Measuring Complex Problem Solving: The MicroDYN approach

Version 1.0 / 27.9.08

Samuel Greiff
Department of Psychology
University of Heidelberg
D-69117 Heidelberg, Germany
+49 6221 54 7613
samuel.greiff@psychologie.uni-heidelberg.de

Joachim Funke
Department of Psychology
University of Heidelberg
D-69117 Heidelberg, Germany
+49 6221 54 7305
Joachim.Funke@urz.uni-heidelberg.de

ABSTRACT

In educational large-scale assessments such as PISA only recently an increasing interest in measuring cross-curricular competencies can be observed. These are now discovered as valuable aspects of school achievement. *Complex problem solving* (CPS) describes an interesting construct for the diagnostics of domain-unspecific competencies. Here, we present MicroDYN, a new approach for computer-based assessment of CPS. We introduce the new concept, describe proper software and present first results. At last, we provide an outlook for further research and specify necessary steps to take in the effort to measure CPS on an individual level.

Keywords

Assessment, problem solving, competencies.

INTRODUCTION

Complex problem solving within dynamic systems has been an area of major interest in experimental research over the last decades (Frensch & Funke, 1995; Funke & Frensch, 2007). Comparatively little research has been conducted about complex problem solving in the context of individual differences even though some efforts have been made (e.g., Beckmann, 1994; Wagener, 2001). However, embedded in the recent development of large-scale assessments in educational settings, cross-curricular competencies such as complex problem solving have been discovered as valuable aspects of school achievement (Klieme, Leutner, & Wirth, 2005).

Starting from a practical point of view, applied implications of complex problem solving are frequently found in everyday life and involve situations comprising of the following characteristics:

- Different variables influence one or more outcomes,
- the underlying system is not static,
- exhaustive information and evaluation of the situation may not be obtained.

Many activities can be described within this formal framework ranging from medical emergencies over evaluating one's monthly expenses to handling ticket machines at train stations.

A first successful approach towards measuring CPS (CPS and dynamic problem solving are identical; we argue that complex problem solving is in itself always dynamic as opposed to analytical problem solving) in a large-scale context was conducted in PISA 1999 (Wirth & Funke, 2005). The finite automaton HEIFI embedded in the context of space travel could explain additional variance in student achievement after controlling for general intelligence. Furthermore, factor analytical results, structural equation models and multidimensional scaling suggested that CPS, analytical problem solving, domain specific literacy and general intelligence are correlated and yet separable constructs with CPS being best separable from the others (Wirth, Leutner, & Klieme, 2005).

These results indicate construct validity and in particular convergent and divergent validity for CPS. However, HEIFI was an ad hoc constructed instrument with questionable psychometric qualities so that measurement range and clear classification remains unclear calling for a properly piloted and validated testing device. A new approach is outlined in this paper and first empirical results are presented. Milestones on the way to measuring CPS are further specified.

THE MICRODYN APPROACH

Despite the awakening interest in individual differences, there is still a substantial lack of well-scrutinized testing devices. Additionally, little agreement on how to measure complex problem solving on an individual level has been reached and sound theoretical foundations to be used as starting points are still rare (Greiff & Funke, 2008b).

Another major shortcoming of complex problem-solving research as it was introduced by Dörner in the 1970s (Funke & Frensch, 2007) is its "one-item-testing". Virtually all devices consist of one large and rather complicated scenario the participant has to work through. At the end either overall performance or various status and process indicators are calculated and evaluated. Thus, CPS instruments are tests which contain exactly one excessive item or at best one bundle speaking in IRT-terms (Embretson & Reise, 2000) if various independent subsystems are considered as some authors do (e.g., Müller, 1993). Other tests allow subjects to explore a given system over a period of time and then ask several questions about this one system. That does not make the answers any less dependent.

Bearing these severe limitations in mind, the question arises how dynamic problem solving could possibly be measured with psychological tests. We assume that individual differences might possibly be detected within the formal framework of linear structural equation systems (LSE-systems), which we call the *MicroDYN approach*. This type of items has been used considerably in experimental research as indicators for problem solving performance (Blech & Funke, 2005). The basic approach, however, is now a different one as outlined below.

Items based on this approach require participants to detect causal relations and control the presented systems. We suppose that the everyday examples mentioned above can be modeled by MicroDYN systems since advanced skills in strategic planning, internal model building and system control are crucial in the specified situations as well as tested within the framework of MicroDYN systems. To solve the severe problem of one-item-testing, various completely independent systems are presented to the subjects (see below).

In summary, we choose to work within the formal framework of linear structural equation systems. The MicroDYN approach may be able to overcome some of the shortcomings mentioned above:

1. The lack of sound theoretical frameworks calls for a different kind of framework, which MicroDYN systems offer formally (theoretical embedment).
2. MicroDYN systems are easily constructed and can be varied in difficulty freely (infinite item pool).
3. A sufficient number of items can be presented (item independency).
4. Many everyday activities can be described by MicroDYN items (ecological validity).

The Items

An example of a typical MicroDYN item is presented in Figure 1. MicroDYN systems consist of exogenous variables, which influence endogenous variables, where only the former can be actively manipulated. Possible effects include main effects, multiple effects, multiple dependencies, autoregressive processes of first order, and side effects, which all can be freely combined.

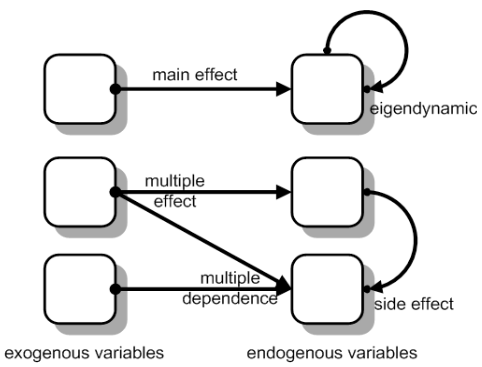

Figure 1: Underlying structure of a MicroDYN item with all possible effects displayed.

Main effects describe causal relations from exactly one exogenous variable to exactly one endogenous variable. If an exogenous variable is involved in more than one main effect, this is labeled a *multiple effect*. Effects on an endogenous variable influenced by more than one exogenous variable are labeled *multiple dependence*. Participants can actively control these three effects as they manipulate the values of *exogenous variables* within a given range. Effects merely incorporated within endogenous variables are called *side effects* when endogenous variables influence each other, and *autoregressive processes* when endogenous variables influence themselves (i.e. growth and shrinkage curves). Participants cannot influence these two effects directly, however, they are detectable by adequate use of strategy. Additionally, all effects may differ in path strength.

Participants face between 8 and 12 of these items each lasting about 6 minutes summing to an overall testing time of approximately one hour including instruction and trial time. The MicroDYN items are minimally but sufficiently complex and at the same time adequately in number. Each item is processed in three stages:

Stage 1, *exploration phase*: Participants can freely explore the system. No restrictions or goals are presented at this time. Participants can reset the system or undo their last steps. A history to trace prior steps is provided. Exploration strategies can thus be assessed.

Stage 2, *drawing the mental model*: Simultaneously (or subsequently) to their exploration, participants are asked to draw the connections between variables as they suppose. This helps in assessing acquired causal knowledge.

Stage 3, *control phase*: Participants are asked to reach given target values on the endogenous variables by entering adequate values for the exogenous variables. During this phase, the practical application of the acquired knowledge is assessed.

Current Research

Up to now little knowledge exits about how MicroDYN systems behave and which attributes cause their difficulty despite their extensive use in experimental research in the last decades. Based on a detailed task-analysis, seven factors are identified as potentially relevant for item difficulty (Table I). Testing these item-characteristics is understood as a first step to competence levels. The research design, first result and a brief discussion are provided below.

Design

We used a within-subject design (n=50) with repeated measures on all factors. An overall of 15 MicroDYN systems was presented, each lasting about 6 minutes (split on two sessions).

Dependent variables

Correctness of mental model: Subjects are asked to draw the connections between variables as they suppose. Better performance is indicated by a higher value on the dependent variable.

Table I: Attributes potentially determining difficulty in MicroDYN systems and their explanation.

	Attribute	Explanation of attribute
(1)	**Quality of effects**	different causal relationships (as depicted in figure above)
(2)	**Quantity of effects**	number of effects (regardless their quality)
(3)	Strength of paths	Specifies strength of an effect (and hence its detectability)
(4)	**Number of variables**	Mere number of exogenous and endogenous variables
(5)	Variable dispersion	Specifies how closely a given number of effects clusters on the variables
(6)	Effect configuration	Order and alignment
(7)	Starting & target values	Self-explaining; target values influence only end. variables

Control performance: After exploring the system extensively, subjects are asked to reach given target values on the endogenous variables as control task (results not yet available).

The factors mainly focused were Quality of effects, Quantity of effects and Number of variables.

Quality of effects: Main effects, multiple effects and side effects were tested against each other as can be seen in Figure 1 (multiple dependencies and eigendynamics were not tested at this stage).

Quantity of effects: Two different quantities (2 vs. 4 effects) were tested against each other. This is outlined schematically in Figure 2.

Number of variables: Systems were constructed equally only differing in number of variables as can be seen from Figure 3.

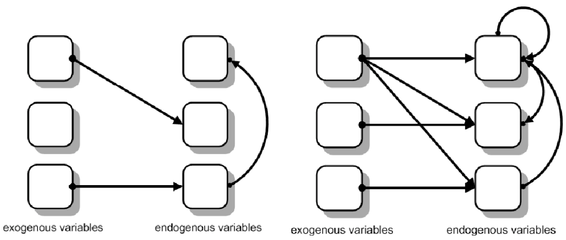

Figure 2: Two items with low resp. high number of effects.

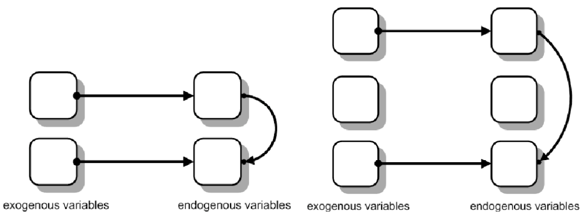

Figure 3: Two Items with 2 resp. 3 exogenous and endogenous variables.

Results

Table II provides an overview of the ANOVA-results. There is a medium strong effect for Number of variables showing that two systems being totally equal the one with more additional (and unnecessary) variables is more difficult. The explained variance is 0,158. A graphical depiction is found in Figure 4.

Table II: ANOVA results for the tested effects.

Independent variable	F	df_{Num}	df_{Denom}	p	Eta^2 (partial)
Number of ex. & end. Variables	*8,650*	*2*	*92*	*0,001***	*0,158*
Quality of effects	*18,270*	*2*	*90*	*0,001***	*0,289*
Quantity of effects	2,290	1	45	>0,10	0,048
Quality x Quantity	0,500	2	90	>0,05	0,011

There is a strong effect for Quality of effects showing that side effects increase difficulty heavily. Multiple effects and main effects do not vary significantly in the dependent variable (contrast not shown); however, multiple effects seem to be slightly easier. This might be due to participants' a priori expectation of a higher likelihood for multiple effects as these occur most frequently in real world settings. The explained variance is 0,289. A graphical depiction is found in Figure 5.

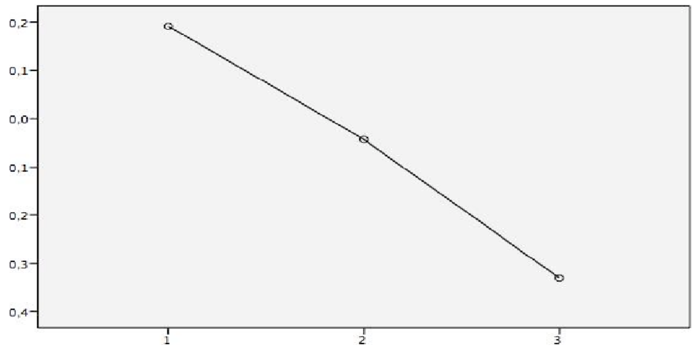

Figure 4: Number of variables. Ordinate: performance. Abscissa: Number of exogenous and endogenous variables (ranging from 2 to 4).

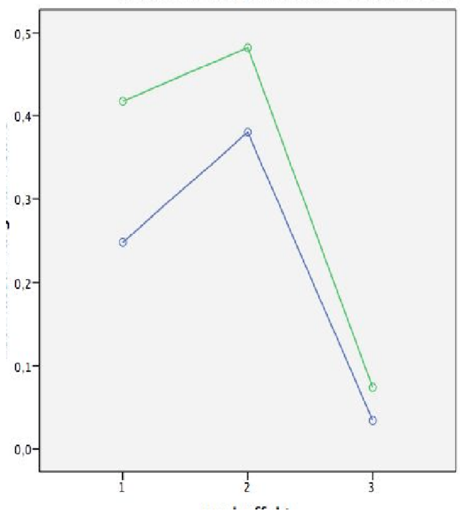

Figure 5: Quality and Quantity of effects. Ordinate: performance; Abscissa: Quality of effects (1=main effect, 2=multiple effect, 3=side effect); light line: 4 effects, dark line: 2 effects.

Surprisingly, items with only 2 effects are not easier than those having 4 effects. Apparently, the opposite might be true even though not statistically reliable. This unexpected result might be due to problems with the dependent variable we chose as outlined below. The explained variance is 0,048 and non-significant. A graphical depiction is found in Figure 5.

There is no interaction between Quality and Quantity of effects. Other interactions were not planned in the design.

Further screening of the data suggests the following effects:

- There is some evidence for problems with the dependent variable. The difference between correctly and incorrectly drawn connections in relation to the total number of correct connections was used and might pose some problems, which can be overcome by more complex indicators. Currently, a simulation study is carried out to decide which indicators represent problem-solving performance best.
- Correctness of mental model and control performance are weakly correlated (averaged r=0.15) suggesting that results might look differently for control performance.
- Subjects have considerable problems detecting side effects and tend to mistake them as two- to four-way multiple effects.
- There are only moderate training effects. As time passes, subjects perform slightly better. However, the training effect is less than half a standard deviation.

Implementation

The programming and development of the software is carried out in close cooperation with the DIPF (Frankfurt, Germany) and SOFTCON (Munich, Germany). The final version will leave considerable freedom to the researcher regarding graphical layout, semantics and item generation.

Currently, the software is in the process of development. It runs stable in a preliminary version. An authoring tool integrated in the open-access platform TAO (Plichart, Jadoul, Vandenabeele, & Latour, 2004; Reeff & Martin, in press) will be released late 2008/early 2009. An up-to-date screenshot is presented in Figure 6.

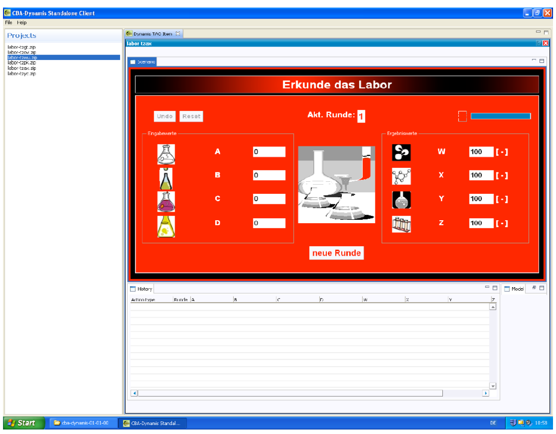

Figure 6: Screenshot of the MicroDYN software.

In the left panel loaded and ready-to-start items are displayed. The dark box is the actual item consisting of exogenous variables on the left and endogenous on the right. Additionally, an elapsed-time meter, a round counter, a reset and an undo-button are available. The history is placed at the page bottom. Here participants can trace their former manipulations and their effects for deeper analysis.

Perspective

Data acquisition for the first experiment finished some weeks ago. Data have been presented recently on two conferences (Greiff & Funke, 2008a, 2008b); in-depth analyses are currently carried out.

There is need for a follow-up study to learn more about item difficulty (i.e. multiple dependencies and eigendynamics have yet not been studied) in MicroDYN systems, which will start within the next weeks. Subsequently, explorative competence levels can be derived and tested in a pilot study. Simultaneously, the existing software is upgraded. The preliminary time schedule is graphically depicted in Figure 7.

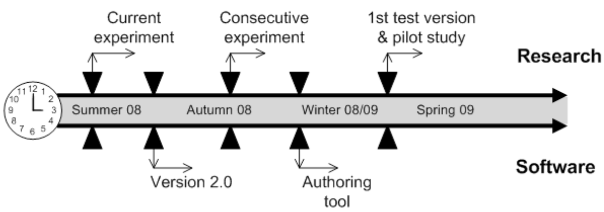

Figure 7: MicroDYN development: Preliminary time schedule until middle 2009.

Not yet incorporated are aspects of strategy and process data. By looking at the way subjects explore a system, different strategies can be identified and evaluated. This promising approach has been widely neglected in psychological diagnostics so far and is a promising field of enhancing prediction in achievement facets. First interesting ideas can be found in Rollett (2007).

The aim of the MicroDYN approach is to provide a well-scrutinized and empirically valid testing instrument for dynamic problem solving, which covers cognitive facets that yet cannot be tested by conventional tests of cognitive ability.

Applicability

If CPS can be nomothetically classified and established as a valid construct it might be relevant in virtually all areas involving prediction or explanation of cognitive performance.

In the context of educational large-scale assessments, a detailed analysis of factors determining difficulty as described yields important information for item construction and is a prerequisite for a formally and theoretically valid testing device for individual competence levels in complex problem solving.

MicroDYN might capture a construct yet not testable in cognitive psychology. Testing subjects on independent items in dynamic and interactive situations looking simultaneously at process and status data opens new doors in prediction of performance in various cognitive constructs such as student achievement.

ACKNOWLEDGEMENTS

The preparation of this paper was supported by grant Fu 173/11-1 from the German Research Foundation (DFG) in the Priority Programme "Models of Competencies for Assessment of Individual Learning Outcomes and the Evaluation of Educational Processes"(SPP 1293).

REFERENCES

Beckmann, J. F. (1994). *Lernen und komplexes Problemlösen. Ein Beitrag zur Konstruktvalidierung von Lerntests*. Bonn: Holos.

Blech, C., & Funke, J. (2005). Dynamis review: An overview about applications of the Dynamis approach in cognitive psychology. Bonn: Deutsches Institut für Erwachsenenbildung. Electronically available under http://www.die-bonn.de/esprid/dokumente/doc-2005/blech05_01.pdf

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

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

Funke, J., & Frensch, P. A. (2007). Complex problem solving: The European perspective - 10 years after. In D. H. Jonassen (Ed.), *Learning to solve complex scientific problems* (pp. 25-47). New York: Lawrence Erlbaum.

Greiff, S., & Funke, J. (2008a). *Schwierigkeiten in Problemlöseszenarien - Was ist das und was macht sie aus?* Paper presented at the AEPF, Kiel, 26th August 2008.

Greiff, S., & Funke, J. (2008b). What makes a problem complex? Factors determining difficulty in dynamic situations and implications for diagnosing complex problem solving competence. In J. Zumbach, N. Schwartz, T. Seufert & L. Kester (Eds.), *Beyond knowledge: the legacy of competence* (pp. 199-200). Wien: Springer.

Klieme, E., Leutner, D., & Wirth, J. (Eds.). (2005). *Problemlösekompetenz von Schülerinnen und Schülern*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Müller, H. (1993). *Komplexes Problemlösen: Reliabilität und Wissen*. Bonn: Holos.

Plichart, P., Jadoul, R., Vandenabeele, L., & Latour, T. (2004). *TAO, a collaborative distributed computer-based assessment framework built on Semantic Web standards.* Paper presented at the AISTA, Luxembourg.

Reeff, J.-P., & Martin, R. (in press). Use of the internet for the assessment of students' achievement. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of competencies in educational settings*. Göttingen: Hogrefe & Huber.

Rollett, W. (2007). *Strategieeinsatz, erzeugte Information und Informationsnutzung bei der Exploration und Steuerung komplexer dynamischer Systeme. Dissertationsschrift.* Braunschweig: Technische Universität Carolo-Wilhelmina.

Wagener, D. (2001). *Psychologische Diagnostik mit komplexen Szenarios. Taxonomie, Entwicklung, Evaluation*. Lengerich: Pabst Science Publishers.

Wirth, J., & Funke, J. (2005). Dynamisches Problemlösen: Entwicklung und Evaluation eines neuen Messverfahrens zum Steuern komplexer Systeme. In E. Klieme, D. Leutner & J. Wirth (Eds.), *Problemlösekompetenz von Schülerinnen und Schülern* (pp. 55-72). Wiesbaden: VS Verlag für Sozialwissenschaften.

Wirth, J., Leutner, D., & Klieme, E. (2005). Problemlösekompetenz - Ökonomisch und zugleich differenziert erfassbar? In E. Klieme, D. Leutner & J. Wirth (Eds.), *Problemlösekompetenz von Schülerinnen und Schülern* (pp. 73-82). Wiesbaden: VS Verlag für Sozialwissenschaften.