Doctoral thesis submitted to
the Faculty of Behavioural and Cultural Studies
Heidelberg University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy (Dr. phil.)
in Psychology

Title of the thesis

Assessment of Problem Solving Skills by means of Multiple Complex Systems – Validity of Finite Automata and Linear Dynamic Systems

presented by
Dipl.-Psych. Andreas Fischer

year of submission
2015

Dean: Prof. Dr. Klaus Fiedler
Advisor: Prof. Dr. Joachim Funke

Table of content

Acknowledgements

First of all, with regard to this thesis, I would like to thank my supervisor Prof. Dr. Joachim Funke, who provided me the opportunity to write this thesis the way I did (in the BMBF project Fu 01JG1062). I deeply appreciate the helpful advice and the creative freedom Joachim offered me throughout all these years. Joachim was always open-minded and ready to discuss each idea objectively, which was deeply inspiring and very helpful in phases of doubt. I also would like to thank my friends and colleagues Dr. Daniel Holt and Dr. Sascha Wüstenberg, who managed to share an office with me for several years each and who contributed indirectly to this thesis by the countless fruitful discussions that helped me clarifying ideas and concepts related to this thesis. Of course this list will always be inconclusive for there are many friends and colleagues who supported me in one way or the other – nevertheless I would like to thank them all. Last but not least – this goes without saying – I am also very grateful to my beloved wife Saskia, my children and the rest of my family for supporting me, believing in me and for making all this possible.

Summary

The assessment of highly domain-general problem solving skills is increasingly important as problem solving is increasingly demanded by modern workplaces (e.g., Autor, Levy, & Murnane, 2003) and increasingly present in international large-scale assessments such as the Programme for International Student Assessment (PISA, e.g., OECD, 2014). This thesis is about the computer-based assessment of problem solving skills based on Multiple Complex Systems (MCS, Greiff, Fischer, Stadler, & Wüstenberg, 2014): The main idea of the MCS approach is to present multiple computer-simulations of “minimally complex” problems (Greiff, 2012) in order to reliably assess certain problem solving skills. In each simulation, the problem solver has to interact with a problem in order to find out (a) how to adequately represent the problem, and (b) how to solve the problem. Up to now, two instances of the MCS approach have been proposed: (1) the MicroDYN approach (based on simulations of linear equation systems) and – more recently, in the second paper of this thesis – (2) the MicroFIN approach (based on simulations of finite state machines). In the current thesis I will elaborate on three research questions regarding the validity (cf. Bühner, 2006) of the MCS approach: (1) its *content validity* with regard to the concept of complex problem solving; (2) the *convergent validity* of different instances of the MCS approach; (3) the *discriminant validity* of the interactive problems of the MCS approach with regard to traditional static measures of reasoning and analytic problem solving skills. Each research question will be addressed in one corresponding paper:

In a first paper (Fischer, Greiff, & Funke, 2012) complex problem solving is defined as the goal-oriented control of systems that contain multiple highly interrelated elements. After reviewing some of the major strands of research on complex problem solving (e.g., research on strategy selection, information reduction, intelligence, or on the interplay of implicit and explicit knowledge in the process of complex problem solving) a theoretical framework outlining the most important cognitive processes involved in solving complex problems is derived. The theoretical framework highlights both interactive knowledge acquisition (*problem representation)* and interactive knowledge application (*problem solution)* as the two major phases in the process of complex problem solving. Both phases are represented in all current instances of the MCS approach.

In a second paper (Greiff, Fischer et al., 2013) the convergent validity of MicroDYN and MicroFIN is investigated (thereby introducing MicroFIN as an alternative to MicroDYN) in order to demonstrate that both instances address the same kind of problem

solving skills. Based on a multitrait-multimethod analysis of a sample of university students (*N* = 339) it is demonstrated that – in addition to method-specific skills – both instances assess a common set of skills (method-general traits) related to (1) representing and (2) solving different kinds of interactive problems. In a regression of science grades on reasoning and the skills assessed by the instances of the MCS approach it is demonstrated that only the method-general *representation* trait and reasoning have substantial unique contributions. Thus, MicroDYN and MicroFIN seem to address a common set of skills and *this* set of skills is relevant for explaining school grades in science classes even beyond reasoning.

In a third paper (Fischer et al., in press) the discriminant validity of the *interactive* MicroDYN test is investigated by relating it to reasoning and traditional *static* measures of Analytic Problem Solving skills (APS) as they were applied in PISA 2003 (OECD, 2004). Besides a common core of problem solving skills addressed by both kinds of tasks (e.g., analyzing complex information about the information given at a certain moment in time) Fischer et al. (in press) expected to find evidence for additional skills that were related to interactive problems only (e.g., systematically generating information and interactively testing hypotheses). Results indicate that MicroDYN shares a lot of variance with APS even after controlling for reasoning in a sample of high-school students (N = 577) and the university student sample (see above). With regard to the explanation of school grades MicroDYN had an incremental value compared to reasoning and APS in the high-school student sample but not significantly so in the university student sample (whereas APS had an incremental value in both samples).

Basically these findings highlight both potential and limitations of the MicroDYN approach in its current form. Current instances of the MCS approach address a small set of problem solving skills reliably, but it takes more than these skills to competently solve complex problems. Implications for future research on the assessment of problem solving skills are discussed.

List of scientific publications regarding the publication-based thesis

I. Manuscript

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

II. Manuscript

Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M., & Martin, R. (2013). A multitrait–multimethod study of assessment instruments for complex problem solving. *Intelligence, 41,* 579-596.

III. Manuscript

Fischer, A., Greiff, S., Wüstenberg, S., Fleischer, J., Buchwald, F., & Funke, J. (in press). Assessing analytic and interactive aspects of problem solving competency. *Learning and Individual Differences*, DOI:10.1016/j.lindif.2015.02.008.

Additional Manuscripts

Fischer, A., Greiff, S., & Funke, J. (under review). The history of complex problem solving. In B. Csapó, J. Funke, & A. Schleicher (Eds.), *The nature of problem solving.* Paris: OECD.

Greiff, S., Fischer, A., Stadler, M., & Wüstenberg, S. (2014). Assessing complex problem-solving skills with multiple complex systems. *Thinking & Reasoning (ahead-of-print),* 1-27.

Greiff, S. & Fischer, A. (2013). Der Nutzen einer komplexen Problemlösekompetenz: Theoretische Überlegungen und praktische Befunde. *Zeitschrift für Pädagogische Psychologie, 27,* 27-39.

Greiff, S. & Fischer, A. (2013). Measuring complex problem solving: an educational application of psychological theories. *Journal of Educational Research Online, 5,* 38-58.

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

1. Introduction

When people are confronted with complex problems there are a variety of strategies and heuristics[1] that can be observed (e.g., Klahr, 2002; Dörner, 1996; Schaub, 2013): Some people systematically gather information on multiple aspects of the situation (as well as the situation's dynamic development over time), generate detailed plans and hypotheses about how the situation may be transformed into a better one, and they scientifically adjust their plans and hypotheses to informative feedback. However, many people prefer less systematic strategies when confronted with complex problems: They focus on irrelevant details, rely on rigid methodism or gather confirmative instead of informative feedback (basically relying on dogmatic beliefs, e.g., Schaub, 2013; Dörner, 1996). Individual differences of this kind have received increasing interest in assessment contexts (cf. Fischer, Greiff, & Funke, under review), especially since "cross-curricular" problem solving skills have been included to complement curricular competencies (like reading, math or science) in large-scale assessments such as the Programme for International Student Assessment (PISA) 2000[2].

Two different kinds of problems have been proposed to measure cross-curricular problem solving skills (cf. OECD, 2014; Wirth & Klieme, 2003; Fischer, et al., in press): (1) *static problems*, which disclose all the information relevant to a solution at the outset and that can be solved by analytically deriving a solution from the information given (*Analytic Problem Solving, APS*); (2) *interactive problems* that demand additional skills because relevant information has to be uncovered in the process of problem solving by dynamically interacting with complex situations (*Complex Problem Solving, CPS*). Skills related to solving static and interactive problems are increasingly demanded by workplaces all over the world as a result of technological developments (e.g., Autor, Levy, & Murnane, 2003; for an overview see OECD, 2014). Correspondingly, in PISA 2012 problem-solving skills were assessed by means of multiple *static* and *interactive* problems in an international

[1] In this context the term "strategy" refers to "a plan –some sort of consciously intended course of action, a guideline (or set of guidelines) to deal with a situation." (Mintzberg, 1987, p.11) According to Gigerenzer and colleagues, "Heuristics are efficient cognitive processes, conscious or unconscious, that ignore part of the information. " (Gigerenzer & Gaissmaier, 2011, p. 451) that can be understood as a certain kind of "strategies that guide information search and modify problem representations to facilitate solutions" (Goldstein & Gigerenzer, 2002, p.75)

[2] In PISA 2000 both static and interactive problems were applied in a German national extension (Klieme, Hartig, & Wirth, 2005). In PISA 2003 static problems were applied on an international level, whereas interactive problems were applied in the national extension (Leutner, Fleischer, Wirth, Greiff, & Funke, 2012). In PISA 2012 both static and interactive problems were applied on an international level for the first time (OECD, 2014).

sample of about 85.000 students from 44 countries and economies around the world (cf. OECD, 2014).

In the current thesis both kinds of problem will be investigated from a psychometric point of view: The focus will be on the *Multiple Complex Systems approach* (Greiff et al., 2014) to assessing interactive problem solving skills (MCS approach, section 2) and its two instances MicroDYN (which is based on dynamic linear equations; see section 2.1) and MicroFIN (which is based on finite state machines; see section 2.2). After elaborating on the problem solving skills that are addressed by the MCS approach (section 2.3) three core papers of this thesis will be presented in order to address three main research questions regarding this approach: (1) its *content validity:* what are complex dynamic problems and which cognitive components and skills are most relevant for solving them (section 3.1; Fischer et al., 2012); (2) its *convergent validity:* how are different instruments for assessing some central interactive problem solving skills – MicroDYN and MicroFIN – related to each other and to school grades as external criteria of problem solving skills in educational contexts (section 3.2; Fischer, Greiff, et al., 2013); and (3) its *discriminant validity:* how are these interactive skills for solving complex dynamic problems different from reasoning and from performance in static problems that can be solved analytically (section 3.3; Fischer, et al., in press). In a final section (section 4) the main findings of this thesis will be summarized, and implications for future research will be outlined.

2. Conceptual Background

Before the core papers of this thesis will be presented in section 3, some of the basic terms and concepts of problem solving research, as well as the MCS approach and its history in psychometric contexts shall be introduced:

Complex and Dynamic Problem Solving. Formally speaking there is a problem if a person has a goal but does not instantly know how to reach it given his or her current state (e.g., Duncker, 1945). For instance, if a person wants to change the language settings of a new smartphone but does not know where to find them, this person has a problem. Characteristically, solving a problem requires conscious thinking about the next steps towards a goal state (e.g., Funke, 2003). According to Fischer et al. (2012) *complex* and *dynamic* problems consist of many highly interrelated elements, which have to be considered simultaneously (i.e., “complexity”, cf. Pollok, Chandler, & Sweller, 2002; Dörner, 1996; Weaver, 1948; Rey & Fischer, 2013) and a series of decisions with subsequent decisions depending on previous decisions as well as on environmental features that may change over time (i.e., “dynamics”, cf., Edwards, 1962; Gonzalez, Lerch, & Lebiere, 2003; Fischer et al., 2012).

For example, consider the problem of managing a corporation: This is a complex problem with multiple highly-interrelated decisions that have to be made simultaneously (e.g., the optimal decision on how many employees to hire – given a certain amount of money – may strongly depend on how much you pay them and upon how many machines you buy for them to work with, e.g., Sager, Barth, Diedam, Engelhart, & Funke, 2011). This set of decisions can be revised dynamically at multiple points in time (e.g., at the end of each month). Each decision can be evaluated with respect to multiple criteria (e.g., short-term and long-term profit), and has to be made under a high amount of "subjective uncertainty" (Osman, 2010) as the causal structure of the problem is not known in detail. Accordingly, the most characteristic features of complex dynamic problems have often been described as (1) complexity, (2) interconnectedness, (3) dynamics, (4) polytely, and (5) intransparency (Funke, 2001; 2003; Fischer et al., 2012).

Figure 1 visualizes the formal structure of a complex system that – although not highly complex[3] – exemplifies these five most characteristic features of complex problems (cf. minimal complexity, Greiff, 2012).

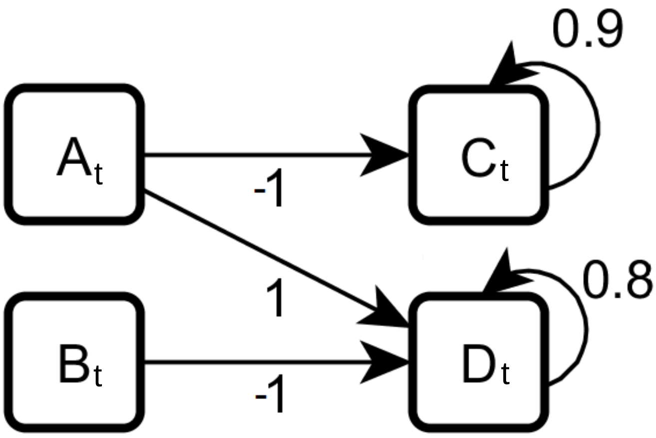

Figure 1. Structure of a minimally complex MicroDYN-task, characteristically unknown to the test subject, with two simultaneous decisions (A and B) and two goal-dimensions (criterion C and D) at multiple points t in time. A has multiple consequences, and criterion D is dependent on both decisions.

Please note that models applied for the purpose of simulation are always simplified representations of reality (Meadows, Randers, & Meadows, 2004). For instance, in a medical context, a structure of this kind may be used to simulate the dynamics of certain

[3] Please note, even systems with less variables and relations have been labeled “complex systems” in the literature, e.g., the sugar factory proposed by Berry & Broadbent (1984) – which can be understood as a MicroDYN task with the goal of regulating the sugar production by hiring and firing workers. The sugar factory is based on the linear equation [$\text{Production}_t = (-1)*\text{Production}_{t-1} + 2*\text{Workers}_{t-1}$ + random shock].

symptoms (C & D), which can be regulated by varying the amounts of two kinds of intervention (A & B). Even if reality is more complex than this model or any other model, a simple linear model like this may allow for valid conclusions about the reality it represents (cf. Vester, 2012; Meadows et al., 2004; Strunk & Schiepek, 2006). Of course, a minimally complex system may never be able to adequately represent all the aspects involved in managing a corporation, regulating climate change or organizing developmental aid (i.e., highly complex problems). Nevertheless it can still be highly interesting from a psychometric perspective (e.g., Funke, 2010): Due to minimal complexity, multiple independent problems can be presented to subjects in a short amount of time (e.g., 12 problems with 5 minutes per problem can be presented in one hour testing time, see Greiff & Fischer, 2013) which allows for highly reliable conclusions about specific problem solving skills (Funke, 2010; Fischer et al., in press). Additionally, the skills required for solving each problem (see section 2.3) can be clearly identified (and varied between problems, e.g., Greiff, Fischer, et al., 2013) which is more difficult with regard to realistic and highly complex problems (see Funke, 2001). To illustrate this point the research tradition, which led to the development of the MCS approach, will be outlined before the MCS approach itself will be elaborated on in more detail.

The history of Complex Problem Solving Research. Empirical research on Complex Problem Solving started in the late 1970’s (Fischer et al., under review), when Dietrich Dörner and colleagues implemented computer simulations of highly complex problems in order to examine problem solving in realistic environments (e.g., Dörner, 1976; Dörner, Kreuzig, Reither Stäudel, 1983). Some of these simulations have also been proposed for assessment purposes (e.g., Dörner, 1986) because their cognitive demands seemed to be fundamentally different from traditional static problems (e.g., Funke 2001; Putz-Osterloh, 1981) and because traditional measures of intelligence seemed to be weak predictors of performance in complex problems (Süß, 1999; Wittmann & Süß, 1999). However, some severe psychometric issues impeded the high potential of complex problems in assessment contexts:

Performance in single highly-complex problems often proved to be unreliable (Süß, 1999) and did not sufficiently correlate with performance in other problems, with intelligence or with external criteria (e.g., Dörner & Kreuzig, 1983; Kluwe, Misiak, & Haider, 1991; Süß, 1999). Moreover, because CPS simulations usually take a lot of testing time, it has been difficult to present more than one problem in order to increase reliability (Greiff & Fischer, 2013). However, as Süß (1999) argues, when CPS performance is aggregated over multiple complex systems, convergent validity could indeed be demonstrated: For instance, Süß (1999) reports data on multiple complex problems (the simulations Tailorshop, PowerPlant, & Learn, cf. Wittmann & Süß, 1999). Each simulation was presented twice in order to increase reliability (with the exception of the simulation Learn, because this simulation took two hours of testing time for a single trial, Wittmann & Süß, 1999). Süß (1999) reports the complex problems to be correlated with each other (r = .22 - .38) and fluid intelligence to be correlated with performance in each CPS simulation (r = .35 - .46) as well as with an aggregate score of CPS performance (the sum of performance measures over all complex problems; r = .56). In a regression of CPS performance on problem-specific knowledge (assessed after an exploration phase, where *"subjects played each simulation for several trial runs to acquire familiarity with the simulation"*, Wittmann & Hattrup, 2004, p. 401) and fluid reasoning both predictors proved to have *unique* contributions for each of the three problems[4]. Findings like these highlight that CPS – aggregated over multiple problems – depends on reasoning, but they also highlight the incremental value of prior knowledge and *knowledge acquisition skills* for the solution of complex problems (Greiff & Fischer, 2013). However, the results of Süß (1999) also demonstrate the high heterogeneity between different complex problems. Additionally the number of CPS simulations applied in this study was rather low compared to traditional and static psychometric tests. The Multiple Complex Systems (MCS) approach presented in this thesis emerged as an answer to these problems (Funke, 2001; Greiff & Fischer, 2013): The idea of the MCS approach (cf. Greiff, 2012) is to present a larger number of less complex dynamic problems (typically up to twelve problems) and to ensure a sufficient amount of homogeneity between problems by means of formal frameworks such as linear equation systems of finite state machines (see below) as proposed by Funke (2001). As a result, current instances of the MCS approach do not address all the skills potentially relevant for different kinds of CPS, but they do address a small set of skills – central to solving a wide range of problems – reliably (cf. Fischer, et al., in press). For a detailed overview concerning the history of Complex Problem Solving, see Funke (2003) or Fischer, Greiff, & Funke (under review).

[4] When problem-specific knowledge and reasoning were controlled for, correlations between CPS tasks (see above) dropped to insignificant values around zero (r = .04 - .15).

The Multiple Complex Systems approach. Modern computer-based assessment of complex dynamic problem solving is increasingly based on the seminal work of Funke (2001) who proposed to use the interaction with computer-simulations[5] of (a) dynamic linear equation systems (the *MicroDYN approach*, proposed by Greiff & Funke, 2010) or (b) finite state machines (the *MicroFIN approach*, proposed by Greiff, Fischer, et al., 2013 in the second paper of this thesis) as indicators of the problem solving skills involved in identifying (i.e., *representing*) and controlling (i.e., *solving*) complex dynamic problems. Both approaches (see sections 2.1 and 2.2) and the problem solving skills involved (see section 2.3) will be outlined in more detail below. To exemplify how these kinds of tasks work in general, let us refer to Cronbach (1961, p.7), one of the early predecessors of the idea of applying multiple complex problems as tasks for assessing complex problem solving abilities:

> "*We inquire about a piece of apparatus with a ring of lights [i.e., dependent variables] and a few pushbuttons [i.e., independent variables]. 'This' he says, 'is an experimental test which permits us to present much longer and more complex tasks than the usual puzzle. It is used to measure abilities of high-level scientific and technical workers; (…) The apparatus is wired so that it follows some simple rules. These rules change with every problem. There are three pushbuttons which turn on and off various combinations of lights. The person's task may be to turn on light number 3 only. He presses the buttons in turn to find out what lights each button controls [i.e., to <u>represent</u> the problem]. For instance, when he presses button 1, lights 3, 4, and 5 go on. When he has all the information, he must find a sequence of actions which will leave only light 3 lit [to <u>solve</u> the problem]." (Cronbach, 1961, p.7)*

Likewise, in the MCS approach, each test comprises multiple minimally complex problems, and each problem has to be represented and solved interactively by a series of inputs and resulting outputs (cf. Fischer et al., in press). According to Funke (2001) the formal frameworks of linear equation systems and finite state machines both allow for (a) designing multiple independent tasks based on a well-defined set of commonalities and for (b) systematically varying item-characteristics related to task difficulty. Using multiple complex systems based on the formal frameworks proposed by Funke (2001) for

[5] In the research literature different labels refer to this kind of simulations (e.g., „Microworlds", „complex problems", "dynamic problems", "interactive problems").

psychometric purposes is regarded as the Multiple Complex Systems approach (*MCS approach*).

As the current thesis is meant to present recent research on both MicroDYN and MicroFIN – the two instances of the MCS approach proposed so far – the specifics of both approaches (see section 2.1 for MicroDYN and 2.2 for MicroFIN) and the problem solving skills that are required for solving problems in both approaches (see 2.3) will be outlined, before the core papers of this thesis will be presented (see 3).

2.1 MicroDYN – Microworlds based on dynamic linear equations

The MicroDYN approach to assess problem solving skills is based on multiple tasks. Each task is an interactive simulation of a linear equation system (e.g., Greiff & Fischer, 2013). For example, one task – as depicted in Figure 2 – may simulate the variables “Motivation”, “Power of throw” and “exhaustion” of a handball team, as well as their linear dependence on the amounts of certain kinds of training (e.g., Training A, B and C). In each task, the test subject can set the amount of each kind of training (i.e., the independent variables at the left side of Figure 2) and watch the resulting changes in each of the other variables (i.e., the dependent variables at the right side of Figure 2).

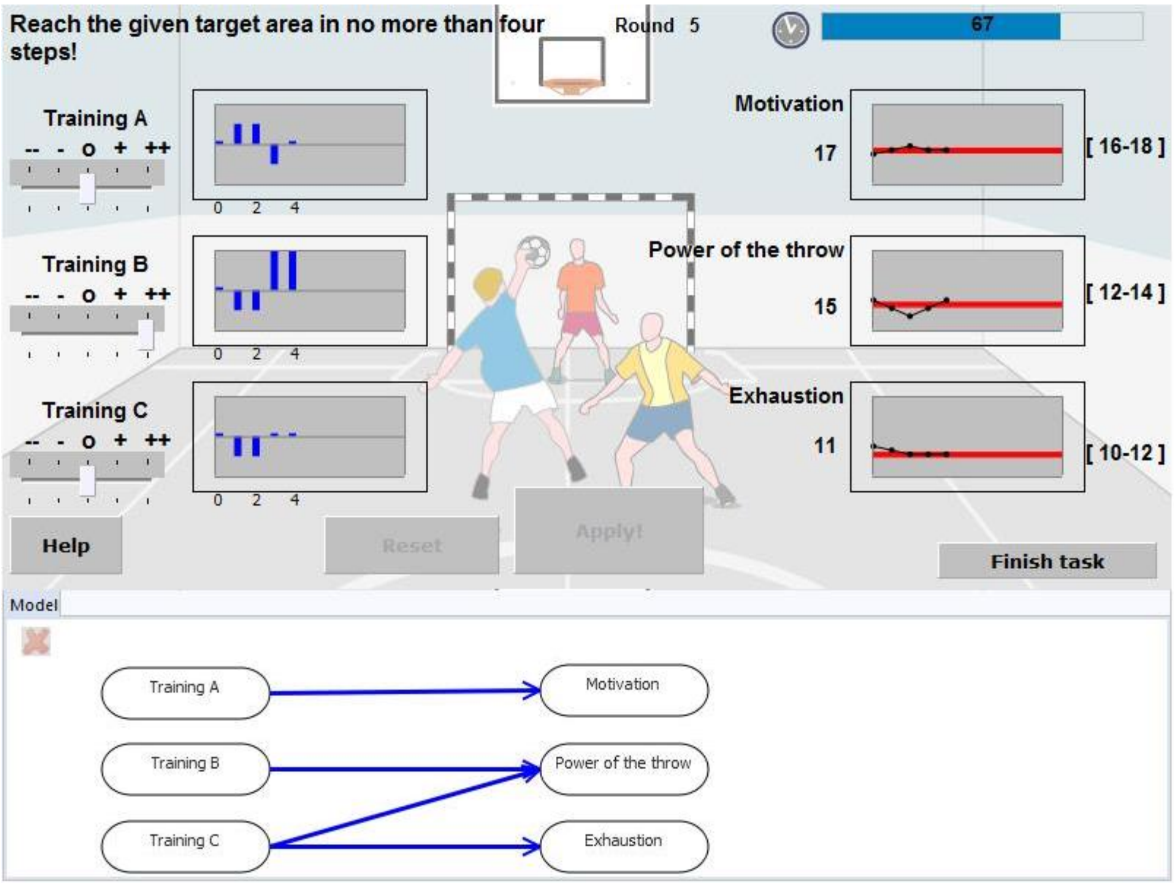

Figure 2. Screenshot of the second phase of a MicroDYN-task. Independent variables (left side of the screen) are labeled with fictional pseudo-words, in order to not trigger any helpful prior knowledge about the problem’s causal structure. In a first phase, a problem solver has to interact with the problem and draw the causal structure in a causal diagram (representation at lower part of the screen). Afterwards, in a second phase, certain goal ranges are shown for each dependent variable (solution at the right side of the screen).

Procedure: At the beginning of each task the test subject does not know how independent and dependent variables are interrelated. Thus, in a first phase of the task, he or she has to find out about the relations by interacting with the simulation. For instance, if the test subject wants to know if Training A is related to Motivation, he or she may increase the amount of Training A (e.g., by setting the corresponding slider to “++”), click on the “Apply”-button and monitor if Motivation changes as a result of this variation. At the end of phase 1, hypotheses about the causal relations between variables have to be drawn in a causal diagram at the bottom of the screen and phase 2 begins. In phase 2 the test subject is confronted with specific goal values for each dependent variable, and has to reach these goal values in less than five turns. If each dependent variable equals its corresponding goal value (or if “apply” was clicked five times or if time is up), the MicroDYN task ends.

Scoring: For each MicroDYN task, performance in each phase is typically assessed based on the correctness of the causal model (*problem representation*) and on the number of goals reached (*problem solution*). For alternative ways of scoring see Greiff (2012).

2.2 MicroFIN – Microworlds based on finite state machines

The MicroFIN approach to assessing problem solving skills is also based on multiple tasks. Each task is an interactive simulation of a finite state machine (Greiff, Fischer et al, 2013). At each moment in time a finite state machine is in a certain state, and inputs to the machine (e.g., pressing buttons or time passing by) can cause a state transition from one state to another (with the set of state transitions being finite, see Buchner & Funke, 1993; Funke & Buchner, 1992). Thus, the machine can only be in a finite number of states and typically either the current state or the most current state transition is related to some form of output (see Funke, 2001, or Buchner & Funke, 1993, for a more detailed description of finite state machines). MicroFIN-tasks are highly heterogeneous in nature, as every well-defined problem space – consisting of goals, states and state transitions regarding a problem (see Newell and Simon, 1972) – can approximately be described in this framework. Most characteristic instances of finite state machines are automatons like washing machines (Figure 3, left), toy cars similar to the “BigTrak” toy (Figure 3, middle) or virtual pets similar to the “Tamagotchi” toy (Figure 3, right). A detailed formal description of these automatons can be obtained by the author of this thesis.

Figure 3. Exemplary finite automatons – simulated in the Finite AUtomaton Simulation Tool (FAUST) – which can be obtained by the author of this thesis.

Procedure: At the beginning of each task the test subject does not know about the transition matrix of the automaton. Thus, in a first phase of the task, he or she has to find out about the consequences (e.g., visual output of the automatons in Figure 3) of each possible input (e.g., a button of an automaton in Figure 3) in different states of the automaton. For instance, if the test subject wants to know how the "RPT"-button in the BigTrak simulation works, he or she may click the "RPT"-buttons in different states of the automaton (e.g., after clicking one of the other buttons) and monitor the state transitions resulting from these inputs. At the end of phase 1, a series of questions about possible state transitions of the automaton are to be answered. In phase 2 the test subject is confronted with a specific goal (e.g., "make BigTrak drive left, up, left") and has to reach this goal using as few inputs as possible. If the goal is reached (or time is up), the MicroFIN task ends.

Scoring: MicroFIN task performance in each phase is typically assessed based on the correctness of the answers to questions about the automaton's state transitions (*problem representation*) and on whether the goal was reached or not (*problem solution*). For alternative possibilities of scoring please refer to Buchner & Funke (1993).

2.3 Problem Solving Skills assessed by means of MicroDYN and MicroFIN

In this section, some of the cognitive and behavioral processes that are *generally* involved in solving MicroDYN and MicroFIN tasks will be elaborated on. Please note, some cognitive and behavioral processes may be considered useful for solving all tasks of this kind (e.g., systematically building and testing hypotheses on the automaton's causal structure; or carefully monitoring the consequences of one's inputs before deciding on the next input) whereas other processes may be involved in solving a specific task but irrelevant to solving other tasks (e.g., regarding the automatons in Figure 3, many facts about toy cars may be helpful in controlling a toy car, but irrelevant for setting up a washing machine). Thus, by definition the former can be assumed to be more general than the latter (cf. Anderson, 2013) even if typically both kinds of processes are closely intertwined and referring to each other (e.g., deciding which hypothesis to build and which aspect to test may be informed by task-specific knowledge).

In this section I will temporarily abstract from task-specific processes and focus on the problem solving skills that are assumed to be generally involved in (and therefore assessed by) both MicroDYN and MicroFIN tasks: For instance, every task in both MicroDYN and MicroFIN requires the *dual-search for hypotheses and information* on how to viably represent the problem (cf. Klahr & Dunbar, 1988; Klahr, 2002), followed by the *search for a solution* in the resulting problem space (cf. Newell & Simon, 1972). Each kind of search happens in a corresponding search space (e.g., a solution is searched within the space of all possible states of the problem that can be reached by state transitions) which is usually assumed to be ordered and constrained by task-specific prior knowledge (e.g., knowledge about the value of certain states or state transitions for reaching one's goals, cf. Damasio, 1994; Newell & Simon, 1972).

In the following sections the dual-search for information and hypotheses as well as the search for a solution (and the corresponding problem solving skills assessed by all kinds of MicroDYN and MicroFIN tasks) will be elaborated on in more detail. A problem solving skill in this context can be considered to be the knowledge on *when* and *how* to apply different strategies (or on how to generate this kind of knowledge) in order to search for adequate representations and solutions to problems effectively (cf. Greiff et al., 2015; Anderson, 1982)[6]. Please note, this conceptualization of the skills MicroDYN and MicroFIN address –

[6] As complex problem solving is a fairly complex process itself (see section 3.1) some researchers have proposed additional search spaces for certain kinds of problems (e.g., Schunn and Klahr, 2002, proposed a space of possible data-representations, and Burns & Vollmeyer (1996) proposed a space of model assumptions).

i.e., the conceptual relation of the MCS approach to prior research on scientific discovery and problem solving – is highly influenced by the studies of Fischer et al. (2012) and Greiff, Fischer et al. (2013) and thus can be considered a contribution of the current thesis (even though this section precedes sections 3.1 and 3.2 due to a better flow of reading).

Search for a Representation. Klahr and Dunbar (1988) proposed to think of scientific discovery and interactive concept formation as a complex[7] problem solving process, which involves the coordinated search in two spaces – (a) the space of hypotheses and (b) the space of experiments that are informative for testing these hypotheses (Klahr, 2002). According to the dual-space model (cf. Klahr, 2002), in order to adequately represent a problem, the problem solver has to (1) *state hypotheses* concerning how different aspects of the problem are interrelated (e.g., based on prior knowledge or based on interactions with the problem), to (2) *test hypotheses* by interacting with the problem (i.e., conducting an experiment on focal aspects of the problem and comparing the problem's state changes with the hypotheses' predictions), and to (3) *evaluate evidence* in order to come to a conclusion (i.e., to accept or reject hypotheses or to initiate further inquiry). Corresponding to the two problem spaces, Klahr and Dunbar (2002, p.57f) highlight two essential problem solving skills that are highly domain-general and relevant for representing MicroDYN and MicroFIN tasks: These skills can be described as *"knowing where to look and understanding what is seen. The first skill – experimental design – involves the design of experimental and observational procedures [such as applying the principle of isolated variation of variables]; the second – hypothesis formation – involves the formation and evaluation of theory."* (Klahr & Dunbar, 2002, p. 57f.)

Search for a Solution. When the problem's structure is sufficiently known, problem solving can be described as the search for a solution in a single problem space (Newell & Simon, 1972). In this case, the problem space can be described by the set of states the problem can be in, the set of operators applicable to the states (i.e., state transitions), and the set of goal states (see section 1.2 for the close relation of problem spaces and finite state machines). Corresponding to this problem space, Newell and Simon (1972) highlight a range of highly domain-general problem solving skills, which can also be applied to MicroDYN and MicroFIN tasks: Depending on the task-specific knowledge, which constrains the problem space, these methods can range from (1) applying a viable series

[7] Please note in this case – and independent of the problem – it is the search space that is complex, i.e., containing multiple interrelated search spaces (Klahr, 2002, p. 27). The idea is based on the *General Rule Inducer* (Simon & Lea, 1974) – a single information processing system accounting for performance in both rule induction and problem solving.

of inputs from memory over (2) means-end-analysis or the planning heuristic, to (3) randomly generating and testing inputs.

MicroDYN and MicroFIN – Comparing Eggs and Oranges? At the outset MicroFIN and MicroDYN seem to address different kinds of problems: The former is about quantitative relations between numerical variables, whereas the latter is about qualitative relations between discrete inputs and states (Greiff, Fischer et al., 2013). Indeed, some problems are more intuitively – and less computationally costly – described as finite automatons, others are more intuitively described as as linear equation systems. In the recent past, this may have fostered a high heterogeneity between tests following the MicroDYN and MicroFIN approach (see sections on MicroDYN and MicroFIN for characteristic examples of each approach). Thus, on the one hand, one may argue that MicroDYN and MicroFIN are likely to address different problem solving skills. On the other hand, the searches for representations and solutions have been described as domain-general and basic problem solving processes (e.g., Klahr, 2002) and as inherently involved in representing and solving both MicroDYN and MicroFIN tasks (Greiff, Fischer, et al., 2013). Therefore, besides all the differences, it seems reasonable to assume a common core of problem solving skills, assessed by both MicroDYN and MicroFIN (in fact this is a major point that will be made in section 3.2 when presenting the study of Greiff, Fischer, et al., 2013).

Additionally, both formal frameworks proposed by Funke (2001) should not be misperceived as mutually exclusive: On a conceptual level, both frameworks are more like different perspectives on complex problems in general, and (although this may seem cumbersome at first) it may foster our understanding of both approaches to note, that both approaches can approximately emulate each other: For instance, Table 1 shows a small excerpt of the transition matrix of a MicroFIN task (i.e., a finite state machine) “emulating” key aspects of a MicroDYN-task[8] (i.e., a linear equation system) as the one shown in Figure 1.

[8] Schoppek (2002) introduced a similar approach to understanding MicroDYN tasks, when he formally described the problem space of MicroDYN tasks by means of a set of states and state transitions. Buchner, Berry, & Funke (1995) demonstrated how to emulate the linear equation of the Sugar Factory simulation within the framework of finite state machines.

Table1. *Excerpt of the transition matrix of a MicroDYN-automaton.* Each line represents a state (see first column) with a certain output (column 2) and a set of state transitions (column 3 to 7) triggered by input events (e.g., buttons labeled “More A”, “Less A”, “More B”, “Less B”, or “Apply”).

State	Output	More A	Less A	More B	Less B	Apply
0	Symptom 1=00% ; Symptom 2=00% ; Medizin A=0 ; Medizin B=0	1	0	3	0	0
1	Symptom 1=00% ; Symptom 2=00% ; Medizin A=1 ; Medizin B=0	2	0	4	1	1
2	Symptom 1=00% ; Symptom 2=00% ; Medizin A=2 ; Medizin B=0	2	1	5	2	11
3	Symptom 1=00% ; Symptom 2=00% ; Medizin A=0 ; Medizin B=1	4	3	6	0	3
4	Symptom 1=00% ; Symptom 2=00% ; Medizin A=1 ; Medizin B=1	5	3	7	1	4
5	Symptom 1=00% ; Symptom 2=00% ; Medizin A=2 ; Medizin B=1	5	4	8	2	113
6	Symptom 1=00% ; Symptom 2=00% ; Medizin A=0 ; Medizin B=2	7	6	6	3	6
7	Symptom 1=00% ; Symptom 2=00% ; Medizin A=1 ; Medizin B=2	8	6	7	4	106
8	Symptom 1=00% ; Symptom 2=00% ; Medizin A=2 ; Medizin B=2	8	7	8	5	215
9	Symptom 1=10% ; Symptom 2=00% ; Medizin A=0 ; Medizin B=0	10	9	12	9	0
10	Symptom 1=10% ; Symptom 2=00% ; Medizin A=1 ; Medizin B=0	11	9	13	10	10

The complete matrix of this automaton contains 1090 states and can be obtained by the author of this thesis. As one can see in Table 1, each state of the automaton is associated with a certain output (e.g., “Symptom C = 10%; Symptom D = 00%; Medicine A = 0; Medicine B = 0”) which visualizes the current level of independent variables (Medicine A & Medicine B) and dependent variables (Symptom C & Symptom D). A state transition is triggered whenever the user makes an input to the automaton (“More A”, “Less A”, “More B”, “Less B”, and “Apply”). The state transition matrix specifies the resulting state for each input: For instance, like specified in Table 1, if the problem solver presses the “Apply”-Button in state 9 (“Symptom 1 = 10%; Symptom2 = 00%; Medicine A = 0; Medicine B = 0”), the resulting state is state 0 (“Symptom 1 = 00%; Symptom 2 = 00%; Medicine A = 0; Medicine B=0”). Of course, the automaton just described is a rather simplistic simulation of a MicroDYN task (and it may well be expanded to cover more input-events per independent variable, more states, and a more impressive visual output), but it serves well as a proof of concept: MicroFIN tasks can emulate MicroDYN tasks with an arbitrary degree of approximation. Of course, MicroFIN, in turn, can also be emulated in a linear equation framework (by using dummy-variables and nonlinear transformations of variables in the linear equation model, cf. Bortz, 2005) but for the sake of brevity this point will not be elaborated on in this thesis.

To summarize, based on commonalities in (a) the kind of skills required to solve both kinds of problems (cf. Greiff, Fischer et al., 2013), and (b) the problems that can be formulated within each approach, besides all potential differences (which will be addressed in the discussion, section 4) there also seems to be a common core of skills underlying performance in both instances of the MCS approach. Consistent with this idea, in the PISA 2012, a global problem solving score is built based on both MicroDYN and MicroFIN tasks (amongst others, cf. OECD, 2014).

3 Research on Tools and Perspectives concerning the MCS approach

Regarding empirical research on the assessment of CPS skills, the MicroDYN approach was applied in a wide range of studies: For instance, the skills assessed by means of MicroDYN tasks (see section 2.3) proved to be incrementally valid for explaining school grades beyond different measures of reasoning (e.g., Wüstenberg, Greiff, & Funke, 2012; Greiff & Fischer, 2013; Greiff, Wüstenberg, Molnár, Fischer, Funke, & Csapó, 2013; Greiff, Kretzschmar, Müller, Spinath, & Martin, 2014), structurally invariant across different school grades (Greiff, Wüstenberg et al., 2013) and, as reported in Greiff (2012), closely related to representation (R^2 = .43, p < .01) and solution (R^2 = .45, p < .01) of a complex dynamic problem called “space shuttle” or “HEIFI” (a finite state machine with 120 states and 20 input-buttons, applied in a national extension of PISA 2000, cf. Wirth & Klieme, 2003).

Thus, MicroDYN evolved to be the method of reference when it comes to assessment of problem solving skills by means of multiple complex systems. MicroFIN, on the other hand, is a more recent approach (Greiff, Fischer et al., 2013; Neubert, Kretzschmar, Wüstenberg, & Greiff, 2014) with a lot of potential for going beyond MicroDYN (see below). Both instances of the MCS approach (MicroDYN and MicroFIN) have been applied in PISA 2012 to assess how students interactively solve minimally complex problems (OECD, 2014).

In this thesis three questions related to the validity of the MCS approach are investigated. Three different studies on the MCS approach are reviewed, emphasizing both potential and limitations of current assessment instruments: (1) The first study defines CPS and proposes a theoretical framework for understanding the main components of CPS (knowledge acquisition and knowledge application); (2) the second study relates different MCS assessment instruments – a traditional MicroDYN test, a MicroFIN test and the test GeneticsLab – to each other. For the first time, this study demonstrates external validity of these instruments on a latent level (with incremental validity for explaining science grades beyond fluid intelligence), and allows for clearly separating both method-specific and method-general aspects of variance; after examining the relations of different

MCS instruments to each other, (3) the third study is about the relation of the *interactive* MicroDYN test to *static* measures of reasoning and Analytic Problem Solving (APS, as it was assessed in PISA 2003). This study proves MicroDYN to substantially complement the traditional and static measures of APS in a high-school student sample, but not significantly so in a university student sample. APS, on the other hand, has substantial unique contributions to predicting school grades beyond both reasoning and MicroDYN in both samples. Thus, the study demonstrates an empirical value of assessing problem solving skills by means of MicroDYN in high-school student samples (as it is done in PISA 2012, for example), but also points at additional (analytic) problem solving skills current instances of MicroDYN do not address.

In all of these studies, the MicroDYN approach proved to allow for highly reliable and sufficiently valid assessment of a narrow set of problem solving skills (see section 2.3). However, results concerning the interactive MicroFIN test in study 2 and the static APS test in study 3 also point towards additional aspects of complex problem solving that are insufficiently addressed by current operationalizations of the MicroDYN approach.

3.1 Review 1: Content Validity (cf. Fischer, Greiff, & Funke, 2012)

Fischer et al. (2012) outlined some of the major strands of research on CPS (e.g., research on strategy selection, information reduction, intelligence, or on the interplay of implicit and explicit knowledge in the process of CPS) in order to derive a theoretical framework for the most important cognitive processes involved in solving complex problems. The framework highlights both problem representation (knowledge acquisition) and problem solution (knowledge application) as the two major phases in the process of CPS. Additionally, the framework is closely related to the five most characteristic features of complex problems:

In a first phase, the problem solver has to acquire a parsimonious and viable representation of the problem: The problem solver is assumed to generate information about the problem (to cope with intransparency), to infer the relevant elements and relations of the system (to cope with interconnectedness) and to increasingly focus on the most relevant aspects only (to cope with complexity). When there is a sufficient amount of knowledge it has to be applied effectively: In this phase, the problem solver is assumed to make a dynamic choice by (a) instantly choosing an action that is known to be effective in the current situation, or by (b) deriving a solution based on the knowledge available (to influence the system's dynamics). Finally, evaluation of the resulting problem state (and its deviation from the current set of goals) may indicate progress, result in further knowledge acquisition[9] or - if the current goal cannot be reached in time or is less important than other goals – result in a change of goals[10] (due to polytely).

Please note the theoretical framework is formulated on a rather abstract level, because many details of CPS depend on attributes of problem and/or problem solver that can hardly be generalized across problems or problem solvers. For instance, some problems are emotionally[11] disturbing or existential (e.g., Baltes & Smith, 1990) while others are not – and problem solvers may differ with regard to the strategies they assume to be efficient. The theoretical framework proposed by Fischer et al. (2012) focuses on the cognitive processes that are relevant for characteristic instances of CPS (cf. Funke, 2003). As Fischer et al. (2012) suggest it may well be expanded to cover (a) the processes involved in specific kinds of CPS, or (b) the processes that will empirically prove to be part of CPS in the future. In its current state, the framework reflects the research results it was based upon – a least common denominator of an interdisciplinary field of research.

The framework explicitly highlights knowledge acquisition as part of the CPS process and has considerably shaped the understanding of MicroDYN and MicroFIN: Before this theory was published, MicroDYN and MicroFIN were assumed to address three different facets of CPS (e.g., Wüstenberg, et al., 2012; Greiff, 2012), but based on the framework of Fischer et al. (2012), knowledge acquisition and knowledge application were increasingly conceptualized as the main facets of CPS that are addressed by both MicroDYN and MicroFIN (whereas information generation and model building were regarded as closely intertwined processes as a result of the dual-search for information and hypotheses involved in knowledge acquisition, see Greiff & Fischer, 2013b). In accordance with the framework, MicroDYN and MicroFIN address problem solving skills related to both knowledge acquisition (i.e., problem representation) and knowledge application (i.e., problem solution). However, the framework also points towards shortcomings regarding the content validity of MicroDYN and MicroFIN, as it highlights heuristics that are relevant for CPS but not applicable to current instances of the MCS approach (e.g., asking an expert). The skills addressed by MicroDYN and MicroFIN may

[9] As Funke (2003) noticed, in problem solving the lack of knowledge that may be revealed during the attempt of problem solving may refer to both means (e.g., the applicability or the consequences of important operators may not be known) and ends (e.g., criteria for goal-achievement may be too unspecific, cf. Dörner, 1976).

[10] for more detailed considerations about the process of goal selection please refer to Dörner et al. (1983)

[11] For a noteworthy attempt to address the complex interaction of motivation, cognition & emotion in human action regulation regarding existential problems (hunger, thirst, pain, etc.) the interested reader may refer to Dörner et al. (2002) or Dörner & Güß (2013)

be central for solving many kinds complex problems (especially in scientific domains, see section 3.2), but the process of CPS seems to depend on different factors as well (Fischer et al., 2012). Additionally certain problem solving skills, related to the problem solver's switching between knowledge acquisition and knowledge application are not addressed by MicroDYN and MicroFIN, due to the artificial separation of phases (see sections 2.1 and 2.2). I will elaborate on these shortcomings of the MCS approach in the discussion (section 4).

3.2. Review 2: Convergent Validity (cf. Greiff, Fischer et al., 2013)

Greiff, Fischer et al. (2013) presented the first empirical study on MicroFIN tasks and their empirical relation to two kinds of MicroDYN tasks (a traditional MicroDYN test and a simplified MicroDYN test with binary input-variables, called "GeneticsLab", cf. Greiff, Fischer et al., 2013). More specifically, Greiff, Fischer et al. (2013) demonstrated that the incremental value of MicroDYN tasks beyond measures of fluid intelligence (which was reported in a lot of studies using a wide range of different measures for fluid intelligence, e.g., Wüstenberg et al., 2012; Greiff & Fischer, 2013; Greiff, Wüstenberg et al., 2013) can also be found on a latent level if different instances of the MCS approach (MicroFIN and two instances of MicroDYN) are used to indicate method-general representation and solution skills. Please note each single homogenous test of CPS skills (e.g., MicroDYN) necessarily addresses a wide range of cognitive skills and abilities to a certain degree (Horn & Masunaga, 2006). By assessing specific skills (e.g., representation and solution skills) by means of multiple heterogenous tests each, it becomes possible to examine the method-general factors (i.e., "traits", related to representing and solving problems in different tests) and how well they are tapped by each test. With regard to the MCS approach Greiff, Fischer, et al. (2013) report sufficient convergent validity of the different tests (with substantial loadings of all measures and trait consistencies mostly between .50 and .60), a moderate relation of both latent traits to a measure of fluid intelligence / reasoning (b = .49-.53; p < .01) and incremental validity of the latent problem representation trait (as measured by different instances of the MCS approach) for explaining science grades (b = .22; p < .01) beyond fluid intelligence. So on a general level all tests seem to address a common core of problem solving skills, and these skills are incrementally valid for predicting school grades.

However, on average, trait consistency was about 50% for both MicroFIN (with method-specificity up to .89) and GeneticsLab (with method-specificity up to .59). This implies, that different tasks of MicroFIN had as much in common with the method-general traits as they had uniquely in common with each other (i.e., different MicroFIN tasks seem to additionally address something beyond the method-general representation and solution skills described in section 2.3 of this thesis) and the same applies for GeneticsLab. For instance, in MicroFIN each automaton represents a different aspect of reality and requires different strategies to be applied. Thus the method-specific aspects of MicroFIN may represent the impact of analogy-building or strategy selection skills as well as different kinds of world knowledge (concerning each of the automatons or concerning finite state machines in general) for instance.

With regard to GeneticsLab these method-specific factors of representation and solution (which may be related to differences in the tasks' Graphical User Interface, see Greiff, Fischer et al., 2013) were correlated with the method-general factors solution and representation (please note the reversed order, cf. Greiff, Fischer, et al., 2013). Concerning MicroFIN, the method-specific factors concerning representation and solution (i.e., the skills that were assessed in addition to the skills that are common to both MicroDYN and MicroFIN) were *not* significantly related to the method-general factors. These findings render MicroFIN especially interesting for future research on additional aspects of CPS. However, for MicroFIN there also was a higher variance in trait and method loadings (compared to the GeneticsLab) which indicates a greater heterogeneity of MicroFIN tasks (see section 2.2). Implications for future research on MicroFIN will be outlined in detail in the discussion of this thesis (section 4).

In addition to introducing MicroFIN and to proving convergent validity between different measures of the MicroFIN and MicroDYN approach, the study of Greiff, Fischer et al. (2013) also was the first study to prove that the skills assessed by means of MicroDYN and MicroFIN were especially related to science grades on a latent level, emphasizing the close relation between both approaches and the process of scientific discovery (see section 2.3). In previous studies, MicroDYN was simply thought of as assessing domain-general problem solving skills, but obviously, the skills required to interactively solve MicroDYN and MicroFIN tasks are more similar to understanding and solving scientific and technical problems (e.g., Scherer & Tiemann, 2012; Cronbach, 1961) than to problem solving in the domain of reading or writing for instance. Please note, these skills – involved in interactively representing and solving complex problems – may nevertheless be applicable to a wide variety of problem domains (such as fixing automatons, dynamically managing corporations, coordinating developmental aid, etc.) and curricula (e.g., biology, physics, chemistry). Thus, the problem solving skills assessed by means of the MCS approach may still be considered as being cross-curricular and highly domain-general, even if they are neither universally applicable nor sufficient for solving every kind of complex problem.

From today's point of view, the major contribution of Greiff, Fischer, et al. (2013) may have been to introduce the MicroFIN approach as an alternative to the MicroDYN approach, and to highlight the heterogeneity inherent in MicroFIN automatons (see Neubert et al., 2014, for a recent replication of these findings in a high school-student sample). Please note, complex problems in real-life are highly heterogeneous (e.g., Funke, 2001) and this heterogeneity has to be carefully addressed by each approach to assessing CPS. We will come back to this point in the discussion (section 4).

3.3. Review 3 Discriminant Validity (cf. Fischer et al., in press)

The third study of this thesis elaborated on the relation of *interactive* MicroDYN tasks to the traditional *static* tasks that were applied in PISA 2003 to assess Analytic Problem Solving (APS). Besides a common core of problem solving skills addressed by both kinds of tasks (e.g., analyzing complex information about the information given at a certain moment in time) Fischer et al. (in press) expected to find evidence for *additional* skills that were related to interactive problems only (i.e., not indicated by static problems and static tests of reasoning). For instance, in contrast to APS, MicroDYN tasks require (1) interactively gathering information in the light of hypotheses and (2) adapting plans and hypotheses to the information uncovered. Corresponding to this interpretation, in PISA 2012 MicroDYN tasks are referred to as "interactive problem solving" and they are meant to complement traditional static problems. Fischer et al. (in press) put this assumption to the test for the first time, by contrasting interactive MicroDYN tasks (indicating CPS skills) with static problems (indicating APS skills) and a static matrices test (indicating logical reasoning), in order to determine (a) the amount of covariance between CPS and APS skills, as well as (b) if there was an incremental value of the interactive MicroDYN tasks compared to static problem solving tasks *and* the reasoning test with regard to explaining external criteria.

With regard to the first question (a), in a sample of university students (n = 339) as well as in a sample of high-school students (n = 577) there was a large overlap between APS and both representing and solving MicroDYN tasks ($r = .73 - .77$). Commonality analyses revealed that a large amount of variance in APS can be explained by variance that is common to representation and solution in MicroDYN, but not common to reasoning (ΔR^2 = .33 - .52). This empirical overlap indicates non-fluid aspects of problem solving competency assessed by both APS and MicroDYN tests (Fischer et al., in press) and may be due to strategic knowledge (Strohschneider & Güss, 1999) on *when* and *how* to apply different search strategies in the process of problem solving (cf. operative intelligence, Dörner, 1996). In a similar vein Horn & Masunaga (2006) report that expert reasoning and expert working memory tasks (with stimuli from a subject's domain of expertise) address different constructs – and evoked different kinds of strategies – than traditional fluid reasoning and working memory tasks. The findings of Fischer et al. (in press) indicate sufficient discriminant validity of MicroDYN with regard to reasoning (r = .21 - .52).

With regard to the second question (b), the APS test proved to be incrementally valid for predicting school grades (compared to MicroDYN and reasoning) in both samples (ΔR^2 = .04-.14), whereas the MicroDYN test was incrementally valid in the high-school student sample (ΔR^2 = .03) but not significantly so in the university student sample. Basically, results indicate that in university student samples APS may address the same problem solving skills, which account for the incremental validity of the problem representation score in MicroDYN beyond reasoning (please note, this incremental validity of MicroDYN has been reported in a wide variety of studies, e.g., Greiff & Fischer, 2013; Wüstenberg et al., 2012; Greiff, Wüstenberg, et al., 2013). These findings demonstrate that there are aspects of problem solving competency that are not sufficiently addressed by MicroDYN. At the same time they question the incremental value of MicroDYN compared to traditional static measures of problem solving competency. Future research is needed to clarify the conditions an incremental value of MicroDYN depends upon. In general, APS seems to be even more closely related to school grades than MicroDYN is.

In summary, MicroDYN has proven to be a highly reliable and valuable tool for assessing a small range of problem solving skills (Funke, 2010) that are closely related to and not substantially different from APS (Fischer et al., in press). The high relations between MicroDYN and APS (beyond reasoning) may be interpreted as convergent validity of both approaches as measures of method-general problem solving skills: For instance, both measures may address skills like systematically building hypotheses based on complex information (i.e., *"understanding what is seen"*, Klahr, 2002, p.57). These skills seem to be helpful for coping with static problems (e.g., the APS test) as well as with interactive problems (e.g., the MicroDYN test). However, the findings reported also render static APS tests an interesting *alternative* to MicroDYN tests for assessing problem solving skills in adult samples (as in this sample there was no incremental value of MicroDYN beyond APS). Consequently, future research should (a) elaborate on the cognitive processes that account for an incremental value of the APS approach beyond the MicroDYN approach in more detail and may also (b) expand MicroDYN to address additional aspects of problem solving competency that are not addressed by current operationalizations of the MicroDYN approach or the APS approach.

4 Discussion and outlook

In the present thesis I elaborated on the Multiple Complex Systems approach to assessing problem solving skills (and its two instances MicroDYN and MicroFIN, see sections 2.1 and section 2.2) in more detail. Major findings and contributions include (1) a clear *conceptual understanding* of complex problem solving and the cognitive processes assessed by current operationalizations of the MCS approach (see sections 2.3 and 3.1), (2) empirical evidence for the *convergent validity* of MicroDYN tasks, MicroFIN tasks and tasks of the GeneticsLab test (see section 3.2) and (3) only limited support for *discriminant validity* of MicroDYN tasks with regard to static tasks of problem solving and reasoning (see section 3.3).

Both instances of the MCS approach seem to address a common set of skills that explains school grades in science even beyond logical reasoning, and there are close relations between the MCS approach and static problem solving tasks (that cannot be attributed to reasoning). With regard to the regression of school grades, an incremental value of the MCS approach (beyond reasoning and static problem solving tasks) was found in the high-school student sample, but not in the university-student sample (Fischer et al., in press). The static problem solving tasks were incrementally valid compared to reasoning and MicroDYN tasks. In summary, the MCS approach still has a lot of potential that waits to be explored in more detail. In the following sections some directions for future and ongoing research on the MCS approach will be outlined.

External Criteria. First of all, research on the MCS approach to assessing CPS should consider a wider range of external criteria for establishing external and construct validity. In many studies on the MCS approach, school grades have been applied as the major external criterion. This may have been a good starting point: School grades are easily assessed and can be conceived of as a highly aggregated measure of a person’s problem solving competency in multiple domains (e.g., the highly general Grade Point Average). Please note, in PISA 2012, problem solving – aggregated over multiple static and interactive tasks – proved to be highly related to school-related competencies like reading (r=.75), mathematics (r=.81) and science (r=.78) on a latent level (OECD, 2014).

However, as this step is taken for all current instances of the MCS approach (e.g., Wüstenberg et al., 2012; Greiff, Fischer et al., 2013), different external criteria may increasingly be considered in order to conclusively prove the validity of MCS instruments as measures of highly-domain general aspects of problem solving competency (cf. Süß, 1999). For instance, in previous studies, MicroDYN tasks proved to be predictive for performance in a complex in-basket simulation (Fischer & Funke, 2013) and different measures of performance in the space shuttle simulation HEIFI (Greiff, Wüstenberg, & Funke, 2012). In a similar vein, MicroDYN proved to explain performance in static problem solving tasks beyond fluid intelligence (as we reported in the third study of this thesis, cf. section 3.3). This approach of validation could be expanded by relating performance in MicroDYN and MicroFIN tasks to behavior and performance in highly-complex and/or high-fidelity simulations of real complex problems. Other external criteria with both a high relevance and a close conceptual relation to domain-general problem solving skills may be self-efficacy (e.g., Sherer, Maddux, Mercandante, Prentice-Dunn, Jacobs, & Rogers, 1982), mastery (e.g., Pearlin & Schooler, 1978), or measures of job performance of managers (or other agreed-upon experts in solving complex problems). External criteria of this kind may also be important for further investigating if MicroDYN tasks have an incremental value compared to static problems in adult samples (cf. section 3.3).

Critical Concerns. Over the past few years, many studies shed light on the MicroDYN approach, on what it measures, and on what it does not measure. Some of these aspects deserve careful consideration: For instance, given the findings of the third study of this thesis (see section 3.3) in university student samples, it may be more informative to apply traditional static problems than to use interactive MicroDYN tasks at least with regard to explaining academic success (Fischer et al., in press). Of course, MicroDYN in turn may be more informative for explaining different criteria (e.g., the interactive HEIFI problem, see above), or act as a criterion for other psychometric tests (similar to the study of Süß, 1999), but this remains a question yet to be answered empirically. Please note, these results do not question the validity of MicroDYN as a measure of problem solving skills, but only its incremental value compared to traditional (static) measures of Analytic Problem Solving (see section 3.3). Future research may demonstrate an incremental value of MicroDYN tasks for explaining other criteria or in other samples, but – due to the large empirical overlap of MicroDYN and static problem solving tasks reported in Fischer et al. (in press) – they may also fail to do so.

From a conceptual point of view MicroDYN tasks bear another shortcoming I would like to address: All MicroDYN tasks can be solved by applying the same narrow set of skills and strategies. Although this can also be seen as a benefit (with regard to reliability of MicroDYN tasks), it comes at a cost: Choosing a strategy adequate to the specific problem at hand (closely related to searching the model space of Burns & Vollmeyer, 1996) – possibly switching strategies between (sub-) problems – may be an interesting skill on its own. This skill may be involved in solving the first MicroDYN task of a test, but in later tasks this search may be heavily constrained by previous tasks. For instance, subjects may simply apply the same strategy again, instead of searching a promising strategy from scratch. Thus the average over performance in multiple MicroDYN tasks does not represent the skills in searching for a strategy well. Instead MicroDYN focuses on know-how concerning the search for information, hypotheses and solutions only (see section 2.3). APS tests (e.g., section 3.3) and current operationalizations of the MicroFIN approach (e.g., section 3.2), on the other hand, *do* require the search for adequate strategies for each single task. Each task within these tests requires different strategies to be applied and different assumptions to be made. From this point of view, it is interesting to note that in PISA 2012, MicroDYN tasks were also mixed with multiple finite automata, in order to assess “interactive problem solving” (OECD, 2014). Consequently, in all these instances (APS, MicroFIN and problem solving in PISA 2012), searching for a strategy adequate for solving the problem at hand, seems to be involved to a larger extend (compared to traditional MicroDYN tests). With regard to MicroFIN, this additional demand may even explain part of the method-specific covariance reported in the second study of this thesis (section 3.2; Greiff, Fischer, et al., 2013). However, Greiff, Fischer, et al. (2013, p.590) also report the method-specific factors of representation and solution in MicroFIN to be related to reasoning (with r=.43 for representation and r=.25 for solution) and they report no unique contributions of any method-specific factors to explaining school grades (when compared to all other method-specific factors and the method-general traits; Greiff, Fischer, et al., 2013, p.591). Additionally, even if MicroFIN captures an interesting additional aspect of problem solving competency, this aspect may be central for solving intelligence tests as well and there may not be a unique contribution beyond reasoning. In this case, one may argue, it may suffice to complement MicroDYN tasks with a separate intelligence test in order to cover the most important general aspects of problem solving competency. However, Greiff, Fischer, et al. (2013, p.589) also report correlations *between* method-specific factors (r = .02-.52). These correlations in turn may have artificially reduced unique contributions of MicroFIN's method-specific factors in the regression model reported. Thus, the assumption of an incremental value of the unique aspects of MicroFIN (which are as large as the commonalities between MicroFIN and MicroDYN) cannot be rejected conclusively. Also, as emphasized above, school grades are not the only relevant criterion for a test of problem solving skills. Thus, the incremental value of MicroFIN's unique demands remains another question to be answered empirically in future studies. In order to make adequate use of the heterogeneity inherent in MicroFIN tasks, it may be fruitful to address this issue.

Embracing Heterogeneity of CPS. As mentioned above, operationalizations of MicroFIN proved to be more heterogeneous in nature than operationalizations of MicroDYN (e.g., Greiff, Fischer, et al., 2013; Neubert et al., 2014; Greiff et al., 2014). This heterogeneity between minimally complex systems could be a benefit for addressing a larger range of skills relevant to solving complex and dynamic problems. In contrast to highly complex problems[12] multiple minimally complex problems can more easily be applied in order to address different well-defined problem solving skills. However, to make adequate use of the heterogeneity of MicroFIN tasks, a *heterogeneous set* of tasks (representing a broad range of demands inherent in different kinds of complex dynamic problems) with multiple *homogenous clusters* (i.e., multiple tasks sharing the same constellation of demands) has to be applied.

For example, one of these clusters may well be described by current operationalizations of the MicroDYN-approach – as problem representation in MicroDYN tasks is a factor that already proved to be reliably indicated by multiple homogenous tasks and predictive for external criteria beyond logical reasoning. Having tasks of this kind – representing one homogenous cluster of tasks – and knowing how they work empirically, is an important first step in research on the assessment of complex and dynamic problem solving. But given the large heterogeneity of MicroFIN tasks (and complex dynamic problems in general, cf. Fischer et al., 2012), these tasks should definitely be complemented by additional clusters, in order to address the full spectrum of demands inherent in different kinds of complex dynamic problems (for an overview see Dörner, 1996).

[12] e.g., the simulations Tailorshop, Power Plant, or Learn – who also proved to be heterogeneous with regard to the cognitive demands they pose on the problem solver (see Süß, 1999).

Examples of how additional (creative) skills can be assessed by means of finite automatons can be found in the PISA 2012 operationalization of problem solving skills (see OECD, 35ff): For instance, one question on a simulated MP3-Player (with three buttons: ◄, ◉, ►) may be stated as follows: “Describe how you could change the way the MP3-player works so that there is no need to have the bottom button (◄). You must still be able to change the type of music, and increase or decrease the volume and the bass level”, OECD, 37). Questions like these may help exploiting additional aspects of the potential inherent in the MCS approach which may be useful to a wider range of assessment contexts.

Of course, from a practitioner’s point of view, the addition of tasks with additional cognitive demands may not be a goal in itself (for it may not be feasible to apply a test that addresses all the demands possibly inherent in every kind of complex problem). However, research on the assessment of CPS has to highlight and address the multitude of demands inherent in different complex problems (e.g., Dörner, 1996), in order to adequately determine (a) the structure of problem solving competency and (b) the best and most valid indicators for each important factor of this structure. Taxonomies of CPS (like the taxonomy proposed by Wagener, 2001), may propose a conceptual starting point for systematically looking for characteristic task demands that may complement the task demands inherent in MicroDYN tasks. For example, other finite state machines may include time-dependent transitions, feedback-loops, random shocks, interactive and nonlinear relations between variables, etc. (cf. Dörner 1996). Addressing this issue in detail is beyond the scope of this thesis, but it should definitely be elaborated on in future studies.

Formal Models of Complex Problem Solving. Cognitive modelling of concrete complex problems (Anderson, 2007) may be applied to further inquire about the individual differences (i.e., person-specificity) and the domain-generality of processes involved in solving concrete complex problems (Fischer et al., 2012). Computational models require specific assumptions about every detail of the process. Thus, specifying models[13] of different problem solving strategies, and applying them to a heterogeneous set of complex problems (e.g., constructed according to the MicroDYN or MicroFIN approach) may reveal information about the degree of domain-generality for each combination of strategies. As we outlined in section 2.3, some cognitive processes (e.g., instance-based learning, Gonzalez, Lerch, & Lebiere, 2003) are applicable and useful in a wide range of complex problem situations, whereas others – e.g., applying the "Vary One Thing At a Time" (VOTAT) strategy (cf. Tschirgi, 1980) – may be useful only in a small subset of problems (e.g., traditional MicroDYN tasks). If the behavior predicted by different cognitive models is compared to participants' behavior (e.g., mouse-clicks while working on a MicroFIN task) it may also be possible to identify individual differences in the process of CPS (instead of just fitting one model to data aggregated over multiple participants). See Scheibehenne, Rieskamp, & Wagenmakers (2013) for a concrete example of how a "cognitive-toolbox"-approach like the one proposed by Fischer et al. (2012) could be rigorously tested on both individual and aggregate levels. Modelling individual problem solving behavior with regard to different complex problems may deepen our understanding of CPS and may allow for systematically going beyond the general level of understanding proposed by Fischer et al. (2012).

Final remarks. After having talked a lot about the MCS approach to assessing important problem solving skills – and how it *differs* from previous approaches – let me end this thesis by relating our conception back to other approaches of CPS research. For even if I am convinced that our approach makes a unique contribution to the assessment of problem solving skills, In my opinion it is not simply "better" (or worse) than other approaches. Rather, I want to emphasize that no approach proposed so far is able to capture the whole picture of Complex Problem Solving and thus it may be wiser to complement different approaches instead of deciding for one single approach only.

For instance, *domain-specific* knowledge (and more general skills like building analogies to prior knowledge, which may be an important aspect of Analytic Problem Solving, cf. section 3.3; Fischer et al., in press) is an important aspect of problem solving in each domain (Fischer et al., 2012). This aspect is not explicitly addressed by current instances of the MCS approach. Additionally, there are a lot of highly domain-general skills that are relevant for many kinds of CPS but not addressed by current instances of the MCS approach (Fischer et al., in press; Dörner, 1996).

In real life problem solving (in every domain) *both* domain-specific and domain-general skills and abilities (e.g., intelligence, cf. Wittmann & Hattrup, 2004) have to be integrated effectively. Also, in real life there is no artificial separation of phases and no explicit instruction that tells a problem solver when to start and when to stop representing the problem. Instead, the skills assessed by the MCS approach have to be integrated in an adaptive way (as described in the first paper of this thesis, in section 3.1). As this kind of

[13] The idea of crossing multiple models with multiple finite state machines evolved in a discussion with Dr. Daniel Holt.

adaptive integration strongly depends on various features (e.g., features of the situation as well as of the problem solver, cf. Fischer et al., 2012) highly complex and realistic problems may be *required* to assess it (e.g., Fischer & Funke, 2013). Again, it may be wiser to complement well-validated highly complex problems with the MCS approach instead of substituting the former with the latter. For instance, Paul Baltes and colleagues (Baltes & Smith, 1990; Baltes & Staudinger, 2000; Staudinger & Baltes, 1996) developed a well-validated approach to assess human wisdom[14] based on performance in complex and fundamental life problems: Baltes and colleagues scored solutions to hypothetical life problems on five criteria: Besides (1) rich declarative and (2) procedural knowledge about the fundamental pragmatics of life (i.e., knowledge of goals and how to reach them, cf. Baltes & Smith, 1990), wise solutions acknowledge complexity by considering (3) multiple relevant contexts (lifespan contextualism) and (4) multiple possible perspectives, goals and values implicit in the problem description (value relativism), as well as (5) the uncertainty resulting from participants' bounded rationality (uncertainty management). Several authors have noticed conceptual relations between these aspects of wisdom and complex problem solving (Dörner, 1986; Staudinger & Baltes, 1996; Maercker, 1995; Baumann & Linden, 2008; Sternberg, 2007) and it seems to me that each of Baltes' criteria is relevant to many instances of problem solving in real life (where problem solvers often have to consider multiple intra- and interpersonal goals, to transcend the initial problem description by exploring various contexts, and to manage the uncertainty inherent in characteristic instances of CPS). Most of these aspects are not sufficiently addressed by current operationalizations of the MCS approach.

The assessment of problem solving skills has always been and will always be a complex endeavor as an adequate selection of assessment instruments strongly depends on a variety of factors. Based on the findings of this thesis I consider the MCS approach a reliable and valid assessment instrument addressing a narrow set of problem solving skills (with an incremental value for explaining science grades and Analytic Problem Solving over and above fluid reasoning). But even if these skills may be central to many instances of CPS, they are not sufficient for every kind of CPS and I want to emphasize that no single reliable measure of skills is likely to address *all the skills* possibly relevant to all

[14] In Western research traditions, "wisdom" can be understood as knowledge and deep understanding of the most important truths – i.e., "knowledge about what is good and right for humans" (Baltes & Freund, 2003, p.251). Baltes & Smith (1990) defined wisdom as a psychological construct to be expert knowledge in the domain of the fundamental pragmatics of life (i.e., knowledge of important goals and how to reach them) that allows for exceptional judgment and advice concerning existential life problems.

kinds of complex problems. As I put it in Fischer et al. (2012, p. 37): “The more we learn about the process of problem solving, the more we have to acknowledge the complexity of both the process and the kind of problems that are involved in realistic problem solving in naturalistic environments.”

References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369-406.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford: Oxford University Press.

Anderson, J. R. (2013). *Kognitive Psychologie [Cognitive psychology]* (7th edition). Heidelberg: Springer.

Autor, D.H., Levy, F. & Murnane, R.J. (2003). The skill content of recent technological change: an empirical exploration. *The Quarterly Journal of Economics, 118*, 1278-1333.

Baltes, P. B., & Freund, A. M. (2003). The intermarriage of wisdom and selective optimization with compensation: Two meta-heuristics guiding the conduct of life. In Keyes, Corey L. M., & Haidt, Jonathan (Eds.), *Flourishing: Positive psychology and the life well-lived (pp. 249-273)*. Washington: American Psychological Association.

Baltes, P. B., & Smith, J. (1990). Weisheit und Weisheitsentwicklung: Prolegomena zu einer psychologischen Weisheitstheorie. *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie, 22(2),* 95-135.

Baltes, P. B., & Staudinger, U. M. (2000). Wisdom: a metaheuristic (pragmatic) to orchestrate mind and virtue toward excellence. *American psychologist*, *55(1),* 122-136.

Baumann, K., & Linden, M. (2008). *Weisheitskompetenzen und Weisheitstherapie: die Bewältigung von Lebensbelastungen und Anpassungsstörungen [wisdom competencies and wisdom therapy: coping with life burdens and adjustment disorders].* Lengerich: Pabst Science Publishers.

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

Bortz, J. (2005). *Statistik: Für Human-und Sozialwissenschaftler [Statistics for human- and social-scientists]* (6th edition). Heidelberg: Springer.

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

Buchner, A., Funke, J., & Berry, D. C. (1995). Negative correlations between control performance and verbalizable knowledge: Indicators for implicit learning in process control tasks? *The Quarterly Journal of Experimental Psychology, 48,* 166-187.

Bühner, M. (2006). *Einführung in die Test-und Fragebogenkonstruktion [Introduction to test and questionaire construction]* (2nd extended edition). München: Pearson.

Burns, B. D., & Vollmeyer, R. (1996). Goals and problem solving: Learning as search of three spaces. In G.W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Meeting of the Cognitive Science Society (pp. 23-24).* Mahwah, NJ: Erlbaum.

Cronbach, L.J. (1961). *Essentials of psychological testing (2nd edition).* New York: Harper & Row.

Damasio, A. R. (1994). *Descartes' Irrtum. Fühlen, Denken und das menschliche Gehirn [Descartes' error. Feeling, thinking and the human brain].* München: List.

Dörner, D. (1976). *Problemlösen als informationsverarbeitung [Problem solving as information processing].* Stuttgart: Kohlhammer.

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

Dörner, D. (1996). *The logic of failure: Recognizing and avoiding error in complex situations.* New York: Metropolitan Books.

Dörner, D., Bartl, C., Detje, F., Gerdes, J., Halcour, D., Schaub, H., & Starker, U. (2002). *Die Mechanik des Seelenwagens [mechanics oft he soul wagon]*. Bern: Huber.

Dörner, D., & Güss, C. D. (2013). PSI: A computational architecture of cognition, motivation, and emotion. *Review of General Psychology, 17*, 297–317.

Dörner, D. & Kreuzig, H.W. (1983). Problemlösefähigkeit und Intelligenz [Problem solving ability and intelligence]. *Psychologische Rundschau, 34,* 185-192.

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

Duncker, K. (1945). On problem solving. *Psychological Monographs*, *58(5)*.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processings. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 4(2),* 59-74.

Fischer, A., & Funke, J. (2013). *Problem solving competency - More than the sum of its parts?*. Presented at the EARLI Conference, München, August 27th-31st, 2013.

Fischer, A., Greiff, S., & Funke, J. (2012). The Process of Solving Complex Problems. *Journal of Problem Solving, 4,* 19-42.

Fischer, A., Greiff, S., & Funke, J. (under review). The history of complex problem solving. In B. Csapó, J. Funke, & A. Schleicher (Eds.), *The nature of problem solving.* Paris: OECD.

Fischer, A., Greiff, S., Wüstenberg, S., Fleischer, J., Buchwald, F., & Funke, J. (in press). Assessing analytic and interactive aspects of problem solving competency. *Learning and Individual Differences, DOI:10.1016/j.lindif.2015.02.008.*

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

Funke, J. (2003). *Problemlösendes Denken [Problem solving thinking].* Stuttgart: Kohlhammer.

Funke, J. (2010). Complex problem solving: a case for complex cognition?. *Cognitive Processing, 11(2),* 133-142.

Funke, J. & Buchner, A. (1992). Finite Automaten als Instrumente für die Analyse von wissensgeleiteten Problemlöseprozessen: Vorstellung eines neuen Untersuchungsparadigmas [Finite automata as instruments for the analysis of knowledge-driven problem solving processes: presentation of a new paradigm]. *Sprache & Kognition, 11,* 27-37.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology, 62, 451-482.*

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychological review, 109(1),* 75-90.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27(4),* 591-635.

Greiff, S. (2012). *Individualdiagnostik komplexer Problemlösefähigkeit [Individual diagnostic of complex problem solving ability].* Münster: Waxmann Verlag.

Greiff, S. & Fischer, A. (2013). Der Nutzen einer komplexen Problemlösekompetenz: Theoretische Überlegungen und empirische Befunde [Usefulness of a complex problem solving competency: Theoretical considerations and empirische results]. *Zeitschrift für Pädagogische Psychologie, 27,* 27-39.

Greiff, S. & Fischer, A. (2013b). Measuring Complex Problem Solving: An educational application of psychological theories. *Journal of Educational Research Online, 5,* 38-58.

Greiff, S., Fischer, A., Stadler, M., & Wüstenberg, S. (2014). Assessing complex problem-solving skills with multiple complex systems. *Thinking & Reasoning (ahead-of-print),* 1-27.

Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M., & Martin, R. (2013). A multitrait–multimethod study of assessment instruments for complex problem solving. *Intelligence, 41,* 579-596.

Greiff, S., & Funke, J. (2010). Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal komplexer Systeme [Systematic reearch on complex problem solving research with minimally complex systems]. *Beiheft der Zeitschrift für Pädagogik, 56*, 216-227.

Greiff, S., Kretzschmar, A., Müller, J. C., Spinath, B., & Martin, R. (2014). The Computer-Based Assessment of Complex Problem Solving and How It Is Influenced by Students' Information and Communication Technology Literacy. *Journal of Educational Psychology, 106(3),* 666-680.

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic Problem Solving A New Assessment Perspective. *Applied Psychological Measurement, 36(3),* 189-213.

Greiff, S., Wüstenberg, S., Molnár, g., Fischer, A., Funke, J., & Csapó, B. (2013) Complex Problem Solving in Educational Contexts – Something Beyond g: Concept, Assessment, Measurement Invariance, and Construct Validity. *Journal of Educational Psychology, 105,* 364-379.

Horn, J. & Masunaga (2006). A merging theory of expertise and intelligence. In K.A. Ericsson, N. Charness, P.J. Feltovich, R.R. Hoffman (Eds.) *Handbook of Expertise and Expert Performance (pp. 587-612).* Cambridge: Cambridge University Press.

Klahr, D. (2002). Scientific Discovery as Problem Solving In D. Klahr (Ed.), *Exploring Science (pp. 21 – 40)*. Massachusetts: MIT Press.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive science, 12(1),* 1-48.

Klahr, D., & Dunbar, K. (2002). A Paradigm for Investigating Scientific Discovery in the Psychological Lab In D. Klahr (Ed.), *Exploring Science (pp. 41 – 61)*. Massachusetts: MIT Press.

Klieme, E., Hartig, J., & Wirth, J. (2005). Analytisches Problemlösen: Messansatz und Befunde zu Planungs-und Entscheidungsaufgaben. Problemlösekompetenz von Schülerinnen und Schülern [Analytic problem solving: measurement approach and results concerning planning- and decision problems]. In E. Klieme, D. Leutner, J. Wirth (Eds.), *Problemlösekompetenz von Schülerinnen und Schülern (pp. 37-54).* Wiesbaden: VS. Wiesbaden: VS.

Kluwe, R.H., Misiak, C., & Haider, H. (1991). The control of complex systems and performance in intelligence tests. In H. Rowe (Ed.), *Intelligence: reconceptualization and measurement (pp. 227-244).* Hillsdale: Lawrence Erlbaum Associates.

Leutner, D., Fleischer, J., Wirth, J., Greiff, S., & Funke, J. (2012). Analytische und dynamische Problemlösekompetenz im Lichte internationaler Schulleistungsvergleichsstudien [Analytic and dynamic problem solving competency in light of international school comparisons]. *Psychologische Rundschau, 63(1),* 34-42.

Maercker, A. (1995). *Existentielle Konfrontation: Eine Untersuchung im Rahmen eines psychologischen Weisheitsparadigmas (Studien und Berichte Nr. 62).* Berlin: Max Planck Institute for Human Development.

Meadows, D., Randers, J., & Meadows, D. (2004). *Limits to growth: the 30-year update.* Vermont: Chelsea Green Publishing.

Mintzberg, H. (1987) 'The Strategy Concept 1: Five Ps of Strategy,' *California Management Review, 30(1),* 11-24.

Molnár, Gy., Greiff, S., Csapó, B. (2013): Inductive reasoning, domain specific and complex problem solving: relations and development. *Thinking skills and Creativity, 9(8),* 35-45.

Neubert, J. C., Kretzschmar, A., Wüstenberg, S., & Greiff, S. Extending the Assessment of Complex Problem Solving to Finite State Automata (2014). *European Journal of Psychological Assessment,* 1-14.

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

OECD (2004). *Problem solving for tomorrow's world. First measures of cross-curricular competencies from PISA 2003.* Paris: OECD.

OECD (2014). *PISA 2012 Results: Creative Problem Solving. Students' skills in tackling real-life problems (Volume V).* Paris: OECD.

Osman, M. (2010). Controlling uncertainty: a review of human behavior in complex dynamic environments. *Psychological Bulletin, 136(1),* 65-141.

Pearlin, L. I., & Schooler, C. (1978). The structure of coping. *Journal of health and social behavior,* 19, 2-21.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and instruction, 12(1),* 61-86.

Putz-Osterloh, W. (1981). Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg [On the relationship between test intelligence and problem solving performance]. *Zeitschrift für Psychologie, 189,* 79-100.

Rey, G. D., & Fischer, A. (2013). The expertise reversal effect concerning instructional explanations. *Instructional Science, 41(2),* 407-429.

Sager, S., Barth, C. M., Diedam, H., Engelhart, M., & Funke, J. (2011). Optimization as an analysis tool for human complex problem solving. *SIAM journal on optimization, 21(3),* 936-959.

Schaub, H. (2013). Der Mensch als Problem und Problemlöser. Der Systemische Ansatz für Führung und Krisenmanagement. In: C. Unger, T. Mitschke , D. Freudenberg (Eds.) *Krisenmanagement – Notfallplanung – Bevölkerungsschutz (pp.193-209)*. Berlin: Duncker & Humblot.

Scheibehenne, B., Rieskamp, J., & Wagenmakers, E. J. (2013). Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological review, 120(1),* 39–64.

Scherer, R., & Tiemann, R. (2012). Factors of problem-solving competency in a virtual chemistry environment: The role of metacognitive knowledge about strategies. *Computers & Education, 59(4), 1199-1214.*

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

Schunn, C. & Klahr, D. (2002). Multiple-Space Search in a More Complex Discovery Microworld. In D. Klahr (ed.), *Exploring Science (pp. 161-199).* Massachusetts: MIT Press.

Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The self-efficacy scale: Construction and validation. *Psychological reports, 51(2),* 663-671.

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

Staudinger, U. M., & Baltes, P. B. (1996). Weisheit als Gegenstand psychologischer Forschung. *Psychologische Rundschau, 47(2),* 1-21.

Sternberg, R. J. (2007). A systems model of leadership: WICS. *American Psychologist, 62(1),* 34-42.

Strohschneider, S., & Güss, D. (1999). The fate of the Moros: A cross-cultural exploration of strategies in complex and dynamic decision making. *International Journal of Psychology, 34(4),* 235-252.

Strunk, G., & Schiepek, G. (2006). *Systemische Psychologie. Eine Einführung in die komplexen Grundlagen menschlichen Verhaltens.* Heidelberg: Spektrum Akademischer Verlag.

Süß, H. M. (1999). Intelligenz und komplexes Problemlösen. *Psychologische Rundschau, 50(4),* 220-228.

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child development, 1-10.*

Vester, F. (2012). *The art of interconnected thinking* (2nd ed.). München: MCB Publishing House.

Wagener, D. (2001). *Psychologische Diagnostik mit komplexen Szenarios.* Lengerich: Pabst.

Weaver, W. (1948). Science and Complexity. *American Scientist,* 36, 536-544.

Wirth, J., & Klieme, E. (2003). Computer-based assessment of problem solving competence. *Assessment in Education: Principles, Policy & Practice, 10(3),* 329-345.

Wittmann, W. W., & Hattrup, K. (2004). The Relationship Between Performance in Dynamic Systems and Intelligence. *Systems Research and Behavioral Science, 21,* 393-409.

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.), *The future of learning and individual differences research: Process, traits, and content (pp. 77-108)*. Washington: American Psychological Association.

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

Declarations in accordance with § 8 (1) b and § 8 (1) c of the regulations for doctoral degrees of the Faculty of Behavioural and Cultural Studies of Heidelberg University.

§ 8 (1) b
I declare that I wrote this thesis independently using only the aids specified and that I have correctly referenced all quotations.
§ 8 (1) c
I declare that I have not submitted this thesis in this or any other form as an examination paper and that I have not submitted it to any other faculty.

Heidelberg, March 2015
Dipl.-Psych. Andreas Fischer

Appendix

I. Manuscript

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

II. Manuscript

Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M., & Martin, R. (2013). A multitrait–multimethod study of assessment instruments for complex problem solving. *Intelligence, 41,* 579-596.

III. Manuscript

Fischer, A., Greiff, S., Wüstenberg, S., Fleischer, J., Buchwald, F., & Funke, J. (in press). Assessing analytic and interactive aspects of problem solving competency. *Learning and Individual Differences*, DOI:10.1016/j.lindif.2015.02.008.

The Process of Solving Complex Problems

Andreas Fischer, Samuel Greiff, and Joachim Funke[1]

Acknowledgements

This research was funded by a grant of the BMBF (FKZ: 01JG1062). We gratefully thank Sascha Wüstenberg and Daniel Holt for their comments.

Correspondence concerning this paper should be addressed to Samuel Greiff, Department of Psychology, Heidelberg University, Hauptstraße 47-51, D-69117 Heidelberg, Germany. Phone: +49 6221 547613. Fax: +49 547273. E-mail: Samuel.Greiff@psychologie.Uni-Heidelberg.de

Abstract

This article is about Complex Problem Solving (CPS), its history in a variety of research domains (e.g., human problem solving, expertise, decision making, and intelligence), a formal definition and a process theory of CPS applicable to the interdisciplinary field. CPS is portrayed as (a) knowledge acquisition and (b) knowledge application concerning the goal-oriented control of systems that contain many highly interrelated elements (i.e., complex systems). The impact of implicit and explicit knowledge as well as systematic strategy selection on the solution process are discussed, emphasizing the importance of (1) information generation (due to the initial intransparency of the situation), (2) information reduction (due to the overcharging complexity of the problem's structure), (3) model building (due to the interconnectedness of the variables), (4) dynamic decision making (due to the eigendynamics of the system), and (5) evaluation (due to many, interfering and/or ill-defined goals).

Keywords

Complex Problem Solving, CPS, Operative Intelligence, Dynamic Problem Solving, Dynamic Decision Making, Expertise

---

[1]University of Heidelberg, Germany

1. Introduction

In times of increasing globalization and technological advances, many problems humans have to face in everyday life are quite complex, involving multiple goals as well as many possible actions that could be considered, each associated with several different and uncertain consequences, in environments that may change dynamically and independent of the problem solvers' actions (Funke, 2003). In order to solve complex problems, people usually have to acquire and to apply knowledge about complex systems concerning the systems' structure and dynamics (Funke, 2001). Examples for Complex Problem Solving (CPS) are easily found, e.g., using unknown complex technical devices (like a new mobile phone, a computer, a vending machine, etc.), managing complex organizations (like corporations or communities) or making predictions in complex environments (like forecasts of the weather, political elections or the stock market, etc.). In research on human problem solving CPS is a matter of interest since the 1970s, when there was a shift of emphasis from simple, static, well-defined and academic problems (like the Tower of Hanoi or items of classical intelligence tests), to more complex, dynamic, ill-defined, and realistic problems (Wenke, Frensch, & Funke, 2005). Since then, research on human problem solving focused on interviewing experts of certain knowledge domains, on studying the effects of expertise on problem solving activities and decision making, or on simulating complex problems[1] based on real systems humans could have to deal with in their daily lives (like planning a day, managing an organization, fire fighting, and so on). Along with more complexity in research on problem solving new questions arose: How does expertise and prior knowledge influence problem solving in complex situations? Are there certain strategies especially useful for coping with complex problems? How is a complex situation represented in the human mind with its restricted capabilities? Which facets of intelligence are most important for solving complex problems? Some of these questions were addressed by different fields of research (e.g., research on problem solving, on expertise, on information reduction, on decision making, and research on intelligence), but in spite of a lot of fruitful research on CPS in these areas, up to now most of this research has been conducted with a focus on empirical data mining rather than theoretical considerations (see Funke, 2010), without a clear-cut definition (see Quesada, Kintsch, & Gomez, 2005) commonly accepted in the scientific community.

The article at hand wants to contribute to the solution of this shortcoming: After summarizing the most important empirical and theoretical contributions to the field, we want to come up with a process theory of CPS based on a formal definition, applicable to the interdisciplinary field. We want to consider (a) what is known about the most im-

[1] Osman (2010) refers to the complex scenarios of this kind as "Complex Dynamic Control Tasks" and points out that these tasks are known in the fields of CPS, dynamic decision making, naturalistic decision making, and process control amongst others.

portant determinants of the process of CPS in the different domains of research (such as *expertise*, *decision making*, and *intelligence*) and (b) how these contributions fit together if viewed under an integrative perspective.

2. What is meant by Complex Problem Solving?

Research on CPS produced a lot of characterizations and operationalizations of complex problems (for an overview see Frensch & Funke, 1995), but up to now there has not been a definition of complex problems commonly accepted in the scientific community (Quesada et al., 2005). There is an ongoing debate about (a) what should be considered complex in CPS and (b) how complexity might be measured in detail (see Quesada et al., 2005 for a discussion).

The definition of CPS proposed and applied in this article is based on the constitutive concepts "complexity", "problem", and "problem solving" which in turn are understood as follows:

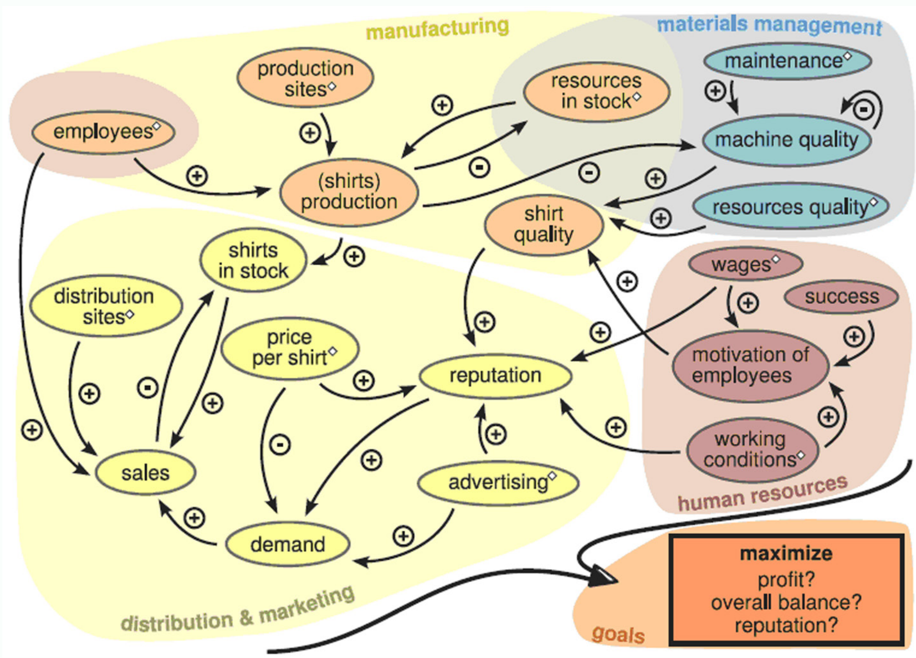

Figure 1. *The structure of the CPS scenario TAILORSHOP, with the positive and negative dependencies between the influential variables. Diamonds represent the participant's control possibilities. (Engelhart, Funke & Sager, 2011)*

1. The *complexity* of a system[2] may be defined as the number of elements and relations of the system (see Funke, 1985). As Dörner (1989) stated, "the complexity of a domain of reality is the higher, the more features there are and the more these features are interdependent" (Dörner, 1989, p. 60, translated by the authors).
2. A *problem* is considered to exist, "when a living creature has a goal but does not know how this goal is to be reached. Whenever one cannot go from the given situation to the desired situation simply by action, then there has to be recourse to thinking" (Duncker, 1945, p.1). Dörner has gone into more detail when he emphasized that "barriers" between the given situation and the desired goal state, i.e., the lack of knowledge, can be further classified according to the amount of (a) ignorance of the means/operators applicable, and (b) lack of concreteness concerning the goal state (see Dörner, 1976, or Funke, 2003).
3. *Problem solving* can be defined as successfully searching for an operation or a series of operations in order to transfer the given actual state of the system to a *goal state* (Newell & Simon, 1972; Dunbar, 1998).

Based on these three concepts, CPS can be defined as a kind of problem solving, with the problem itself (the structure of (a) the external problem representation and/or (b) the mental representation of the problem), or the process of its solution having to be formalized as a set of many highly interrelated elements, i.e., a complex system. According to Halford, Wilson and Phillips (1998) the complexity of relations can be quantified by the number of variables related to each other: For example, the mental representation of a criterion y depending on a predictor x could be expressed as a binary relation r(y,x), whereas a dependency on multiple predictors could be represented as a relation of higher rank, e.g., the ternary relation r(y,x1,x2), and thus would be considered more complex. Structures more complex than quaterny relations are assumed to have to be processed by either conceptual chunking or segmentation in order to not exceed human processing capacity (Halford et al., 1998).

One famous example for CPS—that can be considered complex because the structure of the external problem representation (see Figure1) is to be formalized as a complex system—is the TAILORSHOP (see, e.g., Funke, 2003), a computer simulated scenario of a small organization involved in shirt production. Originally programmed by Dörner in the 1980s on his calculator it was implemented on many platforms and used in a variety of contexts.

In this scenario, the problem solver takes the role of managing a small tailorshop, deciding what actions to take or what information to gather, aiming at the maximization of the capital at the end of each simulated month (a global goal which is dependent on a set of conflicting subgoals). On an abstract level, the structure of the TAILORSHOP scenario is formalized as a complex dynamic[3] system, consisting of many highly interrelated vari-

[2] "A system is understood as an entity functionally closed and separated from its environment, consisting of elements which are in interaction with each other. Systems can be open to processes of exchange with their environment. Depending on the depth of system analysis there can be different hierarchical layers discriminated within systems and heterarchical interactions between systems." (Strunk & Schiepek, 2006, p.102)

ables (see Funke, 2003). In the literature on CPS, it is mostly the structure of the external problem representation that is considered complex. So a problem usually is considered being of a certain complexity, even if it might seem less complex to problem solvers with more expertise (as well as it is considered being of a certain difficulty, independent of the ability of a problem solver). This view is essential in order to understand the research on some of the most noteworthy aspects of CPS: For instance, using parsimonious but viable heuristics (see research on *decision making strategies*) and representations (see research on *information reduction*) are often considered most important for coping with complex problems (see Gigerenzer & Brighton, 2009; Gonzalez & Lebiere, 2005; Klauer, 1993). To efficiently cope with complex problems using adequate heuristics and representations (see research on *intelligence*) the problem solver has to either use or acquire sufficient implicit or explicit knowledge about the problem (see research on *expertise*). Especially when the problem is not presented as a set of nameless and abstract variables, but embedded in a plausible semantic context (like the TAILORSHOP, described above), prior knowledge about the elements to focus on or about the strategies to apply best, helps in reducing the problem space that has to be searched through for a solution to the problem (see research on *human problem solving*).

On the following pages we will review what is known about these most important aspects of CPS, and how it fits together in an integrative process theory of CPS. Therefore we will review and summarize findings of five fields of research that have contributed most to the understanding of CPS: (a) human problem solving, (b) expertise, (c) decision making strategies, (d) information reduction and (e) intelligence.

3. Human Problem Solving

The most general conception of problem solving up to now, which might as well be expanded and applied to CPS, has been Newell's and Simon's (1972) Theory of Human Problem Solving. The theory was proposed to explain findings on simple static and well-defined problems not as complex as the TAILORSHOP, but Newell and Simon already addressed all the aspects necessary to solve problems of arbitrary complexity.

Following the authors, some of the most important aspects of human problem solving may be summarized as follows:

1. Human problem solving starts with constructing an internal representation of the external problem statement, a "problem space" (i.e., a set of possible states of the problem, given the initial state, the applicable operators, and certain goal states). Which operators can be considered applicable might be different for problem solvers of different expertise and intelligence (see Newell & Simon,

[3] A dynamic system is a system, that contains a vector of variables, that is dependent on former states of the same vector, e.g., $\mathbf{Y}(t) = f(\mathbf{Y}(t-1))$ (see Funke, 1985, p.4)

1972).

2. Given an internal representation of the problem, a method for reaching the current goal is being searched for. General searching algorithms (like "hill-climbing", or "means-end-analysis") are distinguished from more domain specific methods (like "take the hammer to get the nail into the wall")
3. Using a method can change the external problem as well as the internal representation. Of course, changes in the environment or the consequences of a method may lead to new (sub-)problems or new possible solutions. Methods also can be aborted when metacognitive processes do interfere. When a method does not lead to a goal state, (1) another method can be tried, (2) the internal representation may be changed, i.e., the problem may be reformulated, or (3) the attempt of solving the problem may be aborted.

When it comes to CPS constructing a parsimonious but viable internal representation is far from trivial (in contrast to the problems Newell and Simon used in their studies, where a correct internal representation is usually assumed to be given). Usually a problem solver has to actively acquire knowledge about the complex problem by systematically interacting with it (see Funke, 2001) as the initial assumptions about the structure of the problem are mostly false or incomplete (Dörner, 1989). Often the problem solver has to define one or more of the problem's components him- or herself based on aspects like prior knowledge (e.g., experience with analogous problems, or generalized schemas for this kind of problems) and features of the task (Novick & Bassok, 2005) and usually building a viable internal representation of a complex problem involves processes like rule induction (Simon & Lea, 1974), generating and testing hypotheses (Klahr & Dunbar, 1988) and causal learning (Buehner & Cheng, 2005).

Fortunately, there are some theories that elaborated on certain aspects of knowledge acquisition in more detail (e.g., explaining when active information generation takes place and how it leads to better representations): Ohlsson (1992) proposed a Theory of Representational Change. When the current problem representation does not cue the operators sufficient to solve the problem, Ohlsson speaks of an "impasse". An impasse can be broken, when the problem representation is changed, as a different problem representation might cue other concepts in long-term memory. Representational change may occur in different ways: (1) Elaboration of (or search for additional) information about the problem; (2) constraint relaxation, i.e., removing inhibitions on what is regarded as permissible; (3) re-encoding, i.e., reinterpreting the problem representation. With his theory of representational change he emphasized the importance of a viable problem representation for solving problems and thus elaborated on an aspect of special importance to CPS. MacGregor, Ormerod, and Chronicle (2001) proposed that changes in the problem representation and in strategy use may occur due to monitoring processes, when the rate of progress is perceived to be too slow to solve the problem in time. According to their theory, it is not as much the impasse, but the perception of an impasse (or even an unacceptable slow-down in progress) that leads to phenomena like restructuring, considering new operators and insight.

Simon and Lea (1974) have further elaborated on Newell's and Simon's (1972) concept of problem space in a way that also proved to be fruitful for the CPS research. They conceptualized the problem space as divided into a rule-space (containing possible rules of the problem) and an instance-space (containing possible states of the problem) with information in each space guiding the search in the other space (see also Klahr & Dunbar, 1988, for an extension and application of the dual-search concept to the complex field of scientific discovery). This conception sheds light on how instances and rules of the problem are explored (i.e., how a solid representation of a complex problem is built) and can be considered fundamental in modeling the influence of knowledge about instances and structural knowledge on problem solving as it is considered by research on the influence of expertise on CPS.

In summary, information processing theories on human problem solving have proposed some useful ideas and assumptions that are most relevant when building a process theory of CPS. E.g., they try to explain when information generation and elaboration takes place, how it leads to viable internal representations (or models) of the problem system, and how the internal representation of the problem determines the solution strategies applicable. Especially the distinction of structural knowledge and knowledge about instances proved to be very fruitful for thinking about the influence of expertise on CPS. The next section will further elaborate on this distinction, and propose the most influential theories on how different kinds of knowledge may influence the process of CPS.

4. Expertise

There is a large quantity of research on differences between experts and novices of a certain knowledge domain concerning the influence of different kinds of domain-specific knowledge on CPS. In fields as different as reading, writing, arithmetic, mechanics, policies, jurisdiction, management, or debugging (for an overview see Sternberg & Frensch, 1991) there has been a lot of research on the processes and kinds of knowledge involved in CPS. What could have seemed to be a turning away from general aspects of problem solving in favor of more domain-specific problem solving strategies nonetheless produced a deep insight in some general effects of expertise on general problem solving.

For instance:

- Experts can (a) apprehend a greater number of elements in working memory and (b) retain these elements for a longer duration, when the elements are part of a meaningful configuration within their domain of expertise (see expertise

wide-span memory; Horn & Blankson, 2005).
- Experts classify problems according to deep features, relevant to the solution, rather than superficial features (Chi, Feltovich, & Glaser, 1981);
- There are differences in the semantic memory of experts, compared to novices regarding, e.g., the associations between concepts (Chi et al., 1981);
- Experts are faster in solving problems if they are asked to do so (Chi, Glaser, & Rees, 1982);
- Experts are more precise if not working under uncertainty (Johnson, 1988);
- Experts seem to have better metacognitive abilities (like self-monitoring) (Larkin, 1983).

Consequently the process of gaining expertise concerning a certain problem via building explicit and implicit representations of the problem at hand was of special interest to the research community and stimulated a lot of interesting theoretical and empirical results on CPS as will be outlined in the next paragraphs.

One of the most influential theories on (1) gaining explicit declarative knowledge and on (2) the effects of expertise on problem solving and learning is John Sweller's *Cognitive Load Theory* (CLT). Sweller (2005) assumed that the human cognitive architecture, in order to efficiently adapt to dynamic environments, consists of (a) a working memory, which is capable of processing (e.g., combining, contrasting or manipulating) two to four elements/chunks simultaneously (see also Halford et al., 1998), and (b) a long-term memory with almost unlimited capacity for chunks of declarative knowledge. To spare working memory capacity for processes relevant to learning (i.e., elaboration and self-explanation) work load has to be as small as possible. In CLT there are three kinds of work load differentiated: (a) *intrinsic load*, resulting from the complexity of the task (dependent on learner's expertise and the interactivity of elements to be processed); (b) *extraneous load*, determined by demands resulting from suboptimal instructional design; (c) *germane load*, resulting from effortful learning and elaboration and leading to schema construction or automation (Sweller, 2005). Schemata are assumed to be stable representations of transient experiences, assumed to (1) guide future recognition (assimilation) of similar experiences, (2) initiate appropriate actions and expectations and/or (3) be accommodated to new experiences if necessary (see von Glasersfeld, 1997). When it comes to problem solving, according to Sweller (2005), an expert can assimilate what seems to be multiple elements to a novice under one single schema to spare work load via working on a chunk of higher order (i.e., a chunk containing chunks) instead of having to work on multiple chunks. Furthermore, schemata are assumed to have an executive function, guiding the problem solving process of experts, whereas novices have to rely on inefficient and more general search strategies causing additional work load. Therefore, gaining expertise about the structure and the dynamics of the problem plays one of the most important roles in solving complex problems, as expertise (a) helps to reduce intrinsic load given a certain interactivity between the elements of the task and (b) is assumed to moderate the usefulness of certain strategies and the effect of problem characteristics (this moderating effect is commonly referred to as the *expertise reversal effect,* see Kalyuga, 2007). The principles derived from CLT have successfully been applied to CPS tasks such as air traffic control and interactive games (for an overview see Osman, 2010).

Whereas CLT has its focus predominantly on explicit declarative chunks of knowledge, other approaches have emphasized the importance of implicit knowledge for CPS. The importance of implicit knowledge in controlling dynamic systems was made clear by Berry and Broadbent (1984) who found that practice and learning did enhance performance, although it did not lead to verbalizable knowledge about the system structure. Broadbent and colleagues examined implicit learning in system control using minimal complex dynamic systems like the SUGAR FACTORY (which is based on the equation Pt = 2*W - Pt-1 + e; where W is the number of workers, P is the amount of sugar produced at a moment in time t, and e is a random error term. (See Berry & Broadbent, 1984). They proposed an *instance-based theory of system control*, claiming that successful interactions with a dynamic system were stored in memory as a kind of "look-up table" of instances, containing information about (a) the perceived state of the system and (b) the input necessary to reach the target level (Broadbent, Fitzgerald, & Broadbent, 1986). Decisions about what action to execute in a given situation can then be based on the instance matching the perceptual properties of the current situation best. A lot of instance-based theories have been proposed since then (e.g. Dienes & Fahey, 1995; Lebiere, Wallach, & Taatgen, 1998; Logan, 1988), each able to reproduce the behavior of participants trying to solve the SUGAR FACTORY. A modern version of the instance-based learning theory within the framework of the cognitive architecture ACT-R was proposed by Gonzalez, Lerch, and Lebiere (2003) in order to explain decision making in complex, dynamic situations especially under uncertainty (Gonzalez & Lebiere, 2005). Gonzalez et al. (2003) assumed that every decision is stored as an instance, i.e., as a chunk of knowledge with slots containing (a) a set of features of the situation, (b) the decision made, and (c) the expected utility of this decision. In the absence of instances similar to the current situation, the decision maker is assumed to rely on simple heuristics for making his or her decision (e.g., random choice). When instances similar to the current situation are retrieved from memory, the decision maker is assumed to rely on the alternative with the highest aggregated utility after she or he has evaluated a certain amount of alternatives (depending on factors like the aspiration level of the decision maker and the perceived urgency of decision, in regard to the time remaining). After a decision was made, the utility-slot of the decision is updated according to the outcome of the decision via a feedback process.

Even though instance-based learning often leads to successful system control (e.g., in systems like the SUGAR FACTORY), it is of limited transferability as it does not involve information about the properties of the system structure (Berry & Dienes, 1993). See Figure 2 for the difference between knowledge about (a) an instance of a system and (b) the structure of the system. Structural knowledge is transferable and allows for building expectations about the consequences of certain decisions and actions in a given situation. It may be action-guiding even in hypothetical situations or in situations never encountered before.

As Schoppek (2002) pointed out, the usefulness of instance knowledge also depends on the size of the problem space, i.e., on the number of possible input- and output-states of a system as well as their relations. He emphasized that additional knowledge about the system structure becomes necessary when larger systems have to be controlled as an

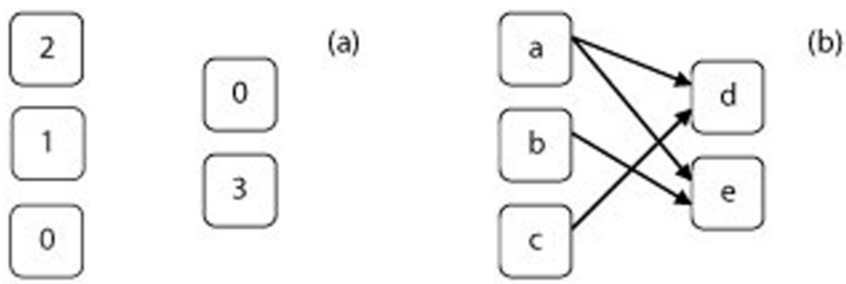

Figure 2. *Visualization of two different kinds of knowledge about an exemplary system containing 5 variables: (a) An instance of the system (consisting of a set of numerical values) and (b) the structure (consisting of the relations between abstract variables).*

instance-based model would require a tremendous amount of input-output-instances to cover a substantial part of the problem state when the system that has to be controlled consists of many input- and output-variables. Structural knowledge may be of use even in situations never seen before. Empirically this assumption proved to be valid: Funke (1993) studied slightly more complex systems (with three input- and three output-variables) and indeed found significant correlations between structural knowledge and control performance (as well as effects of the system's complexity on both measures). Quesada et al. (2005) supplemented Schoppek's view as they mentioned the moderating role of expertise: Experts may be able to have sufficient implicit instance knowledge even about large systems.

The acquisition of structural knowledge about complex systems seems to depend on conscious thought and mental effort (corresponding to the germane load; Sweller, 1988). The acquisition of structural knowledge thus may be fostered by the intention and the opportunity to explore the system before or instead of having to achieve a certain goal (Sweller, 1988; Vollmeyer, Burns, & Holyoak, 1996). Instance knowledge, on the other hand, seems to be acquired without germane load, automatically as a result of practice (Schoppek, 2002). Logan (1988) pointed out that in the absence of implicit knowledge finding the solution to a problem (i.e., the response to a stimulus) requires conscious thought and the application of rules. Only after a vast amount of practice the correct response can be retrieved rapidly and automatically. So in the absence of relevant implicit knowledge there have to be general heuristics and explicit knowledge guiding the course of problem solving (see Gonzalez et al., 2003; Sweller, 2005).

Each interaction with the system may be considered generating an instance that could be stored in memory to implicitly guide future decisions in the face of similar system states—under certain circumstances (factors like time pressure, stress, uncertainty, and high cognitive load may foster the reliance on instance knowledge. See Gonzalez and Lebiere, 2005). In addition to knowledge about instances, systematic strategy use may allow inference of knowledge about the system structure (see section on decision making strategies) which might come in handy under different circumstances (e.g., when trying to reach system states never seen before, maybe due to a large problem space and insufficient expertise).

So after having considered different kinds of knowledge that can be assumed to have an influence on CPS, in order to build a process model of CPS it seems promising to further examine (a) the circumstances determining which kind of knowledge (e.g., structural or instance based) problem solvers usually rely on to make their forecasts, plans and decisions, and (b) what strategy is chosen when no knowledge about the correct solution to a problem is available yet. Answers to this question were proposed in the field of research on decision making and will be reported in the next section.

5. Decision making strategies

Research on decision making has developed a set of decision making strategies containing viable strategies and heuristics for (a) generating relevant information and (b) making good forecasts and decisions in complex environments. When the goal is to specify an input or a series of inputs in order to regulate certain output-variables of a complex system, each possible input vector (e.g., an action in a complex scenario) can be considered an *option*, with several expected *consequences* (e.g., changes in the output variables). Each consequence may have a subjective *utility* and an expected *probability* specific to the current context (i.e., the consequences of an action may be of different perceived use and certainty, dependent on factors like the perceived features of the situation). In complex scenarios there seldom can be an exhaustive evaluation of all possible options and their weighted consequences (due to time pressure and the tremendous amount of variables that would have to be considered). Instead, decisions have to be based on strategies using less information and only a small amount of computation (e.g., by taking the option which has the highest value on the most important consequence – the so-called "take the best"-heuristic). With regard to CPS, it is of special interest to note that simple heuristics like "take the best" or simple tallying can actually achieve higher accuracy in predicting the best outcome than more complex algorithms under certain circumstances – e.g., low predictability of a criterion, combined with small sample sizes relative to the number of available cues, and dependency between cues (Gigerenzer & Brighton, 2009). So when it comes to predicting new observations (instead of just fitting data already observed) sometimes the "less-is-more"-approach holds to be true and it proves to be more accurate to make decisions based on only *one good reason* (i.e., "take the best") than using tallying, multiple regression or even heavy-weight nonlinear strategies like neural networks (Gigerenzer & Brighton, 2009). Therefore, the question is not as much which strategy is the best but which is the best in a certain environment, i.e., under certain conditions.

The applicability and/or the usefulness of some strategies—their ecological rationality (Gigerenzer & Brighton, 2009)—can depend on the existence of prior experiences with the system, on the amount of detailed structural knowledge about the values and weights, on knowledge about the alternatives available, etc. Thus, memory on the one hand constrains the set of heuristics applicable (each long-term and working memory can be considered to constrain what is possible in a certain situation) and on the other hand "selects" heuristics that are likely to yield accurate decisions in a mostly unconscious process (Gigerenzer & Brighton, 2009). Furthermore, the ecological rationality of a heuristic in a given environment is assumed to depend on factors like the *structure* of the environment and feedback, amongst others. According to Rieskamp and Otto (2006), the ecological rationality can be learned by the decision maker via simple reinforcement learning. When goal-oriented decisions are dependent on former decisions and their consequences in an environment that may change both spontaneously or as a consequence of earlier actions, it is commonly referred to as *Dynamic Decision Making* (DDM; Edwards, 1962). Busemeyer (1999) has given an overview of the research on DDM, stating that on the one hand human performance usually can be considered suboptimal, but that on the other hand systematic learning effects were found in almost all of the studies reviewed.

So, during the process of CPS, problem solvers seem to increasingly rely on strategies that are efficient and ecologically rational, i.e., they (1) rely on the correct solution if it is known automatically (*instance knowledge*), elsewise (2) search for a solution based on the current problem representation (*structural knowledge*), or (3) gather new information about the problem (e.g., via random or systematic interaction with the system, via asking an expert, etc.). This conception seems to be consistent with the "Elshout-Raaheim-Hypothesis" (Leutner, 2002), stating that correlations between problem solving and intelligence may be dependent on knowledge about the system in an inverted-U-shaped way (i.e., the correlation may be minimal when prior knowledge is very high or very low as consequently no reasoning is necessary in these cases). In some cases it might even be an option to abandon certain goals due to their unattainability (see Brandtstädter, 2007), or to give up the attempt of a rational solution.

When it comes to gathering information (e.g., when the structural knowledge about the problem proves to be insufficient), some strategies may be especially useful for generating viable structural knowledge about the system. As Vollmeyer et al. (1996) pointed out, systematicity in strategy use allows a problem solver to coherently infer the consequences of single interactions, i.e., to build viable structural knowledge about parts of the system structure. For example, following Tschirgi (1980), to "vary one thing at a time" (while setting the other variables on a constant value like zero)—commonly referred to as the VOTAT-strategy—may be a strategy useful to systematically identify the effects of independent (exogenous) variables on dependent (endogenous) variables in certain scenarios (especially when each exogenous variable was contrasted to the other ones at least one time. Setting the increments of all input variables to a value of zero from time to time may facilitate the detection of eigendynamics and indirect effects). Systematic strategy use and generating (as well as using) structural knowledge might be especially important in complex systems when there is no (or even cannot be) sufficient implicit knowledge about a correct solution of the problem. But as human cognitive resources are limited, even detailed and extensive structural knowledge about all the aspects of a complex system may not be fostering CPS per se as they may overcharge the human working memory. Based on this crucial aspect of complex problems the following section proposes the most influential theories on how and why information reduction is an essential aspect of CPS.

6. Information reduction

As large amounts of knowledge may overcharge human processing capabilities, a most important aspect of coping with complexity is information reduction. Klauer (1993) proposed a theory of information reduction in CPS. Based on the assumption that problem solving involves processes of using certain searching strategies (implicit procedural knowledge) applied to a mental representation of the problem (explicit declarative knowledge) demanding resources of working memory with its limited capacities, Klauer stated, based on his empirical findings, that it was mainly the (declarative) representation of the problem that was reduced in case of capacity overload. Two mechanisms for reducing processing load imposed by complex representations are conceptual chunking and segmentation (Halford, et al., 1998). The consequence is a parsimonious representation.

As Gonzalez and Lebiere (2005) pointed out, the development of effectiveness in DDM involves an increasing selectivity in the use of information, via focusing on relevant features whereas ignoring irrelevant features of the situation. The relevance of features may be determined based on explicit or implicit knowledge. For instance, in their IBLT (see section on expertise) the relevant features of a current situation are assumed to stand out in the recognition process, because they resemble cues in the instances stored in memory. With increasing practice the common features of past instances, similar to the current situation, can be abstracted to guide the attention to the important aspects of the situation (Gonzalez & Lebiere, 2005). This is consistent with the predictions of the chunking/template theories (Chase & Simon, 1973; Simon & Gobet, 1996) and the information reduction hypothesis (Haider & Frensch, 1996).

Newell and Simon (1972) considered a method for planning that emphasized the importance of information reduction regarding irrelevant differences and operators in order to approximately find a way through huge problem spaces. This method consisted of (1) abstracting from details of objects or operators, (2) generating an abstract problem space, (3) searching for a solution in the abstract problem space, and finally (4) trying to map the abstract solution on the concrete problem space with all its details. These considerations are of special importance to CPS, where the abstraction from irrelevant details often is the only way to make adequate forecasts of the system's behavior in spite of the tremendous amount of variables and relations involved. Gaschler (2009, p.5) stated, that "research on information reduction emphasizes practice-related changes of which rather than how information is being processed. Information reduction applies in situations in which tasks contain both relevant and irrelevant information, and denotes a change from a strategy that is based on exhaustive processing of all elements of a task to a strategy that skips the irrelevant task components".

To summarize research on information reduction, in CPS omitting irrelevant task components and finding a parsimonious representation of the problem may enable and foster the search for a solution to a complex problem. Because the search for a solution based on a viable parsimonious model of the problem involves processes like inductive and deductive reasoning, that are commonly subsumed under the concept "intelligence", the next section of this article will review the empirical and theoretical findings on how different aspects of intelligence influence CPS before we will integrate the findings reported so far in a process model of CPS.

7. Intelligence

Theoretically, general intelligence may be defined as "the global capacity of a person to act purposefully, to think rationally, and to deal effectively with his environment" (Wechsler, 1944). Originally, general intelligence as a concept was proposed to explain covariance between a wide range of cognitive tasks, and reasoning as well as problem solving have traditionally been a part of the definition (Sternberg, 1982). Research on intelligence is about the cognitive processes involved in solving tasks and problems and thus may contribute to a profound understanding of CPS. According to the very broad conceptualization of intelligence, it seems quite natural to ask about (1) the amount of variance in CPS performance that can be explained by traditional tests of general intelligence, and about (2) the facets of intelligence that may be most relevant for CPS.

At the beginning of CPS research intelligence surprisingly seemed to be only loosely correlated with performance in complex scenarios (see e.g., Wenke et al., 2005). From today's point of view this lack of evidence in the early days of research on CPS can partially be attributed to the insufficient psychometric qualities of early measures of performance in complex scenarios (e.g., in the TAILORSHOP neither the absolute capital values at the end of each simulated month, nor the changes of capital, but the sum of changes proved to be a reliable and valid measure for CPS performance; for a review, see Danner et al., 2011). According to Danner et al. (2011), using a reliable performance measure revealed substantial correlations of performance in the TAILORSHOP with intelligence measured by Advanced Progressive Matrices ($r=.31$, $p=.001$), job performance rated by supervisors ($r=.19$, $p=.025$), and other measures for CPS performance ($r=.31$, $p<.001$). Süß, Oberauer, and Kersting (1993) also found significant correlations of a TAILORSHOP performance measure with the intelligence facet capacity measured by the BIS test (Jäger, Süß, & Beauducel, 1997). In another influential study, Wittman and Süß (1999) also revealed a substantial effect of working memory capacity on the performance in different CPS scenarios, and the authors also stated, that correlations between the different scenarios became about zero when system-specific knowledge and intelligence were partialed out. According to these findings, working memory capacity and the processes involved in generating system-specific knowledge seem to be the most important facets of intelligence in explaining CPS performance. There is currently an ongoing debate if the generation and application of knowledge in CPS address some facets of general intelligence that are not yet addressed for by traditional intelligence tests (see Wenke et al., 2005).

Generally, traditional intelligence tests, aiming primarily at speed and quality of human symbol processing (i.e., fluid reasoning) as well as working memory capacity, were criticized for their primary focus on the results instead of the process of efficient problem solving behavior (Dörner, 1986). Additionally Horn and Blankson (2005) criticized that there may be more complex "expertise abilities" (Horn & Blankson, 2005, p. 60) different from fluid reasoning, working memory and cognitive speed, which are not adequately addressed for by the tests that are assumed to indicate human intelligence. Putz-Osterloh (1981) stated that the most important differences between the demands of classical tests for measuring intelligence and complex problems were the (1) polytelic situation, the need for an (2) active search for relevant information, for (3) specifying concrete goal states and for (4) choosing productive actions, as well as for (5) a greater relevance of prior knowledge in the latter case. According to this line of argumentation, there are facets of general intelligence that are not yet accounted for by traditional intelligence tests.

With his concept of *operative intelligence* Dörner (1986) emphasized the importance of examining not only speed and precision of some of the basic intellectual processes, but also the more formative aspects of problem solving, for example (1) *circumspection* (e.g., anticipation of future and side effects of interventions), (2) the *ability to organize cognitive operations* (e.g., knowing when to do trial-and-error and when to systematically analyze the situation at hand; when to use exhaustive algorithms and when to rely on heuristics, when to incubate an idea etc.) or (3) the *availability of heurisms* (e.g., being able to build helpful subgoals, to constrain the problem space efficiently). This list of examples is not exhaustive, but it gives an idea of what is meant by the "operative" aspects that are not adequately addressed by traditional intelligence tests but may still be considered relevant for an organized course of intellectual processes (Dörner, 1986). With its explicit focus on gaining and using information and knowledge about the cognitive operations adequate, operative intelligence can be considered one of the most relevant expansions of intelligence as it is measured with current measurement devices:

> 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, and then it is intelligent. (Dörner, 1986, p. 292; translated and emphasized by the authors).

The facets of operative intelligence emphasized in the characterization just given closely resemble the facets most relevant for coping with the characteristic features of complex problems (see Burmeister, 2009; Dörner, Kreuzig, Reither, & Stäudel, 1983; Funke, 1992, 2001, 2003, 2011):

1. the *complexity* of the structure (calling for information reduction),
2. the *interconnectedness* of the variables (calling for building a model of the most relevant effects),
3. the *polytely* of the task (calling for evaluation and for setting priorities),
4. the intransparency of the situation (calling for systematically generating information), and
5. the *dynamics* of the system (calling for Dynamic Decision Making).

These characteristic features of complex problems and the corresponding facets of CPS (see Funke, 2001) can be considered a fruitful starting point for measuring operative intelligence, which in turn might be the most important determining factor of CPS performance. According to Dörner (1986) the most relevant facets of operative intelligence could be measured evaluating and quantifying the "questions" (meaning behavior to explore the system actively generating information) and "decisions" (meaning behavior to control the system goal-oriented) of testees solving complex problems (see the MicroDYN approach presented in Greiff, in press, or Wüstenberg, Greiff, & Funke, in press, for a promising attempt to do so). As Dörner (1986) emphasized, the demands of CPS situations are characteristically intertwined in an inseparable way, and the problem solving process has to be studied as a whole because the parts are interacting with each other and hardly can nor should be examined in isolation:

1. *Information retrieval and information integration*: The problem solver needs a model adequately representing the system and the goal state to aim at. Therefore she or he has to systematically generate, gather, and integrate information to adjust this model to the system.
2. *Goal elaboration and goal balancing*: The problem solver has to specify and substantiate the often vague and global goals she or he wants to achieve. If some specified goals turn out to be contradictory, she or he has to find a satisfying trade-off or balance in only partially reaching the goals.
3. *Action planning and decision making*: The problem solver has to decide what actions to execute, i.e., what decision making strategies to apply (see section on decision making strategies), and which kind of knowledge to rely on (see section on expertise). By forecasting future developments given the system's prior states and her or his own actions she or he can efficiently plan her or his next steps (e.g., chains of consecutive actions with each action building on the results of the previous one).
4. *Self management*: The problem solver may have to face time pressure, stress, and frustration as well as conflicts between his inner values. She or he has to manage these non-cognitive affordances by either changing the system or his own behaviors and habits.

So after these considerations about how efficient CPS may look like and what facets of intelligence may influence the CPS performance we want to proceed by integrating the contributions of all the fields of research mentioned above in a process theory of CPS.

8. Discussion

After reviewing some of the most important fields of research on CPS, and based on the definition given above, we are now going to summarize the interdisciplinary findings in a process theory of CPS, concluding with a short outlook for upcoming research. CPS can be understood as the process of solving problems that have to be considered "complex" (i.e., containing many highly interrelated elements). For instance, every scientist, who wants to describe, explain, and predict a complex system by means of her or his hypotheses (containing a parsimonious but viable subset of all variables possibly relevant) might be facing a complex problem. A mayor of a city as well as a manager of an organization or a policy maker trying to get rid of climate change, each may be considered as having a complex problem to cope with. Trying to make a modern computer do what it is supposed to can turn out to be a complex problem as well as changing certain settings of an unknown mobile phone device. The process of CPS usually consists of different phases: (1) knowledge acquisition and (2) goal-oriented knowledge application (Leutner, Wirth, Klieme, & Funke, 2005). Usually a problem solver switches between these phases in a complex way:

1. At first, the problem solver has to acquire knowledge about the problem.
    a. The problem solver is assumed to explore the system's behavior using a strategy that (a) she or he knows of and (b) seems to be most ecologically rational to her or him (e.g., random or systematic interaction with the system, reading the instructions, asking an expert, etc.).
    b. The exploration leads to (a) knowledge about the system's states and the actions taken (instance knowledge) as well as (b) an internal representation of the problem, containing the most important elements and relations of the system (structural knowledge) which usually is inferred from the instance knowledge.
    c. As the capacity of the problem solver's working memory is limited, the internal representation is object to information reduction. Relations and elements that prove to be less relevant for system control in the course of exploration are assumed to be omitted in order to allow more efficient planning and forecasting.
2. When the problem solver has a certain amount of knowledge about the problem that has to be solved, she or he is assumed to apply the knowledge in order to reach her or his goals.
    a. The problem solver is assumed to use her or his internal representation to make forecasts about the system's dynamics in order to decide (a) if she or he has to intervene and (b) what intervention will have acceptable consequences in the current situation. When the current situation cues the correct intervention immediately (due to instance knowledge), the problem solver is assumed to rely on her or his instance knowledge instead.
    b. Monitoring processes are assumed to detect (a) the progress in solving the problem and (b) the implications of feedback from the environment for the problem representation. When the problem representation proves to be not viable for reaching the goals in time, the problem solver is assumed to either switch back to knowledge acquisition or to change the

goals (depending on factors like the importance of the goals and on the assumed effort of further knowledge acquisition).

This process theory of CPS summarizes what is known about the most important aspects of CPS and is based on the theoretical and empirical contributions of the interdisciplinary field presented in the previous sections. As CPS is a rather abstract concept, further research is needed to specify the process of CPS concerning concrete operationalizations of complex problems (e.g., handling a complex mobile phone may be represented in other ways than regulating an economic system or managing a tailorshop). Concerning this, it seems to be a fruitful approach to build cognitive models of the CPS process (e.g., Schoppek, 2002) in order to develop a deeper understanding of CPS processes taking place in real life.

But even on a more abstract level our theory on the CPS process may be subject to further research. It may be seen as a starting point for further experiments, in order to gradually improve our understanding of what CPS is and how it works (e.g., experimental psychology may further contribute knowledge about variables or interactions with a significant impact on the process of CPS). Psychometrics may contribute to a better understanding of CPS by developing reliable and valid measures for the processes that are assumed to be important for efficient and intelligent CPS (Greiff, in press; Wüstenberg, Greiff, & Funke, in press). Those measurement devices in turn can be used to test process theories on CPS in more detail.

Our conception of CPS is inspired by the pioneering works of Dörner, especially by the concept of operative intelligence (Dörner, 1986) and the considerations of Funke (2001), emphasizing (a) information generation (due to the initial intransparency of the situation), (b) information reduction (due to the overcharging complexity of the problem's structure), (c) model building (due to the interconnectedness of the variables), (d) dynamic decision making (due to the dynamics of the system), and (e) evaluation (due to many, interfering and/or ill-defined goals). 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. The more we learn about the process of problem solving, the more we have to acknowledge the complexity of both the process and the kind of problems that are involved in realistic problem solving in naturalistic environments.

References

Berry, D. & Dienes, Z. (1993). Towards a working characterization of implicit learning. In

D. Berry & Z. Dienes (eds.), *Implicit learning: Theoretical and empirical issues*. Hillsdale, NJ: Erlbaum.

Brandtstädter, J. (2007). *Das flexible Selbst. Selbstentwicklung zwischen Zielbindung und Ablösung* [The flexible self. Self development between attachment and detachment]. München: Elsevier.

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

Broadbent, D., Fitzgerald, P., & Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology*, *77*, 33-50.

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

Burmeister, K. (2009). *Komplexes Problemlösen im Kontext angewandter Eignungsdiagnostik* [Complex Problem Solving in the context of applied diagnostics of suitability]. Universität Greifswald, Philosophische Fakultät.

Busemeyer, J.R. (1999). *International Encyclopedia of the Social and Behavioral Sciences: Methodology*. Mathematics and Computer Science. Amsterdam: Pergamon.

Chase, W. G. & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Chi, M.T.H., Feltovich, P.J., & Glaser, R.(1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152.

Chi, M.T.H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence. Volume 1* (pp. 7-75). Hillsdale, NJ: Erlbaum

Danner, D., Hagemann, D., Holt, D.V., 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.

Dienes, Z. & Fahey, R. (1995). Role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 848-862.

Dörner, D. (1976). *Problemlösen als Informationsverarbeitung* [Problem solving as information processing]. Stuttgart: Kohlhammer.

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

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

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

Dunbar, K. (1998). Problem solving. In W. Bechtel, & G. Graham (Eds.). *A companion to*

Cognitive Science (pp. 289-298). London: Blackwell.

Duncker, K. (1945). *On problem solving. Psychological Monographs*, *58*. American Psychological Association.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human Factors*, *4*, 59-73.

Engelhart, M., Funke, J. & Sager, S. (2011). *A new test-scenario for optimization-based analysis and training of human decision making*. Poster presented at the SIAM Conference on Optimization (SIOPT 2011), May 16-19, 2011, Darmstadtium Conference Center, Darmstadt, Germany.

Frensch, P. & Funke, J. (1995). Complex Problem Solving: *The European perspective*. Hillsdale, NJ: Erlbaum.

Funke, J. (1985). Steuerung dynamischer Systeme durch Aufbau und Anwendung subjektiver Kausalmodelle [Control of dynamic systems via Construction and Application of subjective causal models]. *Zeitschrift für Psychologie*, *193*, 443-465.

Funke, J. (1992). *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung* [Knowledge about dynamic systems: Acquisition, Representation and Application]. Heidelberg: Springer.

Funke, J. (1993). Microworlds based on linear equation systems: a new approach to Complex Problem Solving and experimental results. In G. Strube & K.F. Wender (Eds.), *The cognitive psychology of knowledge* (pp. 313-330). Amsterdam: North-Holland.

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.

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

Funke, J. (2011). Problemlösen [Problem Solving]. In T. Betsch, J. Funke & H. Plessner (Eds.), Denken – *Urteilen Entscheiden Problemlösen* (pp.135-199). Heidelberg: Springer.

Gaschler, R. (2009). *Information reduction as item-general strategy change*. Retrieved, July, 2011, from http://edoc.hu-berlin.de/dissertationen/gaschler-robert-2009-06-30/PDF/gaschler.pdf

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. Topics in Cognitive Science, 1, 107–143.

Gonzalez, C. & Lebiere, C. (2005). Instance-based cognitive models of decision making. In Zizzo, D. & Courakis, A. (Eds.), *Transfer of knowledge in economic decision making*. New York: Palgrave McMillan.

Gonzalez, C., Lerch, F. J., & Lebiere, C. (2003). Instance-based learning in dynamic decision making, *Cognitive Science*, *27*, 591-635.

Greiff, S. (in press). *Individualdiagnostik der Problemlösefähigkeit* [Individual diagnostics of problem solving competency]. Münster: Waxmann. Forthcoming 2012.

Haider, H. & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, *30*, 304–337.

Halford, G.S., Wilson, W.H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental and cognitive psychology. *Behavioral and Brain Sciences*, *21*, 803-865.

Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp.41–68), 2nd ed. New York: Guilford Press.

Jäger, A.O., Süß. H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstrukturtest. BIS-Test, Form 4* [Berlin Intelligence Structure test. BIS-Test, Form 4]. Göttingen: Hogrefe.

Johnson, E.J. (1988). Expertise and decision under uncertainty: Performance and process. In M.T.H. Chi, R. Glaser & M.J. Farr (Eds.), *The nature of expertise* (pp.209-228). Hillsdale, NJ: Erlbaum.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539.

Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*, 1-48.

Klauer, K.C., (1993). *Belastung und Entlastung beim Problemlösen. Eine Theorie des deklarativen Vereinfachens* [Charge and discharge in problem solving. A theory of declarative simplification]. Göttingen: Hogrefe.

Larkin, J.H. (1983). The role of problem representation in physics. In D. Gentner & A. Collins (Eds.), *Mental models* (pp.75-98). Hillsdale, NJ: Erlbaum.

Lebiere, C., Wallach, D., & Taatgen, N. (1998). Implicit and explicit learning in ACT-R. In F.E. Ritter & R.M. Young (Eds.), *Proceedings of the second European Conference on Cognitive Modelling* (ECCM-98) (pp. 183-189). Nottingham, UK: Nottingham University Press.

Leutner, D. (2002). The fuzzy relationship of intelligence and problem solving in computer simulations. *Computers in Human Behavior*, *18*, 685-697.

Leutner, D., Wirth, J., Klieme, E., & Funke, J. (2005). Ansätze zur Operationalisierung und deren Erprobung im Feldtest zu PISA 2000 [Approaches concerning operationalization and their testing in the field test in the context of PISA 2000]. In E. Klieme, D. Leutner & J. Wirth (Eds.), *Problemlösekompetenz von Schülerinnen und Schülern* (pp. 21-36). Wiesbaden: VS Verlag für Sozialwissenschaften.

Logan, G.D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information-processing and insight: A process model of performance on the nine-dot and related problems. Journal of *Experimental Psychology: Learning, Memory and Cognition*, *27*, 176-201.

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

Novick, L.R., & Bassok, M. (2005). Problem Solving. In Holyoak, K.J. & Morrison, R.G. (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 321-350). Cambridge: University Press.

Ohlsson, S. (1992). Information processing explanations of insight and related phenomena. In M.T. Keane & K.J. Gilhooly (Eds.), *Advances in the psychology of thinking*. London: Harvester Wheatsheaf.

Osman, M. (2010). Controlling uncertainty: A review of human behavior in complex dynamic environments. *Psychological Bulletin*, *136*, 65-86.

Putz-Osterloh, W. (1981). Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg [On the relationship between test intelligence and problem solving performance]. *Zeitschrift für Psychologie*, *189*, 79-100.

Quesada, J., Kintsch, W., & Gomez, E. (2005). Complex Problem Solving: A field in search of a definition? *Theoretical Issues in Ergonomic Science*, *6*(1), 5–33.

Rieskamp, J. & Otto, P.E. (2006). SSL: A theory of how people learn to select strategies. *Jorunal of Experimental Psychology: General, 135*, 207-236.

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

Simon, H. A. & Gobet, F. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Science*, *31*, 1–40.

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

Sternberg, R.J. (1982). Reasoning, problem solving and intelligence. In R.J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 225-307). Cambridge: University Press.

Sternberg, R.J. & Frensch, P.A. (Eds.). (1991). *Complex Problem Solving: Principles and mechanisms*. Hillsdale, NJ: Erlbaum.

Strunk, G. & Schiepek, G. (2006). *Systemische Psychologie. Eine Einführung in die komplexen Grundlagen menschlichen Verhaltens* [Systemic Psychology. An introduction in the complex basics of human behaviour]. München: Elsevier.

Süß, H.-M., Oberauer, K. & Kersting, M. (1993). Intellektuelle Fähigkeiten und die Steuerung komplexer Systeme [Intellectual Abilities and the Control of complex systems]. *Sprache & Kognition*, *12*, 83-97.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257-285.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). Cambridge: University Press.

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 specifity and systematicity of strategies on the acquisition of problem structure. *Cognitive Science*, *20*, 70-100.

Von Glasersfeld, E. (1997). *Radikaler Konstruktivismus. Ideen, Ergebnisse, Probleme* [Radical Constructivism. Ideas, Results, Problems]. Frankfurt: Suhrkamp.

Wechsler, D. (1944). The measurement of adult intelligence. Baltimore: Williams & Wilkins.

Wenke, D., Frensch, P. A., & Funke, J. (2005). Complex Problem Solving and intelligence: Empirical relation and causal direction. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and intelligence: Identifying the mechanisms of the mind* (pp. 160-187). New York: Cambridge University Press.

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: American Psychological Association.

Wüstenberg, S., Greiff, S. & Funke, J. (in press). Complex Problem Solving. More than reasoning? *Intelligence*. Forthcoming 2012.

Intelligence 41 (2013) 579–596

Contents lists available at ScienceDirect

Intelligence

journal homepage:

A multitrait–multimethod study of assessment instruments for complex problem solving☆

Samuel Greiff[a,*], Andreas Fischer[b], Sascha Wüstenberg[a], Philipp Sonnleitner[a], Martin Brunner[c], Romain Martin[a]

[a] University of Luxembourg, Luxembourg, Luxembourg
[b] University of Heidelberg, Heidelberg, Germany
[c] Freie Universität Berlin, Berlin, Germany

ARTICLE INFO

Article history:
Received 17 January 2013
Received in revised form 16 July 2013
Accepted 17 July 2013
Available online xxxx

Keywords:
Complex Problem Solving
Reasoning
MicroDYN
Genetics Lab
MicroFIN

ABSTRACT

Recently published studies on Complex Problem Solving (CPS) suggest that assessments of CPS using multiple complex systems are only moderately related to tests of classical cognitive abilities. Further, CPS assessments show incremental validity beyond tests of other cognitive abilities when predicting relevant outcomes. However, these empirical accounts have relied on single CPS assessment instruments. We do not know whether these findings will generalize to the construct level across different CPS assessment instruments. To answer this question, we tested a sample of $N = 339$ German university students who completed three CPS assessment instruments based on multiple complex systems (MicroDYN, the Genetics Lab, and MicroFIN) and the matrices subtest of the Intelligence Structure Test as measure of reasoning. Students further reported their school grades. Analyses including latent multitrait–multimethod models provided support for the conceptualization of CPS as a complex cognitive ability. Results indicated that different CPS assessment instruments showed sufficient convergent validity (with a consistency mostly between .50 and .60). In addition, we found evidence for the divergent validity of CPS from reasoning (reasoning predicted two CPS facets, knowledge and control, $\beta_{KNOW} = .49$ and $\beta_{CON} = .53$, respectively). In the prediction of academic achievement, CPS explained variance in natural science grades after we controlled for reasoning ($\beta_{CPS} = .22$), whereas social science grades were not predicted. Our findings suggest that the validity of CPS generalizes across different measurement instruments.

1. Introduction

Across the last century, the relevance of cognitive abilities has been demonstrated numerous times, and the assessment of cognitive abilities has been a major concern in areas such as education, the economy, public health, and politics. Cognitive abilities have been shown to be related to outcomes such as longevity (Gottfredson & Deary, 2004), individuals' personality (Salas & Cannon-Bowers, 2001), job success (Schmidt & Hunter, 2004), or low crime rates (Herrnstein & Murray, 1994). Results of tests of cognitive performance have been used to promote the selection of students for higher education (Kuncel, Hezlett, & Ones, 2001), to allocate individuals to jobs according to their ability profiles (Autor, Levy, & Murnane, 2003), or to enhance

☆ This research was funded by grants supported by the Fonds National de la Recherche Luxembourg (ATTRACT "ASSKI21") and by the German Federal Ministry of Education and Research (LSA004 and 01JG1062). We are grateful to the TBA group at DIPF (http://tba.dipf.de) for providing the authoring tool CBA Item Builder and technical support.
* Corresponding author at: University of Luxembourg, 6, rue Richard Coudenhove Kalergi, 1359 Luxembourg-Kirchberg, Luxembourg. Tel.: +352 466644 9245.
E-mail addresses: samuel.greiff@uni.lu (S. Greiff), andreas.fischer@psychologie.uni-heidelberg.de (A. Fischer), sascha.wuestenberg@uni.lu (S. Wüstenberg), philipp.sonnleitner@uni.lu (P. Sonnleitner), martin.brunner@isq-bb.de (M. Brunner), romain.martin@uni.lu (R. Martin).

0160-2896/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intell.2013.07.012

cognitive performance by teaching specific strategies (Klauer & Phye, 2008).

After long relying on paper–pencil tests, a shift toward computer-based assessments has recently been initiated. In its earliest implementation about two decades ago, the main purpose of computer-based assessment was to increase standardization and efficiency in testing (Baker & O'Neil, 2002). This practice produced several advantages such as automatic scoring and adaptive testing, but so far, the assessment instrument itself has been limited to a transformation of paper–pencil tests into computer-based tests (Bunderson, Inouye, & Olsen, 1989; Williamson, Bejar, & Mislevy, 2006).

However, if computerized testing is used simply to present computerized versions of paper–pencil tests, the advantages that computers can offer will not be fully utilized (Baker & O'Neil, 2002). That is, computers enable researchers to assess abilities that are not assessable by paper–pencil tests (Kyllonen, 2009) and to develop tasks that interactively respond to examinees' inputs. According to Williamson et al. (2006), the highest added value of using computers in assessment is expected from interactive tasks. Further, Rigas, Carling, and Brehmer (2002) have identified dynamic and interactive task environments as a general source of innovation in cognitive ability testing.

Two major advantages of computer-based assessment—higher efficiency and the inclusion of interactive tasks—were acknowledged by international large-scale assessments such as the Programme for International Student Assessment (PISA; OECD, 2006). In PISA, a major shift from paper–pencil to computer-based test administration was recently implemented, and complex interactive task environments were included in the current assessment cycle (OECD, 2010). For instance, an assessment of problem solving in interactive and dynamically changing task environments to assess Complex Problem Solving (CPS; Funke, 2001; Rigas et al., 2002) was part of the international PISA 2012 survey (OECD, 2010). However, given the short history of computer-based assessment, our knowledge about constructs such as CPS is limited. In this study, we therefore focused on CPS and important questions related to it.

CPS fits into the category of broad cognitive abilities (Funke, 2010), which are viewed as essential for lifelong learning by the OECD (2010). It is assessed in complex simulations (Funke, 2001) that allow for dynamic interactions between examinees and task situations (Raven, 2000; Wirth & Klieme, 2003). This feature makes it impossible to assess CPS without a computer. CPS is usually decomposed into a phase of knowledge acquisition (from here on: knowledge; actively acquiring knowledge about the task; Mayer & Wittrock, 2006) and knowledge application (from here on: control; actively controlling the task; Novick & Bassok, 2005). Recent research on CPS has shown divergent validity with regard to reasoning (e.g., Greiff, Holt, & Funke, 2013; Greiff, Wüstenberg, et al., 2013; Wüstenberg, Greiff, & Funke, 2012) and working memory (e.g., Schweizer, Wüstenberg, & Greiff, 2013) as well as the predictive validity of CPS beyond other cognitive abilities (e.g., Greiff & Fischer, 2013; Wüstenberg et al., 2012). After some controversy with regard to its assessment (e.g., Kröner, Plass, & Leutner, 2005; Wüstenberg et al., 2012), CPS has recently experienced advances in terms of its scalability and psychometric properties. These assessment advances were substantially facilitated by the introduction of two formal frameworks—linear structural equations and finite state automata—and by the introduction of multiple complex systems (MCS; see below; Funke, 2010; Greiff, Wüstenberg, & Funke, 2012).

However, it is not quite appropriate to talk about the construct of CPS when describing these recent results. In fact, the results that we mentioned above were conducted only with single homogenous CPS assessment instruments. If we want to generalize these results to CPS as a construct independent of a particular measurement procedure, we need to use a variety of assessment instruments to measure CPS (Campbell & Fiske, 1959; Eid, Lischetzke, & Nussbeck, 2006).

In order to facilitate our understanding of CPS not only on the level of specific assessment instruments but on the construct level, we applied a combination of different CPS assessments instruments. Specifically, we employed three CPS instruments based on multiple complex systems (Greiff et al., 2012) to address (a) the convergent validity of these instruments by combining them in a multitrait–multimethod (MTMM) approach and (b) their divergent validity by relating CPS on the construct level to reasoning and to academic achievement. To this end, we will first outline the conceptual background behind these two research questions and continue with our presentation of empirical studies. We will conclude by discussing the relevance of CPS and its implications for research on cognitive abilities.

1.1. Research Question 1: Measurement of CPS by different assessment instruments

According to Baker and O'Neil (2002), CPS amplifies the learning of children and adults in a number of formal and informal settings. Further, Mayer and Wittrock (2006) point to the importance of CPS in educational settings aimed at making students better problem solvers. To this end, the OECD (2010) views CPS as a complex cognitive ability that has the interaction between task and examinee (and, thus, computer-based assessment) as a central component. Buchner (1995) defines CPS as:

> The successful interaction with task environments that are dynamic (i.e., change as a function of user's intervention and/or as a function of time) and in which some, if not all, of the environment's regularities can only be revealed by successful exploration and integration of the information gained in that process (p. 14).

In line with this definition, Funke (2001) and Raven (2000) argue that solving complex problems involves a series of complex cognitive operations, and that complex problems can be described by several characteristic features such as complexity, intransparency, interconnectedness, and dynamics. Coping with complex problems further involves monitoring (Osman, 2010) and learning (Leutner, 2002). It requires knowledge about when and how to structure the search for viable hypotheses, informative experiments, and goal-oriented interventions (Klahr, 2002).

For the first time since it was introduced by its founding father Dietrich Dörner (e.g., Dörner, 1986), cognitive theories on CPS and human behavior in computer-simulated environments are rather well-developed (e.g., Dunbar, 1998; Fischer, Greiff, & Funke, 2012; Klahr, 2002). At the same time, translating the concept of CPS into tasks suitable for assessment has proven to be exceptionally difficult (Greiff et al., 2012; Kröner et al., 2005). In general, two approaches for assessing CPS have been developed: microworlds and scenarios based on formal frameworks/multiple complex systems.

Computer-simulated microworlds were implemented as the first CPS scenarios in the 1970s and were developed with the aim of administering task environments with a high resemblance to the real world. However, the goals of producing a reliable measure of CPS and of sufficiently simulating reality with microworlds were quickly overshadowed by measurement issues (e.g., Buchner, 1995; Kröner et al., 2005; Wüstenberg et al., 2012). As a reaction to problems with microworlds, Funke (2001) introduced two formal frameworks: Linear Structural Equation systems (LSE) and Finite State Automata (FSA), which allow for the description of underlying task structures independent of their semantic embedment. In particular, the LSE formalism has been widely adopted by CPS research and has led to the development of a considerable number of tasks.

It was only recently that tasks based on these two formalisms were further augmented by the multiple complex system (MCS) approach (Greiff et al., 2012). Within this expansion, the assessment framework is altered by shortening the time on each task (in early tasks: at least 45 min for one task; in MCS: approximately 5 min) and by administering different tasks with varying difficulty (in early tasks: only one task in one specific context is administered; in MCS: several tasks in different contexts are administered). Hence, examinees work on several independent tasks of varying difficulty and are confronted with an entire battery of CPS tasks in MCS.

Even though results on the validity of CPS mentioned above were largely conducted within the MCS approach, they were gathered by using single assessment instruments. Thus, our knowledge of CPS is limited to specific operationalizations. Given that CPS plays a major role in PISA, showing that different assessment instruments target the same underlying construct and can be used as alternative forms instead of merely accumulating instrument-specific variance is a major concern yet unanswered by empirical research. To this end, the first research question in this study was aimed at evaluating whether assessment instruments based on MCS would indicate one underlying CPS construct (i.e., whether different measures would converge). Thus, we used three different instruments based on MCS (MicroDYN and the Genetics Lab based on the LSE formalism; MicroFIN based on the FSA formalism) to link different CPS instruments. We will now describe the three instruments in more detail.

1.1.1. Multiple complex systems within LSE: MicroDYN and the Genetics Lab

1.1.1.1. MicroDYN. In the first assessment approach to CPS, MicroDYN, examinees are asked to detect quantitative causal relations between input and output variables and to control the underlying system. For instance, within the MicroDYN task handball training (see Fig. 1), different trainings (labeled Training A, Training B, Training C) influence characteristics of a handball team (labeled Motivation, Power of the throw, Exhaustion). While working on the system, examinees face two different phases: In Phase 1, examinees can freely explore the task and are asked to learn how variables are related (3 min). Simultaneously, examinees are asked to identify the most central variable and to draw the connections between variables as they suppose they are (e.g., between Training A and Motivation). In Phase 2, examinees are asked to reach given target values on the output variables (1.5 min). A MicroDYN set consists of approximately 10 tasks that sum to an overall testing time of about 1 h including the instructions.

In line with the two phases, MicroDYN captures two facets of problem solving: acquisition of knowledge (i.e., knowledge; Mayer & Wittrock, 2006) and application of this knowledge (i.e., control; Novick & Bassok, 2005). In MicroDYN, knowledge is assessed by the correctness of the model drawn in Phase 1 (see bottom of Fig. 1), and control is assessed by the ability to reach target values in Phase 2. A detailed description of MicroDYN and its procedure can be found in Greiff et al. (2012).

Recent results suggest validity of the MicroDYN approach. For instance, CPS measured by MicroDYN showed incremental validity in predicting school grades even beyond measures of reasoning (Greiff, Holt, et al., 2013; Greiff, Wüstenberg, et al., 2013; Wüstenberg et al., 2012) and working memory (Schweizer et al., 2013). In the 2012 cycle of PISA, MicroDYN tasks are applied worldwide to measure 15-year-olds' proficiency levels in CPS (OECD, 2010).

1.1.1.2. Genetics Lab. An alternative way to assess CPS within MCS is the Genetics Lab (Fig. 2). Like MicroDYN, the Genetics Lab is based on the LSE framework and quantitatively connects a set of variables. In each task, examinees have to first figure out how physical characteristics of a fictive creature (output variables) are influenced by its genes (input variables; see Fig. 2) and subsequently apply their knowledge. Please note that prior knowledge about genetics does not facilitate performance in the Genetics Lab as labels of variables and causal relations are unrelated to the real world. This also applies to MicroDYN. There, labels for input and output variables are either fictitious or without deep semantic meaning. The Genetics Lab is comprised of 12 independent tasks, which are to be completed in 35 min.

Comparable to MicroDYN, examination of a creature is split into two consecutive phases: a first phase indicating knowledge and a second phase indicating control. In the knowledge phase, the test taker actively manipulates the creature's genes and documents the acquired knowledge in a database (see Fig. 2). In the control phase, given target values on specific characteristics have to be achieved. Both phases are used to derive performance scores about test takers' knowledge and their ability to control the creature. A detailed description of the Genetics Lab is found in Sonnleitner et al. (2012).

In previous studies, performance scores in the Genetics Lab have been found to be reliable and valid measures of CPS. They were shown to have external and construct validity as

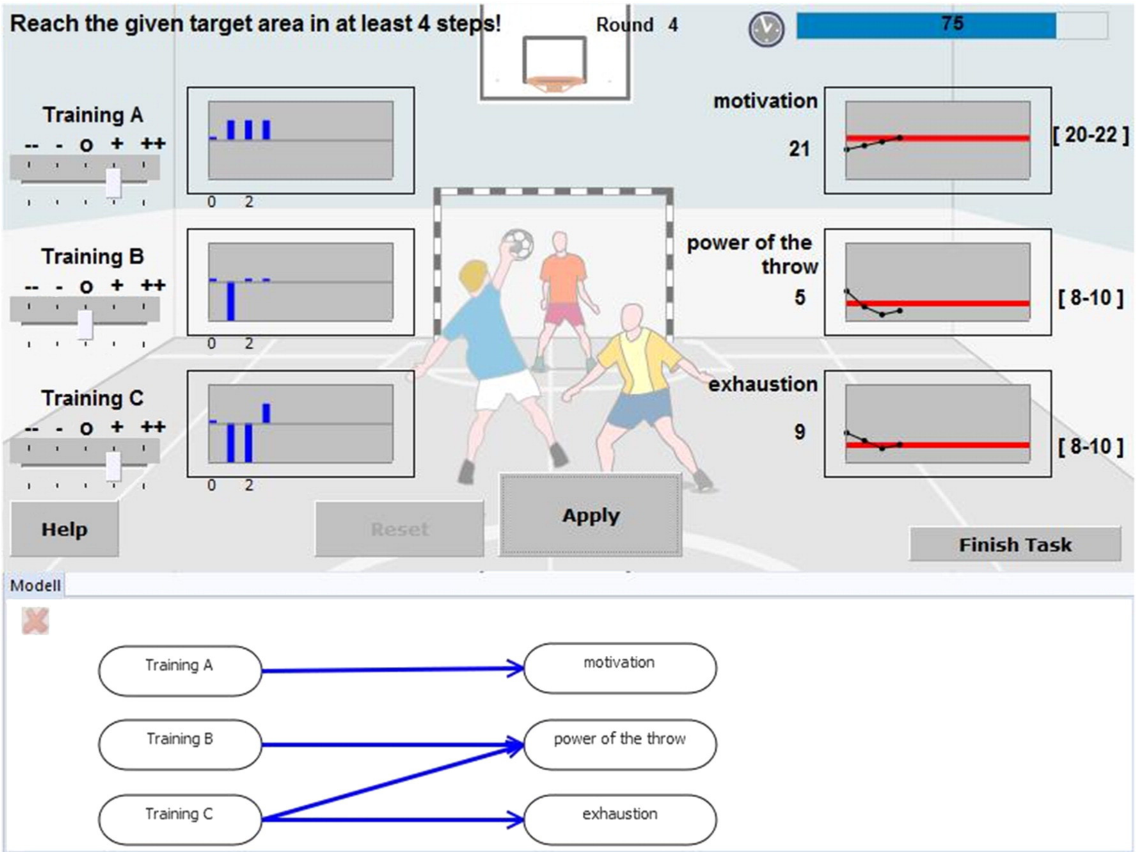

Fig. 1. Screenshot of the MicroDYN task "Handball training." The controllers of the input variables range from "- -" to "++." The current values and the target values of the output variables are displayed numerically (e.g., current value for Motivation: 21; target values: 20–22) and graphically (current value: dots; target value: red line). The correct model is shown at the bottom of the figure (cf. Wüstenberg et al., 2012). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

they showed substantial relations to reasoning ability and school grades (Sonnleitner, Keller, Martin, & Brunner, 2013).

1.1.2. Multiple complex systems within FSA: MicroFIN

MicroFIN within the MCS approach is based on the second formalism of finite state automata. Compared to LSE, FSA relate a set of discrete states of a problem to each other. FSA have been used considerably less in CPS research because developing this type of task is complex and complicated compared to developing LSE tasks such as MicroDYN and the Genetics Lab.

In FSA, examinees are asked to transfer an automaton that can take on a finite set of discrete states from its current state to a goal state by applying a finite set of inputs (Buchner & Funke, 1993). For instance, think of a television device (see Fig. 3) that can take on different states (e.g., "off" or "on"). By pressing buttons on a remote control, the television is transformed into a different state (e.g., it turns "off" when the power button is pushed while the television is on). Even if visualizations of states can semantically imply ordinal relations between states (e.g., Volume 9 may be assumed to be louder than Volume 8), the formal structure of a finite automaton is defined by a finite set of states and qualitative state changes. Thus, in contrast to MicroDYN and the Genetics Lab, the finite state automata used in MicroFIN focus on qualitative connections between variables with different discrete states (Buchner & Funke, 1993).

In the current study, the task shown in Fig. 3 was used to provide interactive instructions to demonstrate the principle of MicroFIN. The instructions were followed by two computer-simulated FSA tasks[1]: a simulated pet and a social situation. The simulated pet was presented in different moods depending on its hunger, fatigue, and boredom. For each of the pet's needs, the set of operators contained one button for performing an intervention (e.g., feeding the pet). The social situation contained two people and a set of objects. Each person could donate each object to the other person, and either person could be happy, neutral, or unhappy, depending on (a) the gifts he or she retrieved, (b) his or her prior mood, and (c) the other person's mood. The overall testing time for these two MicroFIN tasks and the instructions lasted about 15 min.

Like MicroDYN and the Genetics Lab, each FSA consisted of two subsequent phases: First, examinees were instructed to generate knowledge by freely exploring the automaton. At the end of Phase 1, examinees had to answer a set of

[1] Only two FSA tasks were included because the development and implementation of these tasks has proven to be extremely difficult and little is known about their assessment characteristics.

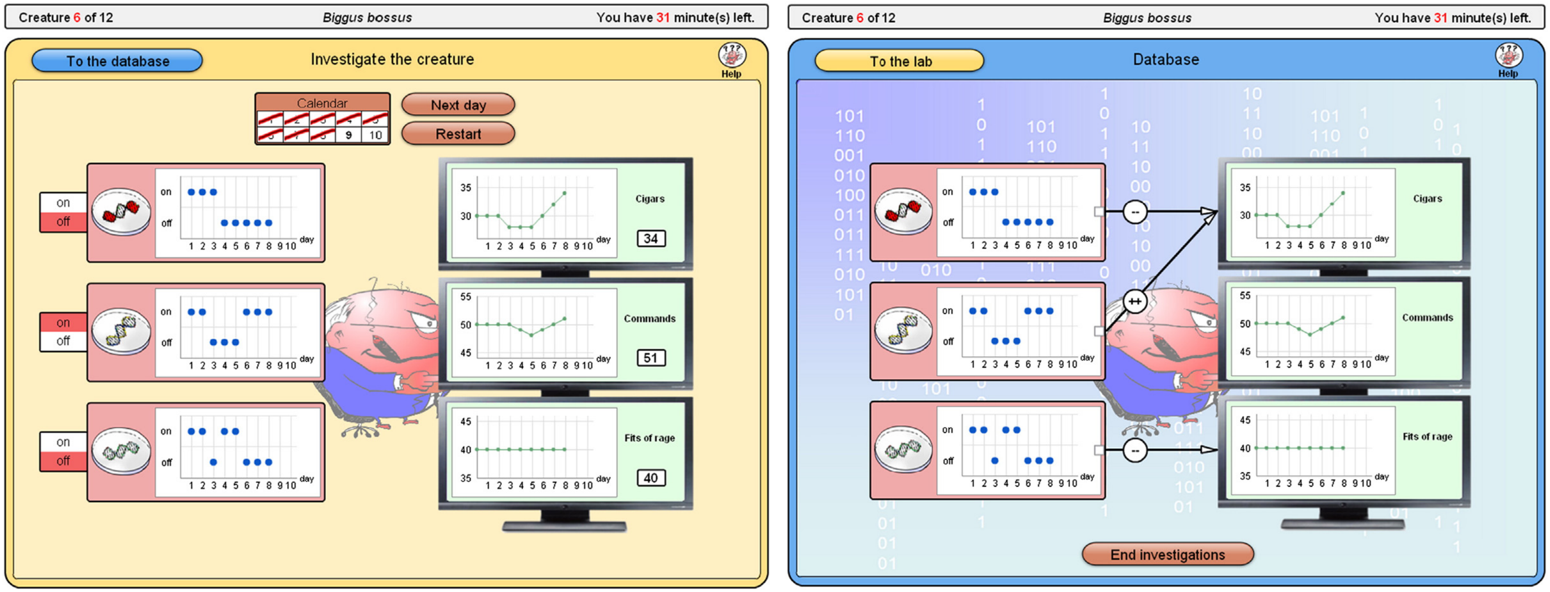

Fig. 2. Screenshot of a Genetics Lab task. Test takers have to actively manipulate genes of fictive creatures (inputs; left part, depicted in red) to find out about their impact on the creatures' characteristics (outputs; left part, depicted in green). The gathered knowledge can be documented by means of a causal diagram in a related database (right picture). Examinees can switch between those interfaces at any time by clicking on the buttons in the upper left part of the screen. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 3. Screenshot of the MicroFIN task "television". Operators are visualized as buttons on a remote control, which can be freely explored by examinees. The current state is visualized as text on the screen.

questions (Buchner & Funke, 1993) indicating their knowledge (about 3 min overall). In the current study, we applied three multiple-choice questions per automaton (e.g., "Which of the following four pictures describes a state of the FSA that results in the depicted state after applying a specific series of operators?"). In Phase 2, examinees had to guide the automaton toward a certain goal (e.g., "Set the TV to Channel 3 and to mute") applying as few steps as possible (about 1 min) to indicate control. For further information on psychometric applications of FSA, see Buchner and Funke (1993).

As an FSA task was used to assess CPS in the German national extension of PISA 2000, knowledge and control of FSA are known to address dynamic aspects of CPS distinct from intelligence and school-related literacy (Wirth & Klieme, 2003). Like MicroDYN, FSA tasks are applied in the international PISA 2012 survey (Greiff, Holt, et al., 2013).

Similar to research on psychometric g, scholars interested in CPS can rely on a number of assessment instruments. With regard to the convergence of microworlds, Wittmann and Süß (1999) reported that the moderate correlations between three microworlds drop to nonsignificance after controlling for reasoning and prior knowledge. Therefore, the authors questioned the relevance of a CPS construct and concluded that CPS was fully explained by reasoning and prior knowledge. Wittmann and Hattrup (2004) reported another study based on the same three microworlds used by Wittmann and Süß (1999). There, a CPS factor extracted from these three microworlds was substantially predicted by measures of general mental ability, reasoning, and creativity, but no analyses were conducted on the uniqueness of a CPS factor after controlling for these influences. On the other hand, Danner, Hagemann, Schankin, Hager, and Funke (2011) showed convergence between two different CPS tests and the incremental validity of CPS. Using a latent state-trait approach, they showed that different CPS microworlds converged even though the overlap in variance between the CPS tasks ranged from only 29% to 44%. However, this shared trait variance predicted supervisory ratings beyond measures of intelligence. Further, Wüstenberg et al. (2012) reported moderate relations between CPS and reasoning and incremental validity for the former beyond the latter when predicting academic achievement using only one CPS assessment instrument based on MCS.

Given this inconsistent state of knowledge on CPS, it is difficult to decide whether CPS assessment instruments sufficiently converge and whether the aforementioned results are related to specific assessment instruments or to CPS as a construct. Research Question 1 is thus posed as follows: "Do computer-simulated CPS assessment instruments show convergent validity or is the assessment of CPS essentially measurement-specific?" .

1.2. Research Question 2: Construct validity of CPS and its relation to reasoning and academic achievement

Some characteristic features of problem solving abilities are part of almost any definition of intelligence (Sternberg & Berg, 1986). However, the translation of these features into measures of intelligence such as reasoning tests is usually limited to basic problem solving abilities. That is, the aforementioned characteristics of complex problems are usually not found in tests of intelligence. Specifically, the concept of CPS is incremental to measures of reasoning insofar as not all of the necessary information is available for CPS at the outset of the problem, active exploration of the system is inevitable, and procedural knowledge has to be used in order to control the underlying system and to account for the feedback received while interacting with it (Wüstenberg et al., 2012). For example, in CPS, the problem solver has to forecast the effects of multiple simultaneous interventions or counteract dynamic changes initiated by the system (Funke, 2001). On the other hand, typical reasoning tasks usually require examinees to draw conclusions to achieve goals but without these specific characteristics of a complex problem (Leighton, 2004). Thus, measures of reasoning (or intelligence in general; Wüstenberg et al., 2012) lack the characteristic features of complex problems. However, empirical findings on the relation between CPS and reasoning have been surprisingly inconsistent. Whereas Putz-Osterloh (1981) reported no relation between a CPS task and a test of fluid intelligence, Kröner et al. (2005) found high correlations between the two and thus concluded that CPS tasks could be used to assess fluid intelligence. However, these inconsistencies are likely to be related to some of the CPS measurement issues mentioned above. Recent findings based on MCS assessment instruments such as MicroDYN, the Genetics Lab, and MicroFIN have indicated that CPS and fluid intelligence are related but separate constructs (e.g., Greiff & Fischer, 2013; Greiff, Holt, et al., 2013; Greiff, Wüstenberg, et al., 2013; Wüstenberg et al., 2012). In this study, we tested for the first time how CPS and reasoning are related to each other on the construct level. After trying to establish the convergence of different CPS assessment instruments in Research Question 1, we aimed to generalize the results regarding the relation between CPS and reasoning to the construct level in Research Question 2.

Further, we evaluated the incremental validity of CPS on the construct level beyond measures of reasoning in predicting academic achievement in the natural and social sciences. We expected that both CPS and reasoning would be more strongly related to grades in the natural sciences than in the social sciences as CPS tasks require cognitive processes that are similar to those required in the natural sciences such as designing experiments or testing hypotheses (e.g., Klahr, 2002; Klahr & Dunbar, 1988). More specifically, the rationale behind

CPS tasks is closely related to the process of scientific discovery (Klahr & Dunbar, 1988). To this end, we interpreted a relation between CPS tasks and science grades as an indication of convergent validity. In the attempt to establish construct validity, we related different CPS assessment instruments in a multitrait–multimethod approach to a measure of reasoning as an indicator of fluid intelligence and to academic achievement. Research Question 2 was thus posed as follows: "How does CPS as captured across different assessment instruments relate to reasoning and to academic achievement?"

To address Research Questions 1 and 2, we used three assessment instruments based on MCS to target different CPS dimensions (i.e., knowledge and control) in order to establish construct validity. Establishing construct validity is achieved by thoroughly investigating research questions that have not been sufficiently addressed in previous studies.

2. Method

2.1. Participants[2]

Participants were 339 German university students (229 female, 92 male, 18 missing gender; mean age: 22.30 years; $SD = 4.02$), who were in their second year of study on average ($M = 1.99$, $SD = 1.69$) and who were enrolled in different undergraduate and graduate programs (57% in social studies, 28% in the natural sciences, and 15% in others). Students could choose between receiving partial course credit or a financial reimbursement of 20 € (approximately $25 US) for their participation. In a small number of cases, the technical delivery platform did not save the data correctly. Participants with this type of missing data were excluded from all analyses. According to the guidelines by Muthén and Muthén (2010), covariance coverage was generally acceptable. Testing took place at the Department of Psychology at the University of Heidelberg, Germany.

2.2. Testing and scoring procedures

Testing lasted approximately 4.5 h and was split into two sessions of 2.5 and 2 h, respectively. In the first session, participants worked on MicroDYN and provided demographic information and their school grades. In the second session, participants worked on the Genetics Lab, on MicroFIN, and on the matrices subtest of the Intelligence Structure Tests as a measure of reasoning (see below). Additional measures that are not relevant for this article were administered at both sessions. The entirely computer-based assessment was administered within the EE4CBA, a stand-alone test-delivery platform. Research assistants who had been thoroughly trained in test administration supervised the sessions.

Detailed descriptions of the three CPS assessment instruments can be found above. These instruments as well as the reasoning test were scored in line with the manuals or as described in the publications, in which the instruments were introduced.

2.2.1. MicroDYN Scoring

2.2.1.1. Knowledge. For knowledge, full credit was given if the model drawn by participants was completely correct, and no credit was given if participants' model contained at least one error.

2.2.1.2. Control. For control, full credit was given if the target values of all variables were achieved, whereas no credit was given if at least one target value was not achieved.

In summary, each task was scored with respect to its two phases, yielding 20 items overall across the 10 MicroDYN tasks.

2.2.2. Genetics Lab scoring

2.2.2.1. Knowledge. For knowledge, an established scoring algorithm (Müller, 1993) was applied combining relational knowledge (i.e., an effect exists or does not) and knowledge about the type and strength of a relation. For computing the global knowledge score, relational knowledge was emphasized by multiplying it by a weight of .75, whereas knowledge about the strength of an effect was weighted by .25 (for further details on this scoring, see Müller, 1993).[3]

2.2.2.2. Control. Participants' control performance was scored with regard to the quality of each applied control step (Keller & Sonnleitner, 2012). Only if a step maximally reduced the distance to the target values was it considered optimal and thus scored with 1 point. As there were three steps per control phase, the maximum score was 3 points.

2.2.3. MicroFIN scoring

2.2.3.1. Knowledge. Knowledge about each automaton was assessed by a total of three questions (see the description of MicroFIN for examples). Answers were scored dichotomously as either right or wrong.

2.2.3.2. Control. Full credit was given if the goal state was achieved with the minimal number of possible interventions. Partial credit was given if the goal was achieved with more interventions than necessary. No credit was given if the goal was not reached.

2.2.4. Matrices subtest of the Intelligence Structure Test

Reasoning was assessed by the matrices subtest of the "Intelligence Structure Test-Revised" (Beauducel, Liepmann, Horn, & Brocke, 2010). The test consisted of 20 2×2 matrices with a figural stimulus in all but one cell. In each matrix, one stimulus was missing, and participants had to choose the missing figural stimulus out of five alternatives. Answers were scored as right or wrong. This test assesses reasoning, which is known to be a good indicator of fluid intelligence (Carroll, 1993).

2.2.5. School grades

Participants were asked to report their final school grades in four natural science subjects (math, physics, chemistry, biology) and four social science subjects (German, history, geography, social studies). School grades ranged from 1 (excellent) to 6

[2] The full correlation matrix including parcels/manifest variables on MicroDYN, the Genetics Lab, MicroFIN, reasoning, and school grades can be downloaded as a supplement from the Elsevier Intelligence homepage.

[3] Please note that in MicroDYN, examinees are only asked to represent their relational knowledge of effects and not the strength of these effects.

(poor) as is usual in German schools. However, when calculating relations to other variables, we reversed the grades so that higher numerical values reflected better performance.

2.3. Statistical analysis

To tackle the first research question on the convergence of different CPS assessment instruments, we specified a latent multitrait–multimethod model with trait and method factors. This model was introduced by Eid, Lischetzke, Nussbeck, and Trierweiler (2003) and Eid et al. (2006). They refer to it as correlated trait correlated method minus 1 (CTC(M-1)) model (see below) and it was developed to tackle questions of measurement consistency. We then used structural equation modeling (SEM; Kline, 2011) to relate the CTC(M-1) model including MicroDYN, the Genetics Lab, and MicroFIN to reasoning and academic achievement to address the second research question. All relevant methods are explained in detail when they appear for the first time in the Results section.

In CTC(M-1) and in SEM, we interpreted path coefficients in the structural parts of the models and factor loadings in the measurement parts. To evaluate model fit, we reported standard model fit indices such as the CFI, TLI, RMSEA, and SRMR (or WRMR if the manifest variables were scored as ordered categorical variables) endorsing the cut-off values recommended by Hu and Bentler (1999). The Weighted Least Squares Mean and Variance adjusted (WLSMV; Muthén & Muthén, 2010) estimator for categorical outcomes was used whenever observed variables with ordered categorical scoring were involved; otherwise, we applied standard maximum likelihood (ML) estimation. Analyses were conducted in Mplus 6.11 (Muthén & Muthén, 2010).

2.4. Research Question 1 results: The measurement of CPS with different assessment instruments

In order to tackle Research Question 1, measurement models for the three CPS assessment instruments based on the MCS approach were derived. Subsequently, a comprehensive CTC(M-1) model (Eid et al., 2003) of knowledge and control as trait factors and the Genetics Lab and MicroFIN as method factors was calculated. MicroDYN was chosen as the reference method in this model, which generally contains one method factor less (i.e., 2) than the number of methods included (i.e., 3).

2.4.1. Measurement models of MicroDYN, the Genetics Lab, and MicroFIN

With regard to measurement models for MicroDYN and the Genetics Lab, we assigned items to parcels by considering the large number of free parameters to be estimated in models with single indicators and acknowledging that previous research indicated 2-dimensional structures for both measures (e.g., Wüstenberg et al., 2012, for MicroDYN and Sonnleitner et al., 2012, for the Genetics Lab). Having empirically based assumptions on dimensionality is a prerequisite for applying parceling (Kline, 2011). Thus, we used the item-to-construct balance recommended by Little, Cunningham, Shahar, and Widaman (2002) to combine items into parcels with each parcel consisting of at least three items. For MicroDYN and the Genetics Lab, a 2-dimensional model with the dimensions knowledge and control showed a good to acceptable fit (for MicroDYN: $\chi^2 = 20.26$; $df = 8$; $p < .001$; CFI = .983; TLI = .967; RMSEA = .069; SRMR = .027; for the Genetics Lab: $\chi^2 = 59.04$; $df = 19$; $p < .001$; CFI = .976; TLI = .965; RMSEA = .083; SRMR = .028) and significantly outperformed a 1-dimensional model (results not reported in detail). Latent correlations between the dimensions were .84 and .85 (both $p < .001$) for MicroDYN and Genetics Lab, respectively.

With regard to the third measure, MicroFIN, no research had empirically indicated its dimensionality and, therefore, no parceling was applied. The six knowledge items (three in each task) and the two control items (one in each task) constituted a measurement model with two CPS dimensions, which showed a correlation of $r = .68$ ($p < .001$) between knowledge and control and exhibited acceptable fit ($\chi^2 = 32.13$; $df = 19$; $p < .05$; CFI = .936; TLI = .906; RMSEA = .047; WRMR = .831[4]). When combining the two dimensions into one, there was hardly any drop in fit ($\chi^2 = 33.54$; $df = 20$; $p < .05$; CFI = .934; TLI = .908; RMSEA = .046; SRMR = .854) and evidence for a substantial difference between the 2- and 1-dimensional models was limited according to the χ^2-difference test (χ^2-difference = 1.71; $df = 1$; $p > .10$).[5] However, we decided to retain the 2-dimensional model to allow for a comparison between MicroFIN and the other two CPS measures in the MTMM analyses. When discussing the results, we need to keep in mind that a more parsimonious 1-dimensional model showed essentially equivalent fit to the 2-dimensional model in MicroFIN.

In sum, for MicroDYN and the Genetics Lab, 2-dimensional models including knowledge and control described the data well and showed good fit. For MicroFIN, results were less consistent, but a structurally identical model with knowledge and control was retained.

2.4.2. CTC(M-1) model of MicroDYN, the Genetics Lab, and MicroFIN

2.4.2.1. Background information on the CTC(M-1) model. According to Kline (2011), specific MTMM models within the SEM approach offer a more systematic way to delineate trait and method effects than the mere inspection of latent MTMM matrices (please refer to the online supplementary materials for the current study's standard latent MTMM matrix). However, a number of models have been proposed in the literature (e.g., Eid et al., 2006; Marsh & Grayson, 1995). Some of these models pose unrealistic restrictions such as uncorrelated method effects, thus allowing only traits to be related to each other (Eid et al., 2003), whereas other models assume both correlated traits and correlated methods. Such correlated trait correlated method (CTCM) models suffer from a number of problems. Specifically, analyses of CTCM models tend to yield inadmissible or unstable solutions (Kline, 2011) as they are not globally identified, and substantive interpretation problems arise when method factors are allowed to correlate, even though this is usually a reasonable assumption (e.g., Kenny & Kashy, 1992).

[4] When the WLSMV estimator for ordered categorical outcomes is used, the WRMR is reported instead of the SRMR.

[5] When using the WLSMV estimator for ordered categorical outcomes, differences in χ^2 values and *dfs* cannot be obtained by directly computing their difference, but follow a nonstandard procedure (Muthén & Muthén, 2010). We applied this procedure for all χ^2 difference tests that included the WLSMV estimation method.

As an alternative, Eid et al. (2003, 2006) introduce a model that is not affected by identification issues. This model contains one method factor less than methods included in the assessment and is labeled the correlated trait correlated method minus 1 (CTC(M-1)) model. That is, one method has to be chosen as the reference method, which constitutes the standard against which the other methods are compared. Eid et al. (2003, 2006) lay out the advantages of this model in detail. In the CTC(M-1) model, a method factor is specified for each trait–method combination except for the standard method, for which no method factors are defined. Thus, each observed variable of the nonstandard methods loads on its respective trait factor and its method factor, whereas each observed variable of the reference method loads on its respective trait factor only. Further, method factors may covary with each other and with trait factors. Method factors of one trait are not allowed to correlate with the respective trait factor (Eid et al., 2006). For instance, the knowledge method factor for the Genetics Lab was not allowed to correlate with the knowledge trait factor.

2.4.2.2. A CTC(M-1) model for MicroDYN, the Genetics Lab, and MicroFIN. For the current analyses, MicroDYN was chosen as the reference method because the existing body of evidence is most comprehensive for MicroDYN (e.g., Greiff et al., 2012; Schweizer et al., 2013; Wüstenberg et al., 2012). Thus, in the CTC(M-1) model with MicroDYN as the reference method, four method factors were specified in addition to the two trait factors knowledge and control: two for the Genetics Lab, constituting the method effect of the Genetics Lab on knowledge and the control parcels in the Genetics Lab, respectively, and two for MicroFIN, specifying the method effects of MicroFIN on the knowledge and control items in MicroFIN. Thus, each of the Genetics Lab and the MicroFIN indicators were specified to load on a trait factor and a method factor. Further, the four method factors were allowed to correlate with each other and with the trait factors of the other trait. For instance, the Genetics Lab method factor for knowledge was allowed to correlate with all other method factors as well as with the trait factor control but not with the trait factor knowledge.

The CTC(M-1) model is depicted in Fig. 4. Please note that not all paths are displayed in Fig. 4. The fit for this CTC(M-1) model was very good (Model 1 in Table 1), and the trait correlation of knowledge and control with regard to MicroDYN as the standard method was high ($r = .83$, $p < .001$), but slightly different from 1. Specifically, if the correlation between knowledge and control was constrained to 1.00, the drop in fit was significant (χ^2 difference = 7.15, $df = 1$, $p < .01$),[6] showing some discriminant validity between traits. The factor loadings of the indicators on the trait and method factors are displayed in Table 2. Further, Table 2 displays the variance components due to trait variation (consistency) and due to method variation (method specificity) for the nonstandard methods. Consistency and method specificity add up to 1.00 (Eid et al., 2003). Reliability estimates reported in Table 2 are defined as the total percentage of variance of an observed variable explained by the factors in the model.

For MicroDYN as the standard method, the consistency was per definition perfect. Considering that items were parceled for MicroDYN and the way reliability was defined, reliability estimates for MicroDYN were satisfactory. Eid et al. (2003) remind us that the CTC(M-1) model tends to underestimate the reliability of observed variables. For the Genetics Lab and MicroFIN, trait consistency and method-specific influences were substantial (Table 2), with consistency indicating convergent validity. Whereas the result pattern for different parcels varied little for the Genetics Lab with just over 50% consistency on average, the variance in the consistency coefficients was much larger for MicroFIN. In fact, MicroFIN showed lower reliability and higher variation in trait and method loadings overall. Knowledge Item 5 for MicroFIN was not substantially associated with the trait or method factors.

2.4.2.3. Relations of method factors with each other and with CPS trait factors in the CTC(M-1) model. CTC(M-1) models yield additional important information about the relations between the method factors and the relations between the method factors associated with one trait (e.g., MicroFIN knowledge) and the trait factor of the other trait (e.g., control).

As displayed in Table 3, method factors of different methods (i.e., the Genetics Lab and MicroFIN) were only loosely related, with coefficients ranging from .02 to .27. That is, there was little evidence that knowledge factors that were specific to GL or MF shared any variance when the common knowledge variance was taken into account. Method factors for different traits within one method, on the other hand (e.g., the Genetics Lab method factors knowledge and control) showed substantial correlations (.39 and .52), indicating some generalizability of method effects across traits. However, when the correlations between trait-specific method factors of one trait were constrained to 1, the χ^2 value decreased significantly (χ^2 difference = 25.94, $df = 1$, $p < .001$, for the Genetics Lab and 4.73, $df = 1$, $p < .05$, for MicroFIN) indicating that the strong assumption of perfect generalizability of method effects across traits had to be rejected (Eid et al., 2006). Thus, the method effects of the Genetics Lab and MicroFIN, respectively, on knowledge and control were different for both traits.

Table 3 further reports the correlations between the method factors of one trait and the other trait factor (e.g., MicroFIN method knowledge and control trait factor). Specifically, the knowledge method factors for the Genetics Lab and MicroFIN were correlated with control, and the control method factors for the Genetics Lab and MicroFIN were correlated with knowledge even though they were not substantial, ranging from .01 to .23. The method factors of one trait were not allowed to correlate with the factor of the same trait (Eid et al., 2003).

Whereas correlations between the method factors for MicroFIN and the trait factors reported in Table 3 were not substantial, the relations between the Genetics Lab method factors and the other trait were positive and significant. Albeit small in magnitude (.17 and .18, $p < .01$), these heteromethod coefficients of discriminant validity showed that the Genetics Lab's specific method effects were related to the traits of the standard method MicroDYN

[6] Throughout this article, a large number of models are evaluated, and most of them are similar versions of each other. We do not report fit statistics for all of these models, but only for the measurement models (in the text) and the main models (Table 1). If the fit for slightly changed models was not reported, the fit indices were generally favorable (i.e., CFI > .95; TLI > .95; RMSEA < .06; WRMR < 1.00; SRMR < .05).

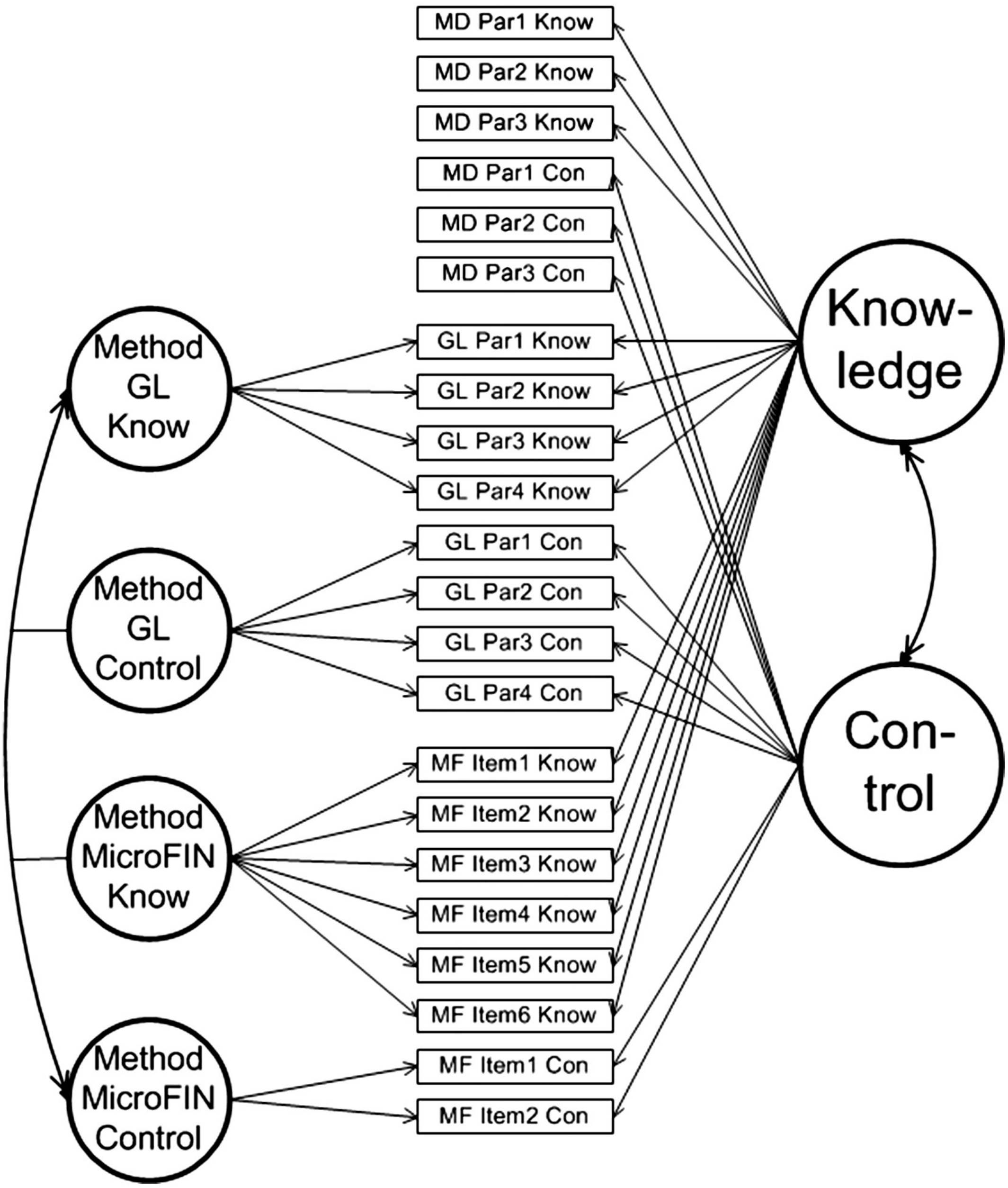

Fig. 4. Conceptual depiction of the CPS CTC(M-1) model for knowledge and control as trait factors with MicroDYN as the standard method and four method factors for the Genetics Lab and MicroFIN. MD = MicroDYN; GL = Genetics Lab; MF = MicroFIN; Par = parcel; Know = knowledge; Con = control. Error terms for endogenous variables are not depicted. Correlations between method factors and trait factors are not depicted. See the text for details.

Table 1
Goodness of fit indices for different models.

Model	χ^2	*df*	*p*	CFI	TLI	RMSEA	CI_{RMSEA} 90%	WRMR
Model 1: CPS CTC(M-1) model	193.66	182	.26	.993	.991	.014	[.000; .028]	.541
Model 2: CPS CTC(M-1) model predicted by reasoning	306.04	270	.06	.981	.978	.020	[.000; .030]	.630
Model 3: CPS CTC(M-1) model predicting academic achievement	369.64	367	.45	.999	.998	.005	[.000; .020]	.642
Model 4: reasoning and CPS CTC(M-1) model predicting academic achievement	513.54	492	.24	.991	.989	.011	[.000; .022]	.707

Note. df = degrees of freedom; CFI = comparative fit index; TLI = Tucker Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval; WRMR = weighted root mean square residual; χ^2 and df estimates are based on WLSMV.

Table 2
Standardized trait and method loadings, consistency, method specificity, and reliability for the CTC(M-1) model.

Method	Trait	Indicator	Trait loading	Method loading	Consistency	Method specificity	Reliability
MicroDYN	Knowledge	Parcel 1	.80**		1.00	.00	.63
		Parcel 2	.71**		1.00	.00	.50
		Parcel 3	.78**		1.00	.00	.62
	Control	Parcel 1	.76**		1.00	.00	.57
		Parcel 2	.67**		1.00	.00	.45
		Parcel 3	.74**		1.00	.00	.54
Genetics Lab	Knowledge	Parcel 1	.59**	.71**	.41	.59	.85
		Parcel 2	.57**	.64**	.44	.56	.74
		Parcel 3	.61**	.60**	.51	.49	.73
		Parcel 4	.65**	.59**	.55	.45	.77
	Control	Parcel 1	.53**	.49**	.54	.46	.52
		Parcel 2	.63**	.56**	.56	.44	.70
		Parcel 3	.53**	.58**	.46	.54	.62
		Parcel 4	.53**	.43**	.60	.40	.46
MicroFIN	Knowledge	Item 1	.37**	.70**	.22	.78	.63
		Item 2	.24**	.55**	.16	.84	.36
		Item 3	.27**	.71**	.13	.87	.57
		Item 4	.36**	.26**	.66	.34	.20
		Item 5	.17*	.03	.97	.03	.03
		Item 6	.38**	.26**	.68	.32	.21
	Control	Item 1	.55**	.25	.83	.17	.37
		Item 2	.24**	.69	.11	.89	.54

* $p < .05$.
** $p < .01$.

after controlling for influences due to the same method. According to Eid et al. (2003), these coefficients are pure discriminant validity correlations for traits not biased by common method influences. For instance, relations between the Genetics Lab method factor control and the trait factor knowledge (.17; Table 3) showed that participants with high knowledge also showed better performance in control in the Genetics Lab beyond what would have been expected due to their control performance in MicroDYN. The same interpretation applies to the Genetics Lab method factor knowledge and the trait factor control (.18). These coefficients, in fact, might indicate that the delineation between the two trait factors, knowledge and control, was not reliable.

Overall, the CTC(M-1) model of the traits knowledge and control and the methods MicroDYN, the Genetics Lab, and MicroFIN allowed an in-depth understanding of trait and method effects when measuring CPS by multiple complex systems. Trait loadings were substantial for MicroDYN and the Genetics Lab and exhibited high variance for MicroFIN. The consistency of the measures was acceptable, but substantial method effects were present as well. Correlations between method effects for the Genetics Lab and MicroFIN with knowledge and control beyond MicroDYN were existent but small in size with a good overall model fit for the CTC(M-1) model. In subsequent analyses to address Research Question 2, reasoning and academic achievement were added to the analyses on the CTC(M-1) model displayed in Fig. 4, and the relations of these constructs to the trait factors knowledge and control as well as to the method factors of the Genetics Lab and MicroFIN were inspected.

2.5. Research Question 2 results: The construct validity of CPS and its relation to reasoning and academic achievement

To establish construct validity, we related the CTC(M-1) model to reasoning and, in a second step, to academic achievement. Additionally, we evaluated the potential of the CTC(M-1) trait and method factors to incrementally predict school grades in the natural and social sciences. In all these

Table 3
Correlations between trait and method factors in the CTC(M-1) model.

		Knowledge	Control	Method GL Knowledge	Method GL Control	Method MF Knowledge	Method MF Control
Trait factors	Knowledge						
	Control	.83**					
Method Genetics Lab	Knowledge	XX	.18**				
	Control	.17**	XX	.52**			
Method MicroFIN	Knowledge	XX	.01	.19*	.23*		
	Control	.23	XX	.02	.27	.39	

Note. GL = Genetics Lab; MF = MicroFIN. XX = correlations are restricted to 0 in the CTC(M-1) model.
* $p < .05$.
** $p < .01$.

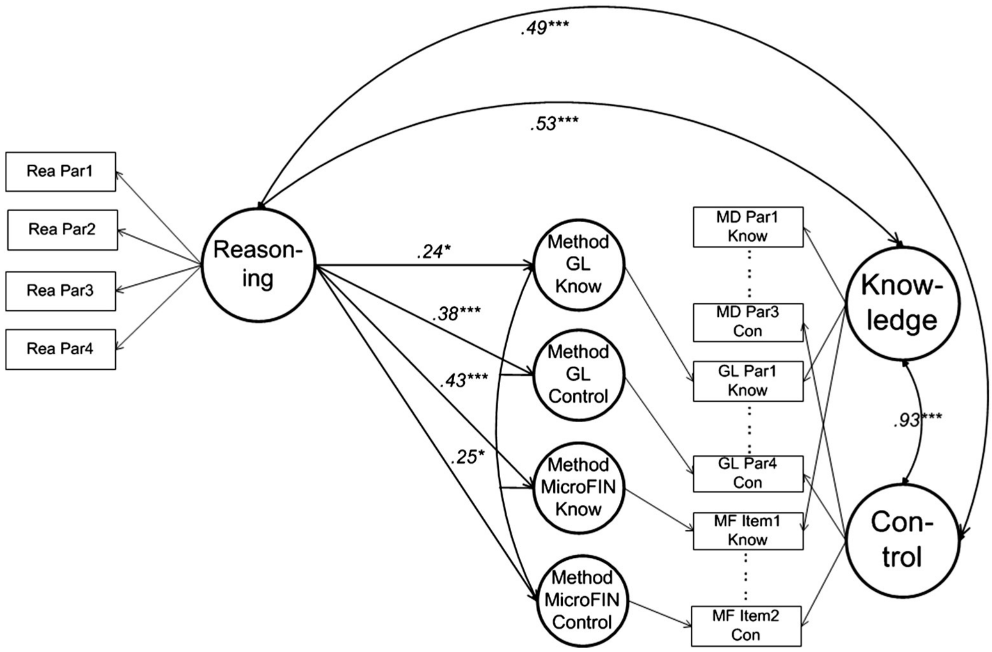

Fig. 5. The CPS CTC(M-1) model predicted by reasoning. Rea = reasoning; MD = MicroDYN; GL = Genetics Lab; MF = MicroFIN; Par = parcel; Know = knowledge; Con = control. Only parts of the CTC(M-1) model are displayed (for all information, see Fig. 4, Tables 2 and 3); parcel loadings for reasoning are reported in the text. Only significant paths depicted. Error terms for endogenous variables are not depicted. *$p < .05$. ***$p < .001$.

analyses, we used the CTC(M-1) model introduced above, checking for relations of reasoning and academic achievement with the trait factors knowledge and control measured by the standard method MicroDYN and with the four method factors of the Genetics Lab and MicroFIN.

2.5.1. Relating MicroDYN, the Genetics Lab, and MicroFIN to reasoning

2.5.1.1. The reasoning measurement model. For the matrices test of the Intelligence Structure Test, the 20 available items were assigned to parcels. The fit for a 1-dimensional model was excellent ($\chi^2 = .31$; $df = 2$; $p > .05$; CFI = 1.000; TLI =1.000; RMSEA = .000; SRMR = .001).

2.5.1.2. Prediction of CPS trait and method factors in the CTC(M-1) model by reasoning. The model fit for the overall structural model in which the two CPS trait factors and the four method factors of the CTC(M-1) model were predicted by reasoning was good (Model 2 in Table 1). Both trait factors, knowledge and control, were substantially predicted by reasoning (Fig. 5), explaining 28% and 24% of the variance, respectively. Even after controlling for reasoning, knowledge and control remained substantially correlated. In fact, after controlling for reasoning, the latent correlation between knowledge and control increased compared to a model without reasoning (from $r = .83$ to $r = .93$, $p < .05$) indicating almost identity between the two dimensions. This surprising result suggests that the somewhat different amounts of reasoning in knowledge and control led to discriminant validity between the two. Specifically, the "reasoning aspects" in knowledge were correlated only to a certain extent with control (the same was true for those in control and their correlations with knowledge). If these aspects were controlled for, only the unique proportions of knowledge and control were correlated, but these—apparently—have more in common than knowledge and control with the "reasoning aspects" included.

Interestingly, both method factors for the Genetics Lab and the knowledge method factor for MicroFIN showed significant relations of moderate size to reasoning (Fig. 5), explaining between 6% and 18%. That is, the Genetics Lab and MicroFIN shared common variance with reasoning beyond MicroDYN. However, the overall amount of variance in the Genetics Lab and MicroFIN explained by reasoning (indirectly through the two trait factors and directly through the method factors) ranged between 22% and 28% of the variance and was comparable to the explained variance in MicroDYN as the standard method.

In summary, discriminant validity between reasoning and CPS was supported in our study: An average of around 25% of the variance in the CTC(M-1) model was predicted by reasoning independent of the specific method (i.e., MicroDYN, Genetics Lab, MicroFIN).

2.5.2. Relating MicroDYN, the Genetics Lab, and MicroFIN to academic achievement

In order to establish the construct validity of CPS, it is crucial to show that it has discriminant validity with reasoning (see Fig. 5), but beyond this, the prediction of relevant external criteria needs to be shown as well.

2.5.2.1. The academic achievement measurement model. Academic achievement was composed of two latent factors: grades in the natural sciences (math, physics, chemistry, and biology) and social sciences (German, history, geography, and social studies). Each of the grades was specified to load on its respective factor. Both school grade factors were substantially correlated at $r = .67$ ($p < .001$). The overall model fit was very good ($\chi^2 = 22.86$; $df = 19$; $p > .10$; CFI = .992; TLI = .989; RMSEA = .026; SRMR = .037) and substantially better than for a 1-dimensional model (results not reported in detail).

2.5.2.2. The prediction of academic achievement by CPS trait and method factors in the CTC(M-1) model. Model fit for a structural CTC(M-1) model with the trait factors knowledge and control and the four method factors predicting the two latent grade factors for the natural and social sciences was good (Model 3 in Table 1). Out of the two trait factors, only knowledge explained substantial amounts of variance in natural science grades (standardized path size: .30; $p < .001$; $R^2 = .09$), whereas control was not statistically relevant ($p > .10$), and social science grades were not predicted at all ($p > .10$). Further, none of the paths regressing grades on the four method factors were significant, thus indicating that the Genetics Lab and MicroFIN did not contain unique variance that was relevant for predicting school grades and that was not already captured by MicroDYN.

2.5.2.3. The prediction of academic achievement by CPS trait and method factors in the CTC(M-1) model controlling for reasoning. In order to evaluate whether CPS would uniquely predict school grades—that is, whether it would exhibit incremental validity beyond reasoning—we evaluated a model in which knowledge and control as trait factors were regressed on reasoning, and only the proportion of variance unrelated to reasoning predicted school grades. This approach corresponds to a latent stepwise regression with reasoning as a first step predictor and CPS as a second step predictor. Relations between the four method factors and academic achievement after controlling for reasoning were not estimated because in the model without reasoning, these relations were already nonsignificant (Model 3 in Table 1).

The final structural model with reasoning and trait factors in the CTC(M-1) model predicting academic achievement showed good fit (Model 4 in Table 1). It is depicted in Fig. 6. As expected, reasoning predicted natural science grades (standardized path size: .28; $p < .001$; $R^2 = .08$) and, weakly, social sciences (standardized path size: .12; $p < .05$; $R^2 = .01$). Those parts of the CPS trait factor knowledge unrelated to reasoning further predicted natural science grades (standardized path size: .22; $p < .01$; $R^2 = .05$) beyond reasoning, whereas control did not yield substantial paths. CPS did not explain social science grades.

As the two trait factors knowledge and control were highly related (around .85) in the CTC(M-1) model and this correlation even increased when controlling for reasoning (to .95; see Figs. 5 and 6), we combined them into one second

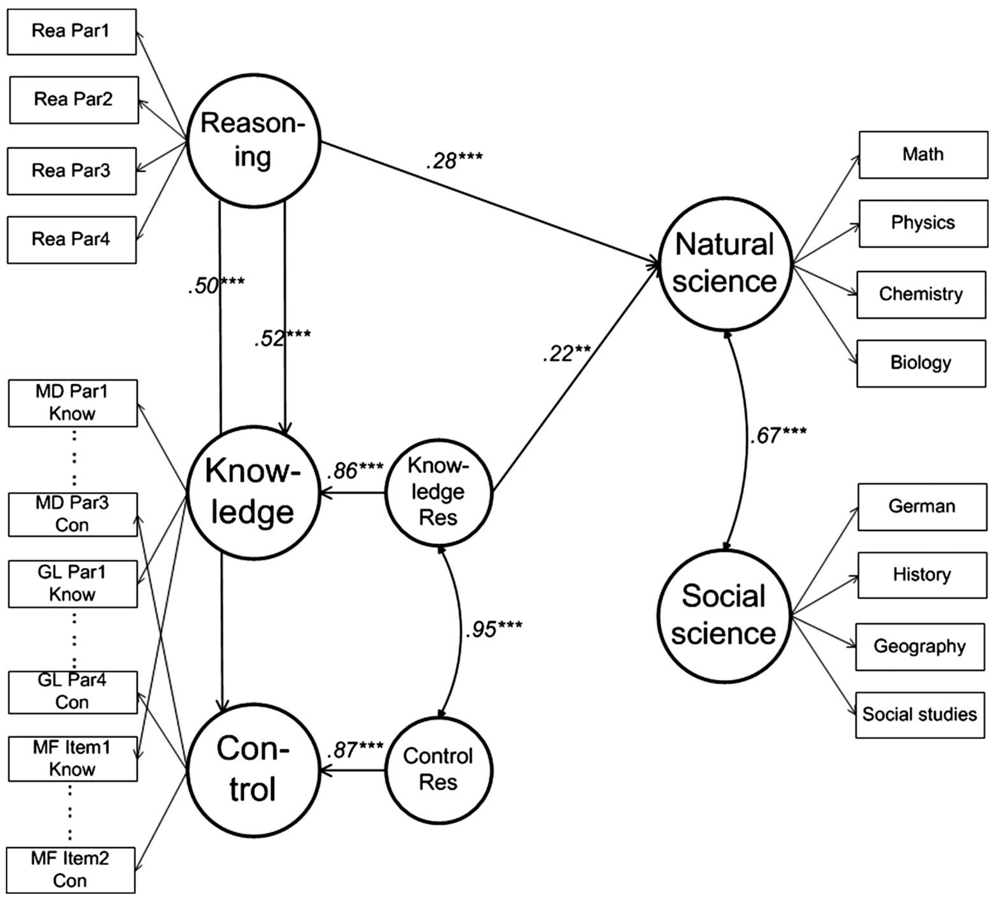

Fig. 6. Reasoning and the CPS CTC(M-1) model predicting academic achievement. Rea = reasoning; MD = MicroDYN; GL = Genetics Lab; MF = MicroFIN; Par = parcel; Know = knowledge; Con = control; Res = residual. Only parts of the CTC(M-1) model are displayed (for all information, see Fig. 4, Tables 2 and 3). Only significant paths are depicted. Method factors are not depicted. Error terms for endogenous variables are not depicted. **$p < .01$. ***$p < .001$.

order CPS factor in order to evaluate whether the incremental validity results would remain essentially the same in an additional model. In this model, reasoning was a good predictor of natural science grades (standardized path size: .28; $p < .001$; $R^2 = .08$) and a weak predictor of social science grades (standardized path size: .14; $p < .05$; $R^2 = .02$). Additionally, CPS in the CTC(M-1) model (this time as a second order factor) showed incremental validity beyond reasoning with a path size that was virtually unchanged (standardized path size: .18; $p < .01$; $R^2 = .03$).

In summary, the analyses for Research Question 2 revealed that the CPS trait factors knowledge and control were moderately related to reasoning and incrementally predicted natural science grades beyond reasoning, additionally explaining approximately 3% to 5% of the variance. The Genetics Lab and MicroFIN as nonstandard methods did not show unique incremental proportions of variance beyond the standard method of MicroDYN.

3. Discussion

The overarching aim of this study was to facilitate our understanding of CPS as a complex cognitive ability. As previous research has largely relied on specific assessment instruments, we administered a variety of different CPS assessment instruments to a sample of German university students. In general, the results of our study provided support for the conceptualization of CPS as a complex cognitive ability. Specifically, the results of Research Question 1 indicated that CPS assessment instruments based on the MCS approach sufficiently converged in a CTC(M-1) model (Main Finding 1). Surprisingly, in contrast to previous research, support from this model for delineating CPS into knowledge and control was limited (Main Finding 2). For Research Question 2, CPS and reasoning were moderately related and CPS predicted academic achievement beyond reasoning (Main Finding 3). Overall, this study provided evidence for the construct validity of CPS both in terms of convergent validity as addressed by Research Question 1 and in terms of construct validity as addressed by Research Question 2.

3.1. Main finding 1: CPS assessment instruments converge sufficiently

The trait loadings of MicroDYN, the Genetics Lab, and MicroFIN were acceptable, ranging from .17 to .80 with generally smaller loadings for MicroFIN. Consistency coefficients, as the percentage of trait loadings that were not attributable to method effects, were substantial for almost all indicators, ranging from .41 to .60 for the Genetics Lab and from .11 to .97 for MicroFIN in relation to MicroDYN as the reference method. In a different research context, Höfling, Schermelleh-Engel, and Moosbrugger (2009) reported consistency coefficients between .17 and .44 as benchmarks, and they considered these to be substantial in their application of a CTC(M-1) model with self-reports and different types of peer reports as method factors. In a recent study on CPS using more than one assessment instrument (albeit not based on MCS but on classical microworlds), Danner et al. (2011) reported consistency estimates between .29 and .44. Comparing these results with our results suggests that the MCS approach may be advantageous in terms of consistency across measures.

Whereas MicroDYN and the Genetics Lab are both based on LSE, MicroFIN is based on FSA, and in our study, MicroFIN was included in only two of the CPS tasks. As a consequence, MicroFIN trait loadings and consistency coefficients exhibited a higher variance and were not satisfactory in part. For instance, one knowledge item on the second MicroFIN task (Item 5) showed a small trait loading of only .17. That is, all three MCS approaches applied in this study—MicroDYN, the Genetics Lab, and MicroFIN—cover somewhat different aspects of CPS with relations being closest between MicroDYN and the Genetics Lab. The OECD (2010) states that the LSE formalism (in MicroDYN and the Genetics Lab) usually leads to a consistent and homogenous set of tasks, whereas the FSA formalism (in MicroFIN) covers a broader variety of tasks (Wirth & Klieme, 2003), which in turn may lower consistency and make scalability more difficult, but which also allows test developers to represent a richer collection of complex problem situations. Bearing this in mind, MicroFIN was "disadvantaged" in this study in a number of ways: fewer items, no parceling, and less empirical experience with CPS assessment through MicroFIN. To this end, it would be premature to certify that MicroFIN offers bad scalability or insufficient fit within the CTC(M-1) model. Instead, we suggest further pursuing MicroFIN as a CPS assessment approach both from conceptual and empirical points of view to fully understand its role and how it may complement LSE assessment approaches.

We conclude that there is a convergence across different CPS assessment instruments, which is a prerequisite for a meaningful application of CPS in international large-scale assessments such as the PISA survey. This availability of alternative assessment instruments touches another important point: training to the test. If, for instance, MicroDYN were used in high-stakes situations such as personnel selection or entrance testing to tertiary education, commercial training programs and guidebooks could quickly appear on the market. In MicroDYN, systematically varying one input variable (e.g., Training A; Fig. 1) while holding the others constant (e.g., Trainings B and C) helps develop an understanding of the causal relations between inputs and outputs and substantially enhances performance (Wüstenberg et al., 2012). Teaching this rather simple strategy may change the nature of the underlying construct assessed with MicroDYN, thus rendering the development of alternative assessment instruments essential for the field of CPS. Our results indicate that the development of alternative assessment instruments such as the Genetics Lab or MicroFIN is indeed possible even though this study can only provide a tentative starting point.

3.2. Main finding 2: Support for delineating CPS into knowledge and control was limited

There was mixed evidence in our study with regard to the number of CPS dimensions. In fact, evidence partly suggested that CPS should not be delineated into knowledge and control. When controlling for reasoning in the CTC(M-1) model, the latent relation between the two dimensions increased from around .85 to .95 (i.e., almost identity), and a second order factor of CPS was as incrementally predictive of academic achievement as knowledge and control separately. Thus, we could not find consistent differences between the two trait factors knowledge and control. On the other hand, a 2-dimensional measurement model showed a clearly better fit than a 1-dimensional model for both MicroDYN and the Genetics Lab. This latter result pattern was expected as the position that knowledge and control were overlapping yet distinct processes has been widely adopted in CPS research and has been backed up by a number of empirical studies (e.g., Wirth & Klieme, 2003; Wüstenberg et al., 2012).

In line with previous research, we expected moderate to high relations between knowledge and control and found only limited evidence for delineating the two at all. An explanation can be found in the rationale underlying the MCS approach. There, the control phase following the knowledge phase (cf. description of MicroDYN, the Genetics Lab, and MicroFIN) is deliberately kept short and does not encompass more than five independent interactions with any specific CPS task to facilitate standardization of the assessment (Greiff et al., 2012) and to prevent examinees from further acquiring knowledge during the control phase (Wüstenberg et al., 2012). However, in this trade-off between standardization and complexity, examinees have limited opportunity to display their ability to interactively regulate a CPS task over longer series of interventions. Thus, control performance naturally relies more strongly on the knowledge acquired in the knowledge phase (Wüstenberg et al., 2012). As a solution, MCS tasks could be extended by presenting several short control phases on each CPS task, enabling examinees to display their full control potential while restricting the acquisition of knowledge in the control phase.

Further, the relation between knowledge and control increased substantially when controlling for reasoning in our study. To this end, both method effects (please remember: method effects in the CTC(M-1) model did not generalize across methods or traits) and different aspects and/or different amounts of reasoning in knowledge and control might have facilitated a spurious empirical difference between the two trait factors. That is, even though reasoning explained knowledge and control roughly to the same extent, different aspects of reasoning (e.g., inductive reasoning for knowledge) might be related to knowledge and control, respectively. After controlling for these aspects by partialling reasoning out, the distinction between knowledge and control vanished, indicating differential influences of reasoning. However, with no straightforward explanation at hand at this stage, we need to further understand the relations of different aspects of CPS in terms of internal validity and in combination with other cognitive abilities such as reasoning. The results reported here should be considered a first starting point.

3.3. Main finding 3: CPS shows incremental validity beyond reasoning

Recent results indicate that CPS is moderately related to other cognitive abilities and exhibits predictive validity beyond them (e.g., Danner et al., 2011; Greiff, Holt, et al., 2013; Greiff, Wüstenberg, et al., 2013). In fact, these empirical findings along with a strong focus on problem solving in education (Mayer & Wittrock, 2006) facilitated the inclusion of CPS in the PISA 2012 survey.

Nevertheless, all previous accounts of the predictive validity of CPS using academic achievement as the criterion have been limited to specific assessment instruments and, thus, lacked generalizability. In this study, we examined relations between CPS and reasoning on the construct level and extended accounts of predictive validity to several CPS assessment instruments (Campbell & Fiske, 1959). In the CTC(M-1) model we tested, knowledge and control were different from reasoning, and CPS exhibited predictive validity beyond reasoning when predicting natural science grades, thus indicating construct validity and convergent validity for CPS. Reasoning predicted social science grades weakly and CPS not at all, which was expected as grades in the social area are generally less well-predicted than in the natural sciences. Beyond this explanation, the finding that CPS predicts natural science grades but does not predict grades in the social sciences is aligned with the conceptualization of CPS. MicroDYN, the Genetics Lab, and MicroFIN all require a systematic approach to understanding causal connections in new situations and, subsequently, they require the application of this knowledge. That is, taking a scientific and systematic approach toward MCS tasks usually results in better performance, and according to Klahr and Dunbar (1988), such an approach can be directly related to problem solving in the context of scientific discovery.

Klahr and Dunbar (1988) describe the scientific acquisition of knowledge as a dual search for hypotheses and information that involves searching for hypotheses to explain information as well as searching for information to test these hypotheses (see also Klahr, 2002). The result of this search is assessed in CPS knowledge scores (i.e., by evaluating causal diagrams in MicroDYN and the Genetics Lab or by asking questions about the causal structure of the problem situation in MicroFIN). Applying the knowledge generated in the subsequent attempt of system control in CPS is essentially the search for a solution to the problem (Newell & Simon, 1972). The effectiveness of searching for information, hypothesis building, and carrying out solutions in the process of CPS may depend heavily on explicit and implicit knowledge about when and how to apply specific strategies such as systematic variation and on the use of heuristics (Kuhn, 2000) such as a means-end analysis, hill-climbing, or random generate-and-test (Klahr, 2002). For instance, if participants want to control a MicroDYN task after having acquired a sufficient amount of knowledge via hypothesis testing, they have to know how to plan interventions and how to use certain heuristics such as a means-end analysis. That is, participants have to abstract from irrelevant details and have to generate a plan for the specific problem at hand (cf. Klahr, 2002; Newell & Simon, 1972). Further, they have to understand that some strategies may be more efficient than others and some heuristics more appropriate than others for a specific complex problem (Klahr, 2002). Thus, principles of scientific discovery (i.e., experimental hypothesis testing) as well as the use of heuristics (i.e., means-end analysis) may play important roles in CPS.

Beyond the ability to use domain-general principles of scientific discovery and the application of strategies and heuristics, learning may play an important role in CPS tasks. In fact, acquiring knowledge about an unknown nontransparent complex system tells us a lot about a person's ability to learn within the testing procedure itself. Subsequently, this novel information needs to be applied in CPS tasks. Flexibly adapting to new situations, penetrating them, and applying new information is closely related to learning and applying what was learned. Both aspects are integral parts of CPS. To this end, CPS tasks are very similar to trainability tests.[7] Typically, trainability tests contain work samples to test behavior that is important for specific jobs (e.g., the job of an electrician) without relying on previous experience in this job (Sackett, 2000). They involve a period of time in which participants can acquire relevant knowledge before performance is assessed (Roth, Buster, & Bobko, 2011), similar to the exploration phase found in CPS tasks (Greiff et al., 2012). Although important differences between CPS tasks and trainability tests exist (e.g., trainability tests clearly apply to a specific content area, whereas this holds for CPS tasks to a lesser extent), the principle of learning central to trainability tests is also found in the concept of CPS and its assessment instruments.

Given that learning and general principles of scientific discovery are integral parts of CPS, we argue that these aspects may be responsible for the added value of CPS. Apart from these explanations, other cognitive and noncognitive constructs such as personality or motivation may be related to the incremental validity of CPS. However, empirical evidence on these issues is scarce, and we strongly encourage future research to address these issues more comprehensively by providing further empirical evidence.

3.4. Limitations

Some limitations in our study need consideration. First, the sample size was small with respect to the number of data points and parameters. Nevertheless, parameter estimates remained stable across analyses, and Yu (2002) reports that valid parameter estimates can be found in SEM for $N < 200$. In addition, the variance in our sample was restricted as only university students were assessed. We can only conjecture about the specific impact of this restriction, but usually restricted variance artificially decreases correlations and path coefficients. We may, therefore, expect higher correlations between CPS, reasoning, and natural science grades in a nonrestricted sample. This restriction in variance may further be responsible for the lack of (substantial) prediction of social science grades, which are usually predicted by measures of cognitive ability. On the other hand, in comparison to this study, Greiff, Wüstenberg, et al. (2013) found similarly sized relations between CPS and reasoning in a sample that covered the full range of cognitive ability. Thus, we conclude that our results need to be interpreted with caution and that future research needs to pay special attention to providing data based on less selective samples. Further, in CTC(M-1) models, one reference method is arbitrarily chosen (MicroDYN in our study). A disadvantage is that results may change depending on which method is used as the benchmark. However, several advantages of the CTC(M-1) model outweigh this disadvantage, and MicroDYN appeared to work well as the reference method as indicated by the nonsignificant paths between the other two methods and academic achievement when evaluating predictive validity. The operationalization of intelligence was rather narrow given that only reasoning was assessed. However, reasoning is considered an excellent marker of fluid intelligence, and Carroll (1993) places it at the core of intelligence. Nevertheless, in line with Wüstenberg et al. (2012), we agree that a more comprehensive operationalization of intelligence would be useful, which, for instance, should relate to the Carroll–Horn–Cattell theory of general mental ability (McGrew, 2009). In fact, limiting intelligence to measures of reasoning is a general shortcoming in studies investigating the relation between CPS and intelligence, and this conceptualization needs to be expanded in future research.

[7] We thank an anonymous reviewer for this comment.

3.5. Conclusion

Complex cognitive abilities were considered relevant long before computers became available in assessment practice to place more complex demands on examinees, but the translation of these abilities into valid assessment instruments was severely constrained by the limitations that were inherent to paper–pencil tests. That is, one could say that there was an "impasse" in cognitive ability testing—a mismatch between the assessment requirements and the operators available to develop it further. As a consequence, researchers had to rely on paper–pencil-based assessments of intelligence despite the limitations associated with them (Rigas et al., 2002). With the introduction of computers into assessment practices, an entirely new world of opportunities both in terms of concepts and assessment instruments opened up (Bunderson et al., 1989), but just a few years ago Williamson et al. (2006) stated that our knowledge about abilities such as CPS had advanced surprisingly little over the last 2 decades. To this end, this study provided evidence for the convergence of different measures of CPS and extended previous results on predictive validity gathered with specific CPS assessment instruments.

On a more general level, there is no doubt that measures of intelligence are relevant predictors of (cognitive) outcomes, but demands in our world are moving rapidly toward increased complexity and, thus, toward an increased demand for CPS (Autor et al., 2003). In other words, cognitive challenges in our lives grow more complex every single day and so do the cognitive abilities we need. Bearing in mind that people must constantly interact with dynamic environments (OECD, 2010), solely using simple task environments to measure cognitive performance does not meet the requirements of modern assessment. The world moves on and the assessment of cognitive abilities such as CPS has to keep pace.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.intell.2013.07.012.

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), 1279–1333.

Baker, E. L., & O'Neil, H. F., Jr. (2002). Measuring problem solving in computer environments: Current and future states. *Computers in Human Behavior*, *18*, 609–622.

Beauducel, A., Liepmann, D., Horn, S., & Brocke, B. (2010). *Intelligence Structure Test.* Oxford: Hogrefe.

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

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

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational Measurement* (pp. 367–407). New York, NY: Macmillan Publishing.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, *54*, 297–312.

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factoranalytic studies.* Cambridge: Cambridge University Press.

Danner, D., Hagemann, D., Schankin, A., Hager, M., & Funke, J. (2011). Beyond IQ. A latent state trait analysis of general intelligence, dynamic decision making, and implicit learning. *Intelligence*, *39*, 323–334.

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

Dunbar, K. (1998). Problem solving. In W. Bechtel, & G. Graham (Eds.), *A companion to cognitive science* (pp. 289–298). London: Blackwell.

Eid, M., Lischetzke, T., & Nussbeck, F. W. (2006). Structural equation models for multitrait–multimethod data. In M. Eid, & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 283–299). Washington, DC: APA.

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait–multimethod methods: A multiple-indicator CT-C(M-1) model. *Psychological Methods*, *8*(1), 38–60.

Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *Journal of Problem Solving*, *4*(1), 19–41.

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

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

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, *13*(1), 1–4.

Greiff, S., & Fischer, A. (2013). Der nutzen einer komplexen problemlöse-kompetenz: Theoretische überlegungen und empirische befunde [Usefulness of Complex Problem Solving competency: Theoretical considerations and empirical results]. *Zeitschrift für Pädagogische Psychologie*, *27*(1), 1–13.

Greiff, S., Holt, D. V., & Funke, J. (2013). Perspectives on problem solving in cognitive research and educational assessment: analytical, interactive, and collaborative problem solving. *Journal of Problem Solving*, *5*, 71–91.

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new measurement perspective. *Applied Psychological Measurement*, *36*, 189–213.

Greiff, S., Wüstenberg, S., Molnar, G., Fischer, A., Funke, J., & Csapo, B. (2013). Complex Problem Solving in educational settings — something beyond g: concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, *105*, 364–379.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve. Intelligence and class structure in American life.* New York: Free Press.

Höfling, V., Schermelleh-Engel, K., & Moosbrugger, H. (2009). Analyzing multitrait–multimethod data. *Methodology*, *5*(3), 99–111.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Keller, U., & Sonnleitner, P. (2012). *Genetics Lab scoring algorithm.* Luxembourg: University of Luxembourg.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait–multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, *112*, 165–172.

Klahr, D. (2002). *Exploring science: The cognition and development of discovery processes.* Cambridge: MIT press.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*, 1–48.

Klauer, K., & Phye, G. (2008). Inductive reasoning: A training approach. *Review of Educational Research*, *78*(1), 85–123.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford.

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

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science*, *9*(5), 178–181.

Kuncel, N., Hezlett, S., & Ones, D. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, *127*, 62–81.

Kyllonen, P. C. (2009). New constructs, methods, and directions for computer-based assessment. In F. Scheuermann, & J. Björnsson (Eds.), *The transition to computer-based assessment* (pp. 151–156). Luxembourg: Office for Official Publications of the European Communitites.

Leighton, J. P. (2004). Defining and describing reason. In J. P. Leighton, & R. J. Sternberg (Eds.), *The Nature of Reasoning* (pp. 3–11). Cambridge: Cambridge University Press.

Leutner, D. (2002). The fuzzy relationship of intelligence and problem solving in computer simulations. *Computers in Human Behavior*, *18*, 685–697.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*(2), 151–173.

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait–multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 177–198). Thousand Oaks, CA: Sage.

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.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10.

Müller, H. (1993). *Komplexes problemlösen: Reliabilität und Wissen [Complex problem solving: Reliability and knowledge].* Bonn, Germany: Holos.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

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 (2006). The Programme for International Student Assessment 2006. Accessed at: http://www.oecd.org/dataoecd/15/13/39725224.pdf

OECD (2010). PISA 2012 field trial problem solving framework. Accessed at: http://www.oecd.org/dataoecd/8/42/46962005.pdf

Osman, M. (2010). Controlling uncertainty: A review of human behavior in complex dynamic environments. *Psychological Bulletin*, *136*, 65–86.

Putz-Osterloh, W. (1981). Über die beziehung zwischen testintelligenz und problemlöseerfolg [On the relation between test intelligence and problem solving success]. *Zeitschrift für Psychologie*, *189*, 79–100.

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

Rigas, G., Carling, E., & Brehmer, B. (2002). Reliability and validity of performance measures in microworlds. *Intelligence*, *30*, 463–480.

Roth, P. L., Buster, M. A., & Bobko, P. (2011). Updating the trainability tests literature on Black–White subgroup differences and reconsidering criterion-related validity. *The Journal of Applied Psychology*, *96*, 34–45.

Sackett, P. R. (2000). Personnel selection. In A. E. Kazdin (Ed.), *Encyclopedia of psychology* (pp. 152–156). New York, NY: Oxford University Press.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology*, *52*, 471–499.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*(1), 162–173.

Schweizer, F., Wüstenberg, S., & Greiff, S. (2013). Do Complex Problem Solving dimensions measure something over and above working memory capacity? *Learning and Individual Differences*, *24*, 42–52.

Sonnleitner, P., Brunner, M., Greiff, S., Funke, J., Keller, U., Martin, R., et al. (2012). The Genetics Lab. Acceptance and psychometric characteristics of a computer-based microworld to assess complex problem solving. *Psychological Test and Assessment Modeling*, *54*, 54–72.

Sonnleitner, P., Keller, U., Martin, R., & Brunner, M. (2013). Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence*, *41*, 289–305.

Sternberg, R. J., & Berg, C. A. (1986). Quantitative integration: Definitions of intelligence: A comparison of the 1921 and 1986 symposia. In R. J. Sternberg, & D. K. Detterman (Eds.), *What is intelligence? Contemporary* *viewpoints on its nature and definition* (pp. 155–162). Norwood, New Jersey: Ablex.

Williamson, D. M., Bejar, I. I., & Mislevy, R. J. (2006). *Automated scoring of complex tasks in computer-based testing: An introduction.* Mahwah, NJ: Lawrence Erlbaum.

Wirth, J., & Klieme, E. (2003). Computer-based assessment of problem solving competence. *Assessment in Education: Principles, Policy & Practice*, *10*(3), 329–345.

Wittmann, W., & Hattrup, K. (2004). The relationship between performance in dynamic systems and intelligence. *Systems Research and Behavioral Science*, *21*, 393–440.

Wittmann, 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, traits, and content determinants* (pp. 77–108). Washington, DC: APA.

Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving. More than reasoning? *Intelligence*, *40*, 1–14.

Yu, C. -Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes.* Los Angeles, CA: University of California.

ARTICLE IN PRESS
LEAIND-01054; No of Pages 8
Learning and Individual Differences xxx (2015) xxx–xxx
Contents lists available at ScienceDirect
Learning and Individual Differences
journal homepage: www.elsevier.com/locate/lindif

Assessing analytic and interactive aspects of problem solving competency

Andreas Fischer [a,*], Samuel Greiff [b], Sascha Wüstenberg [b], Jens Fleischer [c], Florian Buchwald [c], Joachim Funke [a]

[a] University of Heidelberg, Hauptstr. 47-50, 69117 Heidelberg, Germany
[b] University of Luxembourg, 6, rue Richard Coudenhove Kalergi, 1359 Luxembourg-Kirchberg, Luxembourg
[c] University of Duisburg-Essen, Berliner Platz 6-8, 45127 Essen, Germany

ARTICLE INFO

Article history:
Received 9 May 2014
Received in revised form 29 January 2015
Accepted 20 February 2015
Available online xxxx

Keywords:
Interactive Problem Solving
Analytic Problem Solving
Complex Problem Solving
Reasoning
Assessment
MicroDYN

ABSTRACT

This study is about different approaches to assessing Problem Solving Competency (PSC) applied in international large-scale assessments: Analytic Problem Solving (APS) and Interactive Problem Solving (IPS). Based on a university student sample (n = 339) and a high-school student sample (n = 577) we found that both approaches are highly interrelated in both samples, even after controlling for reasoning (R^2 = .33 to .52) indicating that both approaches address a common core of PSC. However, our results also indicate that unique aspects of APS and IPS (beyond each other and reasoning) are explanatory for school achievements in the high-school student sample. However, in the university student sample, only APS has a unique contribution to explaining school achievements (beyond IPS and reasoning) and our findings indicate, that APS – and not interactivity itself – may explain the incremental validity of IPS (beyond reasoning) reported in previous studies. Implications for problem solving research and educational practice are discussed.

1. Introduction

Solving real problems is a complex endeavor: Even the most intelligent persons can fail solving realistic and complex problems, if they don't have important content knowledge or don't know adequate search strategies as well as when to apply them in an adaptive way (cf. Dörner, 1996; Fischer, Greiff, & Funke, under review). This paper is about some of the most important components of Problem Solving Competency (PSC, cf. Fleischer, Wirth, & Leutner, 2014; Greiff & Fischer, 2013a; Wirth & Klieme, 2003) and their interrelations. Problem Solving Competency can be understood as the ability to figure out a solution method for reaching ones goal if no such method is obvious (cf., Duncker, 1945; Wirth & Klieme, 2003), that is, to represent and solve problems in various domains (cf. Bassok & Novick, 2012; Schoppek & Putz-Osterloh, 2003). In international large-scale assessments two different kinds of problems have been proposed for assessing PSC (OECD, 2014):

1) One kind of problem requires a single choice of a solution based on the information given at the outset. A characteristic example for this kind of problem is the problem of finding the shortest path between a set of locations based on a map before actually starting to travel. Problems of this kind can be solved analytically, as all the information required for finding a solution is given at the outset of the problem. We will refer to this kind of problem solving as *Analytic Problem Solving* (*APS*).
2) The other kind of problem requires a series of multiple choices, where later choices can be influenced by the results of previous choices (also known as Dynamic Decision Making, e.g., Gonzalez, Lerch, & Lebiere, 2003). For instance, after starting a travel, the initial plan of which locations to see may be adapted dynamically to unforeseen changes in the situation (e.g., road works on certain paths). In this kind of problem, the problem solver can adapt his or her initial plans and knowledge at multiple points in time, because there is feedback after each interaction with the problem. We will refer to this kind of problem solving as *Interactive Problem Solving* (*IPS*).

* Corresponding author at: Hauptstr, 47-50, 69117 Heidelberg, Germany. Tel.: +49 6221 547301; fax: +49 6221 547273.
E-mail addresses: andreas.fischer@psychologie.uni-heidelberg.de (A. Fischer), samuel.greiff@uni.lu (S. Greiff), sascha.wuestenberg@uni.lu (S. Wüstenberg), jens.fleischer@uni-due.de (J. Fleischer), florian.buchwald@uni-due.de (F. Buchwald), joachim.funke@psychologie.uni-heidelberg.de (J. Funke).

http://dx.doi.org/10.1016/j.lindif.2015.02.008

Both kinds of problems[1] have been proposed to measure PSC, but up to now it has never been tested conclusively, if performance in both measures (APS and IPS) indicates distinct facets of PSC, or if they can be considered to address a common core of PSC (e.g., strategies for analyzing complex problem statements, or for systematically structuring prior knowledge and complex information in a goal-oriented way) sufficiently distinct from logical reasoning (Raven, 2000). In the *Programme for International Student Assessment* (PISA) 2012, both kinds of problems have been used to assess a single underlying PSC factor (OECD, 2014). The studies of Wirth and Klieme (2003) and Scherer and Tiemann (2014) presented first evidence for a multidimensional structure of PSC but they did neither control for reasoning nor analyze external validity of the facets reported.

In the current paper we will clarify the conceptual interrelations of reasoning and PSC and we will present empirical evidence based on two samples (577 high-school students and 339 university students) to demonstrate that APS and IPS address a common core of PSC that cannot be explained by reasoning, and that APS and IPS additionally address unique aspects each, which are important for explaining external criteria beyond reasoning. In the discussion we will focus on findings consistent between samples.

1.1. (Why) PSC is conceptually different from reasoning

It seems obvious that basic logical reasoning (e.g., forming inductive or deductive conclusions based on facts or premises, cf. Carpenter, Just, & Shell, 1990; Mayer, 2011), is closely related to problem solving (Mayer, 2011) and necessarily involved in each valid approach to assess PSC (cf. Greiff & Fischer, 2013a; Wüstenberg et al., 2012). However, in addition to this kind of reasoning PSC also implies a large amount of crystallized[2] abilities (Postlethwaite, 2011), that is, *"the knowledge and language of the dominant culture"* (Horn & Masunaga, 2006, p. 590). More specifically, solving problems in a competent way involves "experimental interactions with the environment" (Raven, 2000, p. 54) and depends on a large base of procedural and declarative knowledge on *how* and *when* to perform different search strategies in order to adequately represent and solve problems (e.g., Dörner, 1996). The importance of crystallized knowledge, especially knowledge about strategies, for PSC has often been emphasized (e.g., Scherer & Tiemann, 2014; Schoppek & Putz-Osterloh, 2003; Strohschneider & Guss, 1999; Tricot & Sweller, 2014) and is a central conceptual difference to basic logical reasoning.[3]

If this claim is correct, each valid operationalization of PSC should prove to be *incrementally* valid, compared to tests of reasoning with regard to external criteria such as academic or occupational success. To our knowledge, it is an open question if common variance between current instances of Analytic and Interactive Problem Solving (e.g., Scherer & Tiemann, 2014) can be attributed to reasoning only.

The present study aims to clarify if both APS and IPS are valid approaches to assessing PSC, that is, if they address "more than reasoning" (Wüstenberg et al., 2012) with regard to explaining (1) each other or (2) school grades (as external criteria of PSC).

1.2. Concept and empirical results concerning Analytic Problem Solving

For a long time, PSC has been assessed by APS tasks, that is, by confronting participants with multiple heterogenous problems each requiring a single solution to be generated analytically (e.g., Boggiano, Flink, Shields, Seelbach, & Barrett, 1993; Fleischer, Buchwald, Wirth, Rumann, & Leutner, under review; Fleischer, Wirth, Rumann, & Leutner, 2010; OECD, 2003). For instance, in PISA 2003 PSC was assessed by a set of multiple problems (OECD, 2003) that required (1) decision making under constraints, (2) evaluating and designing systems for a particular situation, or (3) trouble-shooting a malfunctioning device or system based on a set of symptoms (OECD, 2004, p. 61). All problems were designed to be realistic and refer to "cross-disciplinary situations where the solution path is not immediately obvious and where the literacy domains or curricular areas that might be applicable are not within a single domain of mathematics, science or reading" (OECD, 2003, p. 156; see also Leutner, Funke, Klieme, & Wirth, 2005a,b; Leutner, Wirth, Klieme, & Funke, 2005b).

Empirically, APS is highly correlated to performance in different domains like mathematics ($r = .89$), reading ($r = .82$) and science ($r = .80$) on a latent level (OECD, 2004, p. 55). Due to its broad operationalization APS is also closely related to – but yet empirically distinct from – reasoning ($r = .72$; Leutner, Klieme, Meyer, & Wirth, 2004; $r = .67$, Leutner, Fleischer, & Wirth, 2006; $r = .60$ Scherer & Tiemann, 2014). In general, APS seems to be more strongly related to intelligence and school achievements than IPS is (cf., Leutner et al., 2005a,b; Leutner, Fleischer, Wirth, Greiff, & Funke, 2012; Wirth & Klieme, 2003). To our knowledge there is no study explicitly examining the incremental value of APS over and above measures of reasoning and IPS.

1.3. Concept and empirical results concerning Interactive Problem Solving

IPS tasks are a more recent and computer-based approach to assessing PSC that evolved from research on Complex Problem Solving and Dynamic Decision Making (cf. Fischer, Greiff, & Funke, 2012). The defining feature of IPS is that the problem solver can not only rely on the information given at the outset, but must adapt his or her hypotheses (about how the problem works) and plans (about how to reach one's goals) while interacting with the problem (cf. Fischer et al., 2012; Klahr, 2000). Thus, the IPS approach focuses on effective strategies for searching the spaces of information and hypotheses as well as the resulting problem space (Greiff et al., 2013b). Fig. 1 illustrates an example of a typical interactive problem: This problem is an interactive computer-simulation based on a complex[4] abstract linear equation model (cf. MicroDYN approach, Greiff, 2012; Greiff, Fischer, Stadler, & Wüstenberg, in press). It is about a handball-team, that can be trained by applying different amounts of three different trainings (labeled A, B, & C), with each training possibly influencing motivation, power of throw and exhaustion of the team. The problem has to be solved in two subsequent phases: In a first phase, the problem solver can vary the values of certain input variables (in this case representing the amounts of three trainings, shown on the left side of the screen in Fig. 1), and observe the values of certain output variables (on the right side of the screen in Fig. 1). In this phase, his or her goal is to find out about the causal structure of the simulation and to draw his or her hypotheses into a causal model at the bottom of the screen (*problem representation*, sometimes referred to as knowledge acquisition, see Fig. 1). In a subsequent phase the problem solver is instructed to reach a set of well-defined goals (see the values in brackets in Fig. 1) by

[1] In the literature on complex problem solving (e.g., Funke, 2003; Scherer & Tiemann, 2014) and dynamic decision making (e.g., Edwards, 1962), sometimes APS and IPS have also been referred to as static vs. dynamic decision problems, or as simple vs. complex problems, respectively.

[2] Traditional measures of "crystallized intelligence" are often tests of highly *general declarative* knowledge. They focus on breadth instead of depth of the individual's knowledge base (i.e., they "measure only the elementary knowledge, the beginning [declarative] knowledge, in the various fields of human culture", Horn & Masunaga, 2006, p. 597). The *concept* of crystallized intelligence represents a broader and more diverse range of knowledge (Horn & Masunaga, 2006) — e.g., procedural knowledge as it is tapped by some tests of expertise or PSC, for example.

[3] As a result of these crystallized aspects, PSC can be assumed to be less domain-general than reasoning as well as more prone to training (cf. Scherer & Tiemann, 2014).

[4] Of course one could also simulate even more complex problems containing aspects like negative feedback (e.g., predator–prey-systems, Cushing, 1977; or the sugar-factory-simulation, Berry & Broadbent, 1984), phase transitions, or deterministic chaos (e.g., Verhulst, 1839) within the framework proposed by Funke (2001) but each of these aspects again is likely to address additional or different skills and strategies. Traditional MicroDYN tests seem to reliably address a small set of skills (cf. Greiff & Fischer, 2013a,b; Funke, 2010), that are central for solving a wide range of analytic and/or complex problems.

Please cite this article as: Fischer, A., et al., Assessing analytic and interactive aspects of problem solving competency, *Learning and Individual Differences* (2015), http://dx.doi.org/10.1016/j.lindif.2015.02.008

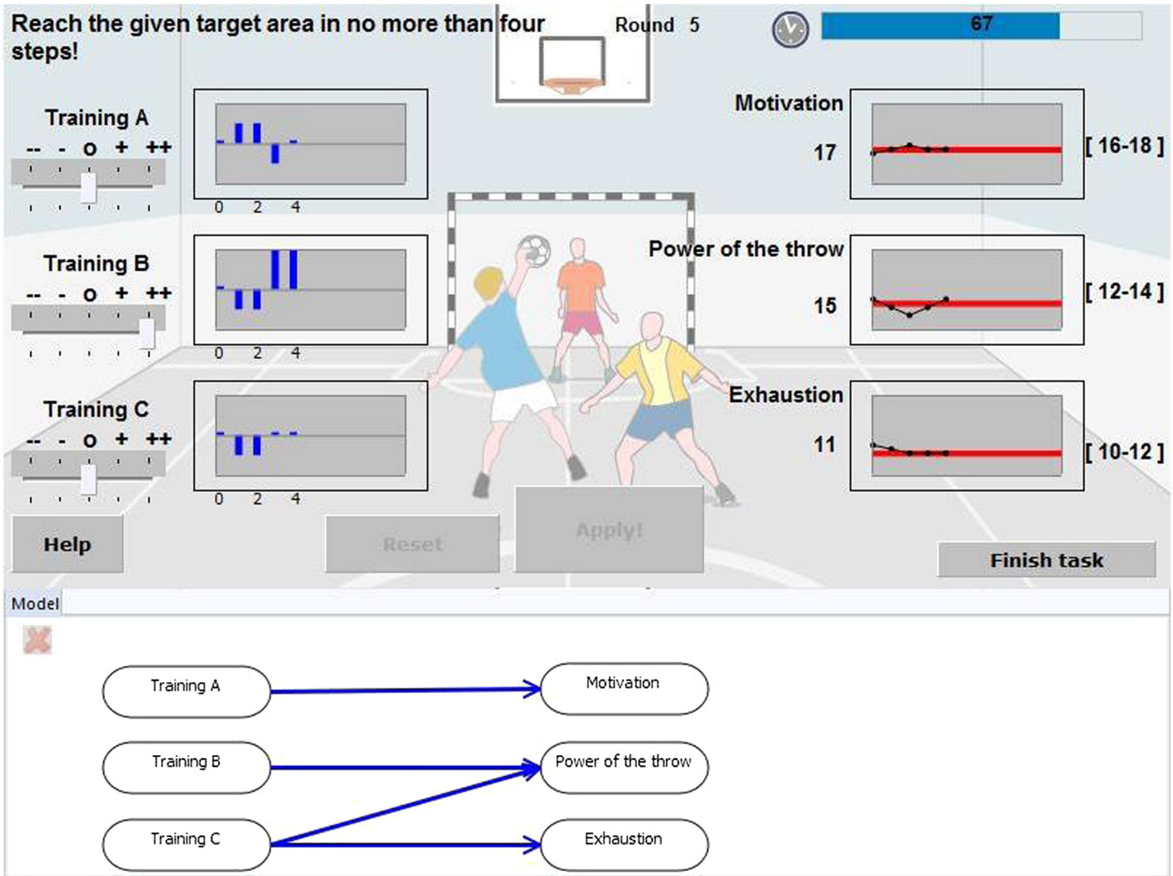

Fig. 1. Screenshot of a MicroDYN simulation. Input-variables (left side of the screen) are labeled with fictional pseudo-words, in order to not trigger any helpful prior knowledge about the problem's causal structure. In a first phase, a problem solver has to interact with the problem and draw the causal structure in a causal diagram (representation at lower part of the screen). Afterwards, in a second phase, certain goal ranges are shown for each output-variable (solution at the right side of the screen).

specifying a series of inputs (*problem solution* sometimes referred to as knowledge application, see Fig. 1).

Recent empirical studies shed light on the aspects of PSC that are assessed within this operationalization of the IPS approach: In the first phase, finding an adequate problem representation seems to primarily depend on applying the control-of-variables strategy, that is, varying one thing at a time ($r = .97$ on a latent level; Wüstenberg et al., 2012), and it seems to indicate a thoughtful application of adequate strategies in the dual-search of hypotheses and information (Greiff et al., 2013b). In the second phase of IPS, finding a solution primarily depends on the application of basal strategies for searching well-defined problem spaces, that is, functional equivalents of means–end analysis (Greiff et al., 2013b; Simon, 1975). The strategies for solution in IPS are highly similar to the ones involved in solving tests of reasoning. Correspondingly, most studies on the incremental validity of the IPS approach demonstrated incremental validity over and above different measures of reasoning *only* for the representation but not for solution in IPS (Greiff & Fischer, 2013a; Greiff et al., 2013b; Wüstenberg et al., 2012).

1.4. Hypotheses

In the current paper, we will test hypotheses regarding two main research questions: (1) do APS and IPS address a common core of Problem Solving Competency that cannot be explained by reasoning? And (2) do APS and IPS additionally address unique aspects of external criteria each?

With regard to the first research question we expect APS to be correlated to both facets of IPS due to a common impact of reasoning (Hypothesis 1a). Furthermore, when regressing APS on both facets of IPS as well as on reasoning, we expect a unique contribution of problem representation in IPS (Hypothesis 1b) due to the additional impact of PSC on both IPS and APS, but we expect no unique contribution of problem solution in IPS because the search for a solution is very closely related to reasoning (Hypothesis 1c).

With regard to the second research question, we will examine the external validity of the IPS and the APS approach: We expect both approaches to be predictive for school grades as external criteria of Problem Solving Competency (Hypothesis 2a). More specifically, when regressing school grades on APS and IPS as well as on reasoning we expect a unique contribution of problem representation in IPS (Hypothesis 2b) because the strategic knowledge indicated by the representation may be important for school grades even beyond APS due to interactive aspects (cf. Wüstenberg et al., 2012). Again, we expect no unique contribution of problem solution in IPS because of its close relation reasoning (Hypothesis 2c). Most of all we expect a unique contribution of APS (Hypothesis 2d), because APS is known to be more closely related to school grades in different domains (cf. Scherer & Tiemann, 2014).

2. Method

2.1. Participants

We assessed two samples: One sample consists of 577 German undergraduate high-school students (age: $M = 14.94$; $SD = 1.29$; gender: 44.2% male, 46.8% female, 9% missing), mainly in the 9th grade (45.1%) but also from 8th grade (26.7%), 10th grade (14.6%) and 11th grade (7.4%) — with 6% missing an entry. Participants in the high-school student sample received 50 Euro for their class inventory. Two samples were chosen to ensure a sufficient amount of generality concerning

Please cite this article as: Fischer, A., et al., Assessing analytic and interactive aspects of problem solving competency, *Learning and Individual Differences* (2015), http://dx.doi.org/10.1016/j.lindif.2015.02.008

our findings (factorial invariance of IPS holds across grade levels 5 and 12, e.g., Greiff et al., 2013c; Scherer & Tiemann, 2014).

The second sample consisted of 339 German university students (age: $M = 22.30$, $SD = 4.02$; gender: 27.1% male, 67.6% female; 5.3% missing) from different fields of study. Most participants of this sample studied social sciences (57%), some participants studied natural sciences (28%) or had a different field of study (15%). Participants in this sample received either 25 Euro for participation or 4 h course credit.

The relations of APS, IPS, reasoning and GPA are original to this paper.[5]

2.2. Materials

2.2.1. Interactive Problem Solving (representation and solution)

The IPS approach to measuring Problem Solving Competency was operationalized by the MicroDYN test of Problem Solving Competency, based on a set of 10 tasks in each sample. According to the detailed description in Section 1.3 of this paper each task consisted of searching for *representation* and *solution* as two subsequent interactive subtasks (for a detailed description of MicroDYN refer to Wüstenberg et al., 2012; Greiff & Fischer, 2013a,b).

For each task, both the correctness of the causal diagram after phase 1 (representation) and the value of output-variables after phase 2 (solution) were scored dichotomously as either (1) *correct* or (0) *incorrect* (cf. Wüstenberg et al., 2012) to indicate strategic knowledge for finding adequate representations and solutions to minimal-complex interactive problems.

2.2.2. Analytic Problem Solving

The APS approach to measuring Problem Solving Competency was based on a set of realistic static problems that were applied in PISA 2003. In the university student sample we applied 4 items (Transit System Q1, Holiday Q2, Course Design Q1 and Freezer Q2) whereas in the high-school student sample we applied 6 items (Cinema Outing Q1, Cinema Outing Q2, Irrigation System Q3, Holiday Q2, Transit System Q1, Childrens' Camp Q1). A detailed description of items can be found in OECD (2004, pp. 59 ff.). The approach is elaborated in more detail in Section 1.2.

For some problems, answers were scored dichotomously as *incorrect* (0) or *correct* (1), for some problems answers were scored polytomously as completely *incorrect* (0), *partially correct* (1), or *correct* (2).

2.2.3. Reasoning (subtest of I-S-T 2000 R; KFT 4-12 + R)

In the university student sample, reasoning was assessed using the matrix subtest of the "Intelligence Structure Test-Revised" (I-S-T 2000 R; Liepmann, 2007). This test consisted of 20 2×2-matrices, each containing a figural stimulus in each but one cell. In each matrix, one stimulus was missing, and participants had to choose the missing figural stimulus out of five alternatives. Answers were scored as right or wrong. Missing values were considered wrong answers.

In the high-school student sample reasoning was assessed by the subscale "figural reasoning" of the KFT 4-12 + R (Heller & Perleth, 2000). This test consists of 23 items requiring to identify the relation of a pair of two figures and to choose one out of five alternatives in order to complete a second pair of figures with the same relation. Answers were scored as right or wrong. Missing values were considered wrong answers.

2.2.4. School grades

Subjects were asked to report their final Grade Point Average (GPA; university student sample) or their last grades in each course (high-school student sample). In the high-school student sample we used a latent factor model to estimate the current GPA. As grades in the German school system range from 1 (*very good*) to 6 (*insufficient*), we reversed GPA, so that higher numerical values indicate better performance.

2.3. Procedure

The university students were tested in two sessions (each lasting about 120 min). In the first session participants worked on the IPS test (about 60 min) and the APS test (about 15 min). In the second session they worked on a set of tests including the reasoning task (about 10 min).

The high-school students were tested in two sessions (45 min each). In the first session participants worked on the IPS test and in the second session on the APS test (about 35 min) and the reasoning tasks (about 10 min).

3. Results

All latent analyses were obtained by using MPlus 5.21 (Muthén & Muthén, 2008), commonality analysis was run in Gnu R by using the yhat package (Rya-Mukherjee et al., 2014). WLSMV-estimators were chosen for the structural equation models with ordinal items (cf. Muthén & Muthén, 2007).

3.1. Measurement models

This section specifies measurement models for each latent construct. With regard to fit indices Hu and Bentler (1999) recommend to use models with Comparative Fit Index (CFI) and a Tucker Lewis Index (TLI) value above .95 and a Root Mean Square Error of Approximation (RMSEA) below .06.

IPS was modeled as a construct with the two correlated factors representation and solution, which fitted the data well (CFI ≥ .955, TLI ≥ .970; RMSEA ≤ .053). APS was modeled as a one-dimensional construct (CFI ≥ .949; TLI ≥ .923; RMSEA ≤ .054) in accordance with the modeling procedure in PISA 2003 (OECD, 2005). Cronbach's Alpha for APS was comparatively low (see Tables 1 and 2) which is consistent with PISA 2003 (OECD, 2005; Fleischer et al., 2014) and with prior studies on APS (Fleischer et al., 2014). Reasoning was also modeled as a one-dimensional construct (CFI ≥ .96; TLI ≥ .988; RMSEA ≤ .046) according to the test manual. In the university student sample each item of the reasoning test was assigned to one of four parcels as described by Greiff et al. (2013b) (CFI = 1.00; TLI = 1.00; RMSEA < .001). In the high-school student sample Grade Point Average was modeled – with slightly suboptimal fit – as a one-dimensional construct indicated by current grades in German, English, Math, Physics, Chemistry, and Biology (CFI = .946; TLI = .953; RMSEA = .166).

Table 1
Latent correlations of both aspects of IPS (representation and solution), APS, reasoning and grade point average (GPA) in the high-school student sample.

	Representation	Solution	APS	Reasoning	GPA
Representation	α = .807				
Solution	.777***	α = .766			
APS	.747***	.729***	α = .543		
Reasoning	.520***	.469***	.562***	α = .914	
GPA	.341***	.226***	.379***	.264***	α = .811

Note. α: Cronbach's Alpha; $n = 577$.
*** $p < .001$.

[5] Other data of the university student sample was published in Greiff et al., (2013b), who studied complex problem solving as a latent construct, with the MicroDYN test being one of multiple indicators of complex problem solving. Part of the MicroDYN- and reasoning-data of a subsample of the high-school student data was published in Frischkorn, Greiff and Wüstenberg (2014) who studied the development of complex problem solving.

Please cite this article as: Fischer, A., et al., Assessing analytic and interactive aspects of problem solving competency, *Learning and Individual Differences* (2015), http://dx.doi.org/10.1016/j.lindif.2015.02.008

Table 2
Latent correlations of both aspects of IPS (representation and solution), APS, reasoning and grade point average (GPA) in university student sample.

	Representation	Solution	APS	Reasoning	GPA
Representation	α = .792				
Solution	.846***	α = .765			
APS	.768***	.763***	α = .462		
Reasoning	.214***	.273***	.395***	α = .820	
GPA	.209***	.206***	.395***	.118*	–

Note. α: Cronbach's Alpha; $n = 339$.
* $p < .05$.
*** $p < .001$.

Table 3
Unique and common commonality coefficients ΔR^2 and the corresponding percent of explained variance (%total) for each predictor in the regression of APS on representation, solution, and reasoning based on the high-school student sample.

Variables	ΔR^2	%total
Unique to representation	0. 051	7. 93
Unique to solution	0. 045	7. 04
Unique to reasoning	0. 031	4. 80
Common to representation, and solution	0. 232	36. 04
Common to representation, and reasoning	0. 031	4. 83
Common to solution, and reasoning	0. 010	1. 60
Common to representation, solution, and reasoning	0. 243	37. 76
Total	0. 645	100. 00

3.2. Structural models

In order to address our first research question (Section 3.2.1) we ran latent regressions of APS on representation, solution, and reasoning for the high-school student sample ($R^2 = .643$; CFI = .966; TLI = .981, RMSEA = .035) and for the university student sample ($R^2 = .672$; CFI = .963; TLI = .975, RMSEA = .038). In order to address our second research question (see Section 3.2.2) we ran latent regressions of GPA on representation, solution and reasoning for the high-school student sample ($R^2 = .172$; CFI = .964; TLI = .979, RMSEA = .035) and the university student sample ($R^2 = .186$; CFI = .963; TLI = .975, RMSEA = .037). Explained variance was highly significant in all models ($p < .01$) and variance-inflation was not indicated (variance-inflation factors below 5 for all predictors, cf. O'Brien, 2007).

Latent correlations between all constructs assessed (see Tables 1 and 2) proved to be positive and substantial for all measures.

Additionally we ran commonality analyses (Nimon, Lewis, Kane, & Haynes, 2008; Rya-Mukherjee et al., 2014) for each of these regression models in order to decompose the explained variance of the criterion into unique and shared contributions for each predictor (Rya-Mukherjee et al., 2014).

3.2.1. Research question 1: does APS share variance with the facets of IPS over and above reasoning?

Our first research question was whether the IPS facets representation and solution were predictive for APS performance over and above reasoning. The regression of APS on representation, solution, and reasoning (see Fig. 2 for the high-school sample) indicated significant unique contributions of representation (ß = .372–.438; $p < .05$) and reasoning (ß = .209; $p < .01$) in both samples. The unique contribution of solution was significant in the high-school student sample (ß = .341; $p < .01$) but not in the university student sample (ß = .334; $p = .10$).

The corresponding commonality analyses (see Tables 3 and 4) revealed that a large amount of the explained variance in APS could be attributed to variance that is common to representation and solution (and *not common to* reasoning) in both samples ($R^2 = .23$–$.43$). *Additionally*, a substantial amount of explained variance in APS could be attributed to variance common to representation, solution and reasoning in both samples ($R^2 = .10$–$.24$).

In summary, we found substantial commonalities between IPS, APS and reasoning (supporting Hypothesis 1a) as well as a unique contribution of IPS representation to explaining APS, beyond IPS solution and reasoning (supporting Hypothesis 1b). We did not find consistent evidence for a unique contribution of IPS solution (partially supporting Hypothesis 1c).

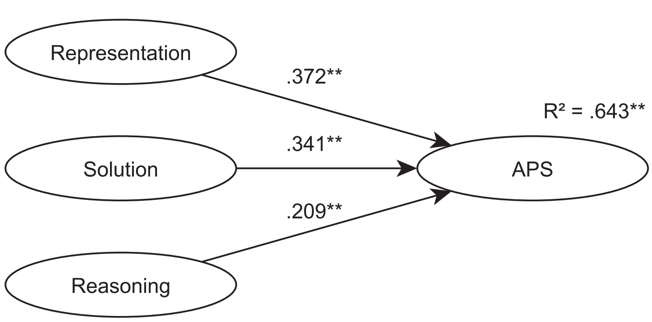

Fig. 2. Unique contributions of IPS representation, IPS solution and reasoning in predicting APS (model b) in the high-school student sample. For correlations between predictors see Table 2. **: $p < .01$.

3.2.2. Research question 2: can GPA be explained by unique aspects of APS or the facets of IPS over and above reasoning?

Our second research question was whether there were unique contributions of APS and the facets of IPS over and above reasoning to predicting school grades as an external criterion of Problem Solving Competency. The regression of GPA on representation, solution, APS, and reasoning (see Fig. 3 for the high-school sample) indicated a significant unique contribution of APS (ß = .336–.644; $p < .05$) and no significant unique contribution of reasoning (ß = −.067–.060; p = .392–.352) in both samples. The unique contributions of representation and solution were significant in the high-school student sample (ß = .241 to −.234; $p < .05$) but not in the university student sample (ß = −.159 to −.133; p = .43 to 51).

The corresponding commonality analyses (see Tables 5 and 6) underlined the importance of APS as they revealed a substantial amount of the explained variance in GPA ($R^2_{Total} = .17$–$.19$) could be attributed to variance that is unique to APS in both samples ($R^2_{APS} = .04$–$.14$). Please note, negative coefficients in Table 5 are not problematic for the analysis: Given the positive correlations among all predictors they indicate statistical suppression effects related to solution in the high-school sample (Rya-Mukherjee et al., 2014).

In summary, we found substantial correlations between GPA and both IPS and APS (supporting Hypothesis 2a) as well as a unique contribution of APS to explaining GPA, beyond IPS and reasoning (supporting Hypothesis 2d). In this regard, we did not find consistent evidence for a unique contribution of IPS representation or IPS solution (partially supporting Hypotheses 2b and 2c).

Table 4
Unique and common commonality coefficients ΔR^2 and the corresponding percent of explained variance (%total) for each predictor in the regressions of APS on representation, solution, and reasoning based on the university student sample.

Variables	ΔR^2	%total
Unique to representation	0. 056	8. 26
Unique to solution	0. 030	4. 45
Unique to reasoning	0. 041	6. 03
Common to representation, and solution	0. 434	64. 19
Common to representation, and reasoning	−0. 003	−0. 45
Common to solution, and reasoning	0. 015	2. 23
Common to representation, solution, and reasoning	0. 103	15. 29
Total	0. 676	100. 00

Please cite this article as: Fischer, A., et al., Assessing analytic and interactive aspects of problem solving competency, *Learning and Individual Differences* (2015), http://dx.doi.org/10.1016/j.lindif.2015.02.008

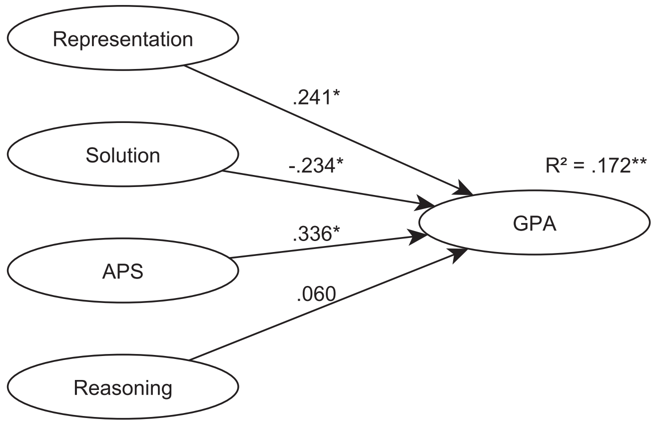

Fig. 3. Unique contributions of IPS representation, IPS solution, APS and reasoning to predicting GPA (model c) in the high-school student sample. For correlations between predictors see Table 1. *: $p < .05$; **: $p < .01$.

Table 6
Unique and common commonality coefficients ΔR^2 and the corresponding percent of explained variance (%total) for each predictor in the regressions of GPA on representation, solution, APS, and reasoning based on the university student sample.

Variables	ΔR^2	%total
Unique to representation	0. 006	3. 33
Unique to solution	0. 005	2. 40
Unique to APS	0. 135	72. 59
Unique to reasoning	0. 004	1. 95
Common to representation, and solution	0. 018	9. 63
Common to representation, and APS	−0. 002	−0. 90
Common to solution, and APS	−0. 003	−1. 44
Common to representation, and reasoning	−0. 001	−0. 69
Common to solution, and reasoning	0. 001	0. 29
Common to APS, and reasoning	0. 001	0. 42
Common to representation, solution, and APS	0. 013	6. 93
Common to representation, solution, and reasoning	−0. 001	−0. 63
Common to representation, APS, and reasoning	0. 001	0. 53
Common to solution, APS, and reasoning	0. 001	0. 37
Common to representation, solution, APS, and reasoning	0. 010	5. 23
Total	0. 186	100. 00

4. Discussion

In the current paper we have outlined and contrasted two different approaches to assessing aspects of Problem Solving Competency: One approach is based on static problems that have to be solved analytically (i.e., APS), whereas the other approach is about dynamic decisions in problem situations which have to be represented and solved interactively (i.e., IPS).

4.1. Research question 1: does APS share variance with the facets of IPS over and above reasoning?

Our first question was whether performance in APS shares variance with the crystallized strategic knowledge assessed by IPS. Indeed, we found APS highly correlated to both facets of IPS, and commonality analysis revealed that – besides a medium to large amount of variance that APS shares with IPS and reasoning (supporting Hypothesis 1a) – a large amount of variance in APS can be explained by the facets of IPS (over and above reasoning). Especially representation in IPS consistently had a significant unique contribution to explaining APS (supporting Hypothesis 1b), which demonstrates the relevance of crystallized strategic knowledge on generating and testing hypotheses (Greiff & Fischer, 2013a,b).

The unique contribution of solution in IPS was significant in the high-school student sample but not in the university-student sample (partially supporting Hypothesis 1c). In this regard we didn't expect to find a unique contribution because of the close conceptual relation between solution and reasoning. However according to the theoretical framework of Fischer et al. (2012), in many cases of IPS it may be more effective to rely on implicit knowledge about inputs that work (instance-based knowledge) than to actually reason about structural knowledge (i.e., reasoning and representation). This kind of knowledge is not addressed in the causal model that indicates representation in IPS (Greiff & Fischer, 2013b). Future studies should elaborate on these incremental aspects of solution in IPS (cf. Fischer et al., 2012) in more detail in order to clarify what they depend upon.

Table 5
Unique and common commonality coefficients ΔR^2 and the corresponding percent of explained variance (%total) for each predictor in the regressions of GPA on representation, solution, APS, and reasoning based on the high-school student sample.

Variables	ΔR^2	%total
Unique to representation	0. 019	10. 73
Unique to solution	0. 019	11. 02
Unique to APS	0. 040	23. 34
Unique to reasoning	0. 002	1. 36
Common to representation, and solution	−0. 013	−7. 27
Common to representation, and APS	0. 031	17. 74
Common to solution, and APS	−0. 013	−7. 79
Common to representation, and reasoning	0. 003	1. 46
Common to solution, and reasoning	−0. 000	−0. 10
Common to APS, and reasoning	0. 010	5. 62
Common to representation, solution, and APS	0. 020	11. 80
Common to representation, solution, and reasoning	−0. 001	−0. 51
Common to representation, APS, and reasoning	0. 017	10. 12
Common to solution, APS, and reasoning	−0. 002	−0. 91
Common to representation, solution, APS, and reasoning	0. 040	23. 39
Total	0. 172	100. 00

4.2. Research question 2: can GPA be explained by unique aspects of APS or the facets of IPS over and above reasoning?

Our second research question was whether IPS and APS can explain school grades and if they have unique contributions compared to each other. Indeed, we found substantial correlations between Grade Point Average and both APS and the facets of IPS (supporting Hypothesis 2a). More importantly, APS had a unique contribution to explaining school grades beyond reasoning and both aspects of IPS in both samples (supporting Hypothesis 2d). This was expected because APS was known to be highly predictive for school achievements in different domains (OECD, 2004; Scherer & Tiemann, 2014), but it has never been proven empirically up to now.

With regard to representation and solution in IPS we found unique contributions in the high-school student sample but not in the university student sample (partially supporting Hypotheses 2b and 2c). With regard to representation in IPS we expected a unique contribution whereas for solution in IPS we did not expect a unique contribution because of its close conceptual relation to deductive reasoning (however, there are theoretical differences, see above). Again, future research is needed to determine the conditions for an incremental value of IPS with regard to the regression of GPA.

4.3. Shortcomings

First, comparisons between samples have to be drawn with caution, as the tests we used differed between samples.[6] Of course they were highly similar in nature and can be assumed to address the same constructs. Nevertheless, we see a potential shortcoming here. Second, the operationalization of school grades may have had different meanings for high-school students (current grades) compared to university students (final Grade Point Average). Future studies should additionally

[6] This was done because of the different levels of competence between the two samples.

investigate concurrent measures of success for university students, to replicate our findings and to further validate measures of IPS and APS (especially as skills and competence may change over time, see Molnár, Greiff, & Csapó, 2013). On the other hand, both shortcomings also highlight the robustness of our findings: Even in spite of these differences, most of our findings were consistent between samples highlighting the generalizability of our conclusions.

Please note, up to now the incremental value of assessing IPS strategies compared to reasoning has been demonstrated using a variety of reasoning tests and school grades (Greiff & Fischer, 2013a; Greiff et al., 2013b,c; Wüstenberg et al., 2012) and the results seem to be independent of different operationalizations as they can be attributed to a latent underlying construct (Greiff et al., 2013b). One may argue, that representation and solution in IPS do not cover the whole range of interactive strategies, but – even if we totally agree with this point – the high proportion of variance in APS explained by representation and solution beyond reasoning ($R^2 = .33$ to .52) indicates the centrality of the strategies assessed for cross-curricular Problem Solving Competency (which is basically the understanding of problem solving in PISA 2003 and PISA 2012).

4.4. Summary and outlook

Our findings clearly indicate close relations between Interactive and APS (beyond reasoning), as well as the incremental value of APS for predicting school grades (beyond IPS and reasoning) in both samples. For the first time, these findings show that APS – just like IPS – requires more than reasoning and thus is a promising measure of Problem Solving Competency for both High-School students and University Students. We did not find consistent evidence for a unique contribution of IPS (beyond APS and reasoning). More specifically, we found an incremental value of representation in IPS in the high-school sample, but not in the university student sample. This finding empirically supports the idea of complementing measures of APS with measures of IPS in school-related assessments, as it was done in PISA 2012 for instance.

In both samples, we found IPS to be closely related to APS beyond reasoning. Thus, both APS and IPS seem to address a common core of Problem Solving Competency (beyond reasoning) and APS seems to be more closely related to school grades than IPS. Our findings indicate that the analytic aspects of Problem Solving Competency assessed by IPS tasks (Scherer & Tiemann, 2014) may account for the incremental value of IPS beyond reasoning reported in previous studies on university student samples (e.g., Wüstenberg et al., 2012).

Implications for educational practice are manifold: In assessment contexts APS may be used to complement traditional assessment instruments (especially when GPA is an intended criterion) and in training contexts PSC maybe fostered by training and teaching strategic knowledge concerning how to acquire and how to apply knowledge or how to analyze evidence in IPS and APS (Scherer & Tiemann, 2014) – and much more easily so than reasoning. Realistic complex problems – from managing a Tailorshop (e.g., Danner et al., 2011) to solving in-basket tasks (e.g., Fischer & Funke, 2013) – shed light on the complexity of Problem Solving Competency itself (cf. Fischer et al., 2012). The tasks presented in the current study may be a first step to assessing certain aspects of PSC reliably, but the full potential of computer-based assessment still waits to be fully exploited.

Acknowledgments

This research was funded by grants from the German Research Foundation (DFG Az. Fu 17314-2) and from the German Federal Ministry of Education and Research (BMBF FKZ 01JG1062). We are grateful to the TBA group at DIPF (http://tba.dipf.de) for providing the MicroDYN authoring tool CBA Item Builder and technical support.

References

Bassok, M., & Novick, L.R. (2012). Problem solving. In K.J. Holyoak, & R.G. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 413–432). New York, NY: Oxford University Press.

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

Boggiano, A.K., Flink, C., Shields, A., Seelbach, A., & Barrett, M. (1993). Use of techniques promoting students' self-determination: Effects on students' analytic problem-solving skills. *Motivation and Emotion*, *17*, 319–336.

Carpenter, P.A., Just, M.A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431.

Cushing, J.M. (1977). Periodic time-dependent predator–prey systems. *SIAM Journal on Applied Mathematics*, *32*, 82–95.

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*, 225–233.

Dörner, D. (1996). *The logic of failure: Recognizing and avoiding error in complex situations.* New York, NY: Perseus Books.

Duncker, K. (1945). On problem-solving. *Psychological Monographs*, *58*(5).

Edwards, W. (1962). Dynamic decision theory and probabilistic information processings. *Human Factors*, *4*(2), 59–74.

Fischer, A., & Funke, J. (2013, August). Problem solving — More than the sum of its parts? *Paper presented at the 15th Biennial Conference of the European Association for Research on Learning and Instruction (EARLI), Munich, Germany*.

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

Fischer, A., Greiff, S., & Funke, J. (2015w). The history of complex problem solving. In B. Csapó, J. Funke, & A. Schleicher (Eds.), *The nature of problem solving*. Paris: OECD (under review).

Fleischer, J., Buchwald, F., Wirth, J., Rumann, S., & Leutner, D. (2015w). Analytical problem solving: Potentials and manifestations. In B. Csapó, J. Funke, & A. Schleicher (Eds.), *The nature of problem solving*. Paris: OECD (under review).

Fleischer, J., Wirth, J., & Leutner, D. (2014). Effekte der kontextuellen Einkleidung von Testaufgaben auf die Schülerleistungen im analytischen Problemlösen und in der Mathematik [Effects of test context on students' achievements in analytical problem solving and mathematics]. *Zeitschrift für Pädagogische Psychologie*, *28*, 217–227.

Fleischer, J., Wirth, J., Rumann, S., & Leutner, D. (2010). Strukturen fächerübergreifender und fachlicher Problemlösekompetenz. Analyse von Aufgabenprofilen. Projekt Problemlösen [Structures of cross-curricular and subject-related problem solving competence — Analysis of task profiles]. *Zeitschrift für Pädagogik*, *56*, 239–248 (Beiheft).

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

Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.

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

Frischkorn, G. T., Greiff, S., & Wüstenberg, S. (2014). The development of complex problem solving in adolescence: A latent growth curve analysis. *Journal of Educational Psychology*, *106*(4), 1007–1020.

Gonzalez, C., Lerch, J.F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*, 591–635.

Greiff, S. (2012). *Individualdiagnostik der komplexen Problemlösefähigkeit [Individual diagnostics of complex problem solving ability]*. Münster: Waxmann.

Greiff, S., & Fischer, A. (2013a). 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*, 27–39.

Greiff, S., & Fischer, A. (2013b). Measuring complex problem solving: An educational application of psychological theories. *Journal for Educational Research Online*, *5*(1), 38–58.

Greiff, S., Fischer, A., Stadler, M., & Wüstenberg, S. (2015s). Assessing complex problem-solving skills with multiple complex systems. *Thinking and Reasoning*. http://dx.doi.org/10.1080/13546783.2014.989263 (in press).

Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M., & Martin, R. (2013c). A multitrait–multimethod study of assessment instruments for complex problem solving. *Intelligence*, *41*, 579–596.

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

Heller, K.A., & Perleth, C. (2000). *KFT 4-12 + R: Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision [Cognitive ability test for grades 4 through 12, revised]*. Göttingen: Beltz Test.

Horn, J., & Masunaga (2006). A merging theory of expertise and intelligence. In K.A. Ericsson, N. Charness, P.J. Feltovich, & R.R. Hoffman (Eds.), *Handbook of expertise and expert performance* (pp. 587–612). Cambridge: Cambridge University Press.

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Klahr, D. (2000). Scientific discovery as problem solving. In D. Klahr (Ed.), *Exploring science. The cognition of development of discovery processes* (pp. 21–40). Cambridge: MIT Press.

Leutner, D., Fleischer, J., & Wirth, J. (2006). Problemlösekompetenz als Prädiktor für zukünftige Kompetenz in Mathematik und in den Naturwissenschaften [Problem solving competence as a predictor of future competencies in mathematics and science]. In M. Prenzel (Eds.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (pp. 119–137). Münster: Waxmann.

Leutner, D., Fleischer, J., Wirth, J., Greiff, S., & Funke, J. (2012). Analytische und dynamische Problemlösekompetenz im Lichte internationaler Schulleistungsvergleichsstudien: Untersuchungen zur Dimensionalität [Analytical and dynamic problem-solving competency from the perspective of studies on academic achievement in international comparison: Analysis of dimensionality]. *Psychologische Rundschau*, *63*, 34–42. http://dx.doi.org/10.1026/0033-3042/a000108.

Leutner, D., Funke, J., Klieme, E., & Wirth, J. (2005). Problemlösefähigkeit als fächerübergreifende Kompetenz [Problem-solving-competency as a cross-curricular key competency]. In E. Klieme, D. Leutner, & J. Wirth (Eds.), *Problemlösekompetenz von Schülerinnen und Schülern* (pp. 7–20). Wiesbaden: VS Verlag für Sozialwissenschaften.

Leutner, D., Klieme, E., Meyer, K., & Wirth, J. (2004). Problemlösen. In PISA-Konsortium Deutschland (Ed.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (pp. 147–175). Münster: Waxmann.

Leutner, D., Wirth, J., Klieme, E., & Funke, J. (2005). Ansätze zur Operationalisierung und deren Erprobung im Feldtest zu PISA 2000 [Approaches for operationalization and their evaluation in PISA 2000 field study]. In E. Klieme, D. Leutner, & J. Wirth (Eds.), *Problemlösekompetenz von Schülerinnen und Schülern* (pp. 21–36). Wiesbaden: VS Verlag für Sozialwissenschaften.

Liepmann, D. (2007). *Intelligenz-Struktur-Test 2000 R: IST 2000 R [Intelligence structure test 2000 (revised)].* Göttingen: Hogrefe.

Mayer, R.E. (2011). *Problem solving and reasoning.* Learning and cognition in education, 112–117.

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity*, *9*, 35–45.

Muthén, L.K., & Muthén, B.O. (2007). *MPlus user's guide*. Los Angeles, CA: Muthén & Muthén.

Muthén, B.O., & Muthén, L.K. (2008). *Mplus (Version 5.1)*. Los Angeles, CA: Muthén & Muthén.

Nimon, K., Lewis, M., Kane, R., & Haynes, R.M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, *40*(2), 457–466.

O'Brien, R.M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, *41*(5), 673–690.

Organisation for Economic Cooperation and Development (OECD) (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills.* Paris: OECD.

OECD (2014). *PISA 2012 results: Creative problem solving: Students' skills in tackling real-life problems (Volume V)*. Paris: OECD Publishing.

Organisation for Economic Cooperation and Development (OECD) (2004b). *PISA 2012 Results: Creative problem solving: Students' skills in tackling real-life problems, Vol. V*, Paris: OECD.

Postlethwaite, Bennett Eugene (2011). *"Fluid ability, crystallized ability, and performance across multiple domains: a meta-analysis."*. PhD thesis, University of Iowa http://ir.uiowa.edu/etd/1255.

Raven, J. (2000). Psychometrics, cognitive ability, and occupational performance. *Review of Psychology*, *7*(1–2), 51–74.

Rya-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D.W., Slotow, R., & Hamer, M. (2014). Using commonality analysis in multiple regressions: A tool to decompose regression effects in the face of multicollinearity. *Methods in Ecology and Evolution*, *5*(4), 320–328.

Scherer, R., & Tiemann, R. (2014). Evidence on the effects of task interactivity and grade level on thinking skills involved in complex problem solving. *Thinking Skills and Creativity*, *11*, 48–64.

Schoppek, W., & Putz-Osterloh, W. (2003). Individuelle Unterschiede und die Bearbeitung komplexer Probleme [Individual differences and the solution of complex problems]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *24*(3), 163–173.

Simon, H.A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, *7*, 268–288.

Strohschneider, S., & Guss, D. (1999). The fate of the Moros: A cross-cultural exploration of strategies in complex and dynamic decision making. *International Journal of Psychology*, *34*(4), 235–252.

Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review*, *26*(2), 265–283. http://dx.doi.org/10.1007/s10648-013-9243-1.

Verhulst, P. -F. (1839). Notice sur la loi que la population suit dans son accroissement. In J.G. Garnier, & A. Quételet (Eds.), *Correspondance Mathématique et Physique*, *Vol. 10*. (pp. 113–121). Brussels (Belgium): Société Belge de Librairie.

Wirth, J., & Klieme, E. (2003). Computer-based assessment of problem solving competence. *Assessment in Education: Principles, Policy & Practice*, *10*, 329–345.

Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving: More than Reasoning? *Intelligence*, *40*, 1–14.

Please cite this article as: Fischer, A., et al., Assessing analytic and interactive aspects of problem solving competency, *Learning and Individual Differences* (2015), http://dx.doi.org/10.1016/j.lindif.2015.02.008