Chapter 6

Interactive problem solving: Exploring the potential of minimal complex systems

by

Samuel Greiff

ECCS unit, University of Luxembourg, Luxembourg.

and Joachim Funke

Department of Psychology, Heidelberg University, Germany.

Interactive problem solving (IPS) is considered an increasingly important skill in professional and everyday life as it mirrors the interaction of a human user with dynamic technical and non-technical devices. Here, IPS is defined as the ability to identify the unknown structure of artefacts in dynamic, mostly technology-rich, environments to reach certain goals. Two overarching processes, the acquisition of knowledge and its application, can be theoretically distinguished and this chapter presents two measurement approaches assessing these processes: MicroDYN and MicroFIN, both of which rely on the idea of minimal complex systems (MICS). Whereas MicroDYN models quantitative connections between elements of a problem (i.e. more and less), MicroFIN models qualitative connections between them (e.g. on/off or white/black). This chapter summarises research on these approaches and discusses the implications for educational assessment.

Introduction

Complexity is an intriguing feature of modern times but one that is difficult to grasp. One of its sources is the increasing impact and the augmented presence of interactive devices in our daily environment. Devices that change dynamically with user interaction have arguably led to a greater amount of interactivity and complexity in today's world. This happens not only in light of newly emerging software tools, which make continuous learning necessary, but also in light of specialised hardware that confronts us with complex interactions: mobile phones, ticket machines, electronic room access keys, copiers and even washing machines now require sequences of interactions to set up these devices and to let them run. To master them, one has to press buttons or use sliders and other continuous controls (such as the accelerator of a car, the volume of an MP3 player or the temperature of a heating system) or even a combination of both. Even in non-technical contexts, humans have to face interactive situations, even though this area is only slightly touched on by research on problem solving – although the Programme for International Student Assessment (PISA) 2015 has made efforts to assess collaborative problem solving. For instance, social interaction may be described as a complex and – through verbal and non-verbal interaction – dynamically changing problem. In a conversation, both continuous changes, such as mood increases or decreases, and discrete changes, such as eye contact or avoidance of it, may occur. In fact, humans experience many problems when interacting with different classes of devices. That is one of the reasons why research on problem solving as an individual ability is of increasing importance and assessment of problem solving skills comes into play in various practical and scientific fields.

When interacting with such devices, the problem solver essentially needs to master two tasks. The first is to understand the device itself, the way it dictates interactions with the user, and the individual skills and abilities required to interact with the device (Dörner, 1986; Funke, 2001). . In general, what a human user has to do in dealing with a complex and interactive device is straightforward: find out how the device works and store this information (knowledge acquisition: find a strategy to build up knowledge and then represent it internally; Mayer and Wittrock, 2006). The second task for the solver is to try to reach the goal state (knowledge application: apply the acquired knowledge to reach a certain goal; Novick and Bassok, 2005). Therefore, these two tasks of knowledge acquisition and knowledge application are of primary interest to understand and assess problem solving. Both are essential in the context of problem-solving research (Funke, 2001; Novick and Bassok, 2005; OECD, 2014; Wüstenberg, Greiff and Funke, 2012).

An interactive problem is one where knowledge acquisition and knowledge application are used to apply non-routine actions to reach a certain goal state (Greiff, Holt and Funke, 2013). Dörner (1996) used an even more detailed approach towards describing interactive problems: according to him, solving a problem demands a series of operations which can be characterised as follows: elements relevant to the solution process are large and manifold (complexity), highly interconnected (connectivity), and changing over time (dynamics). Neither structure nor dynamics are disclosed to the problem solver (intransparency). Finally, the goal structure is not as straightforward as suggested above: in dealing with a complex problem a person is confronted with a number of different subgoals, which are to be weighted and co-ordinated against each other – a polytelic situation.

Each of these attributes of an interactive problem corresponds to one of five demands placed on the problem solver as defined by Dörner (1986): 1) complexity requires reduction of information; 2) connectivity requires model building; 3) dynamics requires forecasting; 4) intransparency requires information retrieval; and 5) polytely requires the evaluation of multiple criteria.

Whereas all five requirements play an important role in interactive problem solving, the first three can be understood as specific dimensions of knowledge acquisition and the last two are subsumed under the overarching process of knowledge application (Fischer, Greiff and Funke, 2012). An assessment needs to either focus on the two overarching processes or on each of the five

subdimensions, but some conceptual background is needed in addition to psychometric demands dictated by scientific standards of assessment (e.g. Funke and Frensch, 2007).

In this chapter, we focus our understanding of problem solving mainly on the interaction of a human user with a dynamic and complex device and the corresponding processes of knowledge acquisition and knowledge application (Fischer et al., 2012; Novick and Bassok, 2005). To label this focus adequately, we propose to use the term interactive problem solving (IPS),[1] which emphasises the characteristic interaction between problem solver and problem and the inherent changes of the problem situation. This term is also used in PISA 2012 (OECD, 2014). In this chapter, after an explanation of the concept in the first section, the second section illustrates how to measure IPS in a psychometrically sound way based on the idea of interacting with several complex systems. The third section presents a review of recent studies based on this concept. Finally, in the discussion we present three open questions and provide tentative answers to them.

Interactive problem solving

What is IPS and how can it be defined? The 21st-century offers the citizens and workers of industrialised countries a technology-rich environment. It is taken for granted that a person is able to use mobile communication, household devices, electronically equipped cars, public transportation and various technologies in the workplace, in short that they can use technologies offering a broad range of applications, which are rapidly changing and which often introduce new features and services to their users. These features and services require interaction with mostly automated environments (such as using credit cards for public transport or paying via phone) that are in principle similar to each other but at the same time context-specific and change from region to region. For example, when travelling around the world, one experiences a lot of different systems to make payments for public transport even if in principle the systems are fundamentally the same (Funke, 2001; Greiff and Funke, 2009). Thus, in concentrating on IPS we focus on a 21st-century skill that is not assessable by traditional paper-and-pencil tests as these are hardly interactive – we focus on a competence that can only be assessed by a person interacting with a dynamically reacting environment in a computer-based assessment setting.

From this observation, IPS is defined as the ability to explore and identify the structure of (mostly technical) devices in dynamic environments by means of interacting and to reach specific goals (compare with OECD, 2014). It is worth commenting on the different components of this definition in turn. First, the term "ability" refers to the fact that IPS develops over a lifetime and that it can be learned and fostered, for instance by specific training or by schooling (Frischkorn, Greiff, and Wüstenberg, 2014; Greiff et al., 2013). Second, the two tasks mentioned ("to identify the unknown structure" and "to reach specific goals") refer to the abstract requirements of 1) system identification and exploration (knowledge acquisition); and 2) system control and steering (knowledge application; Funke, 1993). System identification simply means finding out how a given device such as a new mobile phone works and how it reacts to certain inputs. It is usually preceded or accompanied by a phase of exploration, without any specific targets (Kröner, Plass and Leutner, 2005; Wüstenberg et al., 2012). System control simply means to know how to get what you want, for instance making a phone call with the new mobile phone. This control heavily relies on the knowledge acquired beforehand (Funke, 2001; Greiff, 2012). Third, the notion of an "unknown structure" points to the fact that IPS deals with new situations for which a routine solution is not at hand and that even though prior knowledge may play a role, it is neither essential nor necessary to solve the problem (Greiff, Wüstenberg and Funke, 2012). Fourth, the allusion to "(mostly technical) devices in dynamic environments" refers to an important aspect of this type of problem, namely the interaction between a user and the device. This interaction implies that during the course of interaction the state of the device changes, either depending on the type of intervention or dynamically by itself (Buchner, 1995). The resulting dynamics are in contrast to simple problems that are static in nature such as a

chess problem (Chi, Glaser, and Rees, 1982). Fifth, the term "devices" contrasts the objects involved in IPS to natural objects, (social) events and so on, which are not the focus of IPS even though they can be formally modelled within the definition of IPS. It specifies for example, that IPS does not deal with problems requiring other abilities subsumed under the label emotional intelligence (Otto and Lantermann, 2005). Sixth, "by means of interacting" characterises the mode of exploration: information is generated and retrieved not by reading a manual or by applying content knowledge picked up in school, but by actively dealing with the system. Not explicitly included in the definition is the aspect of metacognition such as monitoring and reflecting one's own problem-solving activities. But as problem solving is not only an ability but also a process, it needs some feedback and control structures that guide the activities. In the definition mentioned above it is implicitly assumed that these control processes accompany the entire problem-solving process (OECD, 2014; Wirth, 2004). IPS is embedded into cultural contexts, which is easily seen by the examples that are largely taken from industrialised countries. This cultural dependency is readily acknowledged even though it is assumed that IPS will be important in any society that wants to move to success in the sense of industrialisation and in the sense of a technological society (OECD, 2014).

Measuring interactive problem solving

With a definition of IPS at hand, the question of how to measure the construct immediately emerges. That is, how can the theoretical concept be adequately translated into empirical scales?

This question is far from trivial – besides the theoretical concept further psychometric and assessment demands need to be thoroughly met. According to Buchner (1995), two approaches are frequently used to measure IPS: 1) computer-simulated microworlds with a touch of real life composed of a large amount of variables (> 1 000 in the case of the famous scenario "Lohhausen" from Dörner, 1980); 2) simplistic, artificial and yet complex problems following certain construction rules (e.g. the DYNAMIS-approach using linear structural equation systems from Funke, 1992). Whereas the first approach uses ad hoc constructed scenarios to demonstrate individual differences, the second approach uses systematically constructed scenarios to demonstrate the effects of system attributes (Buchner, 1995). Both approaches have specific advantages and disadvantages in terms of 1) time used for testing; 2) realism; 3) the underlying measurement model; 4) available data on reliability and validity; 5) comparability between different scenarios; and 6) overall scalability with respect to varying difficulty levels. A detailed description of those is found in Greiff (2012).

Recently, Greiff et al. (2012) and Wüstenberg et al. (2012) proposed the MicroDYN approach as a measurement advancement of the DYNAMIS-approach by Funke (1992, 2001). MicroDYN was developed with a psychometric perspective for use in large-scale assessments (such as PISA) with computer-based test administration (Reeff and Martin, 2008). It contains an entire set of tasks, each of which consists of a small system of causal relations, which are to be explored within 3-4 minutes and afterwards controlled for given goal states. That is, MicroDYN allows for the assessment of the two overarching processes of knowledge acquisition and knowledge application. The main feature of the MicroDYN approach is the search for minimal complex systems (MICS), that is, systems which contain all (or at least most) of the features of a complex system (complexity, dynamics, polytely and intransparency; see Funke, 1991) but at the same time have low values for these parameters compared to the extremely difficult microworlds. From a psychometricians' point of view this approach has several advantages over the microworlds used before, such as maintaining the validity of the tasks and reducing testing time to a minimum (compare also Greiff, Fischer, Stadler and Wüstenberg, 2015, for a detailed description of this approach, which they call the *multiple* complex system (MCS) approach, using the term multiple to stress the number of independent complex problem-solving tasks that are administered, rather than the level of complexity as in minimal complex systems; of note, the terms minimal and multiple complex systems are often used interchangeably).

The MicroDYN approach uses the formalism of linear structural equations (LSEs) to model systems with continuous variables (Funke, 1993). It is described in detail in Greiff et al. (2015) and in Wüstenberg et al. (2012). In this chapter, we argue not only for the use of minimal complex systems within the MicroDYN approach but also to extend the idea from systems with continuous variables to systems with discrete variables (MicroFIN; introduced by Buchner and Funke, 1993). Funke (2001) showed that both formal frameworks – LSEs and finite state automata (FSAs) – are ideal instruments for problem-solving research. Using minimal complex systems with the framework of finite state automata therefore seems a natural extension of the MicroDYN approach and has been introduced in greater detail by already Greiff et al. (2015). But before we go into the details of MicroDYN and MicroFIN, we will briefly explain the guiding philosophy behind the approach of minimal complex systems.

The philosophy behind minimal complex systems

One of the key concepts for the MicroDYN and MicroFIN approach is the overarching idea of "minimal complex systems" (MICS). The starting point for MICS is the assumption that complex systems are needed in problem-solving assessment because their features differ markedly from simple systems (in terms of complexity, connectivity, dynamics, intransparency and polytely; Dörner, 1986; Fischer et al., 2012) and are more than the sum of simple processes (Funke, 2010).

In the initial phase of problem-solving research beginning in the 1970s, it was thought that complex systems should realise a maximum amount of these features of complex problems, or, in other words, the more complex, the more connected, the more intransparent, and the more dynamic a system was, the better it was assumed to capture problem-solving behaviour. The underlying idea was to create computer simulations that were able to resemble reality to a highly detailed degree and thereby, to finally bring reality to the laboratory. So, for example, the famous microworld, Lohhausen (Dörner, 1980), contained over 1 000 variables with a testing time of well over several hours, whereas other less famous scenarios are said to incorporate up to 25 000 variables. The rationale behind this research strategy was to realise complexity in its extreme value, that is, to realise "maximal complex systems" (MACS).

What happened to these MACS systems from an assessment perspective? Despite their potential to realise highly complex systems they were associated with a lot of methodological problems and could not answer fundamental questions satisfactorily (Funke and Frensch, 2007): how to evaluate participants' actions and decisions (search for the optimal intervention); how to separate the effects of individual actions from those inherent in the dynamics of the microworld (one single intervention into a system with 5 000 connected variables can have 5 000 direct consequences and many more indirect effects in the next cycles); how to construct a test that contains more than one task because single tasks lead to dependencies between all actions within this system (Greiff et al., 2015); how to produce multiple tasks without spending too much time on assessment?

Here, basic measurement issues should have come into play early on, but conceptual considerations were always deemed more important than psychometric desiderata, which should form the basis of any test development.

The conception of MICS addresses these unsolved issues with a simple strategy: instead of realising increasingly complex systems (trying to aim for the top level of complexity) with a focus on content validity and psychometrically questionable results, MICS focuses on the lowest level and asks for the minimum value of complexity that is still valid in terms of representing the underlying concept of interactive problem solving. Complexity is a fuzzy term (Fischer et al., 2012; Funke, 2001) – we do not know what the most complex system on earth or even in the universe is.

So, the upper limit of complexity is still open. But, for good reasons, the lower limit of complexity must be somewhere between nothing and a little bit of complexity. Instead of searching for the upper bounds of complexity in MACS, in MICS we concentrate on the lower bound.

This shift in focus has been frequently applied in cognitive assessments; for example, intelligence tests are not meant to reflect real world tasks but to capture the basic processes necessary for mastering these tasks on an abstract level. It brings several advantages for test developers: 1) the time spent on a single scenario is not measured in hours but in minutes, thereby increasing the efficiency of test application; 2) due to the short time for task application, a series of independent tasks can be presented instead of one, thereby increasing reliability and making the set of tasks in principle scalable; 3) because of the existence of a formal framework for task development, tasks can be embedded in arbitrary real-world scenarios that are independent of the system's structure, thereby increasing ecological validity; and, 4) easy, medium and difficult tasks can be presented, broadening the range of difficulty and thereby increasing conceptual validity.

After more than 30 years of research with interactive problems, the measurement perspective now becomes a leading force for innovation. It continues a line of psychometric research started by Wagener (2001) and Kröner et al. (2005) who both tried to develop tests for interactive problem solving with the measurement perspective in mind. The MICS approach presented here is a logical continuation of this approach.

The basic elements of minimal complex systems

The MICS principle can be applied to the two formalisms mentioned above, linear structural equations and finite state automata (see again Greiff et al., 2015 for a detailed description). More specifically, the widely applied principles of LSEs and FSAs in the MACS approach are altered to allow a lower level of complexity by simply reducing the number of elements and connections involved and by reducing the time on each task. Because LSEs and FSAs (Funke, 2001) address different aspects (the first describes quantitative changes in continuous variables, the second qualitative changes in discrete variables), the following section introduces them separately even though it is assumed that they both tap into the same construct (OECD, 2014).

MicroDYN

In MicroDYN, a system's basic elements are the number of exogenous and endogenous variables and the number and type of relations between them. The relation between exogenous and endogenous variables can be qualified by its strength (weight of the relation) and direction (positive or negative). All of these features are used to scale a task's difficulty. In the MACS approach there may be 20 exogenous and endogenous variables with 30 connections between them, whereas the MICS approach gets by with as few as 4 to 6 variables and between 1 and 5 connections, but still yields systems with considerably varying difficulty. That is, even with this amount of formally low complexity, difficult tasks can easily be created (Greiff et al., 2012; Wüstenberg et al., 2012). The variables are labeled either abstractly with "A", "B", and "C" or with semantic meaningful labels such as "water", "wind", and "energy". Figure 6.1 illustrates a MICS system with two exogenous (A, B) and two endogenous (Y, Z) variables.

The subscript of each variable indicates the point in time (t respective $t+1$), the system itself being event driven in discrete steps. From Equation 1 it follows that the value of Y at time $t+1$ is equal to the value of variable A at time t, multiplied by 2. From Equation 2 follows the same logic of computation. The graphical depiction in Figure 1 and the two equations (1) and (2) are equivalent in terms of the underlying causal structure but the diagram is more convenient to understand.

Exploration in the MicroDYN environment requires a careful analysis of the intervention effects: the increase of an exogenous variable could lead to an increase in one or more endogenous variables, or a decrease, or a mixed effect (increase in some variables and decrease in others), or to no effect at all. By carefully designing the sequence of interventions, a meaningful pattern of outputs can be generated and hypotheses about causal effects can be formulated (compare Sloman, 2005). As well as these relations between input variables (A and B in the example) and the output variables (Y and Z), the system may change

by itself, which is expressed by connections between one endogenous variable and another (for example, Y could influence Z) or on itself (for example, Y could influence itself over time), and which also need to be represented. After the problem solver has spent some time gathering and expressing knowledge about the system, externally given goal values need to be reached. Thus, knowledge acquisition and knowledge application, the two main problem-solving processes (Novick and Bassok, 2005) can be assessed. Further developments incorporating the more detailed approach of Dörner (1986) with his five dimensions also exist and allow for a more detailed diagnostic approach (compare Fischer et al., 2012; Greiff, 2012).

Figure 6.1 **Structure of a MICS system with two exogenous and two endogenous variable2**
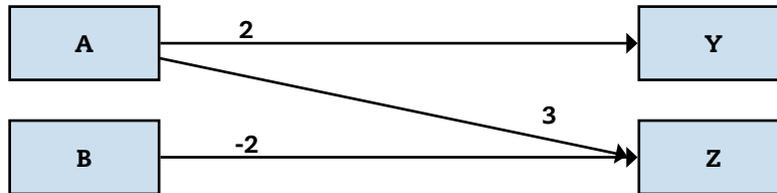
Note: The endogenous variables (A and B) and the exogenous variable are connected with causal paths indicated by arrows; weights for these paths are indicated next tot he arrows.
Source: Funke, J. (2001), "Dynamic systems as tools for analysing human judgement", Thinking and Reasoning, Vol. 7/1, pp. 69-89.

Formally, the system shown in Figure 6.1 can be described by two linear equations:

$$Y_{t+1} = 2 \times A_t \quad (1)$$
$$Z_{t+1} = 3 \times A_t - 2 \times B_t \quad (2)$$

MicroFIN

In MicroFIN, the basic elements are input signals, output signals, states and the state transitions between them. In contrast to MicroDYN, numbers and quantitative values are not important in MicroFIN, but elements of the problem are connected by qualitative transitions between them (Funke, 2001; Greiff et al., 2015). Figure 6.2 illustrates a simple finite state automaton with two input signals and three states (Thimbleby, 2007).

Figure 6.2 **A simple finite state automaton**
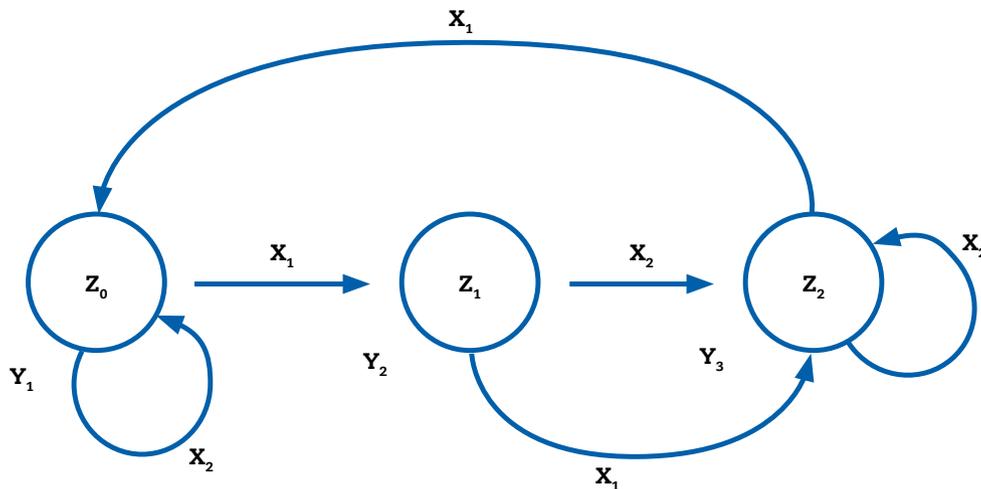
Note: The figure represents a finite state automaton with three states z_0, z_1, and z_2, and two input variables x_1 and x_2, leading to three possible output signals y_1, y_2, and y_3 depending on the reached state.
Source: Funke (2001:78).

Table 6.1 **State transition matrix of a fictitious finite state automaton**

state / output	input	
	x_1	x_2
z_0 / y_1	z_1	z_0
z_1 / y_2	z_2	z_2
z_2 / y_3	z_0	z_2

Note: The table shows the state transition matrix for the finite state automaton represented in Figure 6.2. For each state there exists exactly one corresponding output signal (y_1, y_2, and y_3). The cells of the matrix indicate the next state for the machine given the combination of current state and input signal.

Source: Funke (2001:77).

Formally, the diagram in Figure 6.2 is equivalent to the state transition matrix in Table 6.1. Once again, the graphical representation seems more convenient when representing an FSA. Acquiring knowledge in a MicroFIN environment requires a step-by-step analysis of the state transitions. In principle, at least as many steps are required as there are different state transitions before an FSA is entirely penetrated, whereas in practice, this number can be reduced by detecting hierarchies and analog functions within the automaton. Wirth (2004) suggests first ideas for measuring exploration behaviour, as well as knowledge representation and its application, whereas Neubert, Kretzschmar, Wüstenberg and Greiff (2015) present first empirical results on the validity of MicroFIN as an empirically applicable assessment instrument.

To give an example of an FSA, a digital watch is (in large part) a typical finite state machine. Transitions between different states are mostly of a qualitative nature (e.g. from alarm mode to snooze mode) and the tasks mentioned above and also present in MicroDYN have to be met: knowledge about the system has to be gathered and represented (knowledge acquisition) and a given goal, such as setting the alarm for 7 o'clock in the morning, has to be reached (knowledge application). Whereas MicroDYN tasks yield the advantage of being homogenous and adding up to a narrow but reliable test set (low bandwidth, high fidelity), MicroFIN tasks are closer to real life and more heterogeneous in the skills they test (high bandwidth, low fidelity; Greiff et al., 2015). Besides these differences, there are considerable overlaps between MicroDYN and MicroFIN, which we will elaborate on now.

Common elements of MicroDYN and MicroFIN

MicroDYN and MicroFIN have some common elements with respect to the task subjects have to meet and with regard to the potential diagnostic procedures. From the subjects' point of view, both paradigms require similar activities, at least in principle: the identification of an unknown dynamic device and the subsequent control of this device. Subjects have to set up an exploration strategy which generates information about the system and incorporate this into a mental representation (knowledge acquisition); the subsequent control performance is the result of a goal-directed application of the acquired knowledge (knowledge application). Within these two main processes (Novick and Bassok, 2005), further subprocesses can be distinguished and separate indicators for the problem-exploration stage have yielded promising results (see below). Because of the similarity of the tasks in MicroDYN and MicroFIN, we derive the same diagnostic parameters in both cases. A detailed description of potential indicators and scoring procedures is found in Buchner (1995), Neubert et al. (2015), and Wüstenberg et al. (2012). In general, both tasks tap into the same construct, as empirically shown by Greiff et al. (2013) and Neubert et al. (2015). For the first time, MicroDYN and MicroFIN make psychometric assessment procedures available to capture this construct originally coming from cognitive psychology.

Recent results on interactive problem solving

Empirical research on MICS employing the MicroDYN and MicroFIN approach has gathered momentum over the last couple of years with a number of interesting findings now available. Most of these studies have demonstrated the gain in validity from combining formal frameworks from problem-solving research with a psychometric measurement perspective. Knowledge acquisition and knowledge application as overarching processes can be empirically separated in different populations (e.g. Greiff et al., 2013; Kröner et al., 2005; Wüstenberg, et al., 2012). The correlation between the different problem-solving processes is usually high, but different from unity. Even more important than concurrence between empirical and theoretical structure is that MicroDYN and MicroFIN as well as other measures of problem solving are correlated with and yet distinct from general intelligence (Kröner et al. 2005; Wüstenberg et al., 2012; Sonnleitner et al., 2012) and from working memory (Bühner, Kröner and Ziegler, 2008; Schweizer et al., 2013). Further, IPS predicts school achievement incrementally beyond intelligence, but only if MICS are used rather than MACS, showing that 1) measures of intelligence do not fully overlap with IPS; and that 2) the MICS approach is needed to produce reliable problem-solving scales (compare in detail Greiff et al., 2012, comparing MicroDYN to Space Shuttle; Greiff, Stadler, Sonnleitner, Wolff and Martin, 2015 comparing MicroDYN and MicroFIN to the Tailorshop, another assessment instrument based on MACS). Abele et al. (2012) report that IPS is predictive of building up knowledge in specific domains, which leads to domain-bound problem-solving skills. That is, domain-specific problem solving is largely determined by knowledge in the respective area, but IPS might be a significant prerequisite for gaining this knowledge. In general, MicroDYN and MicroFIN exhibit good psychometric characteristics and promising results with regard to internal and external validity (for further studies see, for instance, Fischer et al., 2015; Frischkorn et al., 2014; Greiff and Wüstenberg, 2014; Wüstenberg, Greiff, Molnár and Funke, 2014). For a more comprehensive example of using MICS for empirical research, see Chapter 8.

Discussion

MicroDYN and MicroFIN focus on the use of minimal complex systems for the assessment of IPS. The existing body of empirical research supports this approach. This final section addresses three open questions. First, can IPS be separated conceptually and empirically from constructs such as intelligence, knowledge or working memory capacity? Second, are knowledge acquisition and knowledge application separate and yet central features of problem solving? And third, have minimal complex systems the same potential for assessing the construct of dealing with complexity as maximal complex systems?

Construct validity

Obviously, problem solving and intelligence are overlapping constructs and there are at least two positions claiming that they are considerably related. First, the ability to solve problems features prominently in almost any definition of human intelligence. Thus, problem-solving capacity is viewed as one component of intelligence (Funke and Frensch, 2007). However, for Dörner (1986), it is the other way around: he considers intelligence as operationalised in classical assessment instruments to be one of several components of problem solving, proposing a theory of operative intelligence combining both constructs on a conceptual level. Second, intelligence is often assumed to be an empirical predictor of problem-solving ability. In a recent meta-analysis, Stadler et al. (2015) corroborate this finding and report that across 47 studies the relation between problem solving and intelligence was moderate (with a mean correlation of 0.43). In summary, the available evidence suggests that the concepts of intelligence and problem solving are moderately related, whereas specific subcomponents of intelligence and problem solving might share additional variance. However, when predicting external criteria, problem solving explains some unique variance in

the criteria. The existing empirical evidence does not speak, however, to the issue of whether subcomponents of intelligence predict subcomponents of problem solving or whether the opposite causal relation holds. Overall, there is still more research needed on the connection between intelligence and problem solving (Wittmann and Süß, 1999) and we are confident that researchers will turn to this issue even more intensively in the future.

Knowledge acquisition and knowledge application

Why should we concentrate on these particular two dimensions as central elements of problem solving? The PISA 2012 framework on problem solving, arguably the largest student assessment worldwide, differentiates four groups of problem-solving processes (OECD, 2014): 1) exploring and understanding; 2) representing and formulating; 3) planning and executing; and 4) monitoring and reflecting. The first three we understand to be mostly equivalent to knowledge acquisition (which includes exploration of the problem situation), and knowledge application, with the main difference that our proposed terms are closer to overt behavior than the PISA terms. The fourth is a meta-component, which is difficult to assess and largely targeted by questionnaires (OECD, 2014; Wirth, 2004).

Another argument why the two dimensions of knowledge acquisition and knowledge application should be seen as central features of problem solving comes from studies with primates. The *Werkzeuggebrauch* (use of tools) of chimpanzees consists in exploring a certain device for later use in a problem situation. Even for humans, *intelligenter Werkzeuggebrauch* (intelligent use of tools) seems to be one of the oldest approaches to problem solving, which Jonassen (2003) now calls the use of cognitive tools (exploration and knowledge acquisition). Whereas nearly 100 years ago the chimpanzees of Wolfgang Köhler (1921) on the island of Tenerife used wooden sticks for reaching out to bananas, we nowadays use (electronic) devices to reach different goals in nearly every area of our life adhering to the application of knowledge. Thus, it may well be that the two main processes, knowledge acquisition and knowledge application, may be augmented by additional processes such as, a successful exploration of the problem as a prerequisite to acquire knowledge. Research will show, which of the competing theoretical conceptions draw the closest picture of the empirical reality (Fischer et al., 2012).

Minimal versus maximal complex systems

Certainly, a microworld such as Lohhausen (Dörner, 1980), with over 1 000 variables, provides a different experience for participants than a small-scale MicroDYN or MicroFIN scenario. The overwhelming amount of information in the case of Lohhausen, the high-stakes situation for the problem solver, the extremely knowledge-rich context and its associated manifold strategies make a substantial difference. For future research on long-term strategy development, large-scale scenarios will have their justification. On the other hand, the use of MicroDYN and MicroFIN tasks offer huge advantages for assessment: reliable measurement, varying difficulty and psychometric scalability, to name just a few. At the same time, the difficulty of the tasks as well as their semantic content can be easily adapted to different target populations, which helps to increase content validity. This flexibility, even though it is accompanied with a narrowing down of the processes and the conceptualisation targeted in the assessment, should encourage us to explore the potential gains of minimal complex systems more deeply.

No doubt, using MICS instead of MACS comes at a cost: the cost of narrowing interactive problem solving down to minimal problems and thereby somewhat impairing the measures' external validity. MACS assessment devices on the other hand are associated with an even larger impairment: the neglect of fundamental assessment principles (like a reference to a "best" solution). That is, when assessing interactive problem solving, a viable compromise needs to be found and considering theoretical and empirical evidence available today, the compromise offered by MICS is one worth while pursuing.

Notes

1   Please note that in the research literature the terms complex, dynamic, interactive and sometimes creative problems are used rather inconsistently and interchangeably. For the sake of clarity, we will consistently use the term interactive problem solving.
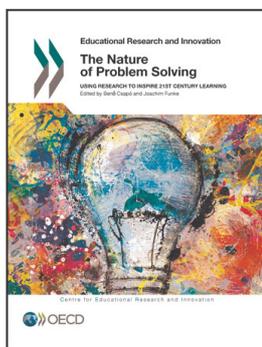
References

Abele, S. et al. (2012), "Dynamische Problemlösekompetenz: Ein bedeutsamer Prädiktor von Problemlöseleistungen in technischen Anforderungskontexten" [The importance of general cognitive determinants in mastering job demands. Some research on the example of dynamic and technical problem solving], *Zeitschrift für Erziehungswissenschaft*, Vol. 15/2, pp. 363-391.

Buchner, A. (1995), "Basic topics and approaches to the study of complex problem solving", in P.A. Frensch and J. Funke (eds.), *Complex Problem Solving: The European Perspective,* Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 27-63.

Buchner, A. and J. Funke (1993), "Finite-state automata: Dynamic task environments in problem-solving research", *Quarterly Journal of Experimental Psychology*, Vol. 46A/1, pp. 83-118.

Bühner, M., S. Kröner and M. Ziegler (2008), "Working memory, visual–spatial intelligence and their relationship to problem solving", *Intelligence*, Vol. 36/6, pp. 672-680.

Chi, M.T.H., R. Glaser and E. Rees (1982), "Expertise in problem solving", in R.J. Sternberg (ed.), *Advances in the Psychology of Human Intelligence, Volume 1*, Erlbaum, Hillsdale, NJ, pp. 7-75.

Dörner, D. (1996), *The Logic of Failure: Recognizing and Avoiding Error in Complex Situations*, Addison-Wesley, Reading, MA.

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

Dörner, D. (1980), "On the difficulties people have in dealing with complexity", *Simulation & Gaming*, Vol. 11, pp. 87-106.

Fischer, A., S. Greiff and J. Funke (2012), "The process of solving complex problems", *Journal of Problem Solving*, Vol. 4/1, pp. 19-42, http://dx.doi.org/10.7771/1932-6246.1118.

Fischer, A., S. Greiff, S. Wüstenberg, J. Fleischer, F. Buchwald and J. Funke (2015), "Assessing analytic and interactive aspects of problem solving competency", *Learning and Individual Differences*, Vol. 39, pp. 172-179, http://dx.doi.org/10.1016/j.lindif.2015.02.008.

Frischkorn, G., S. Greiff and S. Wüstenberg (2014), "The development of complex problem solving: A latent growth curve analysis", *Journal of Educational Psychology*, 106/4, pp. 1007-1020.

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

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

Funke, J. (1993), "Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results", in G. Strube and K.-F. Wender (eds.), *The Cognitive Psychology of Knowledge*, Elsevier Science Publishers, Amsterdm, pp. 313-330.

Funke, J. (1992), "Dealing with dynamic systems: Research strategy, diagnostic approach and experimental results", *German Journal of Psychology*, Vol. 16/1, pp. 24-43.

Funke, J. (1991), "Solving complex problems: Exploration and control of complex systems", in R.J. Sternberg and P.A. Frensch (eds.), *Complex Problem Solving: Principles and Mechanisms,* Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 185-222.

Funke, J. and P.A. Frensch (2007), "Complex problem solving: The European perspective – 10 years after", in D.H. Jonassen (ed.), *Learning to Solve Complex Scientific Problems*, Lawrence Erlbaum, New York, pp. 25-47.

Greiff, S. (2012), *Individualdiagnostik komplexer Problemlösefähigkeit* [Assessment of Complex Problem-Solving Ability], Waxmann, Munster.

Greiff, S., et al. (2015), Assessing complex problem solving skills with Multiple Complex Systems., *Thinking & Reasoning,*Vol. 21, pp. 356-382.

Greiff, S., et al. (2013), "A multitrait-multimethod study of assessment instruments for complex problem solving", *Intelligence*, Vol. 41/5, pp. 579-596.

Greiff, S. and J. Funke (2009), "Measuring complex problem solving: The MicroDYN approach", in F. Scheuermann and J. Björnsson (eds.), *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing,* Office for Official Publications of the European Communities, Luxembourg, pp. 157-163.

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

Greiff, S., M. Stadler, P. Sonnleitner, C. Wolff and R. Martin (2015), "Sometimes less is more: Comparing the validity of complex problem solving measures", *Intelligence*, Vol. 50, pp. 100-113.

Greiff, S. and S. Wüstenberg (2014), "Assessment with microworlds: Factor structure, invariance, and latent mean comparison of the MicroDYN test", *European Journal of Psychological Assessment*, Vol. 30/1, pp. 1-11.

Greiff, S., S. Wüstenberg and J. Funke (2012), "Dynamic problem solving: A new assessment perspective", *Applied Psychological Measurement*, Vol. 36/3, pp. 189-213. http://dx.doi.org/10.1177/0146621612439620.

Greiff, S., et al. (2013), "Complex problem solving in educational settings – Something beyond *g*: Concept, assessment, measurement invariance, and construct validity, *Journal of Educational Psychology*, Vol. 105/2, pp. 364-379, http://dx.doi.org/10.1037/a0031856.

Jonassen, D.H. (2003), "Using cognitive tools to represent problems", *Journal of Research on Technology in Education*, Vol. 35/3, pp. 362-381.

Köhler, W. (1921), *Intelligenzprüfungen an Menschenaffen* [Examining intelligence in primates], Springer, Berlin.

Kröner, S., J.L. Plass and D. Leutner (2005), "Intelligence assessment with computer simulations", *Intelligence*, Vol. 33/4, pp. 347-368, http://dx.doi.org/10.1016/j.intell.2005.03.002.

Mayer, R.E. and M.C. Wittrock (2006), "Problem solving", in P.A. Alexander and P.H. Winne (eds.), *Handbook of Educational Psychology*, Lawrence Erlbaum, Hillsdale, NJ, pp. 287-303.

Neubert, J.C., et al. (2015), "Extending the assessment of complex problem solving to finite state automata: Embracing heterogeneity", *European Journal of Psychological Assessment*, Vol. 31, pp. 181-194.

Novick, L.R. and M. Bassok (2005), "Problem solving", in K.J. Holyoak and R.G. Morrison (eds.), *The Cambridge Handbook of Thinking and Reasoning*, Cambridge University Press, pp. 321-349.

OECD (2014), *PISA 2012 Results: Creative Problem Solving (Volume V): Students' Skills in Tackling Real-Life Problems*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264208070-en.

Otto, J.H. and E.-D. Lantermann (2005), "Individual differences in emotional clarity and complex problem solving", *Imagination, Cognition and Personality*, Vol. 25/1, pp. 3-24.

Reeff, J.-P. and R. Martin (2008), "Use of the internet for the assessment of students' achievement", in J. Hartig, E. Klieme and D. Leutner (eds.), *Assessment of Competencies in Educational Settings*, Hogrefe & Huber, Göttingen.

Schweizer, F., S. Wüstenberg and S. Greiff (2013), "Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity", *Learning and Individual Differences*, Vol. 24, pp. 42-52.

Sloman, S. (2005), *Causal Models. How People Think About the World and its Alternatives*, Oxford University Press.

Sonnleitner, P. et al. (2012), "The *Genetics Lab*: Acceptance and psychometric characteristics of a computer-based microworld assessing complex problem solving", *Psychological Test and Assessment Modeling*, Vol. 54/1, pp. 54-72.

Stadler, M., N. Becker, M. Gödker, D. Leutner and S. Greiff (2015), "Complex problem solving and intelligence: A meta-analysis", *Intelligence*, Vol. 53, pp. 92-101.

Thimbleby, H. (2007), *Press On: Principles of Interaction Programming*, MIT Press, Cambridge, MA.

Wagener, D. (2001), *Psychologische Diagnostik mit komplexen Szenarios: Taxonomie, Entwicklung, Evaluation* [Psychological Assessment with Complex Scenarios: Taxonomy, Development, Evaluation], Pabst Science Publishers, Lengerich.

Wirth, J. (2004), *Selbstregulation von Lernprozessen* [Self-Regulation of Learning Processes], Waxmann, Münster.

Wittmann, W.W. and H.-M. Süß (1999), "Investigating the paths between working memory, intelligence, knowledge, and complex problem solving performances via Brunswik symmetry", in P.L. Ackerman, P.C. Kyllonen and R.D. Roberts (eds.), *Learning and Individual Differences: Process, Trait, and Content Determinants*, American Psychological Association, Washington, DC, pp. 77-108.

Wüstenberg, S., S. Greiff and J. Funke (2012), "Complex problem solving: More than reasoning?", *Intelligence*, Vol. 40/1, pp. 1-14, http://dx.doi.org/10.1016/j.intell.2011.11.003.

Wüstenberg, S., et al. (2014), "Cross-national gender differences in complex problem solving and their determinants", *Learning and Individual Differences*, Vol. 29, pp. 18-29, http://dx.doi.org/10.1016/j.lindif.2013.10.006.