Greiff, Samuel; Fischer, Andreas

Measuring complex problem solving: an educational application of psychological theories

Journal for educational research online 5 (2013) 1, S. 38-58

urn:nbn:de:0111-opus-80196

Journal for Educational Research Online
Journal für Bildungsforschung Online
Volume 5 (2013), No. 1, 38–58

Samuel Greiff & Andreas Fischer

Measuring Complex Problem Solving: An educational application of psychological theories

Abstract
Complex Problem Solving (CPS) is a central topic in modern educational contexts and has received increased interest in educational large-scale assessment studies such as the Programme for International Student Assessment (PISA) and the Programme for the International Assessment of Adult Competencies (PIAAC). Measurement devices up to the present have suffered from a lack of theoretical embedment and low reliability. This article reviews the most important theories of CPS that may be applied to the process of rational test construction. Specifically, the functionalist approach focusing on cognitive processes and the approach of action theory focusing on distinct phases are discussed in their relation to CPS and its assessment. As an example of how to develop a reliable and valid measurement device based on these theories, we propose the development of MicroDYN, which is the operationalization of CPS in PISA 2012.

Keywords
Complex Problem Solving; Functionalism; Action theory; Operative intelligence

Die Messung des komplexen Problemlösens: Eine bildungswissenschaftliche Anwendung psychologischer Theorien

Zusammenfassung
Komplexes Problemlösen hat in den vergangenen Jahren als zentrales Thema in international vergleichenden Bildungsstudien wie dem Programme for International Student Assessment (PISA) oder dem Programme for the International Assessment of Adult Competencies (PIAAC) vermehrt an Bedeutung gewonnen. Demgegenüber steht eine gewisse Vernachlässigung theoretischer

Dr. Samuel Greiff (corresponding author), EMACS unit, University of Luxembourg, 6, rue Richard Coudenhove Kalergi, 1359 Luxembourg-Kirchberg, Luxembourg
e-mail: samuel.greiff@uni.lu

Dipl.-Psych. Andreas Fischer, University of Heidelberg, Department for Cognitive and Experimental Psychology, Hauptstr. 47-51, 69117 Heidelberg, Germany
e-mail: andreas.fischer@psychologie.uni-heidelberg.de

Bezüge sowie einer akzeptablen Reliabilität bestehender Messinstrumente. Dieser Artikel fasst relevante Theorien des komplexen Problemlösens zusammen und zeigt deren Potential im Rahmen eines rationalen Testkonstruktionsprozesses. Dabei werden neben dem funktionalistischen Ansatz, der die Bedeutung kognitiver Prozesse hervorhebt, handlungstheoretische Ansätze mit ihrem Fokus auf separaten Handlungsphasen vorgestellt und in ihren Bezügen zu komplexem Problemlösen diskutiert. Als Beispiel, wie ein reliables und valides Messinstrument vor dem Hintergrund dieser Theorien entwickelt werden könnte, wird das Testverfahren MicroDYN und dessen Entwicklung vorgestellt, das zugleich die Operationalisierung komplexen Problemlösens für PISA 2012 darstellt.

Schlagworte
Komplexes Problemlösen; Funktionalismus; Handlungstheorie; Operative Intelligenz

1. Complex Problem Solving competency

Developing competency in Complex Problem Solving[1] is one of the often claimed but rarely implemented goals of education. As a result measuring this kind of competency is of increasing interest to modern educational assessment (see OECD, 2010). A *complex problem* arises, whenever (a) a person wants to achieve certain goals in a situation that is complex (i.e., containing many highly interrelated elements, see Dörner, 1989) and (b) the causal structure of the situation is not sufficiently known to the person (Fischer, Greiff, & Funke, 2012). In order to solve complex problems the problem solver has to (a) build a parsimonious and viable representation of the most important elements and relations, and to (b) search for a solution based on the representation of the problem (see Novick & Bassok, 2005). Examples for complex problems include (but are not limited to) managing a Tailorshop (Danner et al., 2011) or a Sugar Factory (Berry & Broadbent, 1984), Fire Fighting (Brehmer & Allard, 1991) or Fighting Epidemics (Badke-Schaub, 1993), or understanding and controlling complex technical devices (Buchner & Funke, 1993; Klahr & Dunbar, 1988; Leutner, Wirth, Klieme, & Funke, 2005).

In research on CPS, problems of this kind are usually simulated on computers to examine the processes employed in solving complex problems. As Funke (2001) emphasized, the problems simulated in CPS research differ markedly from the tasks of traditional intelligence tests (and analytical problem solving tasks) with respect to a variety of features.

1 Other terms referring to the same concept as Complex Problem Solving (hereafter CPS), but emphasizing different aspects, are frequently found in the literature, for instance "Interactive Problem Solving" (OECD, 2010), "Dynamic Problem Solving" (Wirth & Funke, 2005), or "Dynamic Decision Making" (Gonzalez, Lerch, & Lebiere, 2003).

> (a) The complexity of the situation and (b) the connectivity between a large number of variables forced the actors to reduce a large amount of information and anticipate side effects; (c) the dynamic nature of the problem situation required the prediction of future developments (a kind of planning) as well as long-term control of decision effects; (d) the intransparency (opaqueness) of the scenarios required the systematic collection of information; (e) the existence of multiple goals (polytely) required the careful elaboration of priorities and a balance between contradicting, conflicting goals. (Funke, 2001, p. 72)

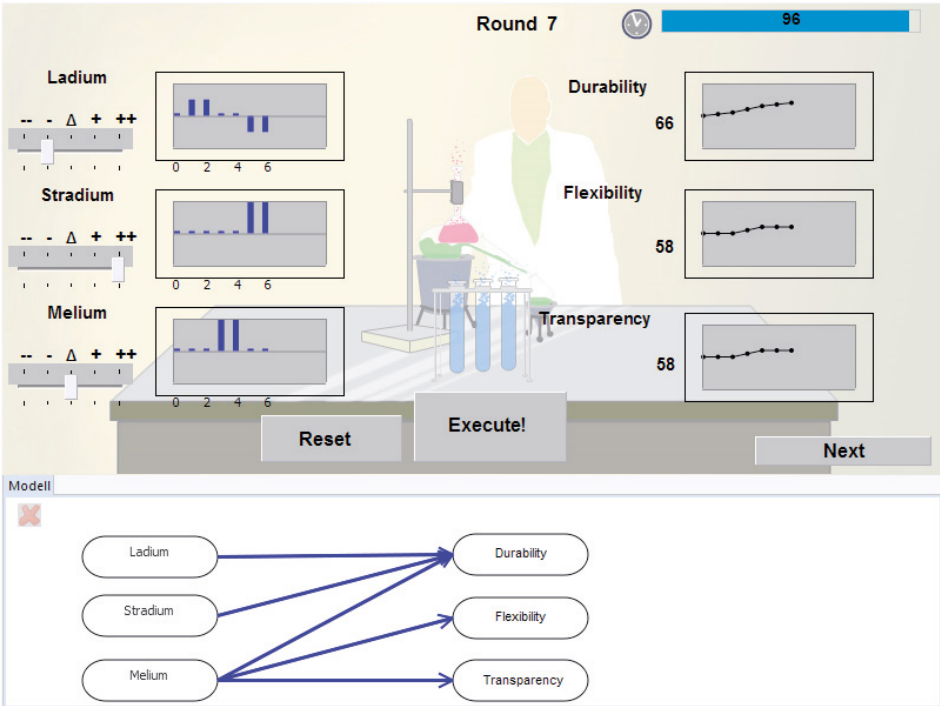
These characteristic features of complex problems are important in a number of educational and occupational contexts. Acquiring and applying knowledge of complex interactions is a major characteristic for problem solving in the natural sciences. In geography or historical science students often have to understand complex matters in order to pass their exams. All of these problems involve "interaction [of the problem solver] with a new system to discover rules that in turn must be applied to solve the problem" (OECD, 2010, p. 15). And although domain-specific prior experience and background knowledge may heavily influence how a problem is represented (Novick & Bassok, 2005), this kind of knowledge is not sufficient to represent or to solve complex problems (Greiff & Fischer, 2013). *Domain-specific* prior knowledge can influence hypothesis building, but it is assumed that there are *domain-general* problem solving strategies and procedures for testing hypotheses (i.e., for representational change) or for searching a solution as well (Fischer et al., 2012; Greiff & Fischer, 2013). The article at hand is about the major theories on these general aspects of CPS, and about how they could be applied to deductive theoretical test construction (Kaplan & Saccuzzo, 2008) in order to develop measurement devices for assessing the most important aspects of CPS competency on a sound theoretical basis.

1.1 CPS in educational assessment

Since tests for CPS competency were introduced in large-scale assessments such as the "Programme for International Student Assessment" (PISA) 2003 (see Leutner, Klieme, Meyer, & Wirth, 2004) the construct proved empirically to be a fruitful extension compared to other competencies such as analytical problem solving competency, general intelligence, reading literacy, science literacy, and mathematics literacy (see Leutner et al., 2005). Wüstenberg, Greiff, and Funke (2012) found CPS competency to be incrementally valid to intelligence with regard to the prediction of school grades ($\Delta R^2 = .06$). Due to the empirical benefit of CPS in PISA and its importance to educational and occupational contexts, a growing interest in an assessment point of view on CPS emerged also in other areas of competency assessment such as the "Programme for the International Assessment of Adult Competencies" (PIAAC) (see Reeff, Zabal, & Blech, 2006).

An illustration of a task for testing important aspects of CPS competency (an item of a MicroDYN task, used as operationalization for CPS in PISA 2012) is depicted in Figure 1. In this task the testee has to explore the linear relations between six quantitative variables. He or she can vary the values of the input variables on the left side of the screen and is shown the resulting changes in the output variables on the right side of the screen after clicking the "Execute"-button. Knowledge about the system structure has to be drawn into a model at the bottom of the screen. Here the testees are to indicate the relations between three compounds and their influences on three outcome measures. After this phase of knowledge acquisition (the phase ends due to a time-out or due to a click on the "Next"-button on the right side of the screen), the testee has to apply the knowledge in order to reach certain goal values for each output variable by varying the values of the input variables.

Figure 1: Exploration in a typical MicroDYN task



1.2 Shortcomings of former research

Even if there are measurement devices that seem to be appropriate operationalizations of CPS at first sight – from managing a Tailorshop (Danner et al., 2011) to landing a Space Shuttle (Leutner et al., 2005) – up to now most of these operationalizations (even the CPS measurement devices used in large-scale assessments

such as the Space Shuttle, Leutner et al., 2005) have merely been ad-hoc constructions taken from experimental research containing time-consuming single items often with low reliability (Funke, 1983) and with a lack of sufficient theoretical embedment of the construct (Greiff, 2012). This has lead to certain problems.
As research on CPS has been predominantly interested in experimental between-group comparisons (i.e., the "item side of the data", De Boeck & Wilson, 2004) and less concerned about the within-group variance or stability (i.e., the "person side of the data", De Boeck & Wilson, 2004), there were large and time-consuming single items called "microworlds" (Kluge, 2008) instead of psychometric tests containing *multiple items*. This kind of "one-item-testing" (Greiff, 2012) often resulted in low reliability coefficients as well as low correlations of CPS tasks with each other (e.g., Wittmann & Süß, 1999) and with external criteria. Even in the case of high reliabilities it has been difficult to determine if measurement devices were addressing (a) the same aspects of CPS, (b) different aspects of CPS or (c) totally different constructs due to their lack of theoretical embedment.

As Dörner (1986) emphasized, the complex structure of the CPS process has to be considered in detail before different aspects of CPS competency can be measured reliably. A proper theoretical embedment would not only allow for measuring global performance in particular complex problems (which might depend on different factors in different operationalizations) but also for a differentiated assessment of facets, which are considered relevant for effective and efficient CPS in general (Greiff, 2012). The call for considering the "person side" of complex problem solving processes in more detail was proposed long ago (see Dörner, 1986), but promising attempts to develop reliable tests based on a sound theoretical basis have been rather scarce (for an overview see Klieme, Funke, Leutner, Reimann, & Wirth, 2001). This is insofar astonishing as there are a lot of functionalist and action theories that may have been used to inform rational test construction regarding the processes and phases most important for CPS.

The article at hand wants to summarize some of the most important theories on CPS that may be used to develop measurement devices for CPS competency. After considering functionalist and action theory approaches in some detail, outlining their overlap and their different contributions to CPS we will propose the development of MicroDYN (the operationalization of CPS in PISA 2012) as an example of how to apply cognitive theories in an educational assessment context in order to produce a reliable and valid assessment instrument for different facets of CPS.

2. Functionalist approach

From the functionalist perspective, mental states are considered to be functional processes, which means they are characterized only by their causal role within a system, relating certain events outside the system (inputs) to corresponding behaviors of the system (outputs) and/or changes of the functional state (see Putnam, 1967). For example, a human being is assumed to employ a certain mental process when it reacts to the sensation of seeing an apple (input) with the behavior of eating the apple (output), and it is most probable that eating an apple will change the mental state (so seeing another apple may lead to a different behavior).

Functionalist theories on human problem solving attempt to explain the complex processes and relations between input and output of human problem solvers using many highly detailed assumptions about the information processing involved. From an assessment point of view these assumptions may shed light on which processes are important for CPS (Fischer et al., 2012) and how these processes may be evoked in a testing situation. Within functionalist theories the human being is often viewed as an analogue to a computer receiving certain input and processing the information in order to return a certain output. The mental domain is generally considered to be a functional organization of a system for processing information (Gadenne, 1996). The mental functions between input and output are described using processes such as attention, perception, storage, transformation, memory organization, and information retrieval. A central concept in functionalist theories on problem solving is the "problem space", a concept proposed by Newell, Shaw, and Simon (1958) in their considerations about a "General Problem Solver". A problem space is assumed to contain (1) the initial state of the problem, (2) a set of operators transforming given states into new states, (3) a set of possible states of the problem, (4) the desired goal-states, as well as (5) additional knowledge available to the problem solver (see Newell & Simon, 1972). All of these aspects are connected with each other on a process level and the problem space determines how the problem can be transformed and when the transformation is finished. Problem solving can be defined as searching the problem space in order to transform the initial state into a goal state (Dunbar, 1998). Many researchers elaborated on the concept of problem space and proposed a set of interacting sub-spaces differing mainly in the entities that are searched for, in order to describe more complex problems such as rule induction (Simon & Lea, 1974), or scientific discovery (Klahr & Dunbar, 1988). These approaches are summarized in the following sections in order to develop ideas of how to measure important aspects of CPS.

2.1 The dual space model

In their dual space model, Simon and Lea (1974) conceptualized problem space as divided in an instance space (containing concrete states of the problem at hand) and a rule space (containing possible combinations of the values and attributes of the objects in the instance space). To illustrate the concept of instance space and rule space, let us assume the owner of a sugar factory wants to maximize the production P of sugar by hiring a certain number of workers W (Berry & Broadbent, 1984). The owner may repeatedly hire a certain number of workers and look for the resulting *instance* of the variables describing his problem (e.g., $W = 600$ workers produce P = 12,000 sugar) in order to see what works and what does not (cf. Instance-Based Learning Theory; Gonzalez, Lerch, & Lebiere, 2003). According to the dual space model, he may also search for a *rule* describing the relation between these variables (e.g., $P_t = 2*W_{t-1}-P_{t-1}$). Of course, the instances generated may influence which rule is considered plausible, and the current rule may influence which instance is looked at next. The usefulness of *instance knowledge* and *rule knowledge* may depend on the size of the system that has to be controlled and the processes necessary to do so. For example, the more complex a system, the less useful it may be to memorize and process single specific instances and the more valuable it may be to focus on general rule knowledge (Fischer et al., 2012; Schoppek, 2002; Vollmeyer, Burns, & Holyoak, 1996). Klahr and Dunbar (1988) applied this dual space approach to the complex process of scientific discovery and to the way scientific experiments are conducted in order to test hypotheses. They assumed an experiment space (corresponding to the instance space) and a hypothesis space (corresponding to the rule space) and postulated a detailed process model of scientific discovery assuming three main components: (1) Searching the hypothesis space for a fully specified hypothesis, (2) testing hypothesis via generating an experiment appropriate to the current hypothesis, and (3) evaluating evidence leading to the acceptation, rejection, or continued consideration of the current hypothesis.

From an assessment point of view both the different kinds of knowledge (instance knowledge and rule knowledge) and the processes involved in the corresponding search may be of interest for measuring CPS competency. Instance knowledge about a complex system could be measured by recognition tasks (e.g., "Is 2-3-2 a possible state of the system?") and rule knowledge could be measured by asking questions about the causal structure between variables (see Funke, 2001). Also systematic use of strategy (Vollmeyer et al., 1996) may be object to evaluation (e.g., it could be evaluated if the testee has systematically varied one variable at a time in order to generate instances and test hypotheses; see Tschirgi, 1980).

Vollmeyer et al. (1996) proposed that setting a specific goal (such as "set the output variable at a value of 2") provokes increased search of the instance space, whereas an unspecific goal (such as "find out about the relations between the variables") stimulates increased processing within the rule space. A similar distinction was made by Klahr and Dunbar (1988) who reported a preference of their subjects for either searching the experiment space (the "experimenters") or the hypothesis space (the "theorists"). Thus, the way goals are formulated in test instruction may influence the kind of processes that are most likely to occur.

2.2 Appreciation of the functionalist approach

The functionalist approach spawned a multitude of fruitful theories and new findings about CPS. Its value for developing CPS measurement devices should not be underestimated. Their assumptions about processes and search spaces involved in CPS could be used to inform test development. In order to test processes relevant to these two spaces, it seems to be appropriate to consider the findings of Vollmeyer et al. (1996), for example by designing problems, consisting of both an exploration and knowledge acquisition process (with unspecific goals) and a separate control process (with specific goal values) regarding one single underlying problem (e.g., Funke, 2001). Yet a lot of research has to be done in order to describe all of the operations and processes taking place within each problem space and their conditions as well as all the interactions possible between problem spaces in detail. Up to now the functionalist approach has not addressed the full amount of the processes important for CPS such as coping with multiple goals (polytely) or the dynamics of systems (Funke, 2003). Nevertheless, it appears that functionalist approaches offer interesting starting points for the development of CPS measurement devices that will be accounted for when we propose the MicroDYN approach in the last sections of this paper.

3. Action theory approach

Action theory has its roots in the considerations Aristotle stated in his "Nicomachean ethics" and was articulated by the sociologist Max Weber (1913). Actions are motivated and intended, embedded in semantic contexts, comprehensible, and goal oriented. They can be *understood* by knowing their reasons, whereas reactive behavior can only be *explained* by knowing the physical and/or physiological causes (Dilthey, 1894). Because action theory emphasizes the importance of goals for a large subset of human behavior, it is of great importance to the psychology of problem solving. This is true especially for CPS, where the existence of multiple goals is a characteristic feature (Dörner, Kreuzig, Reither, & Stäudel, 1983; Funke, 2001).

By specifying the major subgoals of solving complex problems, action theory allows for understanding CPS in terms of different phases. Each subgoal defines a characteristic phase of the overall CPS process consisting of certain types of cognition or behavior (such as hypothesis formation, planning and decision making, and so forth). This separation of subsequent phases was often criticized for being normative or artificial. For example, Dörner and Wearing (1995) pointed out that many persons switch between phases in ways that are difficult to predict. But even if a functionalist *explanation* of an action theory's *understanding* may be possible and desirable (Dörner, 1998; Gadenne, 2004), it is not a necessary condition for applying an action theory of CPS in psychometric contexts. From a psychometric point of view it may even be advantageous to artificially separate the characteristic phases of CPS in order to measure processes relevant for each phase separately (e.g., there could be a phase of hypothesis formation corresponding to a search of the rule space, separated from a phase of planning and decision making corresponding a search of the instance space). Another promising approach to cope with the normativity within action theories was proposed by Dörner and Wearing (1995). They calculated probabilities for the possible transitions between phases and found that these probabilities differed in a systematic way. Successful problem solvers differed from less successful participants, because the former asked less questions about the system's state (such as "Is there a theatre in town?" or "How much is the income tax?") but changed more often to (a) exploration of relations and (b) decision making (Rollett, 2008). From an assessment point of view, the problem solver could be allowed to ask certain questions (about the values of variables and about relations between variables) or to make certain decisions (setting values for the input variables). After the problem solver has chosen among these two options several times, the relative frequencies of transitions to exploration of relations and to decision making may be used as an indicator for good problem solving in different phases.

3.1 Action theory of CPS

One of the major accounts to understanding and explaining the phases of CPS was the *action theory* of CPS proposed by Dörner (1989). Dörner and Wearing (1995) have elaborated on the main components of this action theory and offered a computational *functionalist explanation* of the CPS phases (cf. PSI-theory; Dörner, 1998). The theory may thus be an adequate starting point for identifying the components most relevant for CPS. Dörner and Wearing (1995) identified six characteristic phases for CPS. As a first step of solving a complex problem, there has to be (1) *goal elaboration*, where specific and concrete goals are formulated, and contradictory goals are balanced. While exploring the most important aspects of the system (2) *hypothesis formation* concerning the system's structure is carried out (based on prior knowledge; cf. Novick and Bassok, 2005, and on the data collected by actively intervening in the system; cf. Causal Learning; Buehner & Cheng, 2005). If a problem solver has to cope with a complex system under time pressure, he or she most probably will build *reductive hypotheses*. That is, only the most important effects are considered. Viable reductive hypotheses (a) are easy to handle, (b) provide clear information about how to act, (c) and the pieces of a reductive hypothesis may be true – even if the hypothesis as a whole is incomplete (Dörner & Wearing, 1995). On the basis of these hypotheses (3) *prognosing of the system's dynamics* (the changes in value concerning its variables) takes place. Based on these prognoses, a phase of (4) *planning and decision-making* follows. After the decision has been made, there is (5) *monitoring of the consequences* (the system may change due to the decision of the problem solver or independent of the problem solver's action; i.e. due to "eigendynamics"). From time to time, his or her own information processing can be object to monitoring, in an act of metacognitive (6) *self-reflection* (cf. Wirth & Leutner, 2008).

The theory of Dörner and Wearing (1995) addresses some of the most important phases of CPS. Instead of assuming a correct or complete mental representation of a problem at the beginning of problem solving (that can often be assumed for simple problems, but would overcharge human working memory in the case of CPS; cf. Fischer et al., 2012) the problem solver is assumed to actively build parsimonious models of the complex problem's structure (*hypothesis-formation*) and its dynamics (*prognosing*) before he or she starts planning and decision making. Accordingly, the problem solver could be asked about his representation of the structure (e.g., "Identify the variables with an effect on the output variables!" or "Draw all the connections existing between the variables!") or the dynamics (e.g., "How will the values of the output variables be in two time steps when no further intervention takes place?"). The problem solver's ability to efficiently reduce information about the complex problem (i.e., to build viable *reductive hypotheses*) could be assessed by asking him or her to identify (a) the variables most relevant for the regulation of certain output variables, (b) the variable with the most connections to other variables, or (c) the connection with the highest path coefficient. Another interesting aspect of the theory is the phase of self-reflection, in which the problem solver is assumed to think about his or her own way of building hypotheses, gathering information or making decisions. These general metacognitive capabilities may be assessed via questionnaires (Wells & Cartwright-Hatton, 2004). Wirth and Leutner (2008) referred to the processes involved in planning, executing and evaluating one's own learning processes as self-regulated learning and summarized methods for assessing this kind of competency.

Based on his action theory of CPS, Dörner (1986) elaborated on the *operative* aspects of intelligence that are necessary to efficiently cope with the major demands of complex problems. They are assumed to be (1) generation and integration of information, (2) elaboration and balancing of goals, (3) making plans and decisions, as well as (4) self-management (aspects normally labeled as "flexibility", "foresight", "circumspection", "systematic behavior", and so forth; Dörner, 1986, p. 290). This conceptualization of intelligence in a complex problem solving situation is called *operative intelligence*, as it emphasizes not only speed and accuracy of information processing, but also operative aspects of efficient CPS across different phases:

> Intelligence in a problem solving situation turns out to be being able to collect information, to integrate and structure information goal-oriented, to make prognoses, to plan and to make decisions, to set goals and to change them. To achieve all this, an individual has to be able to produce an organized series of information processing steps, flexibly adapting these steps to the demands of the situation, then it is intelligent. (Dörner, 1986, p. 292; translated by the authors)

These considerations of Dörner (1986, 1989; Dörner & Wearing, 1995) are especially appealing to rational test construction concerning CPS, as they are explicit about different phases that have not yet been accounted for by traditional intelligence tests, and that may efficiently be tested using complex dynamic scenarios (Dörner, 1986). Dörner's action theory (e.g., Dörner & Wearing, 1995) gives a picture of the characteristic phases of CPS, but has some shortcomings with regard to the cognitive processes involved. Combined with some of the functionalists' ideas presented above, his action theory may be an adequate starting point for developing a measurement device for important aspects of CPS competency.

3.2 Appreciation of the action theory approach

According to Funke (2003), the central concept of intention, elaborated on in the action theory approach, appears to enrich problem solving theories that usually assume goals and intentions to be a defining part of problems. Action theory describes different phases of problem solving (usually the number of phases ranges between 4 and 6) on a general level, marking important phases without specifying concrete structural units of the problem solving process or their conditions, whereas functionalist theories target cognitive processes involved in CPS on a more fine grained level. From an assessment point of view, each of the phases in action theory can be assumed to require different cognitive abilities of the problem solver (for instance, elaborating on goals may put other demands on the problem solver than executing a plan). Each of the phases could be used to derive items that measure corresponding and important facets of problem solving competency. On the other hand, it is important to notice that the phases proposed by action theories, in contrast to functionalist approaches, are usually rather normative and not as much explanatory. Individual problem solvers are known to switch between phases in ways that are hardly predictable (see Dörner & Wearing, 1995, p. 70; Schaub & Reimann, 1999, p. 171). Consequently, when assessing CPS based on action theory approaches, it seems reasonable to address this issue for example by designing items that artificially split the CPS process into different phases. For each phase there could be corresponding tasks (e.g., based on the ideas proposed in the last sections). The rather integrative approaches of Dörner and Wearing (1995) and Dörner (1986) allow for referring to the broad empirical basis of both functionalist and action theorist ideas, within *one* coherent theoretical framework. In the last section of this article we will elaborate on how an illustration of a test for different facets of CPS can be derived based on the theories outlined above – coherent with the ideas proposed in the sections on functionalist and action theory approaches.

4. Measuring Complex Problem Solving competency

The competency to explore, understand, and regulate complex dynamic systems is of growing importance to modern educational assessment contexts. We have outlined the most important theories of CPS and we have developed ideas about how they may inform test construction. In order to demonstrate how the ideas presented above could be implemented in an actual measurement device of CPS competency, we will now propose how MicroDYN (Greiff, 2012; Greiff & Funke, 2010), as the operationalization of CPS competency in PISA 2012, was developed on the basis of the theories proposed above. After reviewing the theories of CPS, a deductive theoretical approach (Kaplan & Saccuzzo, 2008) seems possible and appropriate rather than following an inductive empirical approach. Deductive test construction begins with a definition of the construct and its facets. Then, for each facet there has to be a homogenous set of items. Statistical item analysis may be used to ensure the homogeneity and discriminant power of the items.

4.1 A definition of the construct

CPS consists of (a) building a viable representation and (b) finding a way to reach certain goals (Novick & Bassok, 2005) in a situation that contains a large number of elements that are highly interrelated. The two phases of (a) building a viable representation and (b) finding a way to reach certain goals can be defined through the eyes of both the *functionalist approach* and *action theory*. According to the *functionalist approach* of Vollmeyer et al. (1996), which views CPS on a more detailed level, building a viable representation provokes increased search for *rule knowledge*, whereas finding a concrete solution provokes increased search for *instance knowledge*. Following the *action theorist* perspective of Dörner (1986) a complex problem solver has to be capable of (1) systematically gathering information, of (2) integrating the most relevant information, and of (3) building a mental model of the system structure (in order to represent the problem in a parsimonious and viable way; Fischer et al., 2012). Furthermore the person has to be capable of (4) making prognoses, plans, and decisions, and of (5) setting and balancing goals (in order to find a solution; Fischer et al., 2012). Funke (2001) has related these phases of CPS competency (or operative intelligence) to the characteristic features of complex problems: (1) intransparency of the situation, (2) complexity of the structure, (3) interconnectedness of variables, (4) eigendynamics, and (5) multiple and/or ill-defined goals.

Thus, the conceptualization of CPS competency as five characteristic approaches involved in the search for representation and solution (see Table 1) nicely integrates the considerations of different theoretical aspects.

Table 1: The five phases of MicroDYN and their relation to the five characteristic features of CPS within the processes of representation and solution

Cognitive process	Characteristic feature	MicroDYN phase
Representation	Intransparency of the situation	Information generation
	Complexity of the structure	Information reduction
	Interconnectedness of variables	Model building
Solution	Polytely of the task	Goal elaboration and balancing
	Dynamics of the system	Prognosing, planning and decision making

A test for CPS competency should consist of multiple CPS tasks. For each task there should be at least one indicator per facet of CPS competency in order to assess all relevant facets in each problem solving task. Following Funke (2001) CPS tasks can be formulated within the framework of linear structural equation systems. When each task is designed to be a complex system, multiple CPS tasks can be combined to form a test for all the facets of CPS competency. This approach of measuring CPS competency using *multiple* tasks, each formalized within the framework of *linear* structural equation systems as a *complex system*, is called MicroDYN (Greiff, 2012; Greiff & Funke, 2010). The next sections will elaborate on both the tasks and the indicators used within MicroDYN.

4.2 An infinite pool of CPS tasks

Funke (1987, 2001) proposed the formal framework of linear structural equation systems for describing complex problem tasks. In his research, he identified (a) *knowledge acquisition* (corresponding to problem representation) and (b) *knowledge application* (corresponding to the search for a solution) as two main demands on the problem solver (Funke, 2001). Inspired by the *functionalist approach* and the concept of problem spaces as well as by *action theory* and the concept of characteristic phases, he suggests targeting CPS tasks mainly in two stages. Firstly, a problem solver, who is only shown the values of input and output variables (but not the underlying structure of the system), has to specify a series of input values in order to identify the system's structure (the problem solver may draw his or her model of the causal structure between the variables in a causal diagram). Secondly, the problem solver has to specify a series of input values in order to reach given target values (see Figure 2 for an example within a MicroDYN task). In this phase there is a specific goal, whereas in the first part there is the unspecific goal to explore the system. Following Vollmeyer et al. (1996) this is considered to provoke increased exploration of rule space in the first phase and increased exploration of instance space in the second phase with regard to the cognitive processes targeted.

Figure 2: Knowledge application in a typical MicroDYN task containing three input variables (on the left side of the screen), three output variables (on the right side of the screen), target ranges for each output variable (within the edged brackets), and a model of the system's structure (on the bottom of the screen)

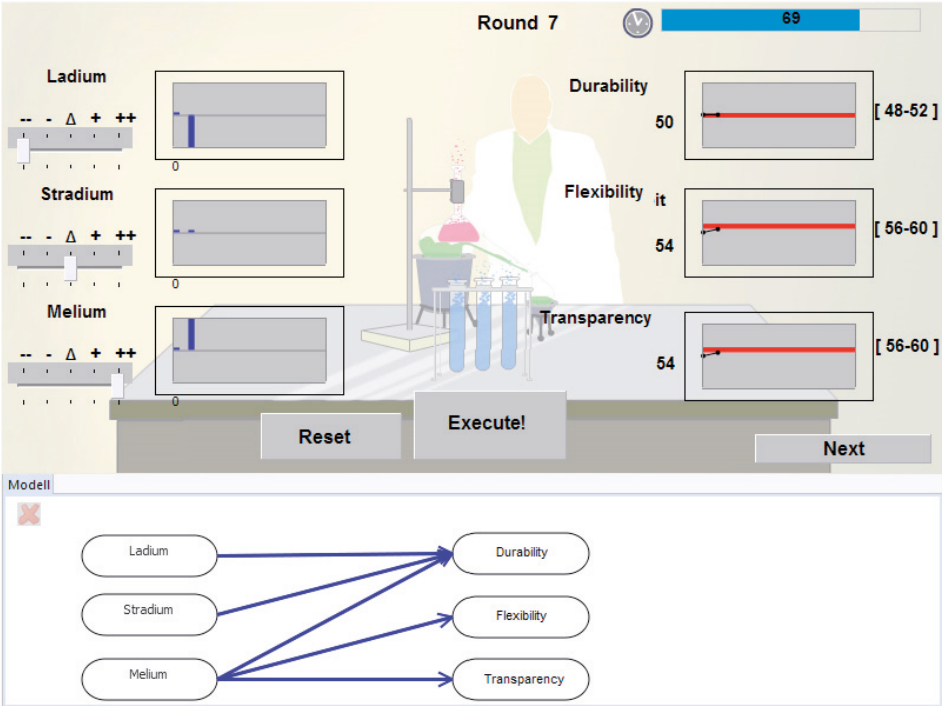
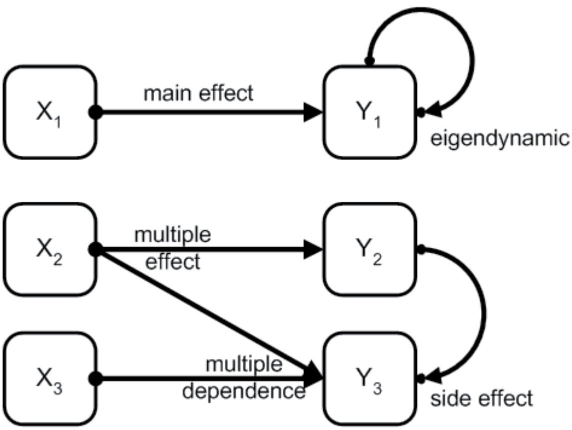


Figure 3: Different kinds of effects possible in a system containing three input and three output variables (cf. Greiff, 2012)



The framework of linear structural equation systems defines an infinite pool of CPS tasks with each task having only one optimal solution. Within these tasks different kinds of effects can be distinguished (see Greiff, 2012). Relations between input and output variables are called main effects, relations between different output variables are called side effects, and a relation of an output variable to itself (a special kind of side effect) is called an eigendynamic. In complex systems, there are usually multiple effects and multiple dependencies per variable. See Figure 3 for an example of the different kinds of effects.

In the next section we will show how to develop indicators for the facets of CPS competency based on the linear structural equations just described (Funke, 2001).

4.3 Five aspects of CPS

Based on the considerations proposed throughout this paper indicators for each facet of CPS competency can be developed. Up to now, Greiff (2012) has empirically validated first indicators for the three facets "information generation" (based on the log-files, systematic series of inputs are quantified), "model building" (the correctness of the causal diagram drawn by the testee is evaluated) and "making prognoses, plans, and decisions" (the difference between target values and achieved values is quantified for all values of output variables). For all three facets, reliable indicators were proposed in Greiff (2012). Specifically, "information generation" was scored in three categories. A systematic variation of input variables (i.e., only single inputs are changed and the others remain constant; cf. Figure 3) combined with idle rounds (i.e., no inputs are changed at all), which were helpful for detecting side effects and eigendynamics, reflected a high level of systematic exploration and were given full credit. If only systematic variation was applied without idle rounds, partial credit was assigned and nonsystematic variation of input variables resulted in no credit being assigned. For the two other facets, "model building" and "making prognoses, plans, and decisions" a binary score (correct and incorrect) proved to be the most reliable indicator. However, the choice of a specific indicator and a specific algorithm to derive this indicator is flexible in the MicroDYN tasks, which is not bound to a specific way to reflect performance. In fact, a number of different measures have been applied in CPS research over the last years (for an overview cf. Greiff, 2012) and depending on the purpose at hand any one of them can be applied in the approach described in this article.

Valid indicators for the other two facets of CPS competency ("information reduction" and "goal elaboration and balancing") are the object of current development and empirical validation. For instance, the problem solver could be asked (a) to identify the most relevant – or the least relevant – variable in the system before he or she has to draw the causal diagram (information reduction), or (b) to decide which goals to aim at in a situation where target values cannot be achieved for every variable (goal elaboration and balancing).

Regarding the indicators for “information generation”, “model building” and “making prognoses, plans, and decisions”, Greiff (2012) reports evidence for separating different dimensions of CPS competency and for its external validity in educational contexts. For example, a three-dimensional model (CFI = .98; TLI = .99; RMSEA = .06) based on data from a university student sample fitted the data significantly better than a one-dimensional model (CFI = .94; TLI = .95; RMSEA = .10), whereas correlations between the indicators were high ($.75 <= r <= .77$), but still reliably different from 1 (Greiff, 2012). On the other hand, some recent research has indicated that “information generation” is directly related to “model building” indicated by a very high latent correlation ($r > .90$) between the two facets (e.g., Schweizer, Wüstenberg, & Greiff, 2013; Wüstenberg et al., 2012). Thus, a systematic way to explore a complex problem may be directly related to establishing a correct mental representation of the problem and the question whether the theoretically derived phases are also separable on an empirical basis requires further research efforts.

With regard to external validity, “model building” predicted school grades ($r = .64$; $p < .01$) in Greiff (2012) and the two facets of the complex problem Space Shuttle (Leutner et al., 2005) also correlated with the corresponding facets of MicroDYN ($r >= .39$; $p < .05$). CPS assessed through MicroDYN proved to be reliably measured (Cronbach’s $\alpha > .80$) and clearly separable from intelligence, as CPS accounted for an additional amount of variance in predicting school grades ($\Delta R^2 = .06$; $p < .05$; Wüstenberg et al., 2012). A similar pattern was reported by Schweizer et al. (2013). There, CPS was separable from working memory and again accounted for additional variance in predicting school grades to a significant extent. This pattern of results (i.e., separability of CPS to other cognitive constructs and incremental validity beyond them when predicting relevant criteria such as academic achievement or supervisory ratings) has been replicated in a number of studies (e.g., Danner et al., 2011; Greiff & Fischer, 2013; Greiff et al., in press) using a heterogeneous set of samples drawn from different populations overall supporting the empirical validity of MicroDYN. In summary, MicroDYN is based on the action theory of CPS with different phases (Dörner, 1986, 1989; Dörner & Wearing, 1995), and the dual space model of Klahr and Dunbar (1988), and seems to allow for measuring three of the most important aspects of operative intelligence in a reliable and valid way by using multiple linear structural equation systems (Funke, 2001) as CPS tasks (Greiff, 2012). These aspects are embedded into the cognitive processes of knowledge acquisition and knowledge application which were elaborated within the *functionalist* approach. In this, MicroDYN combines the general point of view taken by *action theory,* which postulates different phases in CPS and the *functionalist approach,* which informs us on the cognitive processes taking place during the CPS phases.

5. Conclusion and outlook

This article wanted to demonstrate how theories on CPS could be applied in order to develop measurement devices for the competency to solve complex problems. We (1) reviewed theories on CPS from an assessment point of view, (2) explicated some of the most important implications and ideas for measuring CPS on a theoretical level and (3) outlined the development of MicroDYN as an example of how to apply CPS in educational assessment. Tests such as MicroDYN may allow for a proper assessment of students' capabilities to understand and apply knowledge of complex matters, a competency which is important in most modern school subjects (consider, for instance, the complex causes of historical events, the complexity of political systems and their functioning, the laws of physics, and so forth). As acquiring and applying knowledge of complex systems is important to a broad range of problems in daily life, it is of considerable importance to develop solid measurement devices that can be applied in educational contexts and in large-scale assessments such as PISA or PIAAC.

As our review of theories has shown, there are a lot of possibilities to assess the different aspects that may be considered relevant for CPS (and a lot of aspects that could be considered relevant as well). MicroDYN, which itself is still object to further development and validation, should not be conceived as the only possible solution to this issue. Each of the theories proposed above may inform the development for relevant indicators or whole measurement devices. However, MicroDYN may be seen as prove for the feasibility to develop measurement devices for CPS on a theoretical basis and for the potential to combine theoretical approaches on different levels of abstraction into a comprehensive assessment instrument. By further developing MicroDYN in line with empirical results, and by doing empirical research on the theories MicroDYN was based on, MicroDYN is a good example of how test development and empirical research on CPS could stimulate each other.

In contrast to classical intelligence tests (which are primarily concerned with the quality and speed of basal cognitive abilities; see Dörner, 1986) complex problems propose a different set of characteristic demands and requirements that have to be met by the problem solver (see Table 1). Therefore, CPS is an interesting construct, which is important to psychometricians (Dörner, 1986; Greiff, 2012; Funke, 2001). However, only time will tell if CPS proves to be a stable and valuable construct that is going to be further developed and applied in competency assessments and educational contexts in the near future.

Nevertheless, it seems to be an important step in the research on CPS to develop reliable and valid measurement devices with sufficient theoretical embedment, that may both (a) be adapted to the newest empirical findings of the field, and (b) allow for additional research on the theories it was based on.

Acknowledgements

This research was funded by grants supported by the German Research Foundation (DFG; Fu 173/14-2), the Fonds National de la Recherche Luxembourg (ATTRACT „ASSKI21“), by the European Union (290683; LLLight in Europe), and by the German Federal Ministry of Education and Research (01JG1062 & LSA004).

References

Badke-Schaub, P. (1993). *Gruppen und komplexe Probleme: Strategien von Kleingruppen bei der Bearbeitung einer simulierten Aids-Ausbreitung* [Groups and complex problems: Strategies of small groups working with a simulated Aids epidemic]. Frankfurt: Peter Lang.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quaterly Journal of Experimental Psychology, 36A*, 209–231.

Brehmer, B., & Allard R. (1991). Real-time dynamic decision making: Effects of task complexity and feedback delays. In J. Rasmussen, B. Brehmer, & J. Leplat (Eds.), *Distributed decision making: Cognitive models for cooperative work* (pp. 319–334). Chichester: Wiley.

Buchner, A., & Funke, J. (1993). Finite state automata: Dynamic task environments in problem solving research. *Quarterly Journal of Experimental Psychology, 46A*, 83–118.

Buehner, M. J., & Cheng, P. W. (2005). Causal learning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 143–168). Cambridge: University Press.

Danner, D., Hagemann, D., Holt, D., Bechthold, M., Schankin, A., Wüstenberg, S., & Funke, J. (2011). Measuring performance in a Complex Problem Solving task: Reliability and validity of the Tailorshop simulation. *Journal of Individual Differences, 32*, 225–233.

De Boeck, P., & Wilson, M. (Eds.). (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York, NY: Springer.

Dilthey, W. (1894). Ideen über eine beschreibende und zergliedernde Psychologie [Ideas concerning a describing and analyzing psychology]. *Sitzungsberichte der Königl. Preußischen Akademie der Wissenschaften zu Berlin, 2. Halbband 1894, pp. 1309-1407*. Berlin.

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

Dörner, D. (1989). *Die Logik des Misslingens. Strategisches Denken in komplexen Situationen* [The logic of failure. Strategic thinking in complex situations]. Hamburg: Rowohlt.

Dörner, D. (1998). *Bauplan für eine Seele* [Blueprint for a soul]. Reinbeck: Rowohlt.

Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (1983). *Lohhausen: Vom Umgang mit Komplexität* [Lohhausen: About coping with complexity]. Bern: Huber.

Dörner, D., & Wearing, A. T. (1995). Complex Problem-Solving: Towards a (computersimulated) theory. In J. Funke & P. Frensch (Eds.), *Complex Problem Solving – The European perspective* (pp. 65–99). Hillsdale, NJ: Erlbaum.

Dunbar, K. (1998). Problem solving. In W. Bechtel & G. Graham (Eds.), *A Companion to Cognitive Science*. London: Blackwell.

Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *Journal of Problem Solving, 4*, 19–42.
Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor? [Some statements about problems of problem solving research: Is test intelligence a predictor?]. *Diagnostica, 29*, 283–302.
Funke, J. (1987). Experimentelle Analyse des Umgangs mit computersimulierten dynamischen Systemen [Experimental analysis of handling complex dynamic systems]. In U. Schaarschmidt (Ed.), *Neue Trends in der Psychodiagnostik. Band 1: Grundlagen und Aufgaben, Testtheorie und Testmethodik, Diagnostik und Experiment* (pp. 162–168). Berlin: Psychodiagnostisches Zentrum der Humboldt-Universität Berlin.
Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning, 7*, 69–89.
Funke, J. (2003). *Problemlösendes Denken* [Problem solving thinking]. Stuttgart: Kohlhammer.
Gadenne, V. (1996). *Bewußtsein, Kognition und Gehirn. Einführung in die Psychologie des Bewußtseins* [Consciousness, cognition and brain. Introduction to the psychology of consciousness]. Bern: Huber.
Gadenne, V. (2004). Intentionalität, Bewusstsein und die anti-reduktionistische Bewegung [Intentionality, consciousness and the anti-reductionist movement]. In C. Annerl (Ed.), *Philosophie der Psychologie* (pp. 72–103). Bern: Huber.
Gonzalez, C., Lerch, F. J., & Lebiere, C. (2003). Instance-based learning in dynamic decision making, *Cognitive Science, 27*, 591–635.
Greiff, S. (2012). *Individualdiagnostik der Problemlösefähigkeit* [Individual diagnostics of problem solving competency]. Münster: Waxmann.
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*, 1–13.
Greiff, S., & Funke, J. (2010). Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal komplexer Systeme [Systematic research on complex problem solving competency using minimal complex systems]. *Zeitschrift für Pädagogik, 56* (Beiheft), 216–227.
Greiff, S., Wüstenberg, S., Molnar, G., Fischer, A., Funke, J., & Csapo, B. (in press). Complex Problem Solving in educational settings – Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*.
Kaplan, R. M., & Saccuzzo, D. P. (2008). *Psychological testing: Principles, applications, and issues*. Belmont, CA: Wadsworth.
Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1–48.
Klieme, E., Funke, J., Leutner, D., Reimann, P., & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Resultate aus einer Schulleistungsstudie [Problem solving as cross-curricular competency. Conception and first results of a study on school performance]. *Zeitschrift für Pädagogik, 47*, 179–200.
Kluge, A. (2008). Performance assessments with microworlds and their difficulty. *Applied Psychological Measurement, 32*, 156–180.
Leutner, D., Klieme, E., Meyer, K., & Wirth, J. (2004). Problemlösen [Problem solving]. In M. Prenzel, J. Baumert, W. Blum, R. Lehrmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost, & U. Schiefele (Eds.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (pp. 147–176). Münster: Waxmann.

Leutner, D., Wirth, J., Klieme, E., & Funke, J. (2005). Ansätze zur Operationalisierung und deren Erprobung im Feldtest zu PISA 2000 [Approaches to operationalization and its testing in the field trial to PISA 2000]. In E. Klieme, D. Leutner, & J. Wirth (Eds.), *Problemlösekompetenz von Schülerinnen und Schülern* (pp. 21–36). Wiesbaden: VS.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem-solving. *Psychological Review, 65*, 151–166.

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

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.

OECD – Organisation for Economic Co-operation and Development. (2010). *PISA 2012 problem solving framework (draft for field trial)*. Paris: OECD.

Putnam, H. (1967). The nature of mental states. In D. M. Rosenthal (Ed.), *The Nature of Mind* (pp. 197–203).

Reeff, J.-P., Zabal, A., & Blech, C. (2006). *The assessment of problem solving competencies. A draft version of a general framework*. Bonn: Deutsches Institut für Erwachsenenbildung.

Rollett, W. (2008). *Strategieeinsatz, erzeugte Information und Informationsnutzung bei der Exploration und Steuerung komplexer dynamischer Systeme* [Strategy usage, generated information and used information in the exploration and control of complex dynamic systems]. Münster: LIT.

Schaub, H., & Reimann, R. (1999). Zur Rolle des Wissens beim komplexen Problemlösen [On the influence of knowledge to complex problem solving]. In H. Gruber, W. Mack, & A. Ziegler (Eds.), *Wissen und Denken. Beiträge aus Problemlösepsychologie und Wissenspsychologie* (pp. 169–191). Wiesbaden: Deutscher Universitäts-Verlag.

Schoppek, W. (2002). Examples, rules, and strategies in the control of dynamic systems. *Cognitive Science Quarterly, 2*, 63–92.

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.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 105–127). Hillsdale, NJ: Erlbaum.

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51*, 1–10.

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science, 20*, 75–100.

Weber, M. (1913). Über einige Kategorien der verstehenden Soziologie [On some categories of an understanding sociology]. *Logos. Internationale Zeitschrift für Philosophie der Kultur, 4*, 253–294.

Wells A., & Cartwright-Hatton, S. (2004). A short form of the Metacognitions Questionnaire: Properties of the MCQ-30. *Behaviour Research and Therapy, 42*, 385–396.

Wirth, J., & Funke, J. (2005). Dynamisches Problemlösen: Entwicklung und Evaluation eines neuen Messverfahrens zum Steuern komplexer Systeme [Dynamic problem solving: Developing and evaluating a new measurement device for controlling complex systems]. In E. Klieme, D. Leutner, & J. Wirth (Eds.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (pp. 55–72). Wiesbaden: VS.

Wirth, J., & Leutner, D. (2008). Self-regulated learning as a competence. Implications of theoretical models for assessment methods. *Journal of Psychology*, *216*, 102–110.
Wittmann, W. W., & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge and Complex Problem Solving: Perfomances via Brunswik-symmetry. In P. L. Ackermann, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait and content* (pp. 77–108). Washington, DC: American Psychological Association.
Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex Problem Solving. More than reasoning? *Intelligence, 40*, 1–14.