Chapter 25
Dynamic Problem Solving: Multiple-Item Testing Based on Minimally Complex Systems

Joachim Funke and Samuel Greiff

Abstract Problem solving and thinking are important issues in contemporary research. With the advent of educational assessment, problem solving has been identified as a cross-curricular competence that plays an important role in educational and in occupational settings. Our research is connected to previous activities in the field of dynamic problem solving. On the basis of Dörner's "Theory of Operative Intelligence", we developed assessment instruments (called MicroDYN and MicroFIN) that allow for psychometrically acceptable measurements in the field of dynamic problem solving. MicroDYN is an approach based on linear structural equation systems and requires from the problem solver the identification of input-output connections in small dynamic systems with varying degrees of complexity. MicroFIN is an approach based on finite state automata and requires from the problem solver the identification of transitions of state in small simulated devices, within a variety of backgrounds. Besides developing of the test instruments, we checked the construct validity in relation to intelligence and working memory in a series of studies with pupils, students, and workers. Also, the internal relations between different facets of the global construct "dynamic problem solving" were analyzed.

Keywords Complex problem solving • Educational assessment • Dynamic decision making • MicroDYN • PISA 2012

J. Funke (✉)
Heidelberg University, Heidelberg, Germany
e-mail: funke@uni-hd.de

S. Greiff
University of Luxembourg, Esch-sur-Alzette, Luxembourg
e-mail: samuel.greiff@uni.lu

D. Leutner et al. (eds.), *Competence Assessment in Education*, Methodology of Educational Measurement and Assessment, DOI 10.1007/978-3-319-50030-0_25

25.1 Introduction

Problem solving research has changed its focus over the last 40 years. After the seminal paper of Dietrich Dörner (1980), which proposed a move from static to dynamic problems, a lot of research has been initiated in that area (for an overview, see: Frensch and Funke 1995; Sternberg and Frensch 1991), delivering new insights into phenomena such as intellectual emergency reaction (Dörner 1997) or the connection between emotion and complex problems (Barth and Funke 2010; Spering et al. 2005).

Our research goals are connected to this "young" tradition: (1) modeling of problem solving competencies based on Dörner's theoretical approach, (2) development of computer-based assessment instruments that allow for the measurement of different levels of proficiency and different facets of problem solving, and (3) empirical tests of the newly developed instruments within a context of educational assessment.

Our research started with questions resulting from basic research in problem solving but as the process developed (due to our collaboration with OECD on the PISA 2012 problem solving assessment), questions related to the applicability of our competence measurement in the context of large-scale assessments also became important.

This chapter presents information on all three issues in an overview format; some parts of this chapter have already been published, with more detailed information, in various publications (e.g., Fischer et al. 2012; Funke 2010, 2012; Greiff and Fischer 2013a, b; Greiff and Funke 2009; Greiff et al. 2013a, b; Greiff and Neubert 2014; Greiff et al. 2012; Wüstenberg et al. 2012; Wüstenberg et al. 2014).

25.2 Modeling of Problem Solving Competencies

In textbooks (e.g., Mayer and Wittrock 2006), problem solving is defined as cognitive processing directed at transforming a given situation into a goal situation when no obvious method of solution is available. This is very similar to the traditional definition of Duncker (1935), in his famous paper on the topic translated by Lynne Lees (Duncker 1945, p. 1): "A problem arises 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". Seventy years later, the definition of the problem situation has not changed substantially. What has changed drastically is the type of problem used in problem solving research: instead of static problem situations we now use dynamic situations that change in response to interventions and to time.

The Transition from Static to Dynamic Problems Dietrich Dörner (1975) was—independently of, but in line with, Donald Broadbent (1977), Andrew MacKinnon and Alex Wearing (1980)—convinced that the psychology of problem solving had to analyze how people deal with dynamics, intransparency, polytely, connectivity, and complexity as defining characteristics of problem situations. This is an issue mostly ignored in previous problem solving research that focused on static problems. But dynamic situations have tremendous consequences for the problem solver: they require the anticipation of future developments and of the short- and long-term consequences of decisions. The intransparency of a problem situation requires active information search and information generation, to gain transparency. The polytelic goal structure requires balancing goals that might compete with each other (antagonistic versus synergistic goals). The connectivity of a given system requires anticipation of even small, unintended side effects of interventions that in the end might adversely influence the intended main effects. The complexity of the problem situation requires reduction of information, so that limited cognitive resources ("bounded rationality" in the sense of Simon 1959) can deal with it.

The transition from static to dynamic problem situations was a turning point in problem solving research. The dynamics and complexities of everyday life problems, as well as those of societal challenges, became subject to theories and to empirical work (Dörner 1997; Frensch and Funke 1995; Sternberg and Frensch 1991; Verweij and Thompson 2006; Zsambok and Klein 1997). "Dynamic decision making" (Brehmer 1989) and "naturalistic decision making" (Klein 1997) were among the labels for the new movement. With his concept of *Operative Intelligence*, Dörner (1986) emphasized the importance of examining not only the 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 *regulate* 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, and so forth), or (3) the availability of *heuristics* (e.g., being able to build helpful subgoals, to constrain the problem space efficiently). It turns out that dynamic problems require these competencies in a greatly different way than static problems, which rely mainly on deduction.

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 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 consist of being able to collect information, to integrate and structure goal-oriented information, 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—only then can it be considered intelligent (Dörner 1986, p. 292).

Table 25.1 The five facets and their relation to the five characteristic features of dynamic problem solving within the processes of representation and solution[a]

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

[a]Modified from Greiff and Fischer (2013b, p. 50)

A central premise of our research approach is its competence orientation (Weinert 2001). According to Klieme and Leutner (2006), competencies are defined as context-specific cognitive dispositions that are needed to successfully cope with certain situations or tasks in specific domains. In our case, we address the competence of dealing with problem situations from different domains that are complex, intransparent at the start of action, and that change their state over time.

The Five Facets of Dynamic Problems The facets of operative intelligence emphasized in this characterization closely resemble the facets of complex dynamic problems (Dörner 1997; Dörner et al. 1983; Funke 1992, 2001) that are most relevant for coping with these characteristic features: (1) the *complexity* of the structure (requiring information reduction), (2) the *intransparency* of the situation (requiring systematically generating information), (3) the *interconnectedness* of the variables (requiring building a model of the most relevant effects), (4) the *polytely* of the task (requiring goal elaboration and for setting priorities), and (5) the *dynamics* of the system (requiring planning and dynamic decision making). Table 25.1 shows these five facets and connects the first three of them to the representation of the problem solving situation (system exploration), whereas the last two are connected to solution approaches (system control).

These characteristic features of dynamic problems and the corresponding facets of *dynamic problem solving* (DPS; see Funke 2001) can be considered a fruitful starting point for measuring operative intelligence, which in turn might be the most important factor determining DPS performance. In the next section we present our ideas for assessing these facets of DPS with the help of computer-based assessment instruments.

25.3 Development of Computer-Based Assessment Instruments

Especially in the assessment of interactive, dynamic problem solving, much progress has been made in recent years. With the help of formalisms such as MicroDYN (problem situations based on linear structural equation systems, LSE approach) and MicroFIN (problem situations based on finite state automata, FSA approach), large-scale assessments such as the Programme for International Student Assessment (PISA; see e.g., OECD 2013) from the Organisation for Economic Cooperation and Development (OECD) have been directed to these competencies that will play an important role in the twenty-first century.

Why are these formalisms so helpful in designing assessment instruments? The answer lies in the fact that on the basis of some elementary building blocks, one can develop arbitrarily complex systems with different semantic embeddings. Figure 25.1 illustrates the modules that were used in our item construction: main effect, multiple effect, multiple dependence, eigendynamic, and side effects, describe different (and arbitrary) relations between an arbitrary number of inputs and an arbitrary number of output variables.

With the help of the building blocks shown in Fig. 25.1, one can design a large universe of MicroDYN systems, starting with a trivial 1x1 system and changing to infinitely complex NxM systems (N, M being the number of input and output variables, respectively) that have to be explored and controlled by our subjects. The building blocks of finite state automata are even simpler: they consist of states and transitions between states. One can build arbitrary complex MicroFIN systems that represent machineries with very different types of behavior (see the examples given by Buchner and Funke 1993). Behind the development of MicroDYN and MicroFIN stands the concept of minimal complexity, which has to be explained first.

The Concept of Minimal Complexity Inspired by ideas from Dörner, but coming from a psychometric perspective, Greiff and Funke (2010) introduced the following idea: rather than increasing problem complexity more and more, to start with *minimally complex systems*: that is, systems that are at the lower end of complexity.

The starting point of this concept is the idea that complex systems are needed in problem-solving research because their features differ markedly from simple static systems (in terms of complexity, connectivity, dynamics, intransparency, and

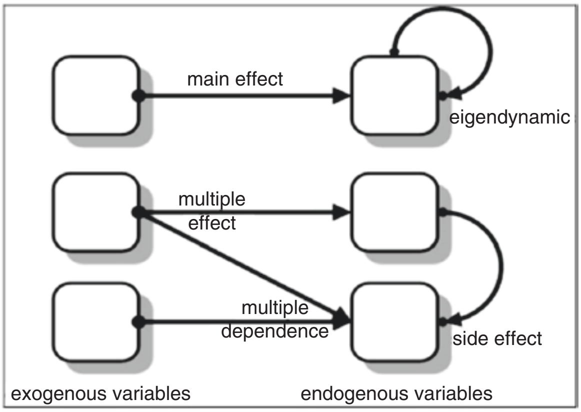

Fig. 25.1 Underlying elementary structure of a MicroDYN item displayed some possible effects between exogenous (input) and endogenous (output) variables (from Greiff and Funke, 2009, p. 159): The modules that were used in our item construction were main effect, multiple effect, multiple dependence, eigendynamic, and side effect

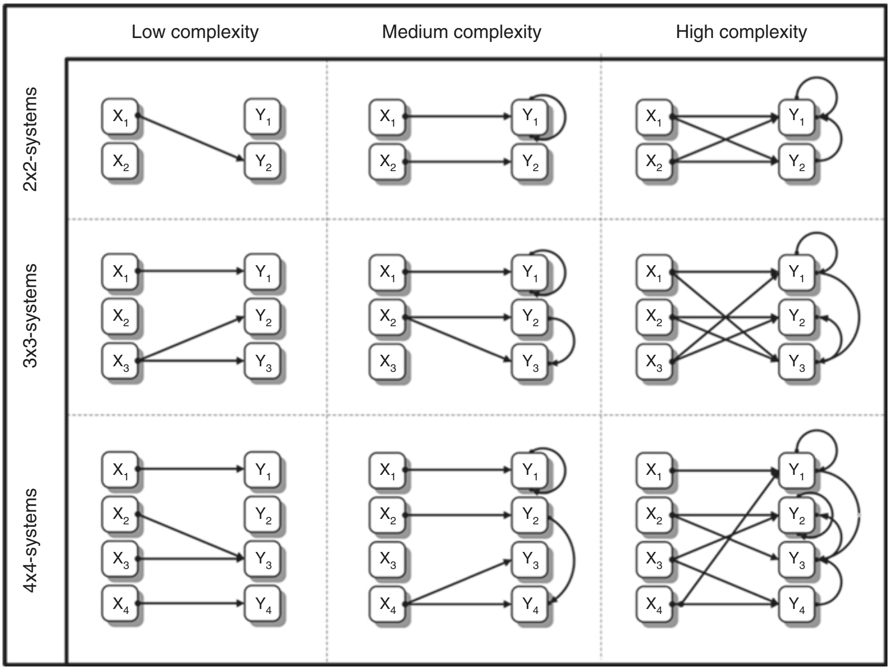

Fig. 25.2 Example of two independent complexity manipulations: (**a**) number of input and output variables (increasing from 2 to 4), (**b**) number of connections (increasing from 1 to 12)

polytely) and their solution requires not simply the addition of simple processes (Funke 2010). The conception of minimally complex systems uses a simple strategy: instead of realizing more and more complex systems (trying to reach for the greatest complexity) with questionable content validity, it instead seeks the minimum complexity. Complexity is a very unclear term—the upper limit of complexity is still open and yet, the lower limit of complexity must be somewhere between nothing and a small degree of complexity. Instead of searching for the upper bounds of complexity, we concentrate on the lower limits and introduce "complexifying elements"—to use a term introduced by MacKinnon and Wearing (1985, p. 170). Figure 25.2 illustrates two types of complexity manipulations for MicroDYN items, as described in Greiff and Funke (2009, p. 160).

This shift in focus to the perspective of minimally complex systems has some advantages for developers of psychometric tests, which can be characterized by the following four points: (1) the time spent on a single scenario is measured not in hours but in minutes, thereby increasing the economies of test application; (2) due to the short time required for item application, a series of items can be presented, rather than one-item testing, thereby increasing reliability; (3) because of our use of formalisms, arbitrary semantic embeddings become feasible, thereby increasing ecological validity; and, (4) a broad range of difficulty levels can be addressed, thereby increasing conceptual validity, as shown by Greiff et al. (2012).

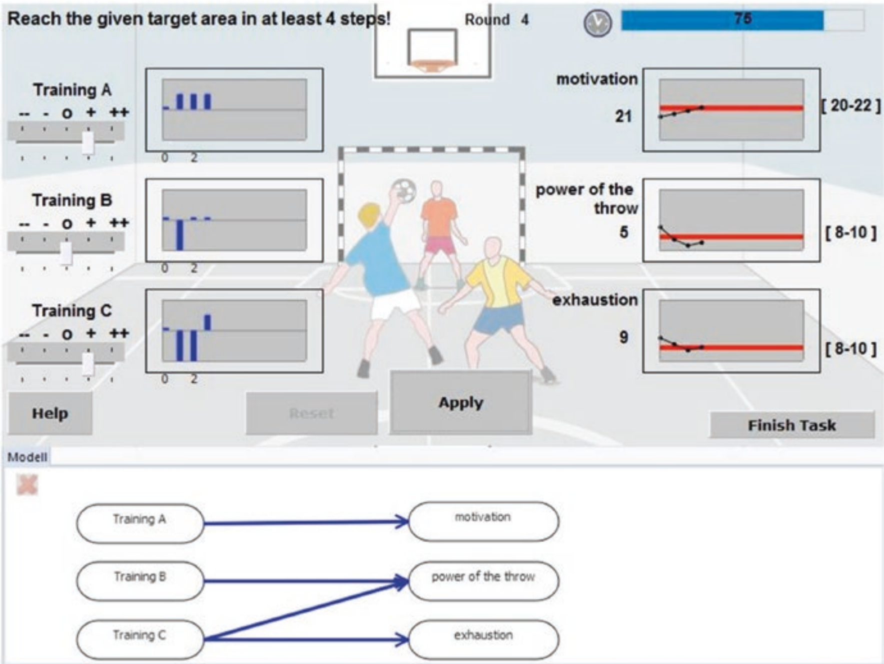

Fig. 25.3 Screenshot of the MicroDYN-item "handball training" (knowledge application phase). The controllers of the input variables (*upper left part*) range from "– –" to "++". The current values and the target values are displayed numerically and graphically (*upper part right*). The correct causal model is presented in the *lower part* (From Wüstenberg et al. 2012, p. 5)

What was the task for the subjects in our experiments? Firstly, a problem solver, who is only shown the values of the input and output variables (but not the underlying structure of the system), had to specify a series of input values in order to identify the system's structure (the problem solver could draw his or her model of the causal structure between the variables in a causal diagram). Secondly, the problem solver had to specify a series of input values in order to reach given target values (see Fig. 25.3 for an example within a MicroDYN task). In this phase ("rule application"), there is a specific goal for controlling the system, whereas in the first part ("rule identification"), there is the unspecific goal of exploring the system and drawing a causal model of the assumed relations ("rule knowledge").

The procedure in the MicroFIN task was very similar: First, participants had to explore the given automaton by pressing the available buttons and seeing what happens. After some time exploring self-selected state-transitions, in the second phase the task is to reach a specified goal state in the machine from a given state, with the least number of button presses. Figure 25.4 illustrates the interface for the "MP3-Player" developed for the PISA 2012 Study.

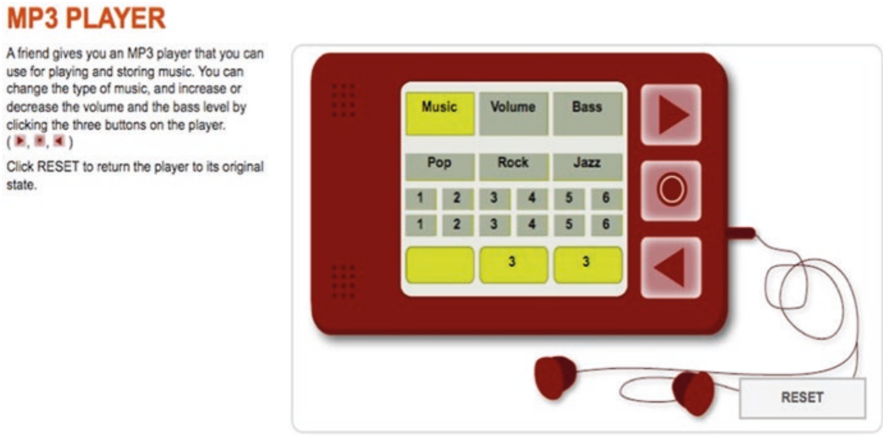

Fig. 25.4 MicroFIN item "MP3 Player" as an interactive problem-solving item example in PISA 2012. By pressing the buttons to the *right*, the MP3 player's state changes (indicated by the highlighted fields) (Version adapted from Greiff et al. 2013a, p. 78)

On the basis of the two formalisms, a large number of items (both for MicroDYN and MicroFIN) with different difficulty levels were developed and used in our studies.

Multiple Item Testing On the basis of the formal mechanisms of LSE and FSA, an additional feature of our approach comes into play, called multiple item testing. The idea comes from psychometrics and entails multiple items instead of single item testing. It is very easy to construct a universe of independent LSE and FSA tasks, each with varying degrees of difficulty. This procedure increases the reliability of measurement, compared to a situation in which one final data point after a long sequence of decisions is taken as a measure of performance (as is done, for example, in the standard procedure for the computer-simulated Tailorshop; see Danner et al. 2011).

Disadvantages of MicroDYN and MicroFIN The use of minimally complex systems and multiple item testing also has some disadvantages, the most important being the fact that dealing with complexity, in the sense of uncertainty management, is lost completely. In some cases, only main effects between three input and three output variables had to be identified—there were neither indirect effects nor delayed effects or goal conflicts. One could argue that—if no eigendynamic or side effects are implemented—these MicroDYN measurements mostly reflect the competence of the VOTAT strategy ("vary one thing at a time") or, in the phrasing of Klahr (2009) the CVS ("control of variables strategy"), but the set of strategies for dealing with complex systems is much larger. If broader strategies are to be assessed, different task requirements other than the identification of linear systems are needed. This has to do with the next point, stimulus sampling.

Stimulus Sampling of Problems For assessment purposes, a large item universe is needed. That is one of the advantages of formal systems (Funke 2001) such as linear structural equation systems or finite state automata. The disadvantage of using these formalisms is the restricted range of problems that follow all the same model. Subjects are confronted with changing semantics, but the deep structure of the problems does not change: one has to deal with linear combinations or with state transitions. After a short time, the problem situations become routine and the assessment runs the risk of no longer addressing problem-solving behavior. How long are subjects in those assessment situations problem solvers, and when do they learn from experience? Fiedler (2011, p. 166) warns against the consequences when stimulus sampling is not done broadly, namely, that "findings may reveal more about the stimuli chosen than the persons being tested".

Learning Effects A last problem of multiple-item testing consists in the fact that some generalizable strategies (e.g., VOTAT) can be learned during work on the first item of an item bundle, thus making the following items of that bundle easier, because problem-solving behavior changes from production to reproduction. Whereas learning within more complex tasks such as Tailorshop is part of the game, in a multiple item situation it could be a disadvantage, and would need to be controlled (see Funke 2014a).

25.4 Empirical Tests of the Newly Developed Instruments

During the active phase of our project, in cooperation with our partner institutions, we ran empirical tests of the newly developed instruments for the assessment of complex problem solving (CPS) based on multiple-item testing with MicroDYN. These tests addressed the following areas:

- Measurement model: What is the internal structure of our assumed competencies? Is it possible to identify the three postulated facets of (1) rule identification (adequateness of strategies), (2) rule knowledge (generated knowledge) and (3) rule application (ability to control a system)?
- Predictive and incremental validity: Do our constructs have validity in predicting external criteria like school grade point average (GPA), and is incremental prediction beyond IQ scores possible?
- Differences with respect to age, gender, culture: Is the data pattern with respect to differential variables (like the mentioned ones) plausible?

To answer these questions, some larger data collections at school were initiated by our research group: (1) school studies at the Heidelberg area, (2) school studies with a research group at Szeged University, and (3) school studies with a research group at Helsinki University. Reports about two of these data collections will be presented here in short (technical details can be found in the following publications: a paper from Wüstenberg et al. (2012) on the measurement model and on predictive and incremental validity (the Heidelberg School Study), and the paper from Wüstenberg et al. (2014) on individual differences with respect to age, gender, and cultural background [German-Hungarian School Comparison Study]).

Wüstenberg et al. (2012) analyzed the internal structure and construct validity of the newly developed MicroDYN items. The computer-based CPS test, with eight MicroDYN items, and the Raven Advanced Progressive Matrices, as traditional test of reasoning, were given to a sample of $N = 222$ university students.

Measurement model: Data analysis based on structural equation models showed that a two-dimensional model of CPS, including rule knowledge and rule application, fitted the data best. In this study, rule identification could not be established as a third facet on its own. Empirically, there was no difference between the two facets of rule identification and rule knowledge.

Predictive and incremental validity: Reasoning predicted performance in rule application only indirectly, through its influence on rule knowledge: This indicates that learning during system exploration is a prerequisite for controlling a system successfully. Also, MicroDYN scores explained variance in GPA even beyond reasoning, showing the incremental validity of our items. Our conclusion: MicroDYN items predict real life criteria such as GPA and therefore, measure important aspects of academic performance that go beyond reasoning.

Wüstenberg et al. (2014) analyzed cross-national and gender differences in complex problem solving. Six MicroDYN items were applied to a sample of 890 Hungarian and German high school students attending eighth to eleventh grade.

Differences with respect to gender and culture: Multi-group confirmatory factor analyses showed that measurement invariance of MicroDYN scores was found across gender and nationality. In regard to latent mean differences it showed that, on average, males outperformed females and German students outperformed Hungarian students. The main reason for these results was the comparatively poor performance of Hungarian females. Log files of process data showing the interaction of participants with the task illustrate that Hungarian females used the VOTAT strategy less often; as a consequence, they achieved less knowledge acquisition. A detailed log-file based analysis of such differences is therefore helpful for a better understanding of data from cross-national comparisons. We expect that such process analyses can also be helpful in better understanding group differences (between nations, gender, etc.) in large-scale assessments like PISA.

Summarizing: As can be seen from the empirical tests, our MicroDYN test development produced reliable data that were able to predict indicators like GPA, beyond IQ scores. Also, differential effects with respect to age, gender, and culture were mostly in line with our expectations and underline the usefulness of the new instruments for such comparisons.

25.5 Educational Application: PISA 2012

In educational contexts, measures of problem solving are useful if one is interested in cross-curricular competencies. The PISA 2012 definition of problem-solving competence is as follows:

> Problem-solving competency is an individual's capability to engage in cognitive processing to understand and resolve problem situations where a method of solutions is not immediately obvious. It includes the willingness to engage with such situations in order to achieve one's potential as a constructive and reflective citizen. (OECD 2013, p. 122)

In the PISA 2012 computer-based problem-solving assessment, with about 85,000 students from 44 countries and economies, over one half of the tasks were *interactive*. Examples of interactive problems encountered in everyday life include discovering how to use an unfamiliar mobile telephone or automatic vending machine. These PISA tasks were developed with the background described in this article; they were constructed on the basis of proposals from our Heidelberg research group.

PISA's *interactive* problems are intransparent (i.e., there is undisclosed information), but not necessarily dynamic or highly complex. *Static* problems are those in which all the information necessary to solve the problem is disclosed to the problem solver at the outset; by definition they are completely transparent.

Students' answers to the 42 problem-solving tasks in the assessment allowed the assignment of students into one of seven proficiency levels, including one that contained the students who performed below the first, and lowest, of six described proficiency levels. At the highest level, students should be able to do the following:

> At Level 6, students can develop complete, coherent mental models of diverse problem scenarios, enabling them to solve complex problems efficiently. They can explore a scenario in a highly strategic manner to understand all information pertaining to the problem. The information may be presented in different formats, requiring interpretation and integration of related parts. When confronted with very complex devices, such as home appliances that work in an unusual or unexpected manner, they quickly learn how to control the devices to achieve a goal in an optimal way. Level 6 problem-solvers can set up general hypotheses about a system and thoroughly test them. They can follow a premise through to a logical conclusion or recognize when there is not enough information available to reach one. In order to reach a solution, these highly proficient problem-solvers can create complex, flexible, multi-step plans that they continually monitor during execution. Where necessary, they modify their strategies, taking all constraints into account, both explicit and implicit. (OECD 2013, p. 122)

What are the educational and political consequences of this assessment? The OECD (2013, p. 122) report formulates:

> that today's 15-year-olds who lack advanced problem-solving skills face high risks of economic disadvantage as adults. They must compete for jobs in occupations where opportunities are becoming rare; and if they are unable to adapt to new circumstances and learn in unfamiliar contexts, they may find it particularly difficult to move to better jobs as economic and technological conditions evolve.

Training and teaching of problem-solving skills therefore becomes a task for schools.

25.5.1 Two Additional Issues: Optimization and Causal Diagrams

Use of Modern Mathematical Optimization Techniques As we have shown, important progress can be expected if the course of problem solving is evaluated quantitatively. Rather than merely evaluating the final solution, the concurrent evaluation of stepwise decision-making promises additional new insights, which can be achieved with the help of modern techniques of mixed-integer nonlinear optimization, as demonstrated by Sager et al. (2011) with the business scenario "Tailorshop". For that scenario, a process performance indicator can be computed under the label of "what is still possible": an indicator that shows the optimal solution at each point in time (during the round-based proceeding through the task), given all previous decisions and actions. For example, even for a subject who has played ten rounds of unsuccessful decision-making, there is still an optimal score for the last two rounds if, from now on, only the best decisions are made. This indicator allows a much more precise evaluation of a subject's solution path, compared to traditional indicators that measure the available money at the end of each round.

Causal Diagrams To measure knowledge acquisition by means of causal diagrams is a standard procedure in assessment procedures, and is used within MicroDYN. It leads to reliable measures of knowledge about causal relations, but it also has some disadvantages: On the one hand, considering causal connections between system variables stimulates thinking about causality that otherwise might not have been possible (see Blech and Funke 2006). On the other hand, Griffiths and Tenenbaum (2009, p. 670) point to an "inherent limitation in the expressive capacity of graphical models", due to the fact that they cannot discriminate between different types of causal entities or different functional relationships between variables, such as conditional links. Progress is needed, with respect to other ways of assessing structural knowledge. One has to be aware of the fact that this kind of mind-mapping turns out to be a secondary task that needs additional resources besides the identification task (see also Eseryel et al. 2013).

This issue relates also to an old question concerning implicit and explicit modes of knowledge about systems (see Berry and Broadbent 1988). Knowledge acquisition processes go for rule-abstraction, whereas knowledge application might be driven more by instance-based decision making (Gonzalez et al. 2003). Therefore, the question of adequate measurement of acquired knowledge is still open (also, learning curves would be helpful, to describe the process of acquisition in more detail).

25.6 Future Developments

Future developments could run along different, promising lines of research—we will explain two of them in more detail: (1) concerning the unit of analysis, an extension of complex problem-solving activities from the individual to the social dimension might occur, and (2) concerning methods, a more process-oriented use of log files resulting from computer-based assessments might reveal more process information. Further ideas for future research are described in Funke (2014b).

From the Individual to the Social Dimension The steep rise of communicative and team tasks in modern society (Autor et al. 2003) makes it evident that there is an inherently social aspect in any type of learning or problem solving (Lee and Smagorinsky 2000). To this end, collaborative problem solving—following Greiff et al. (2013a, p. 81)—is to be incorporated into an international large-scale assessment for the first time. In the PISA 2015 assessment framework (OECD 2012), collaborative problem solving is tentatively defined as “the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution” (p. 7). In keeping with previous efforts to define collaborative problem solving (e.g., Griffin et al. 2011; Morgan et al. 1993; O’Neil et al. 2003), collaboration and problem solving are seen as correlated but sufficiently distinct dimensions. That is, for problem solving, the cognitive processes of interactive problem solving in the PISA 2012 framework will be retained, whereas a new assessment of social and collaborative skills will be added in the PISA 2015 framework.

Process-Oriented Use of Log files To quote Duncker (1945, p. 1, in italics in the original) once again: “How does the solution arise from the problem situation? In what ways is the solution of a problem attained?”, is an important question in understanding the process of complex problem solving. To get answers on this old question, log files are promising a new era of process research (Schulte-Mecklenbeck and Huber 2003; Zoanetti 2010). Behavioral and process data of problem-solving patterns are now partly implemented in the PISA scoring procedures, and are directly connected to the emerging field of educational data mining, in which experimental and psychometric methods are applied to large educational data sets (Rupp et al. 2012). The promises of log-file analyses have been explored in recent work (see Goldhammer et al. 2014; Kupiainen et al. 2014) that gives deeper insights into problem-solving processes.

Optimistic Outlook Summarizing recent developments in problem-solving research under the auspices of what Stellan Ohlsson has correctly labeled the “Newell-Simon paradigm”, Ohlsson (2012, p. 117) wrote:

> In summary, Newell and Simon’s first concept of generality, codified in the General Problem Solver, failed as a psychological theory because it is not true: there is no single problem solving mechanism, no universal strategy that people apply across all domains and of which every task-specific strategy is a specific instance. Their second concept of generality initiated research on the cognitive architecture. The latter is a successful scientific concern with

> many accomplishments and a bright future. But it buys generality by focusing on a time band at which problem solving becomes invisible, like an elephant viewed from one inch away.

This pessimistic statement (specific problem solving research vanishes and ends up in general assumptions on cognitive architectures) is not our point of view. Within this priority program funded by the German Research Foundation, we have delivered some new ideas for psychometric sound assessment of problem solving (multiple item testing based on minimally complex systems from LSE and FSA formalisms). The competencies needed for these tasks are derived from Dörner's theory of operative intelligence. The measurement invariance, latent mean comparisons, and other psychometrically relevant data are documented in international large-scale studies beyond PISA (e.g., Wüstenberg et al. 2014). Therefore, as we have tried to show in this chapter, at least with respect to the assessment of problem solving competencies, some progress has been made in recent research activities, and will also be made in the future.

Acknowledgments The work reported here would not have been possible without the generous funding from the German Research Foundation (DFG) over so many years. Since 2007, continuous support (Fu 173/11, Fu 173/13, Fu 173/14) within the Priority Program (SPP 1293) "Competence Models for Assessing Individual Learning Outcomes and Evaluating Educational Processes" helped us to develop theory as well as assessment instruments, and to collect data. Additional, related funding from the German Ministry of Science (BMBF) supported these developments (FKZ 01JG1062: "Dynamisches Problemlösen als fachübergreifende Kompetenz", 2011–2014 to JF, and FKZ 012LSA004: "Validation and development of individual assessment devices to capture Dynamic Problem Solving", 2012–2015, to JF, together with Frank Goldhammer and SG). Also, DIPF (Frankfurt) and SoftCon (Munich) were extremely helpful in supporting the development of the ItemBuilder software.

Thanks are due not only to funding agencies but also to individual people. Andreas Fischer, Julia Hilse, Daniel Holt, André Kretzschmar, Jonas Neubert, and Sascha Wüstenberg are in the first line of acknowledgement; many Heidelberg student assistants are in the second line, our participants in the data collections in the third line. Colleagues from around the world have discussed our concepts with us intensively: Ray Adams, Esther Care, Beno Csapo, Michel Dorochevsky, John Dossey, Frank Goldhammer, Coty Gonzalez, Art Graesser, Patrick Griffin, Jarkko Hautamäki, Eckhart Klieme, Sirkku Kupiainen, Detlev Leutner, Romain Martin, Reinhold Nickolaus, Barry McCrae, Gyongyver Molnar, Markku Niemivirta, Magda Osman, Ray Philpot, Dara Ramalingam, Heiko Rölke, Andreas Schleicher, Alexander Siegmund, Philipp Sonnleitner, Robert Sternberg, David Tobinski, Mari-Pauliina Vainikainen, Kathrin Viehrig, Joachim Wirth—to mention only some of them. Thanks to all of you!

References

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*, 1279–1333. doi:10.1162/003355303322552801.

Barth, C. M., & Funke, J. (2010). Negative affective environments improve complex solving performance. *Cognition & Emotion, 24*, 1259–1268. doi:10.1080/02699930903223766.

Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology, 79*, 251–272. doi:10.1111/j.2044-8295.1988.tb02286.x.

Blech, C., & Funke, J. (2006). Zur Reaktivität von Kausaldiagramm-Analysen beim komplexen Problemlösen [On the reactivity of causal diagram assessments in complex problem solving]. *Zeitschrift für Psychologie, 117*, 185–195. doi:10.1026/0044-3409.214.4.185.

Brehmer, B. (1989). Dynamic decision making. In A. P. Sage (Ed.), *Concise encyclopedia of information processing in systems and organizations* (pp. 144–149). New York, NY: Pergamon.

Broadbent, D. E. (1977). Levels, hierarchies, and the locus of control. *Quarterly Journal of Experimental Psychology, 29*, 181–201. doi:10.1080/14640747708400596.

Buchner, A., & Funke, J. (1993). Finite-state automata: Dynamic task environments in problem-solving research. *The Quarterly Journal of Experimental Psychology Section A, 46*, 83–118. doi:10.1080/14640749308401068.

Danner, D., Hagemann, D., Holt, D. V., Hager, 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. doi:10.1027/1614-0001/a000055.

Dörner, D. (1975). Wie Menschen eine Welt verbessern wollten und sie dabei zerstörten [How people wanted to improve the world and destroyed it]. *Bild der Wissenschaft, 12*, 48–53.

Dörner, D. (1980). On the difficulty people have in dealing with complexity. *Simulations and Games, 11*, 87–106. doi:10.1177/104687818001100108.

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

Dörner, D. (1997). *The logic of failure: Recognizing and avoiding error in complex situations*. New York, NY: Basic.

Dörner, D., Kreuzig, H. W., Reither, F., Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität* [Lohhausen. On dealing with uncertainty and complexity]. Bern: Huber.

Duncker, K. (1935). *Zur Psychologie des produktiven Denkens [Psychology of productive thinking]*. Berlin: Springer.

Duncker, K. (1945). On problem-solving. *Psychological Monographs, 58*(5), 1–113.

Eseryel, D., Ifenthaler, D., & Ge, X. (2013). Validation study of a method for assessing complex ill-structured problem solving by using causal representations. *Educational Technology Research and Development, 61*, 443–463. doi:10.1007/s11423-013-9297-2.

Fiedler, K. (2011). Voodoo correlations are everywhere—Not only in neuroscience. *Perspectives on Psychological Science, 6*, 163–171. doi:10.1177/1745691611400237.

Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *The Journal of Problem Solving, 4*(1), 19–42. doi:10.7771/1932-6246.1118.

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

Funke, J. (1992). Dealing with dynamic systems: Research strategy, diagnostic approach and experimental results. *German Journal of Psychology, 16*, 24–43.

Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning, 7*, 69–89. doi:10.1080/13546780042000046.

Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing, 11*, 133–142. doi:10.1007/s10339-009-0345-0.

Funke, J. (2012). Complex problem solving. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 682–685). Heidelberg: Springer. doi:10.1007/978-1-4419-1428-6_685.

Funke, J. (2014a). Analysis of minimal complex systems and complex problem solving require different forms of causal cognition. *Frontiers in Psychology, 5*, 739. doi:10.3389/fpsyg.2014.00739.

Funke, J. (2014b). Problem solving: What are the important questions? In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 493–498). Austin: Cognitive Science Society.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology, 106*, 608–626. doi:10.1037/a0034716.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27*, 591–635. doi:10.1016/S0364-0213(03)00031-4.

Greiff, S., & Fischer, A. (2013a). Der Nutzen einer komplexen Problemlösekompetenz: Theoretische Überlegungen und empirische Befunde [The value of complex problem-solving competence: Theoretical considerations and empirical results]. *Zeitschrift für Pädagogische Psychologie, 27*(1), 27–39. doi:10.1024/1010-0652/a000086.

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., & Funke, J. (2009). Measuring complex problem solving: The MicroDYN approach. In F. Scheuermann (Ed.), *The transition to computer-based assessment: Lessons learned from large-scale surveys and implications for testing* (pp. 157–163). Luxembourg: Office for Official Publications of the European Communities.

Greiff, S., & Funke, J. (2010). Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal komplexer Systeme [Systematic research on complex problem solving by means of minimal complex systems]. *Zeitschrift für Pädagogik, Beiheft, 56*, 216–227.

Greiff, S., & Neubert, J. C. (2014). On the relation of complex problem solving, personality, fluid intelligence, and academic achievement. *Learning and Individual Differences, 36*, 37–48. doi:10.1016/j.lindif.2014.08.003.

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement, 36*, 189–213. doi:10.1177/0146621612439620.

Greiff, S., Holt, D. V., & Funke, J. (2013a). Perspectives on problem solving in educational assessment: analytical, interactive, and collaborative problem solving. *The Journal of Problem Solving, 6*, 71–91. doi:10.7771/1932-6246.1153.

Greiff, S., Wüstenberg, S., Holt, D. V., Goldhammer, F., & Funke, J. (2013b). Computer-based assessment of complex problem solving: Concept, implementation, and application. *Educational Technology Research and Development, 61*, 407–421. doi:10.1007/s11423-013-9301-x.

Griffin, P., McGaw, B., & Care, E. (Eds.). (2011). *Assessment and teaching of 21st century skills*. New York, NY: Springer.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*, 661–716. doi:10.1037/a0017201.

Klahr, D. (2009). "To every thing there is a season, and a time to every purpose under the heavens": What about direct instruction? In S. Tobias & T. M. Duffy (Eds.), *Constructivist theory applied to instruction: Success or failure?* (pp. 291–310). New York, NY: Taylor & Francis.

Klein, G. (1997). The current status of the naturalistic decision making framework. In R. H. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress: Emerging theories and applications* (pp. 11–28). Aldershot: Ashgate.

Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG [Competence models for assessing individual learning outcomes and evaluating of educational processes. Description of a new DFG priority program]. *Zeitschrift für Pädagogik, 52*, 876–903.

Kupiainen, S., Vainikainen, M.-P., Marjanen, J., & Hautamäki, J. (2014). The role of time on task in computer-based low-stakes assessment of cross-curricular skills. *Journal of Educational Psychology, 106*, 627–638. doi:10.1037/a0035507.

Lee, C. D., & Smagorinsky, P. (Eds.). (2000). *Vygotskian perspectives on literacy research: Constructing meaning through collaborative inquiry*. Cambridge: Cambridge University Press.

MacKinnon, A. J., & Wearing, A. J. (1980). Complexity and decision making. *Behavioral Science, 25*, 285–296. doi:10.1002/bs.3830250405.

MacKinnon, A. J., & Wearing, A. J. (1985). Systems analysis and dynamic decision making. *Acta Psychologica, 58*, 159–172. doi:10.1016/0001-6918(85)90005-8.

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: Erlbaum.

Morgan, B. B., Salas, E., & Glickman, A. S. (1993). An analysis of team evaluation and maturation. *Journal of General Psychology, 120*, 277–291. doi:10.1080/00221309.1993.9711148.

O'Neil, H. F., Chuang, S., & Chung, G. K. W. K. (2003). Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education: Principles, Policy & Practice, 10*, 361–373. doi:10.1080/0969594032000148190.

OECD (Organisation for Economic Co-operation and Development). (2012, April). *PISA 2015 field trial collaborative problem solving framework*. Paper presented at the 33rd PISA Governing Board meeting, Tallinn, Estonia.

OECD (Organisation for Economic Co-operation and Development). (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: Author.

Ohlsson, S. (2012). The problems with problem solving: Reflections on the rise, current status, and possible future keywords. *The Journal of Problem Solving, 5*(1), 101–128. doi:10.7771/1932-6246.1144.

Rupp, A. A., Nugent, R., & Nelson, B. (2012). Evidence-centered design for diagnostic assessment within digital learning environments: Integrating modern psychometrics and educational data mining. *Journal of Educational Data Mining, 4*, 1–10.

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*, 936–959. doi:10.1137/11082018X.

Schulte-Mecklenbeck, M., & Huber, O. (2003). Information search in the laboratory and on the web: With or without an experimenter. *Behavior Research Methods, Instruments, & Computers, 35*, 227–235.

Simon, H. A. (1959). Theories of decision making in economics and behavioural science. *American Economic Review, 49*, 253–283. doi:10.2307/1809901.

Spering, M., Wagener, D., & Funke, J. (2005). The role of emotions in complex problem-solving. *Cognition & Emotion, 19*, 1252–1261. doi:10.1080/02699930500304886.

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

Verweij, M., & Thompson, M. (Eds.). (2006). *Clumsy solutions for a complex world: Governance, politics and plural perceptions*. Houndmills: Palgrave Macmillan.

Weinert, F. E. (Ed.). (2001). *Leistungsmessungen in Schulen [Assessment in schools]*. Weinheim: Beltz.

Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving—More than reasoning? *Intelligence, 40*, 1–14. doi:10.1016/j.intell.2011.11.003.

Wüstenberg, S., Greiff, S., Molnár, G., & Funke, J. (2014). Cross-national gender differences in complex problem solving and their determinants. *Learning and Individual Differences, 29*, 18–29. doi:10.1016/j.lindif.2013.10.006.

Zoanetti, N. (2010). Interactive computer based assessment tasks: How problem-solving process data can inform instruction. *Australasian Journal of Educational Technology, 26*, 585–606.

Zsambok, C. E., & Klein, G. A. (Eds.). (1997). *Naturalistic decision making*. Mahwah: Erlbaum.