

Perspectives on Problem Solving in Educational Assessment: Analytical, Interactive, and Collaborative Problem Solving

Samuel Greiff,¹ Daniel V. Holt,² and Joachim Funke²

Abstract

Problem solving has received broad public interest as an important competency in modern societies. In educational large-scale assessments paper-pencil based analytical problem solving was included first (e.g., Programme for International Student Assessment, PISA 2003). With growing interest in more complex situations, the focus has shifted to interactive problem solving (e.g., PISA 2012) requiring identification and control of complex systems. In the future, collaborative problem solving represents the next step in assessing problem solving ability (e.g., PISA 2015). This paper describes these different approaches to assessing problem solving ability in large-scale assessments considering theoretical questions as well as assessment issues. For each of the three types of problem solving, the definition and understanding of the construct is explained, items examples are shown together with some empirical results, and limitations of the respective approach are discussed. A final discussion centers on the connection of cognitive and differential psychology within educational research and assessment.

Keywords

problem solving, large-scale assessment, PISA, analytical problem solving, interactive problem solving, collaborative problem solving, complex problem solving

¹University of Luxembourg. Please direct correspondence to samuel.greiff@uni.lu.

²University of Heidelberg, Germany.

Introduction

All life is problem solving. This simple title of one of Karl Popper's (1999) later volumes emphasizes the importance of and the frequency with which our daily lives are peppered with small and large problems: a new kind of software introduced at work, road construction blocking our weekly trip to the gym, a difficult interaction with a new colleague, a scientific problem—the list could be extended indefinitely. But what constitutes a problem from a research perspective, and when do we need to apply our problem-solving skills? According to Mayer (2003), a problem occurs when in any given state, a goal state needs to be reached, and there is no routine method of solution available. The subsequent process of transforming the given state into the desired goal state is defined as problem solving (Lovett, 2002) in which a phase of establishing a representation of the problem (knowledge acquisition; Klahr & Dunbar, 1988) is usually followed by the implementation of a solution process (knowledge application; Novick & Bassok, 2005). Within experimental and cognitive psychology, a large body of studies on problem solving has accumulated (cf. Jonassen, 2007; Mayer & Wittrock, 2006). Problems in some domains such as mathematics (e.g., Daniel & Embretson, 2010), the natural sciences (e.g., Dunbar & Fugelsang, 2005), or technology (e.g., Baumert, Evans, & Geiser, 1998) may require domain-specific problem-solving skills (Sugrue, 1995) that are usually considered analytical (i.e., all information needed to solve the problem is available at the outset; Wirth & Klieme, 2003). Besides analytical problem solving in specific domains, problem solving may involve complex general mental processes that are not bound to specific domains (Funke, 2001; Sternberg, 1995). According to Novick, Hurley, and Francis (1999), these general mental processes are important in a number of settings because they result in general and abstract representation schemas, which are more useful for understanding the structure of novel problems because these general schemas are not contaminated by specific content (Holyoak, 1985).

If Popper is correct that problem solving is everywhere in our lives, then independent of the underlying conception of problem solving as domain-specific or general, problem solving as a construct—even though it originated from cognitive and experimental psychology—has high relevance for educational and assessment perspectives in particular. In fact, according to Mayer and Wittrock, enhancing students' problem-solving capacity is one of educational psychology's greatest challenges and is a major demand placed on any educational institution. Bearing this in mind, it is not surprising that educational large-scale assessments (LSAs) around the world have recently identified problem solving as a core domain that complements classical literacy concepts in school subjects. More specifically, one of the most prominent LSAs, the Programme for International Student Assessment (PISA; OECD, 2009), decided to include assessments of problem-solving abilities in 2003, 2012, and 2015. PISA is a cross-sectional

study of 15-year-old high school students across all member states of the Organization for Economic Cooperation and Development (OECD) and a number of associated countries (totaling over 70 participating countries). It is one of the largest educational assessment programs worldwide, testing approximately half a million students in 3-year cycles and reporting average performances on several literacy scales. Thus, it provides an international benchmark that can be used to compare educational systems. In PISA 2003, the assessment of Analytical Problem Solving (APS)¹ was aligned with a number of different disciplines including mathematics, science, commerce, and literature in line with the domain-specific research mentioned above. However, the majority of these problems were located in the areas of mathematics and science. In the PISA 2012 cycle, by contrast, computer-based tests of Interactive Problem Solving (IPS) focusing on domain-general and content-free aspects of problem solving were administered; these were aligned with a more general and less domain-bound understanding of problem solving. As not only complex mental skills such as problem solving, but also teamwork and communication are becoming increasingly important in modern societies (Autor, Levy, & Murnane, 2003), the upcoming PISA 2015 assessment will include measures of Collaborative Problem Solving (ColPS), thus extending the previous cognitive emphasis on the social aspects of problem solving such as interaction and communication by substantially connecting problem solving to the research area of collaborative learning (e.g., Engelmann, Tergan, & Hesse, 2010).

The focus of this paper lies on these different conceptions of problem solving within PISA. In a way, these conceptions represent research efforts from different communities (Domain-Specific and Analytical Problem Solving in PISA 2003, Interactive Problem Solving in PISA 2012, and Collaborative Learning in PISA 2015), which have until now functioned independently of each other and have yielded few interdisciplinary contributions. To this end, we have observed considerable differences in the approaches to problem solving in PISA 2003, 2012, and 2015, albeit they are all housed under the common umbrella of problem solving. By reviewing and reflecting on the three problem-solving concepts and by evaluating them from an integrative perspective, we try to connect cognitive experimental research and educational assessment into a joint and comprehensive understanding, thus bridging the gap between experimental psychology and assessment in education as well as between different types of problem solving. Thus, this paper is not aimed at facilitating a specific theory or definition of problem solving, but rather at showing how a construct such as problem solving can be understood in different ways at different points in time. Specifically, we will review the understanding of problem-solving concepts endorsed in PISA, illustrate the items, and show the potential contribution of relating cognitive problem-solving research to recent contributions from educational large-scale assessments.

Analytical Problem Solving in PISA 2003

Definition and Understanding

In 2003, problem solving was included in the PISA survey for the first time. Before then, PISA had emphasized narrowly defined ability domains related to disciplinary subjects commonly found in school curricula, such as mathematics, sciences, or reading. The motivation behind extending the range of abilities assessed was the recognition that problem solving is an important cross-curricular skill with high real-world relevance. The PISA 2003 framework explicitly stated that: “The processes of problem solving . . . are found across the curriculum” and “educators and policy makers are especially concerned about students’ competencies of solving problems in real-life settings” (OECD, 2003, p. 154). Moreover, an increasing number of empirical studies have suggested that problem solving may represent an ability domain that can be at least partly delineated from basic cognitive ability and from content knowledge in disciplinary domains such as mathematics and science (e.g., Frensch & Buchner, 1999; Leutner, Fleischer, Wirth, Greiff, & Funke, 2012; Wüstenberg, Greiff, & Funke, 2012). Supporting this assumption, the German national option of PISA found that although German students showed average performance in disciplinary assessments, they scored higher in problem solving ability than other countries (Leutner, Klieme, Meyer, & Wirth, 2004).

Although the PISA 2003 framework acknowledged that there is no comprehensive definition of problem solving (cf. Frensch & Funke, 1995), the working definition described problem solving as “an individual’s capacity to use cognitive processes to resolve real, cross-disciplinary situations where the solution path is not immediately obvious” (OECD, 2003, p. 156). The cognitive processes involved were subdivided into two main branches labeled problem-solving processes and reasoning skills. Reasoning represented the ability to draw valid conclusions from given information or to transfer a solution strategy to similar problems. It was broken down further into the domains of analytical, quantitative, analogical, and combinatorial reasoning. The branch of problem-solving processes consisted of additional abilities required for problem solving, such as understanding and representing the problem (knowledge acquisition), finding solutions (knowledge application), reflecting progress, and communicating the results. Problem representation and finding a solution matched the similar distinction made by Novick and Bassok (2005), as described in the introduction. Reflection and communication were added as part of the initial PISA concept; however, they were largely dropped from the actual assessment conducted later on.

Item Examples

The development of item formats for LSAs is not an easy task. Although there is abundant research on problem solving from a cognitive perspective, work on transferring this

research into item formats suitable for psychometric testing has only recently begun. For international large-scale assessments, the requirements are particularly high because items need to be easily and objectively scoreable, data collection must be economical with large samples, and problems need to be culture-fair and easy to translate. Furthermore, as part of the PISA assessment, all 15-year olds need to be able to understand the items independent of the curriculum in which they are enrolled.

For the PISA 2003 survey, the problems were chosen from the areas of decision making, system analysis, and fault finding and were presented in real-life contexts from school and work situations or personal life (OECD, 2003). They were selected mainly to capture the domain of analytical reasoning. The decision problems required the problem solver to choose among transparently presented alternative options, the system-analysis problems to understand the structure of a complex system with various interrelated items, and the fault-finding tasks to find out why a system is not performing as expected using causal understanding. For illustration, Figure 1 shows an example of a fault-finding task. To solve this problem, the problem solver has to integrate the verbal and pictorial aspects from the problem description and form an appropriate causal model of how the pump works in order to diagnose the problem.

Empirical Results

A detailed analysis of the data for analytical problem solving in the national German extension study of PISA 2003 showed that a three-dimensional structure was adequate for describing problem-solving performance for the problems used in that study (Leutner et al., 2012). These three dimensions corresponded to the predefined item types (decision making, analyzing systems, and fault finding).

Limitations and Open Questions

The example in Figure 1 illustrates the real-world nature of the problems used and how they require the integration of different knowledge domains to find a solution. However, it also highlights the short-comings of the pen-and-paper approach used in PISA 2003: In a real-world setting, most problem solvers would have been likely to interactively try out different options (based on hypotheses about how the system works or on trial-and-error; Klahr, Triona, & Williams, 2007; Novick & Bassok, 2005) to see how the system responds. This in turn may have narrowed down the number of possible causes of the problem and guided the search further. For obvious reasons, this kind of dynamic interaction in real-world settings is not possible with pen-and-paper testing. It was therefore suggested in the 2003 framework that computer-based testing might provide new opportunities for introducing this type of problem; thus, computer-based testing is included in the wave of the PISA 2012 assessments and is the subject of the next section. Furthermore, although the PISA framework deliberately emphasized the information-processing perspective on

Figure 1. Example of an analytical problem-solving item (fault finding) as used in PISA 2003. The accompanying questions require the problem solver to, for example, explain how the movement of the valves enables the operation of the bicycle pump or what may be possible reasons for the failure of air to come from the hose.

Jane had some trouble with her bicycle tyre pump yesterday. She repeatedly pulled up and pushed down on the handle of the pump, but no air came out of the hose. She wanted to find out what was wrong, so she looked in the box where the pump was kept and found a piece of paper with the following information on it.

When the handle-piston assembly is pulled up, air passes through the inner valve and fills the space between the piston and outer valve. When the handle-piston assembly is pushed down, the inner valve closes and the piston forces the air beneath the piston out through the outer valve.

Bicycle tyre pump

problem solving and listed a number of processes involved, the study itself made little use of process analyses.

Interactive Problem Solving in PISA 2012

Definition and Understanding

To overcome the conceptual limitations associated with pen-and-paper testing encountered in PISA 2003 and to make use of process data captured by computer-generated log files, one of the major shifts from PISA 2003 to PISA 2012 was the move toward computer-administered interactive problems, for which students can test different ideas for solving the problem in simulated environments. Interactive problem solving² is characterized by the dynamic interaction between a problem solver and the problem to generate and integrate information about the problem. That is, whereas all relevant information is available at the outset in APS, this information needs to be actively generated in IPS. To this end, the PISA 2012 framework states that problem solving takes place:

When encountering real-world artefacts such as ticket vending machines, air-conditioning systems or mobile telephones for the first time, especially if the instructions for use of such devices are not clear or not available. Understanding how to control such devices is a problem faced universally in everyday life. In these situations it is often the case that some relevant information is not apparent at the outset. (OECD, 2010, p. 18)

The move away from Analytical Problem Solving (see previous section) was motivated by the desire to adequately represent the complexity of our modern world and by the opportunity to simulate this complexity offered by computer-based assessment. In fact, computer-based assessment is able to go substantially beyond the pen-and-paper assessments that were employed in PISA 2003. More specifically, one of the sources of complexity is the increase in dynamic and interactive situations in our daily environments (Autor et al.; Funke, 2001; Greiff, 2012). Not only do software interfaces and their rapid change make continuous learning necessary, but also the way that specialized hardware confronts us with complex interactions: Mobile phones, ticket machines, electronic room access, copiers, and even washing machines now require sequences of interactions to set up these devices and to make them run. The common denominator of these examples is that a problem solver needs to actively interact with any kind of technical or nontechnical system, thereby generating the new information that is necessary to proceed successfully toward building a problem representation and carrying out a goal-directed solution process. However, the targeted type of dynamic situation is by no means limited to technical devices and can be extended even to social situations (cf. Collaborative Problem Solving in the next section).

To understand the underlying skill sets that problem solvers need to apply, a detailed understanding of the conception of the problem, how it drives the interactions, and how it places specific demands on the problem solver would be helpful. This, in turn, leads directly to the individual skill sets required to solve a problem: Finding out how the system under question works (i.e., exploration: finding a strategy to build up knowledge; i.e., a representation) and trying to move toward a given goal (i.e., control: applying the acquired knowledge to reach a certain goal; i.e., to solve the problem). Therefore, the two main tasks, knowledge acquisition (goal: representation of the problem space; Klahr & Dunbar, 1988) and knowledge application (goal: solution of the problem; Novick & Bassok, 2005) are found in IPS as well. In fact, knowledge acquisition and knowledge application are apparently the common denominators in all conceptualizations of problem solving presented in this article. However, compared to APS, knowledge acquisition and knowledge application in IPS involve additional dynamic components and take place in interactive environments (Frensch & Funke, 1995). The decomposition of the underlying cognitive processes in PISA 2012 distinguishes four problem-solving processes: exploring and understanding, representing and formulating, planning and executing, and evaluat-

Figure 2. MicroFIN item “MP3 Player” published as an item example of IPS in PISA 2012. By pressing the buttons to the right, the MP3 player changes its state (indicated by the highlighted fields).

en-GB Programme for International Student Assessment 2012

MP3 PLAYER

A friend gives you an MP3 player that you can use for playing and storing music. You can change the type of music, and increase or decrease the volume and the bass level by clicking the three buttons on the player. (▶, ◻, ◀)

Click RESET to return the player to its original state.

Question 1: MP3 PLAYER CP043Q03

The bottom row of the MP3 player shows the settings that you have chosen. Decide whether each of the following statements about the MP3 player is true or false. Select "True" or "False" for each statement to show your answer.

Statement	True	False
You need to use the middle button (◻) to change the type of music.	<input type="radio"/>	<input type="radio"/>
You have to set the volume before you can set the bass level.	<input type="radio"/>	<input type="radio"/>
Once you have increased the volume, you can only decrease it if you change the type of music you are listening to.	<input type="radio"/>	<input type="radio"/>

ing and reflecting. The first two processes can be seen as subcomponents of knowledge acquisition, whereas the other two represent subcomponents of knowledge application.

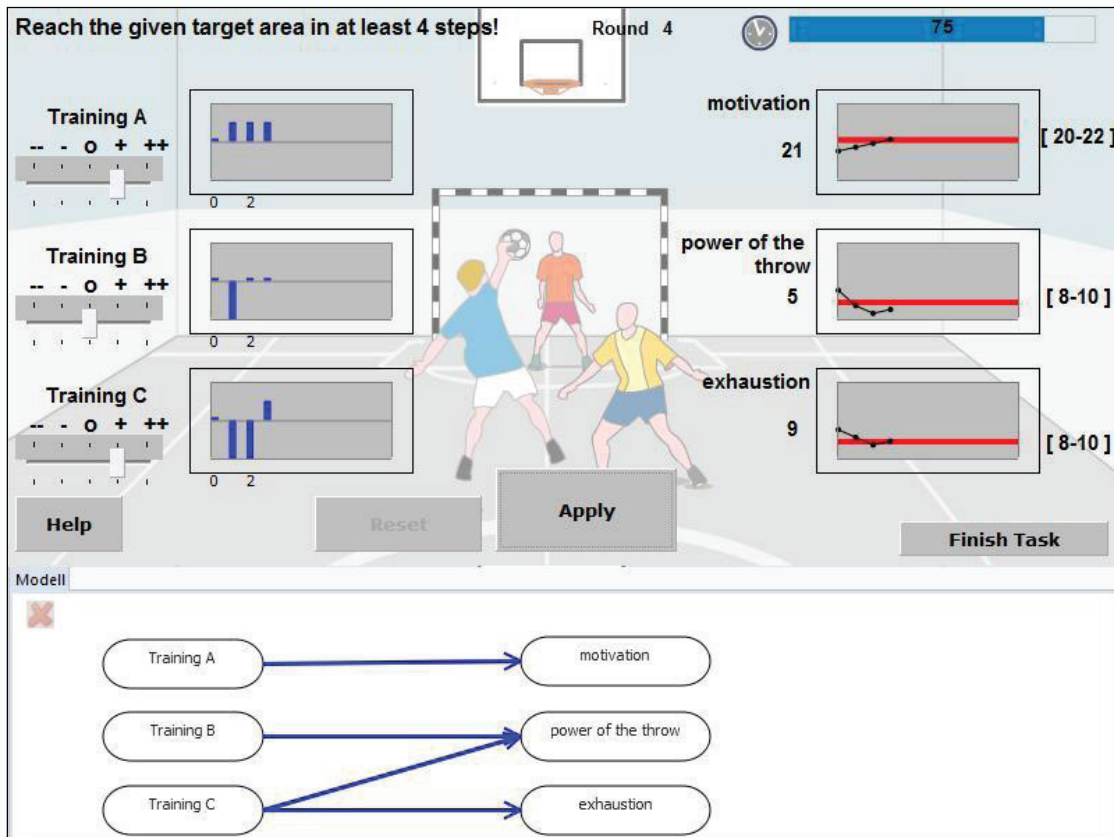
Item Examples

Item examples for interactive problems in line with PISA 2012 are given in Figures 2 and 3, both presuming the use of computers for test administration.

As the generic framework that underlies the item development approach for IPS in PISA 2012, the MicroDYN and MicroFIN approaches (Greiff & Funke, 2009; Greiff, Wüstenberg, & Funke, 2012), based on the formalism of finite state machines and linear equation systems (Funke, 2001), were developed from a psychometric perspective. These two formalisms allow for a systematic construction of problems with varying difficulty and nearly arbitrary semantic embedding, thus enabling the collection of large item samples, which have been used considerably in experimental problem-solving research (Funke, 2001).

PISA employs an entire series of problems as displayed in Figures 2 and 3, consisting of systems that have to be explored within three to four minutes and afterwards controlled to reach given goal states. The main feature of these items is the search for minimally complex systems, that is, systems that at the same time contain all (or at least most) of the features of a complex system (complexity, dynamics, polytely, intransparency; see Funke,

Figure 3. Screenshot of the MicroDYN item “Handball Training.” Problem solvers first have to build a representation and then reach given target values. On the left side: The controllers of the three input variables range from “- -” (value = -2) to “+ +” (value = +2). To the right: The current values of the three output variables are displayed numerically, and the target values of the output variables are displayed graphically and numerically. In the bottom part, the relations between the input and output variables are represented by a causal diagram (Wüstenberg et al., 2012).



2010) but have the lowest values on these parameters. From a psychometrician’s point of view, this ensures the validity of the items and keeps the subjects’ burden of being tested at a minimum level (Greiff et al., 2012).

Empirical Results

Besides data from PISA 2012, with about 200,000 data points to be expected at the end of 2013 (computer-based assessment of 15-year-old pupils from more than 40 countries, with more than 5,000 subjects per country as an optional assessment within PISA 2012), the conceptual delineation of problem-solving ability into representational and solution components has recently been empirically supported using interactive problem-solving tasks (Wüstenberg et al., 2012). Furthermore, studies have shown the capacity of IPS to predict relevant criteria such as academic achievement (Greiff & Fischer, 2013; Schweizer,

Wüstenberg, & Greiff, 2013; Wüstenberg et al., 2012) or supervisor ratings (Danner et al., 2011) beyond measures of intelligence.

Limitations and Open Questions

The shift to IPS comes along with new issues for research, such as analyzing the huge amounts of process data that become available in log files. Whereas for analytical problems in PISA 2003, only the final solution of the problem-solving process was available, the process of exploration becomes an issue for further analyses in interactive problems (e.g., Schulte-Mecklenbeck, Kühberger, & Ranyard, 2010; Zoanetti, 2010). In fact, behavioral and process data of problem-solving patterns are now partly implemented in the PISA scoring procedures and directly connected to the emerging field of educational data mining, in which experimental and psychometric methods are applied to large educational data sets (cf. Rupp, Nugent, & Nelson, 2012).

From a conceptual perspective, one may question whether minimally complex items indeed exhibit sufficient complexity and, therefore, are able to demonstrate all aspects of interactive problem solving equally well. As Fischer, Greiff, and Funke (2012, p. 37) wrote:

In unison with Dörner we want to emphasize that in order to develop a sufficient understanding of the problems humans have to face in their everyday lives, research on problem solving has to further elaborate on complex problems, with both a large amount of possible actions for the problem solver, and a lot of uncertain and surprising consequences in naturalistic environments.

Or, in other words, the complexities of naturalistic environments are sometimes much larger than realized in the IPS items administered in PISA 2012 (this, of course, is true for any assessment). However, the tension between psychometric reliability and external validity—between a psychometrician’s perspective and a phenomenon-driven perspective—cannot be entirely resolved, but has arrived at an acceptable compromise in PISA 2012. Whereas problem solving became interactive in PISA 2012 and moved from a pen-and-paper assessment in 2003 to a computer-based assessment in 2012, in the upcoming PISA cycle in 2015 the practice of problem solving will experience another extension toward the inclusion of problem solving in teams.

Collaborative Problem Solving in PISA 2015

Definition and Understanding

Motivated by the rapidly increasing number of tasks carried out in teams (Brannick & Prince, 1997) and the recently obtained promising results of problem-solving assess-

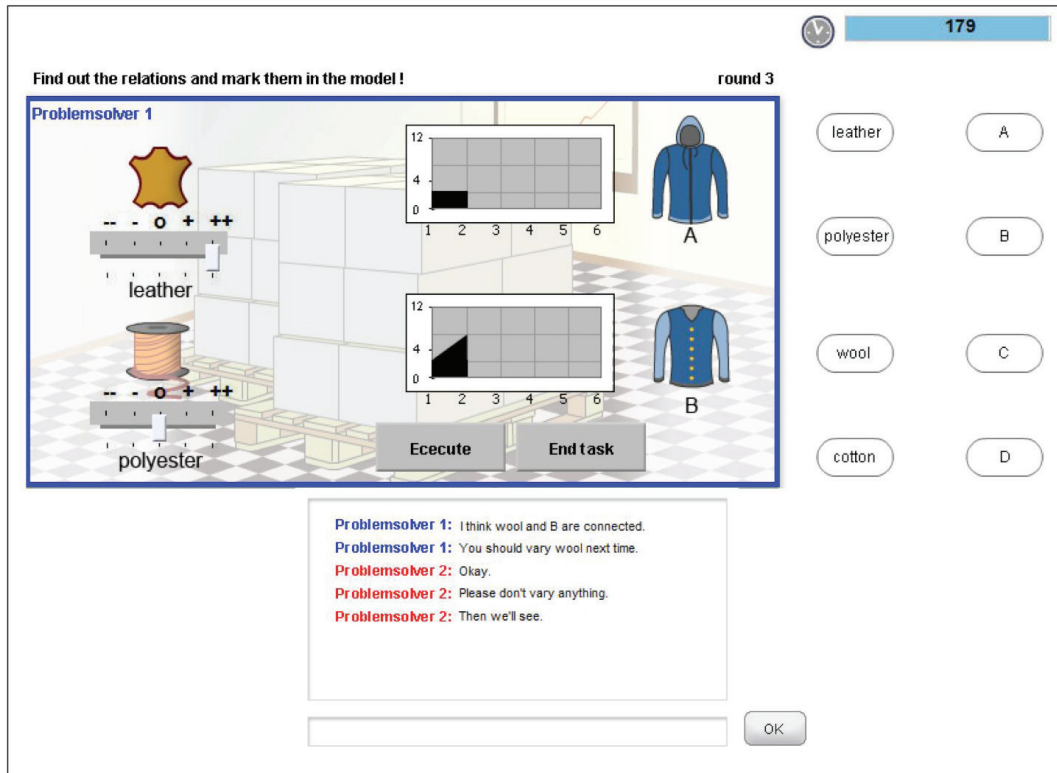
ments from LSAs (e.g., PISA 2003 and 2012), Collaborative Problem Solving as an additional domain will be included in the PISA 2015 survey. By doing so, the interaction between a problem solver and a task—a central feature of IPS for PISA 2012 (OECD, 2010)—will be extended to interactions between several problem solvers. Thus, the steep rise of communicative and team tasks in modern society (Autor et al., 2003) will be acknowledged and Vygotsky's view that there is an inherent social nature to any type of learning or problem solving (Lee & Smagorinsky, 2000) will be incorporated into an international LSA for the first time. In the PISA 2015 assessment framework (OECD, 2012), ColPS is tentatively defined as "the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution" (p. 7). In line with previous efforts to define ColPS (e.g., Griffin, McGaw, & Care, 2011; Morgan, Salas, & Glickman, 1993; O'Neil, Chuang, & Chung, 2003), collaboration and problem solving could be considered to be correlated but sufficiently distinct dimensions. That is, for problem solving, the cognitive processes of IPS in PISA 2012 will still be included (see previous section), whereas a new assessment of social and collaborative skills, which are associated with noncognitive skills (Greiff, 2012), will be added. Although the exact nature of these noncognitive skills has yet to be specified, the understanding of collaboration within the Assessing and Teaching 21st Century Skills initiative (Griffin et al., 2011) constitutes a reasonable starting point. There, participation and cooperation, perspective taking, and social regulation jointly form the collaborative-social dimension of ColPS (Griffin et al., 2011), and the first empirical results indicate that—in principle—these skills may be accessible to measurement (P. Griffin, personal communication, May 2012).

Item Examples

Different types of collaborative settings may elicit different types of behavior, and an LSA with various practical constraints needs to focus on the most essential types of interaction and problem solving. To this end, the psychometric approach initiated in IPS for PISA 2012 (see previous section) is complemented by interaction between problem solvers as shown in Figure 4, in which a potential candidate for a collaborative item is displayed.

Contextualized within a business setting, a problem solver has to understand jacket production at a local factory while a colleague is responsible for production in a second factory. Only by working together and by mutually exchanging individual knowledge (a) can the optimal distribution of resources be explored (exploring and understanding), represented, and communicated (representing and formulating), (b) can the jacket production be optimized (planning and executing), and (c) can a successful business be established (evaluating and reflecting). Whereas these processes, which were borrowed from IPS in PISA 2012 (OECD, 2010), are readily separated during testing—a necessity for maintaining standardized control over the assessment situation and for the one-

Figure 4. Screen mock-up of the collaborative item “Tailorshop”. In the upper middle part, input variables can be manipulated. The model is represented on the right side by a causal diagram. Standardized communication with a virtual agent or a real problem solver is carried out in the lower middle part.



dimensional measurement of single skills—this does not hold for aspects of collaboration: Even though the illustration of communication in Figure 4 is highly standardized, it simultaneously involves aspects of participation and cooperation, perspective taking, and social regulation (Griffin et al., 2011). To this end, the major challenge in any assessment of Collaborative Problem Solving will be to select tasks that can be used to assess specifically targeted aspects of problem solving and collaborative behavior.

Empirical Results

As a comprehensive and widely acknowledged definition of ColPS is currently unavailable, prior research in the area of collaborative learning has focused on team processes (O’Neil, Chung, & Brown, 1997), interaction analyses and team knowledge (Cooke et al., 2003), shared knowledge across teams (Engelmann et al., 2010), or situational group judgement (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001) within experimental settings or conducted assessments based on self-reports (O’Neil et al., 2003) with less concern being placed on the individual assessment issues of collaboration necessitated in LSA. For instance, Engelmann et al. (2010) showed that when spatially

distributed (i.e., computer-supported) collaboration is applied, a heightened awareness of the knowledge and information of the other groups' members reduces the number of interaction problems and increases the quality of the overall result. However, with little experience from a measurement perspective, with primarily qualitative and experimental operationalizations of collaboration at hand, and with a considerably shorter lead time for accumulating relevant findings in comparison to Analytical and Interactive Problem Solving, the assessment in PISA 2015 is well-advised to primarily target problem-solving skills and, additionally, to develop items that carefully extend the demands of problem solving to group settings and their specific requirements without relying too much on the collaborative-social dimension.

Limitations and Open Questions

The introduction of Collaborative Problem Solving as an even more far-reaching extension of classical concepts into PISA reflects the importance that scholars and educationalists attribute to the concepts of problem solving and collaboration in teams. Notwithstanding its relevance, the underlying construct and its assessment have been only vaguely contoured at this point in time. The OECD as the conveyer of PISA seems aware of the issues that are associated with theoretically and empirically delineating ColPS and other constructs (e.g., intelligence and domain-specific problem solving) and is carefully considering the educational and political implications potentially associated with such an assessment. Besides these substantial issues, various conceptual problems need to be solved before ColPS is implemented in PISA 2015. One of them alludes to the question of whether problem solvers should interact with artificially simulated agents (human-agent) or real students located at another computer (human-human). Whereas a broad spectrum of agents could be incorporated into the assessment from a technical perspective and would allow for standardized control over the assessment situation, the external validity of this approach has not been verified. Human-human interactions, on the other hand, are high in face validity, but they are difficult to control and to match in an LSA setting. In the PISA 2012 assessment of IPS, an acceptable compromise between different practical constraints and conceptual considerations was found. For ColPS, a framework that incorporated a number of assessment desiderata was published (OECD, 2012). It considered different constraints (e.g., testing time, technical feasibility, and so forth) that could shape the assessment context (e.g., to ensure sufficient control over the testing situation, a human-agent approach was chosen instead of a human-human approach). Given the complexity of ColPS and the desire to integrate cognitive and social assessments, this consolidation comes along with a number of challenges. Apart from the specific application of ColPS within the PISA survey, the responsibility of the further elaboration of Collaborative Problem Solving and the integration of its disparate desiderata within a comprehensive assessment framework will fall on the researchers active in this area.

Bringing Together Cognitive Research and Educational Assessment

Since PISA 2003, problem solving has come a long way from a pen-and-paper-based assessment of Analytical Problem Solving to a dynamic assessment of Interactive Problem Solving in PISA 2012 and is advancing even further toward an integrative assessment of collaboration and problem solving in PISA 2015. The conceptual and psychometric advances implemented in recent years have motivated this development, but it has also been considerably fostered by technical innovations and by the introduction of computer-based assessments into international LSAs. At the same time, empirical research has yielded a number of studies on the conceptual delineation of different types of problem solving and on the convergent and divergent validity of different problem-solving approaches.

To this end, we need to acknowledge that problem solving is not a consistent field of research even though the definitions of problem solving in PISA have a lot in common. This situation is clearly reflected by the different assessment instruments found in the PISA cycles over the last decade. However, besides the differences mentioned, there is considerable overlap with regard to the cognitive processes that have been targeted (e.g., the notion of knowledge acquisition and knowledge application is found in all conceptualizations of PISA) and with regard to the intention to move beyond the mere assessment of domain-specific abilities in the context of an educational large-scale assessment. To further deepen our understanding of problem solving—be it embedded into a specific content domain (OECD, 2003), as an individual transversal skill (OECD, 2012), or in collaboration with others (OECD, 2015)—further research needs to address the theoretical understanding and the empirical side of problem solving. In order to make some suggestions for this facilitation, we will next describe how bringing together educational assessment and cognitive science, in which problem-solving research is rooted, may benefit both sides and the field of problem solving in general. Originally, research on problem solving emerged in experimental cognitive psychology (cf. Jonassen, 2007), and a strong link between educational assessment and cognitive psychology has yet to be established despite the potentials inherent in such integration. We see several ways in which the cooperation between the disciplines of cognitive psychology and educational assessment can be further extended in the future. For instance, open questions in assessment could be addressed by experimental laboratory studies, whereas log data provided by computer-based assessment in LSAs may prove valuable for understanding cognitive processes and behavioral patterns.

Advantages of Psychometric Studies

Problem solving has long been a staple of experimental cognitive research, and cognitive psychology is therefore in a good position to inform test development by providing problem-solving paradigms, cognitive process models, and detailed task analyses that

may be used in test construction. However, just as test development benefits from the insights of cognitive psychology, the development of problem-solving tasks can be informed by psychometric analysis. For example, Wittman and Süß (1999) used three different computer-based scenarios of IPS and analyzed their experimental data using a structural equation modeling approach. The results showed that the measurement reliabilities of the tests employed were generally low but that a latent variable approach unveiled clear relations between problem solving and reasoning ability that had not been visible before. Whereas we do not share Wittmann and Süß's (1999) specific conclusions that interactive problem-solving skills can be reduced to a combination of reasoning ability and domain knowledge, we agree with the more general point that this area of research will benefit in particular from an integration of experimental psychology and assessment approaches, particularly within large-scale assessments. The method of structural equation modeling and latent modeling in general, which can help to produce a more detailed understanding of what the psychometric components of problem solving are and how they relate to other mental abilities, requires large samples in order to yield reliable results. This is hard to achieve in laboratory-based studies, but large-scale assessments can easily provide a sufficient number of data points, which opens up new avenues for validating assumptions about the structure of the problem-solving process derived with experimental methods.

As an example, one could build on earlier experimental and conceptual work by Dörner (1986) and Funke (2003) and begin with a five-dimensional model for Interactive Problem Solving. These dimensions could consist of system exploration, information reduction, model formation, control considering dynamic change, and prioritization of goals. Although well-founded from a cognitive process perspective, empirical results from educational large-scale assessments resulted in mounting evidence that in fact a two-dimensional description of the problem-solving process (knowledge acquisition and knowledge application in line with Novick & Bassok, 2005) was just as appropriate and even more parsimonious when used as a description of problem-solving ability (e.g., Greiff & Fischer, 2013; Schweizer et al., 2013; Wüstenberg et al., 2012). This finding will in turn guide our future cognitive models of how to understand problem solving from a process perspective.

Advantages of Computer-Based Testing

As described above, the use of computers in testing allows the field to move toward interactive problem solving, involving features such as controlling complex systems or interacting with technological artifacts. It also provides a basis for a controlled approach to collaborative problem solving. Computer-based testing has another advantage, which as of now has not been fully leveraged: Beyond mere summary result scores (i.e., final performance), computer-based testing produces a detailed record of the interaction between the problem solver and the problem, down to the level of single mouse clicks. These data provide a rich trace of the process of problem solving, which in turn may be

used to analyze the strategies used for problem solving by individual problem solvers (e.g., trial-and-error or strategic and analytical approaches), as well as how certain approaches to problem solving are emphasized by different educational systems. To make use of this new source of data, methods for handling the amount of detailed data provided by LSAs will need to be devised. In the simplest case, this can mean an automated test of the presence of particular solution strategies, but more sophisticated approaches using machine learning methods, network analysis algorithms, or cognitive and educational data mining in general may yield further insights into how participants handle specific problems.

New Types of Data Analyses

Whereas structural equation modeling provides insight into the structure of the components of the problem-solving process, item response theory shifts the focus to the level of individual items and their characteristics, which again is interesting from a cognitive and educational perspective (e.g., Embretson & Reise, 2000). For example, using latent class analysis, it is possible to find out whether certain subgroups of participants apply different strategies in solving a problem and how large these classes are. This information is relevant from a diagnostic and educational perspective, but can at the same time advance cognitive process research by uncovering the variety of solution processes that may be involved. So far, this aspect of individual differences in strategies and underlying abilities is rather underrepresented in cognitive research on problem solving, which often assumes one fairly homogeneous approach to problems and largely treats individual differences in strategy as measurement error. The rich data provided by LSAs may help to turn a part of this error into useful information.

Final Implications

In summary, we feel that the cooperation between the different subdisciplines of psychology to produce measurement procedures for large-scale assessments has been successful but limited so far. Problem-solving paradigms were adapted from experimental research for psychometric testing, and basic concepts regarding the cognitive processes involved provided a theoretical basis for item construction. The findings from the application of psychometric assessment in turn have yielded information about the structure of the different cognitive abilities and components of the problem-solving process and have helped to improve the measurement characteristics of problem-solving tasks. In the future, we see the potential to leverage the rich process data generated by computer-based assessments of problem-solving skills and to investigate task characteristics and solution strategies on an even more fine-grained level than has been possible before, thus strengthening the ties between large-scale assessments and cognitive experimental problem-solving research even further and meeting the public demand expressed by politicians, educationalists, and stake holders to learn more about the nature of problem solving.

Acknowledgments

This research was funded by a grant of the German Federal Ministry of Education and Research (LSA004), the German Research Foundation (DFG Fu 173/14-2) and by the European Union (290683; LLL'ight in Europe). We would like to thank two anonymous reviewers for their helpful comments.

Notes

1. Please note that in PISA 2003, the term *problem solving* was officially used. However, in research, the term Analytical Problem Solving is usually used to indicate that all relevant information needed to solve a problem is available at the outset (e.g., Wirth & Klieme, 2003) as in PISA 2003. In this article, we use the term Analytical Problem Solving to describe the assessment of problem solving in PISA 2003.
2. The terms Interactive Problem Solving, Complex Problem Solving, and Dynamic Problem Solving are used synonymously in research (Fischer et al., 2012). In this paper, we consistently use the term Interactive Problem Solving in accordance with PISA terminology.

References

- Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, *118*(4), 1279–333. <http://dx.doi.org/10.1162/003355303322552801>
- 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. [http://dx.doi.org/10.1002/\(SICI\)1098-2736\(199811\)35:9<987::AID-TEA3>3.0.CO;2-P](http://dx.doi.org/10.1002/(SICI)1098-2736(199811)35:9<987::AID-TEA3>3.0.CO;2-P)
- 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: Lawrence Erlbaum.
- Cooke, N. J., Kiekel, P. A., Salas, E., Stout, R., Bowers, C., & Cannon-Bowers, J. (2003). Measuring team knowledge: A window to the cognitive underpinnings of team performance. *Group Dynamics: Theory, Research and Practice*, *7*(3), 179–219. <http://dx.doi.org/10.1037/1089-2699.7.3.179>
- Daniel, R. C., & Embretson, S. E. (2010). Designing cognitive complexity in mathematical problem-solving items. *Applied Psychological Measurement*, *35*(5), 348–64. <http://dx.doi.org/10.1177/0146621609349801>

- Danner, D., Hagemann, D., Holt, D. V., Hager, M., Schankin, A., Wüstenberg, S., & Funke, J. (2011). Measuring performance in dynamic decision making: Reliability and validity of the Tailorshop simulation. *Journal of Individual Differences, 32*(4), 225–33. <http://dx.doi.org/10.1027/1614-0001/a000055>
- 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. 705–25). New York: Cambridge University Press.
- Dörner, D. (1986). Diagnostik der operativen Intelligenz [Assessment of operative intelligence]. *Diagnostica, 32*(4), 290–308.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Engelmann, T., Tergan, S. O., & Hesse, F. W. (2010). Evoking knowledge and information awareness for enhancing computer-supported collaborative problem solving. *Journal of Experimental Education, 78*(2), 268–90. <http://dx.doi.org/10.1080/00220970903292850>
- Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *Journal of Problem Solving, 4*(1), 19–42.
- Frensch, P. A., & Buchner, A. (1999). Domain-general versus domain-specificity: A binary concept and its impact on psychological theory and research. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 137–72). Cambridge, MA: MIT Press.
- Frensch, P. A. & Funke, J. (Eds.). (1995). *Complex problem solving: The European perspective*. Hillsdale, NJ: Lawrence Erlbaum.
- Funke, J. (2001). Dynamic systems as tools for analysing human judgment. *Thinking & Reasoning, 7*(1), 69–89. <http://dx.doi.org/10.1080/13546780042000046>
- Funke, J. (2003). *Problemlösendes Denken* [Problem solving and thinking]. Stuttgart: Kohlhammer.
- Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing, 11*, 133–42. <http://dx.doi.org/10.1007/s10339-009-0345-0>
- Greiff, S. (2012). Assessment and theory in complex problem solving: A continuing contradiction? *Journal of Educational and Developmental Psychology, 2*(1), 49–56. <http://dx.doi.org/10.5539/jedp.v2n1p49>
- Greiff, S. (2012). From interactive to collaborative problem solving: Current issues in the Programme for International Student Assessment. *Review of Psychology, 19*(2), 111–21.
- Greiff, S., & Fischer, A. (2013). Der Nutzen einer komplexen Problemlösekompetenz: Theoretische Überlegungen und empirische Befunde [Usefulness of complex problem solving competency: Theoretical considerations and empirical results]. *Zeitschrift für Pädagogische Psychologie, 27*(1), 1–13. <http://dx.doi.org/10.1024/1010-0652/a000086>
- Greiff, S., & Funke, J. (2009). Measuring complex problem solving: The MicroDYN approach. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 157–163). Ispara, Italy: European Commission.

- Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement, 36*(3), 189–213. <http://dx.doi.org/10.1177/0146621612439620>
- Griffin, P., McGaw, B., & Care, E. (Eds.). (2011). *Assessment and teaching of 21st century skills*. New York: Springer.
- Holyoak, K. J. (1985). The pragmatics of analogical transfer. *The psychology of learning and motivation: Advances in research and theory, 19*, 59–87. [http://dx.doi.org/10.1016/S0079-7421\(08\)60524-1](http://dx.doi.org/10.1016/S0079-7421(08)60524-1)
- Jonassen, D. H. (Ed.) (2007). *Learning to solve complex scientific problems*. New York: Lawrence Erlbaum.
- Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*(1), 1–48. http://dx.doi.org/10.1207/s15516709cog1201_1
- 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*(1), 183–203. <http://dx.doi.org/10.1002/tea.20152>
- Lee, C. D., & Smagorinsky, P. (Eds.) (2000). *Vygotskian perspectives on literacy research: Constructing meaning through collaborative inquiry*. Cambridge, UK: Cambridge University Press.
- Leutner, D., Fleischer, J., Wirth, J., Greiff, S., & Funke, J. (2012). Analytische und dynamische Problemlösekompetenz im Lichte internationaler Schulleistungsvergleichsstudien [Analytical and dynamic problem solving from an international educational studies perspective]. *Psychologische Rundschau, 63*(1), 34–42. <http://dx.doi.org/10.1026/0033-3042/a000108>
- Leutner, D., Klieme, E., Meyer, K., & Wirth, J. (2004). Problemlösen [Problem Solving]. In PISA-Konsortium Deutschland (Ed.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland—Ergebnisse des zweiten internationalen Vergleichs* (pp. 147–75). Münster: Waxmann.
- Lovett, M. C. (2002). Problem solving. In D. Medin (Ed.), *Stevens' handbook of experimental psychology: Memory and cognitive processes* (pp. 317–62). New York: Wiley. <http://dx.doi.org/10.1002/0471214426.pas0208>
- Mayer, R. E. (2003). *Learning and instruction*. Upper Saddle River, NJ: Prentice Hall.
- 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: Lawrence Erlbaum.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*, 730–40. <http://dx.doi.org/10.1037/0021-9010.86.4.730>
- Morgan, B. B., Salas, E., & Glickman, A. S. (1993). An analysis of team evaluation and maturation. *Journal of General Psychology, 120*(3), 277–91. <http://dx.doi.org/10.1080/0021309.1993.9711148>

- Novick, L. R., & Bassok, M. (2005). Problem solving. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 321–49). New York: Cambridge University Press.
- Novick, L. R., Hurley, S. M., & Francis, M. (1999). Evidence for abstract, schematic knowledge of three spatial diagram representations. *Memory and Cognition*, *27*(2), 288–308. <http://dx.doi.org/10.3758/BF03211413>
- O'Neil, H. F., Chuang, S., & Chung, G. K. W. K. (2003). Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education*, *10*(3), 361–73. <http://dx.doi.org/10.1080/0969594032000148190>
- O'Neil, H. F., Chung, G. K. W. K., & Brown, R. S. (1997). Use of networked simulations as a context to measure team competencies. In H. F. O'Neil (Ed.), *Workforce readiness: Competencies and assessment* (pp. 411–52). Mahwah, NJ: Lawrence Erlbaum.
- OECD. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Retrieved from <http://www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/33694881.pdf>
- OECD. (2009). *Chapter 1. Programme for international student assessment: An overview*. Retrieved from <http://www.oecd.org/dataoecd/13/30/48578439.pdf>
- OECD. (2010). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD. <http://dx.doi.org/10.1787/9789264190511-6-en>
- OECD. (2012). *PISA 2015 field trial collaborative problem solving framework*. Presented at the 33rd PISA Governing Board meeting in Tallinn, Estonia.
- Popper, K. R. (1999). *All life is problem solving*. London: Routledge.
- Rupp, A. A., Nugent, R., & Nelson, B. (2012). Evidence-centered design for diagnostic assessment within digital learning environments: Integrating modern psychometrics and educational data mining. *Journal of Educational Data Mining*, *4*, 1–10.
- Schulte-Mecklenbeck, M., Kühberger, A., & Ranyard, R. (Eds.) (2010). *A handbook of process tracing methods for decision research*. New York: Taylor & Francis.
- Schweizer, F., Wüstenberg, S., & Greiff, S. (2013). Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences*, *24*, 42–52. <http://dx.doi.org/10.1016/j.lindif.2012.12.011>
- 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: Lawrence Erlbaum.
- Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem-solving ability. *Educational Measurement Issues and Practice*, *14*(3), 29–35. <http://dx.doi.org/10.1111/j.1745-3992.1995.tb00865.x>
- Wirth, J., & Klieme, E. (2003). Computer-based assessment of problem solving competence. *Assessment in Education: Principles, Policy & Practice*, *10*(3), 329–45. <http://dx.doi.org/10.1080/0969594032000148172>

- Wittmann, W. W., & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem solving performances via Brunswik symmetry. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait and content determinants* (pp. 77–108). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/10315-004>
- Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving: More than reasoning? *Intelligence*, *40*(1), 1–14. <http://dx.doi.org/10.1016/j.intell.2011.11.003>
- Zoanetti, N. (2010). Interactive computer based assessment tasks: How problem solving process data can inform instruction. *Australasian Journal of Educational Technology*, *26*(5), 585–606.