2010). Thus, at the current state of knowledge only a preliminary understanding of ColPS in PISA 2015 with a link to the assessment of IPS in 2012 is feasible. The few existing definitions of ColPS have in common that (a) they postulate the existence of a group consisting of at least two individuals, (b) they assume that there is a problem and a shared goal, which needs to be achieved, and (c) not only cognitive but also social and communication skills need to be used to solve the problem (e.g., O'Neil et al., 2003; Salas, Dickenson, Converse, & Tannenbaum, 1992).

Attempts to assess ColPS are as rare as comprehensive definitions. They tend to treat problem solving and collaboration as different dimensions each being composed of separate processes. For instance, O'Neill et al. (2003) measure ColPS using the process model of Salas et al. (1992) and the teamwork model published by the Center for Research on Evaluation, Standards, and Student Testing (CRESST; Morgan, Salas, & Glickman, 1993). Based on their broad understanding of ColPS as "problem solving activities that involve interactions among a group of individuals" (O'Neil et al., 2003, p. 361) the authors include adaptability, coordination, decision making, interpersonal skills, leadership, and communication/social regulation as collaborative dimensions and content understanding, problem solving strategies, and self-regulation, as problem solving dimensions into their model. Hsieh and O'Neil (2002) as well as Schacter, Herl, Chung, Dennis, and O'Neil (1999) present ideas on tasks and first empirical results, which underline the difficulties involved in combining collaboration and problem solving within one assessment.

How can this narrow body of research inform the PISA 2015 assessment of ColPS and how could such an assessment in general look like? Currently, the PISA 2015 ColPS Expert Group meets on a regular basis further refining the framework, but no papers have been published so far. Thus, the following section can only give an idea of where the assessment is headed without relying on a published framework.

Generally, the extension of individual IPS to ColPS could assume different shapes. First, in line with O'Neil et al. (2003), additional processes of collaboration could be included by simply adding collaborative dimensions to the process model introduced for IPS (OECD, 2010). For instance, exploring a system, a genuine problem solving dimension, could be preceded by a process of communication, in which leadership skills have to be expressed to derive a common plan, a collaborative dimension. That is, the members of a group would have to direct orders and to derive a shared understanding of how the problem is solved best and only then engage individually in the problem solving process of exploring and understanding (Salas et al., 1992).

However, the assumption of separate processes for problem solving and collaboration taking place consecutively may not be realistic as ColPS involves both collaboration and problem solving at all stages and simultaneously. In other words, a second option would be to embed problem solving into the collaborating group or dyad and not to separate them from each other. To this end, ColPS could be understood as a two-dimensional construct with problem solving being the first and collaboration the second dimension each being composed of different subprocesses. An example of such a framework is displayed in Figure 3. There, the four IPS processes from the PISA 2012 survey and three collaborative processes, leadership, interaction, and communication/social regulation on the basis of O'Neil et al. (2003) are combined into 4x3 cells.

Even though straight out empirical, an assessment of ColPS needs to fill each of these cells with content and then derive items addressing them. More specifically, a comprehensive assessment of ColPS would yield overall measures and measures on each of the 12 cells, but – in this framework – not separately on the different problem solving and collaboration processes. It is important to note that even though cells are assessed separately, this does not necessi-

Collaboration → ↓Problem Solving	Leadership	Interaction	Communication & Social Regulation
Exploring & Understanding		Apply strategies	Assign tasks
Representing & Formulating			
Planning & Executing			
Monitoring & Reflecting			

Figure 3. A suggestion for the understanding of collaborative problem solving in the PISA 2015 assessment. Problem solving processes in rows are combined with collaboration processes in columns. Exemplary, two cells are displayed. Empty cells are to be filled during framework development.

tate separate tasks for each cell as several cells could be assessed within one task.

As an example, let's consider the problem solving process exploring and understanding and the collaborative processes interaction and communication/social regulation. The two resulting cell combinations are displayed in Figure 3 and could be labelled as *apply strategies* and *assign tasks*. That is, in the first cell apply strategies, the problem solvers have to communicate with and convince each other which strategic approach to take concerning the exploration of the system, build consensus about the next steps and about the variables to vary when applying strategies. The second cell, assign tasks is mandatory in group problem solving, whereas non-existent in individual problem solving. More specifically, the problem solvers have to engage into processes of social regulation by clarifying hierarchies, asking each other for help, acting according to their duties, and so on to explore and understand the problem situation. A translation into items could be achieved by extending existing concepts from PISA 2012. For instance, the MP3 Player in Figure 1 is easily upgraded to a collaborative setting: Imagine two MP3 Players controlled by different persons and without a manual available (Figure 4). One specific problem could be to exchange a song via Bluetooth in a joint effort each problem solver having control over one MP3 Player only. Problem solvers have to decide who is doing which part of the problem (e.g., finding out were the Bluetooth is turned on or where the song that is to be exchanged is stored) and which strategies to apply (e.g., searching the MP3 Player guided by a hypothesis or randomly), that is, to assign tasks and apply strategies concerning the exploration of the task thereby assessing some of the cells in Figure 3.

Extending the concept of IPS in PISA 2012 in the way described in order to include collaboration is at the current stage still vague and does not rely on empirical evidence. Apparently, the OECD is well aware of these obstacles and yet wiling to meet them trying to provide an innovative measure of ColPS despite unsettled issues of the construct. There are even more challenges and crucial decisions awaiting framework and test developers when dealing with ColPS (cf. Graesser, Jeon, & Dufty, 2008; O'Neil et al., 2003): Will real interaction between pairs or even larger groups of persons be possible considering practical constraints posed on LSA or will there be human agent interactions simulating group problem solving keeping control of standardization but accepting a loss of external validity? What happens when a person gets matched with a problematic other person, which, for instance, may be mute or overly assertive? Which processes are included into the assessment and how many aspects can be measured in a meaningful way? What are the specific operationalisations? That is, how can concepts be meaningfully translated into items, and how can, in turn, these items be meaningfully translated to individual scores? How to deal with cultural issues and individual communication patterns? Is there a way to reliably score complex sequences of communication or is it necessary to heavily restrict user inputs by using standardized forms of communication? Are there defensible psychometrics to handle the kind of data produced? Will groups be composed of two members keeping the assessment straightforward or will larger groups be included acknowledging that real world interaction is seldom limited to pairs? And, last but not least, what are the technological logistics involved in computer-mediated communication?

A group of experts from different fields busily works on solving these issues juggling with practical, political, and scientific desiderata. Arguably, the PISA 2012 problem solving processes will be maintained but additional collaboration processes will be included, where substantially altered or even completely new items will be employed. Whether this yields in valid results representing important real world challenges remains to be seen. Towards the end

Figure 4. Extension of the MP3 Player example to a collaborative item, which has to be solved in a joint effort.

of this paper, I will now consider some of the challenges and the potential of specific impacts in a broader context.

OUTLOOK

This article introduced IPS and ColPS in the context of the PISA survey, a well-established international LSA, by (a) reviewing the theoretical background behind IPS and ColPS, (b) presenting items and item ideas based on this background, (c) summarizing existing empirical literature to the extent it exists, and (d) outlining obstacles yet to be solved.

Even though of scientific and practical relevance, one could argue that this article comes premature. In fact, it will be several years from now until results on ColPS are available to the public. Having said this, one (a) needs to take into account what happened to the PISA 2003 assessment of analytical problem solving (PISA-Konsortium Deutschland, 2004). There, results were received with great interest, but nobody knew what implications or interventions to derive from them. Teachers and educationalists were literally swamped with findings on analytical problems and argued that – despite its unquestioned importance – this construct was none of their concern and within a few months results just trickled away. One lesson learned from this is that research needs to provide specific recommendations on how to handle a certain pattern of results and what implications they yield once results are published. Another (b) reason to focus IPS and ColPS at this point in time is their mode of delivery, the computer. Both assessment within LSA and schooling will increasingly rely on computers. This, in turn, will face people uncomfortable with technological sophistication, so-called technophobics, with severe problems in a number of contexts including educational settings. According to Weil and Rosen (1995), there is a substantial amount of people with high levels of technophobia across countries, whereas knowledge on development and fostering of technophobia is scarce. Considering the impact of LSA and disadvantages technophobics may experience in the assessment situation, now seems by no means too early and research will have to give answers on (a) how to integrate cross-curricular skills into domain-specific concepts of teaching and on (b) how to comprehensively deal with the matter of technophobia.

No doubt, when PISA 2012 results are published in 2013 and PISA 2015 results in 2016 they will be met with large interest - at least by some participating countries. Interestingly, the public attention given to PISA differs considerably across them. Whereas a public outcry was heard throughout Germany and the United States when the results on PISA 2000 were first published and both countries performed worse than expected, other countries have been taking little notice of the PISA results and their accompanying implications over several cycles. That is, whereas a number of countries dismiss results from PISA as essentially irrelevant, quite a few other countries have used the PISA survey to successfully rejuvenate their educational system and to improve their quality of education. Both, IPS and ColPS have the potential of further adding to the understanding of how educational systems function and offer opportunities to discover which improvements may be needed in a specific educational system. Besides all the unsolved issues mentioned throughout this article, participating countries should not let this opportunity pass by without grasping it.

REFERENCES

Abele, S., Greiff, S., Geschwendtner, T., Wüstenberg, S., Nickolaus, R., Nitschke, A., & Funke, J. (2012). Die Bedeutung übergreifender kognitiver Determinanten für die Bewältigung beruflicher Anforderungen [The role of general cognitive determinants in mastering job requirements]. *Zeitschrift für Erziehungswissenschaft, 15*(2), 363-391.

Baumert, J., Evans, R. H., & Geiser, H. (1998). Technical problem solving among 10-year-old students as related to science achievement, out-of-school experience, domain-specific control beliefs, and attribution patterns. *Journal of Research in Science Teaching, 35*(9), 987-1013.

Baxter, G. P., & Glaser, R. (1998). Investigating the cognitive complexity of science assessments. *Educational Measurement: Issues and Practice, 17*(3), 37-45.

Blech, C., & Funke, J. (2005). *Dynamis review: An overview about applications of the Dynamis approach in cognitive psychology.* Bonn: Deutsches Institut für Erwachsenenbildung.

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

Buchner, A. (1995). Basic topics and approaches to the study of complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 27-63). Hillsdale, NJ: Erlbaum.

Bühner, M., Kröner, S., & Ziegler, M. (2008). Working memory, visual–spatial intelligence and their relationship to problem-solving. *Intelligence, 36*(4), 672–680.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the Control of Variables Strategy. *Child Development, 70*(5), 1098-1120.

Daniel, R. C., & Embretson, S. E. (2010). Designing cognitive complexity in mathematical problem-solving items. *Applied Psychological Measurement, 35*(5), 348-364.

Dörner, D. (1986). Diagnostik der operativen Intelligenz [Assessment of operative intelligence]. *Diagnostica, 32*(4), 290-308.

Dunbar, K., & Fugelsang, J. (2005). Scientific thinking and reasoning. In K. L. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 705-725). New York, NY: Cambridge University Press.

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

Frensch, P. A., & Funke, J. (Eds.). (1995). *Complex problem solving: The European perspective.* Hillsdale: Erlbaum.

Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning, 7*(1), 69-89.

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

Funke, J., & Frensch, P. A. (2007). Complex problem solving: The European perspective. 10 years after. In D. H. Jonassen (Ed.), *Learning to solve complex scientific problems* (pp. 25-47). New York, NY: Erlbaum.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., & Brenwald, S. (2009). *Highlights from TIMSS 2007.* Jessup, MD: Institute of Education Sciences.

Graesser, A. C, Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes, 45*, 298-322.

Greiff, S. (2012*). Individualdiagnostik der Problemlösefähigkeit* [Diagnostics of problem solving ability on an individual level]. Münster: Waxmann.

Holyoak, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 59-87). New York, NY: Academic Press.

Hopmann, S. T., Brinek, G., & Retzl, M. (Eds.). (2007). PISA zufolge PISA [PISA according to PISA]. Berlin: LIT.

Hsieh, H. L., & O'Neil, J. F. (2002). Types of feedback in a computer-based collaborative problem solving group task. *Computers in Human Behavior, 18,* 699-715.

Jürges, H., & Schneider, K. (2006). *Age at school entry and teacher's recommendations for secondary school track choice in Germany.* Accessed at: http://www.vwl.uni-freiburg.de/iwipol/faculty_seminar/age_at_school_entry.pdf

Klahr, D., Triona, L. M., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching, 44*, 183–203.

Klieme, E. (2004). Assessment of cross-curricular problem solving competencies. In J. H. Moskowitz & M. Stephens (Eds.), *Comparing learning outcomes. International assessment and education policy* (pp. 81-107). London: Routledge Falmer.

Köller, O., Baumert, J., & Bos, W. (2001). TIMSS - Third International Mathematics and Science Study. In F. E. Weinert (Ed), *Leistungsmessungen in Schulen* (pp. 269-284). Weinheim: Beltz.

Kröner, S., Plass, J. L., & Leutner, D. (2005). Intelligence assessment with computer simulations. *Intelligence, 33*(4), 347-368.

Leutner, D., Klieme, E., Meyer, K., & Wirth, J. (2004). Problemlösen [Problem solving]. In PISA-Konsortium Deutschland (Eds.), *PISA 2003* (pp. 147-175). Münster: Waxmann.

Markman, A. B. (1999). *Knowledge representation*. Mahwah, NJ: Erlbaum.

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.

Morgan, B., Salas, E., & Glickman, A. S. (1993). An analysis of team evaluation and maturation. *Journal of General Psychology, 120*(3), 277-291.

Nagy, P. (2000). The three roles of assessment: Gatekeeping, accountability, and instructional diagnosis. *Canadian Journal of Education, 25*, 262-279.

Novick, L. R., & Bassok, M. (2005). Problem solving. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 321-349). Cambridge, NY: University Press.

Novick, L. R., Hurley, S. M., & Francis, M. (1999). Evidence for abstract, schematic knowledge of three spatial diagram representation. *Memory and Cognition, 27*, 288-308.

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

Organization for Economic Co-operation and Development. (2009a). *Chapter 1: Programme for International Student Assessment: An overview*. Accessed at: http://www.oecd.org/dataoecd/13/30/48578439.pdf

Organization for Economic Co-operation and Development. (2009b). *PISA 2009 results. What students know and can do.* Accessed at: http://www.oecd.org/dataoecd/10/61/48852548.pdf

Organization for Economic Co-operation and Development. (2010). *PISA 2012 field trial problem solving framework*. Accessed at: http://www.oecd.org/dataoecd/8/42/46962005.pdf

Organization for Economic Co-operation and Development. (2011). *PISA exemplary item MP3 Player*. Accessed at: cbasq.acer.edu.au

PISA-Konsortium Deutschland. (Eds.). (2004). *PISA 2003*. Münster: Waxmann.

Pólya, G. (1971). *How to solve it*. Princeton, NJ: University Press.

Raven, J. (2000). Psychometrics, cognitive ability, and occupational performance. *Review of Psychology, 7*, 51–74.

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

Schacter, J., Herl, H. E., Chung, G. K. W. K., Dennis, R. A., & O'Neil, H. F. (1999). Computer-based performance assessments: A solution to the narrow measurement and reporting of problem solving. *Computers in Human Behavior, 15*, 403-418.

Sjoberg, S. (2007). *PISA and real life challenges: Mission impossible?* Accessed at: http://folk.uio.no/sveinsj/Sjoberg-PISA-book-2007.pdf

Sonnleitner, P., Brunner, M., Greiff, S., Funke, J., Keller, U., Martin,...Latour, T. (2012). The Genetics Lab. Acceptance and psychometric characteristics of a computer-based microworld to assess Complex Problem Solving. *Psychological Test and Assessment Modeling, 54*(1), 54-72.

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.

Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem-solving ability. *Educational Measurement Issues and Practice, 14*(3), 29-36.

Weil, M. M., & Rosen, L. D. (1995). A study of technological sophistication and technophobia in university students from 23 countries. *Computers in Human Behavior, 11*(1), 95-133.

Wirth, J. (2004). *Selbstregulation von Lernprozessen* [Self-regulation of learning processes]. Münster: Waxmann.

Wirth, J., & Leutner, D. (2008). Self-regulated learning as a competence. Implications of theoretical models for assessment methods. *Journal of Psychology, 216*, 102-110.

Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex Problem Solving. More than reasoning? *Intelligence, 40*, 1-14.