Jean-Paul Reeff, Anouk Zabal, Christine Blech

The Assessment of Problem-Solving Competencies

A draft version of a general framework

Deutsches Institut für Erwachsenenbildung

Online im Internet:

URL: http://www.die-bonn.de/esprid/dokumente/doc-2006/reeff06_01.pdf

Online veröffentlicht am: 31.07.2006

Stand Informationen: 15.07.2006

Dokument aus dem Internetservice texte.online des Deutschen Instituts für Erwachsenenbildung

http://www.die-bonn.de/publikationen/online-texte/index.asp

Dieses Dokument wird unter folgender creative commons-Lizenz veröffentlicht:

http://creativecommons.org/licenses/by-nc-nd/2.0/de/

Abstract

Jean-Paul Reeff, Anouk Zabal, Christine Blech (2006): The Assessment of Problem-Solving Competencies. A draft version of a general framework

This paper provides an outline for a general framework for the assessment of adults' problem-solving competencies with special attention paid to the requirements of large-scale assessments. The framework can be used as an input to ongoing discussions regarding international comparative studies of adult competencies such as the "Programme for an International Assessment of Adults' Competencies" (PIAAC) currently being prepared by the OECD, and may also be used in the context of national studies in this area. Problem solving is understood as a general dimension, strongly related to and partially overlapping with general mental abilities. The problem-solving abilities focused on within this framework clearly target those aspects that are modifiable by learning. The framework encompasses static analytical problem solving as well as dynamic problem solving, and a rationale for instrument development with the intent to yield a hierarchic scale. It emphasizes the importance of the context in which both the concepts as well as the instruments are embedded.

Authors

Jean-Paul Reeff is a Luxembourg-based independent consultant in education and training. He has a strong focus on assessment and evaluation, as well as on technology solutions for education.

Anouk Zabal is a researcher at the German Institute for Adult Education. She works in the area of educational research and assessment and has a background in cognitive psychology.

Christine Blech has recently finished her diploma in Psychology at the University of Heidelberg.

The Assessment of Problem-Solving Competencies

A draft version of a general framework

Jean-Paul Reeff, Anouk Zabal & Christine Blech

The work on this framework was sponsored by a grant of the German Federal Ministry for Education and Research (BMBF) to the German Institute for Adult Education (DIE). It was part of a feasibility study on adults' competencies assessment in the context of a national research strategy and the OECD „Programme for the International Assessment of Adults' Competencies" (PIAAC).

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This draft framework was the result of collaborative work of an international panel of problem solving experts. We would like to thank the following individuals for their contributions to this paper:

Problem Solving Expert Panel

Prof. Dr. Beno Csapo (Szeged University, Hungary)
Prof. Dr. Peter Frensch (Humboldt University, Germany)
Prof. Dr. Joachim Funke (Heidelberg University, Germany)
Prof. Dr. Detlev Leutner (Essen University, Germany)
Prof. Dr. Minoru Ohtani (Kanazawa University, Japan)
Dr. Joachim Wirth (Essen University, Germany)

Associated Problem Solving Experts

Prof. Dr. John Dossey (Emeritus Distinguished Professor of Mathematics, USA)
Prof. Dr. Eckhard Klieme (German Institute for International Educational Research)

Technology-based Assessment Experts and IT Experts

Dr. Thibaud Latour (CRP Henri Tudor Luxembourg)
Prof. Dr. Romain Martin (Luxembourg University)
Patrick Plichart (CRP Henri Tudor Luxembourg)
Prof. Dr. Alexander Repenning (University of Boulder, Colorado, USA)

Furthermore, our special thanks to the study director Dr. Dieter Gnahs (DIE) for initiating this effort and for his constant support, and Hella Huntemann and Norbert Lenartz (DIE) for their valuable assistance.

I. Introduction

Both educational research and research on human capital have in recent years increasingly focused on the assessment of basic skills and competencies. Although the measurement of competencies is well-established in psychological and educational research, it is only in the last decade that the large-scale assessment of competencies – especially in an international context – has become more prominent. Here substantial progress was made in the measurement of reading literacy and numeracy which are perceived as being fundamental competencies for employment and participation in society. There is a wide-spread consensus that these are but two of a number of crucial skills, one of them being problem-solving abilities. While there is abundant scientific work on problem-solving processes and their measurement, the large-scale assessment of problem-solving competencies is a relatively new endeavour.

This paper provides an outline for a general framework for the assessment of adults' problem-solving competencies with special attention paid to the requirements of large-scale assessments. The framework can be used both in the context of national studies on adults' competencies and as an input to ongoing discussions on international comparative studies of adults' competencies such as the "Programme for an International Assessment of Adults' Competencies" (PIAAC) currently being prepared by the OECD.

The development of this framework follows the rationale of the "draft framework" development for Information and Communication Technology (ICT) literacy in the context of the "Programme for International Student Assessment" (PISA) (Lennon, Kirsch, von Davier, Wagner & Yamamoto, 2003). Based on a first draft, a small group of international experts cooperated in order to define the essential parts of the framework in a common effort. The development work started off from the PISA and the "International Adult Literacy and Life Skills Survey" (ALL) problem-solving frameworks and expanded these systematically. This framework outlines the measurement domain and sketches possible item prototypes. Special emphasis is placed on a general rationale for technology-based assessment (TBA) which substantially extends the range of possible measurement instruments for problem solving in a large-scale perspective. This paper reflects the outcome of this process and can be regarded as a first step in a systematic development effort.

II. Overview

Policy papers and scientific documents consistently rank problem-solving abilities among the most important life skills. The fast pace of technological and societal change we are confronted with increasingly requires problem-solving skills at very different levels. Actually, it is obvious that pure knowledge-driven strategies for successfully coping with the complex challenges of our daily private and professional life will fail due to the sheer amount of available knowledge. Sound strategic knowledge and adequate problem-solving abilities are needed to organize complex information and to deal with potentially competing goals.

Very different communities acknowledge the importance of efficient problem-solving competencies. Labour market experts, human resources managers and vocational education and training experts point to such skills as essential key qualifications (Didi, Fay, Kloft & Vogt, 1993, Hunt, 1995, Binkley, Sternberg, Jones & Nohara, 1999, Baethge, Achtenhagen, Arends, Babic, Baethge-Kinsky & Weber, 2005). The "foundation skills", identified in the U.S. Department of Labour (1991), are essentially related to problem solving. Within an education context, both teachers and curriculum designers often ask for problem-solving oriented teaching and learning settings. On a large scale, the TIMSS study and above all the subsequent TIMSS-Video study shed light on the efficiency of a problem-solving oriented teaching and learning approach in mathematics and science (Stigler, Gallimore & Hiebert, 2000). It is therefore not surprising that problem solving plays a crucial role within PISA, where from the very beginning the discussions on cross-curricular competencies focused on problem solving. The assessment of problem-solving skills was implemented in the first PISA cycle in Germany (Klieme, Leutner & Wirth, 2005) and was an international option in the second PISA cycle (OECD, 2004a, OECD, 2004b). Moreover, a closer look at the frameworks and the instruments related to reading literacy, mathematical literacy and scientific literacy actually reveals a clear problem-solving centered rationale in all three domains. While the above mentioned large-scale surveys provide substantial information on students' problem-solving abilities, the situation in the field of adults' competencies is slightly different. The IALS (International Adult Literacy Survey) instruments are not as problem-solving oriented as the PISA reading literacy instruments. It is only the recent ALL study that provided more problem-solving oriented frameworks for prose and document literacy and numeracy. Moreover, ALL provides a direct assessment of problem solving as an international option (Murray, Clermont & Binkley, 2005; Statistics Canada/OECD, 2005).

Definition

A general and widely accepted definition of problem solving is (from Reeff, 1999, p. 48):
Problem solving is (goal directed) thinking and action in situations for which no routine solution procedure is available. The problem solver has a more or less well-defined goal, but does not immediately know how to reach it. The incongruence of goals and admissible operators constitutes a problem. The understanding of the problem situation and its step-by-step transformation, based on planning and reasoning, constitute the process of problem solving.

Scientific research and recent findings from large-scale surveys show that it is useful to distinguish between static problem situations and dynamic problem situations. In the first case, all problem-related information is available upfront, and the problem situation does not change over time. In the case of dynamic problems, the problem state changes and develops with time and through the respondents' actions. So actually, dynamic problems are much closer to typical everyday-like problem situations – in a dynamic problem situation some aspects of the problem may aggravate over time whereas others may simply disappear. Static problems are suitable for measuring analytical problem-solving skills; dynamic problems tap a mixture of analytical, complex and dynamic problem-solving abilities. It is obvious that some of these abilities are related to general mental abilities as described over decades in the scientific literature.

General structure

Based on these preliminary remarks we propose a general framework for measuring problem-solving competencies. Problem solving is understood as a general dimension, strongly related to and partially overlapping with general mental abilities. The problem-solving abilities focused on within this framework clearly target those aspects that are *modifiable by learnin*g.

The framework encompasses static analytical problem solving as well as dynamic problem solving, and a rationale for instrument development with the intent to yield a hierarchic (a priori-defined) scale. It emphasizes the importance of the context in which both the concepts as well as the instruments are embedded. For example, two important context dimensions of problem solving could be ICT-rich environments and social environments. This would provide the possibility to define ICT-Literacy (understood as an individual's problem-solving competencies in an ICT-rich environment) and collaborative problem solving (understood as individual problem-solving abilities in a social context) within the problem solving domain. Finally, such a framework structure has the potential for the future integration of other domains in an overarching framework for the assessment of competencies, with general mental abilities and problem-solving abilities as general dimensions.

<table>
<tr><td colspan="7">General Mental Abilities
- -
Problem-Solving Abilities (modifiable)</td><td rowspan="3">CONTEXT: Social context/ ICT-rich environment, etc.</td></tr>
<tr><td colspan="2">Reading Literacy</td><td>Numeracy</td><td colspan="2">Problem Solving</td><td>ICT Literacy</td><td>…</td></tr>
<tr><td>Prose Literacy</td><td>Document Literacy</td><td></td><td>Static/ Analytical Problem Solving</td><td>Dynamic Problem Solving</td><td></td><td></td></tr>
</table>

Figure 1: General structure for an integrated framework

As illustrated in Figure 1, such an approach would have the advantage of enabling us to tackle the measurement of problem-solving competencies in three different ways:

1. It is possible to directly assess problem-solving abilities by implementing static and dynamic problem-solving situations in a general context (frequently by using everyday-like situations), which doesn't require "domain-specific" a priori knowledge. This is basically the strategy followed in ALL for the area of analytical problem solving in static situations, as well as the strategy followed by PISA for both types of problem situations. Thus, problem solving here has the same status as the other domains.
2. The approach described above can be extended by translating it into specific contexts of special interest, such as social contexts or ICT-rich environments. In this case scales for collaborative problem solving and ICT literacy could be derived from specifically embedded problem-solving tasks.
3. On a more global level, the concept of general mental abilities and problem-solving abilities could serve as a general dimension for structuring and integrating different competency domains. This would enhance the conceptual coherence of the different competency domains and provide guidance towards a more economic assessment rationale and more effective instrument development strategies.

Static and dynamic problem solving

Both analytical problem solving in static situations and dynamic problem solving have their roots in decades of well-established basic research (Funke, 2003) and are quite well understood. As already mentioned, analytical problem solving was successfully implemented in PISA and ALL. The results of the dynamic problem-solving assessment that was introduced in the German national extension to PISA 2000 are also extremely promising. Actually, recent analyses of these data shed light on the relation between the different PISA domains and provide – beyond theoretical reasons – a data-based rationale for the chosen approach.

Two aspects should be emphasized:

1. Problem solving can be identified as a separate domain.
2. Within problem solving, dynamic problem solving can be clearly distinguished. Furthermore, it can also be discriminated both from intelligence and the other domains and sub-domains.

Together with the fact that dynamic problem solving is becoming increasingly important for everyday private and professional life, these results provide a strong argument for including dynamic problem solving in large-scales surveys of adult competencies such as PIAAC.

Beyond these results, first analyses of the ALL data show that there is a relation between analytical problem-solving skills on the one hand and wages and professional positions on the other hand (Clermont, 2005, personal communication). These results coincide with findings by Schmidt and Hunter (2004) that “GMA [General Mental Ability] correlates above .50 with later occupational level, performance in job training programs and performance on the job” (p. 171). The approach proposed here will enable researchers and policy makers to have even more accurate information on the distribution of problem-solving abilities and provide them with more accurate information on which to base political decisions and interventions (e.g. planning education and training programs, shaping life-long learning strategies, etc.).

Recommendations and perspectives

Based on scientific, logistical and political evidence, and having considered both the recurrent discussion of the relevant issues in PISA and ALL as well as diverse feasibility concerns, we recommend the following general strategy for contemporary large-scale programs on the assessment of adult competencies such as PIAAC or national assessment programs: Start with a combined approach of strategies 1 and 2 (see section “general structure” above) for a first wave. For this first wave only ICT-rich environments would be implemented. At an operational level, some of the conceptual work as well as instruments developed in the area of ICT literacy could be incorporated. Because of the importance of social contexts for the measurement of collaborative problem solving, we recommend focused research activities in this domain in order to develop potential instruments that could be then used in a second wave.

The third possibility mentioned above is more strategic in nature and points towards the possibility of using problem solving as an overarching dimension for the entire assessment. There may be design considerations that require reporting on one unifying scale, and problem solving as one of the most basic human cognitive processes is the natural candidate for this.

At an operational level, we propose to base instrument development efforts on two types of instruments that are well established and have been successfully implemented in large-scale surveys:

1. The analytical problem-solving assessment tools (e.g. “projects” as used in ALL)
2. Dynamic problem-solving assessment tools (e.g. DYNAMIS, finite state automata, as used in PISA Germany)

Although the analytical problem-solving items have to date been administered as a paper-and-pencil test, it is recommended to transfer these to a technology-based platform. This seems quite feasible, however it would be necessary to empirically verify how comparable the two versions actually are. Dynamic assessment tools are necessarily technology-based. Since the existing tools cover a large part of the relevant aspects of the domain and proved to function well within international large-scale assessments, and furthermore we anticipate no problems with using these tools with an adult population, we recommend to continue these development efforts and integrate these instruments into one coherent technology-based platform (e.g. Martin, Latour, Burton, Busana & Vandenabeele, 2005, Plichart, Jadoul, Vandenabeele & Latour, 2004). After these steps, other challenges related to test-based assessment such as adaptive testing could then be addressed.

III. Static analytical problem solving

Problem solving in large-scale assessments became a topic during the early OECD Network A discussions on assessing cross-curricular competencies. Early work by Trier, Peschar and van der Poele (OECD Network A) prepared the ground for further discussions while developing PISA. A European research network that worked on the development of new assessment tools for problem solving came up with first proposals on how to assess problem solving in large-scale surveys (Reeff, 1999). Starting from this pioneer work, the management of the "International Adult Literacy and Life Skills Survey" commissioned the further development of these tools for the ALL study. Within PISA, parallel work on implementing a problem-solving assessment in the second study cycle was pursued.

Static problem solving in ALL

ALL "aims at assessing a broad range of skills important in everyday life and essential for social, professional and economic success" (Murray et al., 2005, p. 196). Problem solving was considered one of the crucial life-skills. Logistical and financial considerations led the management to a paper-and-pencil assessment that resulted in a conceptual focus on analytical problem solving. The ALL problem solving approach, its implementation and some first results will be briefly described and taken as a starting point for a more general framework. Parts of this outline are taken from the ALL problem-solving framework (Reeff, Zabal & Klieme, 2005) that provides the full picture of the ALL analytical problem-solving approach.

The ALL framework is based on a very general definition of problem solving (Hunt, 1994; Mayer, 1992; Mayer & Wittrock, 1996; Smith, 1991) that was quoted in a previous section. It gives a very broad definition of problem solving as a *cognitive process* that underlies the transformation of a non-routine problem situation in order to reach a certain goal. More concretely, analytical problem solving is regarded as the core of such a goal-directed cognitive process. It encompasses the use of content-specific and general knowledge, rules and strategies, and meta-cognition. A person's analytical problem-solving competency may be indicated by his or her performance in identifying a problem, searching for relevant information and integrating it into a coherent problem representation, evaluating the problem situation with respect to given goals and criteria, devising a plan – i.e. an ordered sequence of appropriate actions – and monitoring its execution. Thus, analytical problem solving as it is defined here is closely related to reasoning ability and to the analytical subcomponent in Sternberg's triarchic theory of intelligence.

The cognitive processes that are activated in the course of problem solving are diverse and complex, and they are likely to be organized in a non-linear manner. Among these processes, the following components may be identified:

1. Searching for information, and structuring and integrating it into a mental representation of the problem ("situational model").
2. Reasoning, based on the situational model.
3. Planning actions and other solution steps.
4. Executing and evaluating solution steps.
5. Continuous processing of external information and feedback.

Baxter and Glaser (1997) present a similar list of cognitive activities labelled “general components of competence in problem solving”: problem representation, solution strategies, self-monitoring, and explanations. Analytical problem solving in everyday contexts, as measured by the ALL problem-solving instrument, mainly focuses on the components 1 to 3, and to a certain extent on 4.

The problem itself can be characterized by different aspects:

- The context can reflect different domains, which may be of a theoretical or a practical nature, related to academic situations or to the real world. Problems can have varying degrees of authenticity.
- The scope of a problem can range from working on limited, concrete parts of a task to planning and executing complex actions or evaluating multiple sequences of actions.
- The problem can have a well-defined or an ill-defined goal, it can have transparent (explicitly named) or intransparent constraints, and involve few independent elements or numerous interconnected ones. These features determine the complexity of the problem.

How familiar the context is to the target population, whether the problem involves concrete tasks or complex actions, how well the goal is defined, how transparent the constraints are, how many elements the problem solver has to take into account and how strongly they are interconnected – all these features will determine the level of problem-solving competency that is required to solve a certain problem. The empirical difficulty, i.e. the probability of giving a correct solution, will depend on the relation between these problem features and the subjects’ competency level.

The types of problems chosen for the ALL study were static in nature, with well-defined and transparent goals, moderately complex and with a relatively low degree of interconnectivity. The problems only required everyday a-priori knowledge, and were designed to be suitable for different cultures.

In order to implement analytical problem solving in ALL, the so-called project approach was chosen (cf. to the ALL framework for a motivation of this choice). In order to transform contextualized real-life problems into test items, the project approach put forth in the ALL framework used different problem-solving phases as a dimension along which to generate the actual test items. These steps in the process of problem solving have been frequently described as follows (cf. Polya, 1945/1980):

1. Define the goal.
2. Analyze the given situation and construct a mental representation.
3. Devise a strategy and plan the steps to be taken.
4. Execute the plan, including control and – if necessary – modification of the strategy.
5. Evaluate the result.

These stages correspond to the results of research on vocational training and job analyses within educational research and applied psychology that have been described as a part of the so-called “complete action” approach. Extensive analyses of very different jobs (different professions with varying types of work places) indicate that new forms of labour organization require people to perform more complex operations that go beyond mere routine. Nowadays, even production workers and office clerks are required to master complex tasks requiring integrative skills. Complete actions include different steps such as planning, executing and evaluating. The basic structure of the model of complete action is thus fully compatible with the above mentioned normative process model for problem solving – action steps are similar to problem-solving steps. The model of complete action has been successfully applied to curriculum development, assessment, and certification reforms in various professions in both Germany and Luxembourg (Hensgen & Blum, 1998; Hensgen & Klieme, 1998). The main idea is that both training tasks and also test problems should include all or most elements of a complete action.

The ALL project approach uses this complete action model to establish the underlying structure of the problem-solving test. The different action steps define the course of action for an "everyday" project. One or more tasks or items correspond to each of these action steps. The respondents thus work on the individual tasks that have been identified as steps that need to be carried out as a part of their project (e.g., "planning a family reunion"), and these tasks can vary in complexity. Embedding the individual tasks in an action context yields a high degree of context authenticity. Although they are all part of a comprehensive and coherent project, the individual tasks are designed so that they can be solved independently of one another. This is necessary in order to fulfil the requirements of the underlying measurement model.

The projects describe the problem situation, i.e. the project is introduced, the respondent's role is specified, and the required steps that need to be worked through as a part of the project are listed. The following example illustrates such a project (quoted from Reeff et al., 2005, p. 210-212):

The project is about "Planning a trip and a family reunion". In the introductory part of the project, the respondent is given the following summary describing the scenario and overall problem:

"Imagine that you live in City A. Your relatives are scattered throughout the country and you would like to organize a family reunion. The reunion will last 1 day. You decide to meet in City B, which is centrally located and accessible to all. Since you and your relatives love hiking, you decide to plan a long hike in a state park close to City B. You have agreed to be responsible for most of the organization."

The respondent is then given a list of steps he or she needs to work through, in this example the following list:

- *Set the date for the reunion.*
- *Consider your relatives' suggestions for the hike.*
- *Plan what needs to be done before booking your flight.*
- *Answer your relative's questions about traveling by plane.*
- *Book your flight.*
- *Make sure your ticket is correct.*
- *Plan the trip from City B to the airport.*

The first task of this project "Set the date for the reunion" is a good example of a typical problem-solving task and is now shown as it would appear in a test booklet.

Example task: Set the date for the reunion

! ***The family reunion should take place sometime in July.***

You asked all your relatives to tell you which dates would be suitable. After talking to them, you made a list of your relatives' appointments during the month of July. Your own appointment calendar is lying in front of you. You realize that some of your relatives will have to arrive a day early in order to attend the family reunion and will also only be able to return home on the day after the meeting.

Please look at the list of your relatives' appointments and your own appointment calendar.

List of your relatives' appointments in July 1999

Henry	Karen	Peter	Janet	Anne	Frank
Vacation in City E beginning on July 26; Appointment on July 11	Every day of the week is okay except Thursdays and on July 16	Business appointments on July 2, July 13, and between July 27 and 29	Doesn't have any appointments	Unable to attend reunion on July 5, July 20, or July 24	Has to be away sometime during the 1st full week in July on business, but will find out the exact dates shortly before

Henry, Karen, and Peter could arrive on the same day as the reunion whereas Janet, Anne, and Frank can only arrive on the afternoon before and return home on the day after the reunion.

Example task (cont.)

Your appointment calendar for July 1999

July 1999

Thurs.	1	Meeting with David
Fri.	2	
Sat.	3	
Sun.	4	
Mon.	5	
Tue.	6	
Wed.	7	
Thurs.	8	
Fri.	9	
Sat.	10	Hike in City C
Sun.	11	
Mon.	12	
Tue.	13	
Wed.	14	
Thurs.	15	
Fri.	16	
Sat	17	
Sun.	18	
Mon.	19	
Tue.	20	
Wed.	21	
Thurs	22	
Fri.	23	
Sat.	24	
Sun.	25	
Mon.	26	
Tue.	27	
Wed.	28	Vacation
Thurs	29	Vacation
Fri.	30	Vacation
Sat.	31	

Question 1. Which of the following dates are possible for the family reunion?

Please select all possible dates.

a	July 4
b	July 7
c	July 14
d	July 18
e	July 25
f	July 29

The example task gives a first indication of item structures and formats. The tasks typically start off with a short introduction to the situation, followed by varying types and amounts of information that need to be worked through. In the example task, in order to set the date for the family reunion, the respondent needs to process, compare and integrate the information provided in the list of the relatives' appointments, including the addendum to this list, and their own appointment calendar. Here the information is mostly textual and in form of tables. The answer format is a multiple choice format with more than one correct response alternatives, although the number of correct response alternative is not specified.

Proficiency levels

In ALL four levels of problem-solving proficiency were postulated and empirically validated:

1. Content-related reasoning
2. Evaluating
3. Ordering/Integrating
4. Critical Thinking

These four levels can be described as follows (quoted from Reeff et al., 2005, p. 202):

Level 1:
At a very elementary level, concrete, limited tasks can be mastered by applying content-related, practical reasoning. At this level, people will use specific content-related schemata to solve problems.

Level 2:
The second level requires at least rudimentary systematic reasoning. Problems at this level are characterized by well-defined, one-dimensional goals; they ask for the evaluation of certain alternatives with regard to transparent, explicitly stated constraints. At this level, people use concrete logical operations.

Level 3:
At the third level of problem-solving proficiency, people will be able to use formal operations (e.g. ordering) to integrate multidimensional or ill-defined goals, and to cope with non-transparent or multiple dependent constraints.

Level 4:
At the final and highest level of competency, people are capable of grasping a system of problem states and possible solutions as a whole. Thus, the consistency of certain criteria, the dependency among multiple sequences of actions and other "meta-features" of a problem situation may be considered systematically. Also, at this stage people are able to explain how and why they arrived at a certain solution. This level of problem-solving competency requires a kind of critical thinking and a certain amount of meta-cognition.

The results of the first assessment wave are reported on this scale (Statistics Canada/OECD, 2005). Thus the project approach to measure problem-solving competencies was successfully implemented in an international large-scale comparative study.

Problem solving in PISA

As there are only few and rather preliminary analyses available from the international ALL data set, we recur to data from the German national extension to PISA to better locate static, analytical problem solving as compared to other competencies as well as other aspects of problem solving. Germany, in PISA 2000, added an important national extension to the international core assessment and, within this extension, problem solving had a most prominent role. Different types of problem-solving instruments were used, among them computer-based instruments for dynamic problem solving. The manifold results of this study can be found in Klieme et al. (2005). Three major results are important here:

1. As in ALL, the project approach was successfully implemented within the PISA context.
2. The data show that problem solving does measure *different* competencies than those measured within the core of PISA.
3. Dynamic problem solving in particular is especially distinct and discriminates "new" aspects of problem solving.

Figure 2 shows the relationship between different competencies. The important information is the relative location of competencies to each other, and their distance to a general (intelligence) factor.

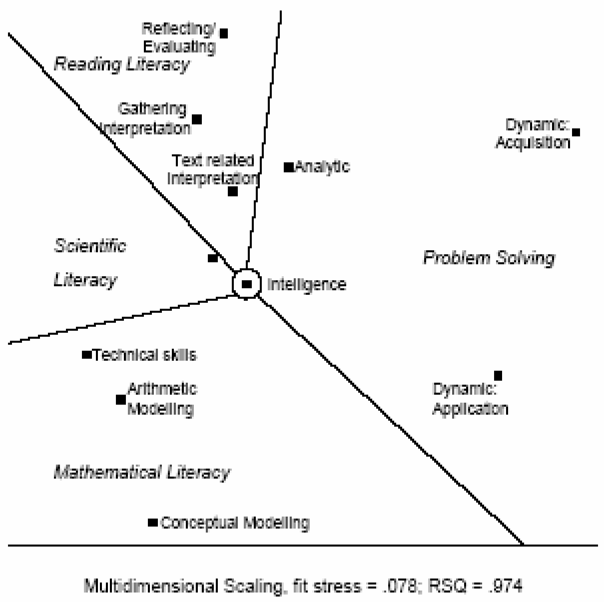

Figure 2: Empirical radex structure of PISA competency domains (from Klieme et al., 2005, p. 78, English translation)

Obviously, the dynamic aspect of problem solving would make a valuable contribution towards further completing a map of relevant competencies. This is consistent with results from long-term research on problem solving (Funke, 2003).

Therefore, we propose to systematically continue the development of the ALL problem-solving framework and instrumentation to enable the integration of dynamic problem solving. Within the paradigm of dynamic problem solving, several approaches and various instrument types have been used over the last 30 years. This framework describes the so-called DYNAMIS paradigm and related instruments in order to illustrate this approach. This does not preclude the use of other approaches like finite state automata or instruments developed for the assessment of ICT literacy.

IV. Dynamic problem solving

Parts of this section are taken from the more comprehensive DYNAMIS-Overview by Funke and Blech in annex A that provides a more extensive account of the DYNAMIS approach.

The use of computer-simulated scenarios in problem-solving research has become increasingly popular during the last 25 years (for a representative collection of papers see, e.g., the two editions from Sternberg & Frensch, 1991, and Frensch & Funke, 1995). This new approach to problem solving seems attractive for several reasons. In contrast to static problems, computer-simulated scenarios provide the unique opportunity to study human problem-solving and decision-making behaviour when the task environment changes concurrently to subjects' actions. Subjects can manipulate a specific scenario via a number of input variables (typically ranging from 2 to 20, in some exceptional instances even up to 2000), and they observe the changes in the system's state in a number of output variables. While exploring and/or controlling a system, subjects have to continuously acquire and use knowledge about the internal structure of the system.

Research on dynamic systems was motivated partly because traditional IQ tests turned out to be weak predictors in non-academic environments (see Rigas & Brehmer, 1999). Computer-simulated “microworlds” were thought to be “ecologically valid”. Simulations of (simplified) industrial production (e.g., Moray, Lootsteen, & Pajak, 1986), medical systems (e.g., Gardner & Berry, 1995), or political processes (e.g., Dörner, 1987) have the appeal of bringing "real world tasks" to the laboratory. Brehmer and Dörner (1993) argue that these scenarios avoid both laboratory limitations and the difficulties associated with field studies because the scenarios are relatively realistic while allowing for systematic control of influential factors.

The DYNAMIS approach

Several everyday activities require the regulation and control of relevant quantitative variables, for example regulating the car speed while driving, or controlling a CAD-machine. In many situations (e.g. technical, economical, ecological) it is necessary to really understand the system before goal-oriented action is at all possible. One scientifically popular way to represent system information is to use the general linear model, at least for systems with quantitative variables (cf. Stevens, 1992). The use of linear structural equation systems as a tool for problem-solving research was introduced by Funke (1985) under the name of DYNAMIS (which was the name of the first software shell for this type of simulations).

How can such a linear DYNAMIS model be used as a tool for analysing decision making and problem solving? Subjects are told that they have to deal with a system that consists of a given number of exogenous and endogenous variables. The exogenous variables can be directly manipulated by the respondent and they (can) indirectly influence the endogenous variables, which cannot be manipulated directly. The general task is (a) to find out how the exogenous and endogenous variables are related to each other, and (b) to control the variables in the system so that they reach certain goal values. Normally, these two subtasks, system identification and system control, are separated experimentally and consist of two separate steps within one task (see Funke, 1993). The basic structure of a simple linear DYNAMIS system with four variables is shown in Figure 3.

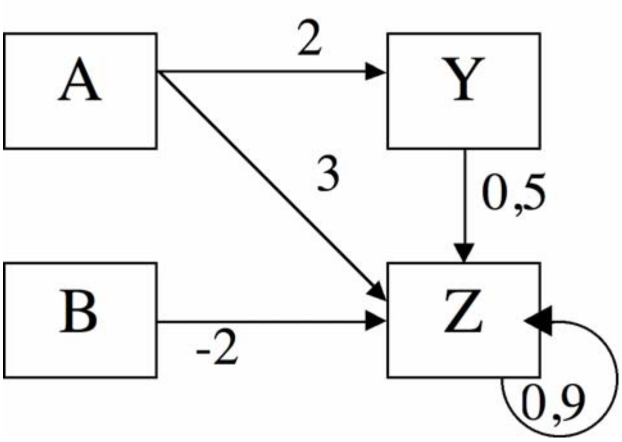

Figure 3: Structure of a simple linear DYNAMIS system with two input variables A and B as well as two output variables Y and Z. Variables are represented as boxes, connections between them are marked by weighted arrows. (Adapted from Vollmeyer & Funke, 1999, p.213)

In the system illustrated in Figure 3, the variables A and B represent the exogenous variables that have a direct effect on the endogenous variables Y and Z. The numbers on the arrows represent the weight with which the respective exogenous variables affect the endogenous ones. The system is described formally by two equations (one for each endogenous variable):

$$Y_{t+1} = 2 * A_t \qquad (1)$$
$$Z_{t+1} = 3 * A_t - 2 * B_t + 0{,}5 * Y_t + 0{,}9 * Z_t \qquad (2)$$

In these equations, the indices t and t+1 represent the actual state of the system; the system state can change in discrete time steps (= cycles).

Some systems also allow for “indirect effects” by permitting endogenous variables to influence each other (in Figure 3, the effect from Y to Z). However, such an effect can only be identified by manipulating the exogenous variable A. Variable A actually has two effects, a larger (“main effect” on Z) and a smaller one (“side effect” on Y). Also, the endogenous variables can influence themselves (see variable Z in Figure 3), thus representing something which is referred to as “eigendynamic” – this is reflected by a constant increase or decrease of this variable independent of other influences.

The status of the system variables is normally shown on the screen. Usually, the system’s history is also presented, i.e. the history of a given number of previous cycles. The structure of the system, however, is not displayed since it is explored during the first phase.

There are many possibilities to construct linear systems with a full range of effects of the kind described above, and identification and control of such DYNAMIS systems can actually prove to be quite difficult. The two main experimental tasks are knowledge acquisition and knowledge application.

Task Demand 1: Knowledge Acquisition

The term “knowledge acquisition” (system identification) describes a complex learning situation during which the subject has to find out details about the connectivity of the variables and their dynamics. The structural aspects of the system (= connectivity) cannot be easily separated from the dynamic aspects because the system itself can only be analysed interactively over time.

In the DYNAMIS situation, this identification problem requires an identification strategy, i.e. a certain way of manipulating the exogenous variables and analysing the consequences (in terms of values of the endogenous variables) in order to derive the causal structure of the system or at least to generate hypotheses about this structure that could be subsequently tested. Identification of system relations can occur at different levels: (a) as identification of the existence or non-existence of a relation, (b) as specification of a direction, (c) as specification of qualitative aspects of this (either positive or negative) relation, and (d) as the exact quantitative specification of the weight of this relation.

Task Demand 2: Knowledge Application

The term “knowledge application” (system control) refers to the application of previously acquired knowledge in order to reach a certain goal state within the system. The goal specifications are usually pre-defined by the task designer.

In the DYNAMIS situation, there are two sub-goals in knowledge application: First, to transform a given state of the endogenous variables (by means of an input vector into the vector of goal values), and second, to keep this goal state on a stable level since in a dynamic system the goal state – once reached – may actually disappear if “eigendynamics” come into effect.

Task characteristics

Using the DYNAMIS model as a starting point, various characteristics of the tasks that relate to the difficulty can be defined by the deep structure of the “system” used:

1.) $X_1 \rightarrow Y_1$	Direct simple effect: exogenous (controllable) variable X_1 affects endogenous (not controllable) variable Y_1
2.) $X_2 \rightarrow Y_1$ $\searrow Y_2$	Multiple effects: exogenous variable X_1 affects endogenous variables Y_1 and Y_2
3.) $X_2 \rightarrow Y_1$ $X_3 \nearrow$	Multiple dependencies: endogenous variable is determined/influenced by multiple exogenous variables
4) Eigendynamics	An endogenous variable changes by itself over time

5) Side effects	An endogenous variable is affected by another endogenous variable
6) Number of variables	The number of endogenous and exogenous variables; the controllability of the system may depend on the ratio of these two
7) Number of relations	Number of simple effects, number of multiple effects
8) Random components	“Adding noise” by adding an additive effect to an endogenous variable in a non-systematic way
9) Delayed effects	Time delay of effects; these effects blur or confuse the general picture
10) Time window size	Number of exploration cycles
11) Transparency	Amount of information explicitly provided about the structure of the system (can range from no information to a fully transparent structure)

The surface structure or the semantic embedding of the scenario is on a different level than the other task characteristics (which describe the deep task characteristics, see above). Semantic embedding can make problems harder or easier, but in an unpredictable manner. For example, elaborate domain knowledge may actually be an obstacle if the “system” doesn’t fit individual assumptions that are based on the problem solver’s individual knowledge base.

As previously mentioned, a person dealing with a complex dynamic system basically has to accomplish two tasks: First, they have to find out how the system works. Second, they have to reach and maintain a certain goal state of the system. In order to understand the processes involved in complex problem solving, it is necessary to focus on both: subjects’ performance in identifying the system (measure of knowledge acquisition), as well as on subjects’ control performance (measure of knowledge application). The following paragraphs illustrate how acquisition and application of knowledge depend on each other. Some indicators of the type and amount of gained knowledge as well as the quality of control performance will also be outlined (once again, a detailed description can be found in Blech & Funke in annex A).

Measuring acquisition of knowledge

Regarding the acquisition of knowledge, the far most common experimental indicator is the structural score derived from subjects’ causal diagrams (see, e.g., Funke, 1985). In studies by Beckmann (1994) or Schulz (2003), for example, the graphical format (arrows in diagrams) is either complemented or replaced by verbal and/or numerical elements. Beckmann recorded participants’ structural knowledge by means of sequential verbal questions after each experimental trial. Subjects were first asked very general questions such as whether they had found out anything new about any relation in the dynamic system. Depending on the subjects’ answers more specific questions followed, for example concerning the strength of the relation between two defined variables. The verbal answers were then translated into a graphical format, i.e. into arrows and symbols between exogenous and endogenous variables implemented in the system's graphical user's interface (see Beckmann, 1994).

This method of causal diagrams has been criticised due to its potential reactivity since this way of recording knowledge may actually draw subjects' attention to causal relations (see Kluwe, 1988). However, this method has two major advantages. First, causal diagrams allow differentiated analyses of three levels of system identification: the identification of the existence or non-existence of a relation, the identification of a relation's direction, and the identification of a quantitative weight indicating the strength and direction of a relation. Second, measures based on causal diagrams have proved to be highly reliable, especially the indicator of the quality of knowledge acquisition that was extensively validated by Müller (1993). Quality of knowledge acquisition is defined by the weighted difference of ratios of correct and false answers relative to the maximum number of correct and false answers (see also Funke, 1992).

Another approach to measuring structural knowledge is Preußler's "pair-task" (Preußler, 1996, 1997, 1998): Two variable names are presented to subjects, and the subject's task is to decide whether or not a relation between these variables exists. According to Preußler, the pair-task is considered less reactive than the method of causal diagrams. However, this method provides only limited information about the knowledge acquired, namely whether relations between certain variables exist or not.

Causal diagrams and the "pair-task" aim at recording abstract structural knowledge about causal relations in a system. It is also possible to examine acquired knowledge at a more concrete level, closer to the application of knowledge. Schoppek (2002) for example emphasises the distinction between structural knowledge and input-output knowledge that "represents specific input values together with the corresponding output values" (Schoppek, 2002, p. 64). Accordingly, in order to consider this type of specific knowledge as well, subjects may be asked to predict the resulting outcome states given certain input states and interventions. The measure for the "quality of prediction" is calculated from the subject's predictions and the actually resulting outcome states (see Funke, 1992, Funke & Müller, 1988). Another way to infer input-output knowledge is to present specific system's states which either correspond to dynamic situations that actually occurred, i.e. target situations, or to additionally constructed distractor situations which resemble target situations but have never been encountered by subjects (Preußler, 2001). Subjects are requested to assess whether a presented situation is "old", i.e. a target, or whether it is completely "new", i.e. a distractor. Based on the correct and false recognition answers, knowledge scores can be calculated.

Even though the latter measures could be regarded as less reactive than causal diagrams, problem solvers confronted with these tasks are conscious that they are being asked about their acquired explicit structural knowledge. There are also ways of measuring the implicit knowledge that avoid this problem by using a lexical decision task (cp. Preußler, 1996). The relevant measure here is based on reaction times.

More recently, Wirth (2004, 2005) proposed more elaborate measures for describing the knowledge acquisition phase. His $\log_{(or)}$ measure describes, for defined time intervals, whether the problem solver focuses more strongly on the identification or on the integration of information.

Measuring control performance

Recent experimental studies generally measure control performance via the quality of system control. This measure calculates mean logarithm deviations of the system's observed states from the defined goal states (see, e.g., Müller, 1993). Müller's validating analyses showed that the quality of system control is just as reliable as the quality of system identification. A measure found in older studies is the mean absolute deviations from goal values (e.g., Funke & Müller, 1988). In one case (Görn, Vollmeyer & Rheinberg, 2001) a very simple indicator of control performance was used, namely the number of accomplished goals.

This list is far from being exhaustive, but illustrates the wide variety of options from which adequate measures for dynamic problem solving can be chosen. The final choice will have to take into account overall design constraints and relations with other competency domains included in the direct assessment.

V. Specific contexts

The ALL approach aimed at choosing everyday contexts for the problem-solving scenarios that required little previous knowledge. It seems useful to recommend a similar approach for dynamic problem-solving scenarios, in order to disentangle problem-solving abilities from knowledge-based decision-making. Nevertheless there are reasons to briefly discuss some specific contexts, both in the light of a short-term implementation and a mid-term development process. Social contexts have a specific role in collaborative problem solving. ICT-rich environments may serve to emphasize an ICT literacy dimension within a problem-solving framework.

Social Contexts

Collaborative problem solving is considered one of the crucial problem-solving aspects in today's life. The concept of collaborative problem solving is discussed in various ways in the scientific literature, but there are only few efforts to develop or implement a large-scale assessment of collaborative problem-solving competencies (e.g. Kunter, Stanat & Klieme, 2005). The expert panel working on this framework proposes to focus on collaborative problem solving as *an individual's ability to solve static and dynamic problems in a social context*. Such a preliminary definition lays the ground for integrating collaborative problem solving into the general framework of an assessment of adults' competencies based on individual data. At an instrument design level, dynamic problem-solving scenarios could easily be constructed to include real or simulated "actions" of other individuals. While instrument *construction* is likely to be rather straightforward, it will be much more difficult to develop related *measures* that fulfil the standards of a large-scale comparative study. In order to do this we recommend doing substantial research and development work on these issues in the next few years.

ICT-rich Environments

The discussion on ICT literacy and its conceptual link with problem solving has not yet come to an end. At a practical level it seems easy to include problem-solving components into traditional ICT literacy tasks. Nevertheless, at this stage it is not clear how ICT literacy could serve as a unifying dimension to integrate important aspects of problem solving. Integrating a meaningful ICT literacy component into a problem-solving framework seems the more feasible and expandable alternative in a long-term program on the assessment of adults' competencies. We propose a two-step procedure to cope with both short-term requirements as well as a sustainable long-term development. A first step would be to emphasize ICT-rich environments when constructing – static and dynamic – problem-solving scenarios as described above. A second step could then be to use problem solving as an overarching dimension to fully integrate other competency domains (cf. next chapter).

Some examples illustrate how ICT-rich environments can be used in constructing problem-solving scenarios:

1. At the lowest level of ICT literacy (or rather computer familiarity), basic tools can be added to traditional problem-solving scenarios in order to facilitate the

problem-solving process. Notepads, spreadsheets or elementary databases could be provided as additional tools to the subject solving the problem.

2. A static problem-solving scenario like “Apartment hunting” could be transposed into a web environment where active web search is required to solve the problem. Within the “Family Reunion” scenario as illustrated above, coordination and communication between family members could be done via e-mail or sms.
3. Dynamic problem-solving scenarios are particularly well suited to cope with ICT-rich environments. They can easily tap this kind of environment as illustrated for example by Finite State Automata (FSA). FSA are by definition ICT “objects”: Handling simulated ticket machines, video recorders, ATMs or other devices brings us directly into the ICT environments we are looking for. Furthermore, many web tasks can be easily decomposed into an FSA scenario. Dynamic tasks like buying something at an e-auction easily fit into the overall framework of dynamic problem solving.
4. Finally, at the highest level of competency, simulation tools can be used to better understand a problem-solving situation and to design more efficient solutions for complex problems.

The basic principle in this approach would be to use the construction rationale of problem-solving instruments in order to derive a scale for problem solving AND information on ICT literacy. This would be an intermediate step as compared to the more far-reaching ambition of the next section.

VI. Problem solving as an overarching dimension

The ALL experience shows that static analytical problem solving as a separate dimension can be successfully implemented within an international large-scale comparative study. Similar evidence comes from national studies and smaller projects. The German PISA 2000 extension shows that beyond static analytical problem solving, dynamic problem solving can also be implemented successfully, as was to be expected from previous basic research. The results also show that there are very good reasons to focus on these dynamic aspects in future assessments. These results and the overview given in the previous chapters clearly indicate that a theoretically and empirically well-founded implementation of a problem-solving assessment as a separate domain is possible for a large-scale assessment of adult competencies such as for example PIAAC.

The previous chapter outlines how one could expand such a traditional approach to other contexts. It also shows how problem solving and ICT literacy could be conceptually linked. This latter rationale can be expanded to other domains.

A closer look at the PISA and ALL frameworks reveal that all existing frameworks (reading literacy, mathematical literacy, scientific literacy, prose and document literacy and numeracy) are clearly problem-solving centered. This even holds for the tentative ALL frameworks for teamwork and practical cognition. The available documents on ICT literacy for PISA and PIAAC do not only show aspects of problem solving, but actually, they use the problem-solving terminology for very essential aspects in their frameworks. Therefore, an effort should be made to conceptually link or even integrate these different domains. While many different stakeholders have requested such an effort from a political point of view, recent empirical results support the idea to base this effort on a problem-solving dimension. Actually, beyond the mere fact that other frameworks contain substantial problem-solving elements at a conceptual level, the PISA Germany 2000 results and basic research provide reasons for both having a separate problem-solving assessment AND a general dimension at the centre of different competencies (cp. Figure 2 on page 18).

Being aware that the assessments of different domains include substantial problem-solving elements, Figure 2 shows a specific position of dynamic problem solving as related to other competencies and a central position of “intelligence” or a “general mental ability” at the very heart of all competencies. A discussion of the relation between such general factors and problem solving is far beyond the scope of this paper. However, based on scientific research, the expert panel, recommends an approach that considers problem-solving competencies as the modifiable part of the underlying dimension “general mental ability”. Figure 1 on page 9 shows this situation and illustrates the different possible approaches that can be derived.

The column under “Problem-Solving Abilities (modifiable)” illustrates the traditional approaches from PISA and ALL: Different separate assessment domains, to be implemented following political and scientific considerations and priorities. The shaded problem-solving part links to the most elementary aspect of this framework, i.e. the recommendation to focus on static and dynamic problem solving to be implemented as a separate problem-solving assessment.

Embedding this approach in an ICT-rich environment shows the potential of linking a problem-solving dimension with other domains. It is possible to generalize this integrative idea to all other competencies. Based on a thorough analysis of the problem-solving elements in other frameworks, and on reflections concerning a separate problem-solving assessment as described earlier in this framework, a problem-solving rationale could be systematically expanded into different domains. Domains would then be considered as contexts that require both general problem-solving abilities AND domain-specific knowledge and competencies.

The expert panel recommends starting off with a separate assessment of static and dynamic problem solving, optionally in ICT-rich environments. The panel also recommends to invest substantial efforts in the overall integrative potential of problem solving as one unifying dimension and to lay the ground for a more comprehensive and sustainable reporting on adults' competencies. If design considerations require reporting on one unifying scale, both basic research and experience with different competency domains make problem solving a promising candidate for such an ambition.

VII. Technology-based assessment

The assessment of dynamic problem solving requires the use of computer-based scenarios. In the experimental studies that triggered the basic research in the field of problem-solving research during the last 30 years, the technology challenges could be overcome rather quickly. In the context of a large-scale study, these problems become much more important and diverse. In parallel to developing this framework, a substantial effort has been made to provide sustainable solutions for a technology-based assessment in general and more specifically for the technology-based assessment of problem-solving competencies. This effort yielded results at three different levels.

First, we developed a preliminary prototype authoring tool for DYNAMIS-like scenarios. The main goal here was to give non-programmers a tool that they can use to quickly develop dynamic problem situations for assessments. The authoring tool generates a parameter file to be read by a prototype simulation engine that executes the scenario. The engine can also be used to test most of the simulations mentioned in the overview by Blech & Funke in annex A.

Furthermore, the framework team wanted to come up with a sustainable solution for generating a large variety of computer-based problem-solving scenarios. Such a solution would also foster the collaborative development of instruments as it is typically required in large-scale national or international studies. The solution proposed in this framework is based on long-term research and development work by Alexander Repenning from the University of Boulder, Colorado. It has its roots in the work on multi-agent systems and was commercialized subsequent to a project supported by the U.S. National Science Foundation. *"AgentSheet" introduced the use of graphical rewrite rules to program the behavior of agents. As a form of end-user programming graphical rewrite rules are pairs of before/ after pictures edited through demonstration by users. For instance, the behavior of a train agent to follow train tracks is programmed by simply demonstrating an example of how the train moves to the right. AgentSheets evolved and introduced additional end-user programming paradigms including programming by analogous examples, and tactile programming. (Repenning, n.d.)* There are several benefits related to this solution, two important practical ones being:

1) All major problem-solving instruments can be easily constructed with this tool. It even offers options that go far beyond traditional instrument design and make full use of agent-based systems.
2) The design of instruments can be carried out by non-programmers. Thus it gives assessment experts and content experts the opportunity to design first outlines of assessment instruments without having to involve programmers. This allows to replicate the successful collaborative item development process that we know e.g. from PISA.

More information on AgentSheets and its context can be found under:

http://www.cs.colorado.edu/~ralex
http://www.cs.colorado.edu/~ralex/Portfolio.pdf
http://agentsheets.com/

At this stage we cannot yet give a final recommendation for AgentSheets, but it should be clear that the product fits the needs we have identified so far, and that it should at least serve as a benchmark for choosing another product.

The user-friendly and flexible authoring of computer-based instruments will be crucial for the success of any large-scale effort in technology-based assessment. Nevertheless, even the best solutions to this problem will not serve to overcome a series of other crucial challenges in the whole process of a technology-based assessment. As a consequence we propose an overarching generic architecture for computer-based assessment that covers all relevant aspects of an assessment. The solution proposed goes back to research work done by the University of Luxembourg and the Luxembourg “Centre de Recherche Public Henri Tudor”. The architecture aims at laying the ground for a very generic technology-based assessment platform that has already been implemented under the name of TAO (the French acronym for “computer-based assessment”). TAO is by now a well-validated research prototype that is currently being reengineered to meet industrial standards. TAO is implemented as an open-source product. This will give full control and ownership of the product to participants in a large-scale survey when it comes to the concrete implementation of an assessment, and thus secures participants’ (country) investments.

Figure 4 shows the global TAO architecture (from Plichart et al., 2004, p. 3):

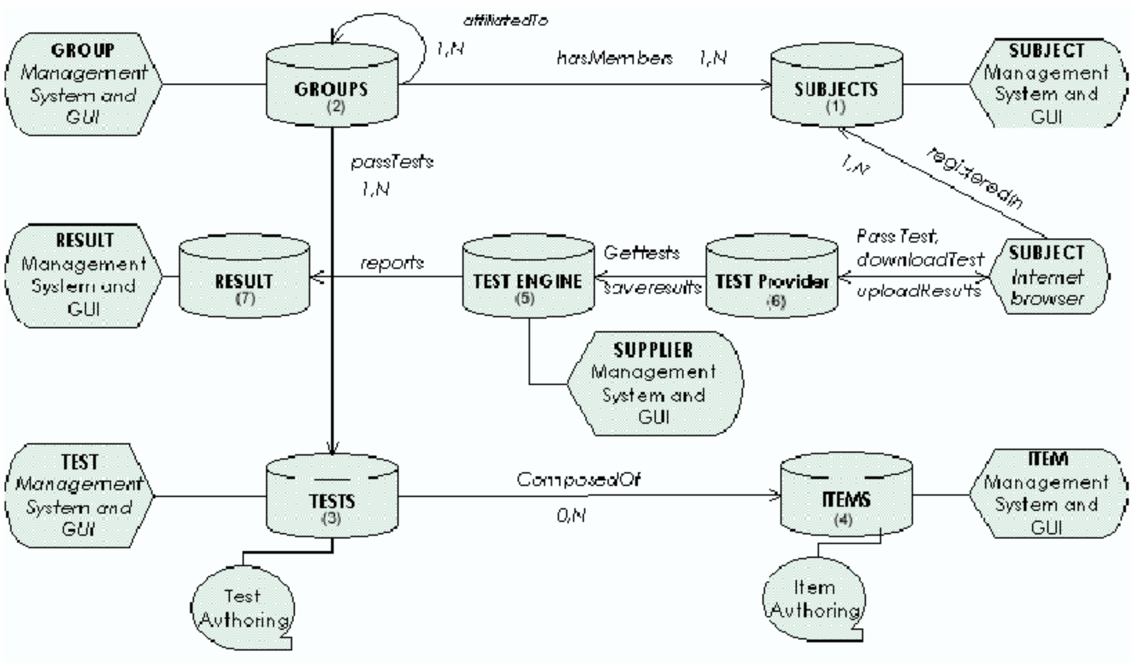

Figure 4: TAO architecture

Plichart et al. (2004) describe the architecture in the following way (pp. 2-3):
Five different fundamental sub-domains can be distinguished in the global testing problem, corresponding to five independent data domains. Following this view, the platform has been split into five types of specialized modules (fig. 1). Each of these

specialized modules covers a specific knowledge domain. The data management modules implement two fundamental functionalities: i) modelling the domain by creating hierarchies of categories characterised by a set of properties, and ii) populating the model by instance data. In the TAO framework, the five different modules are dedicated to the management of i) subjects (1) who are expected to pass tests, ii) groups of subjects (2) that are assigned one or more tests, iii) tests (3) consisting in a collection of items, iv) items (4) roughly corresponding to questions, and v) results (7). In addition to these modules, the test engine (5) is in charge of collecting all the necessary data distributed over all modules in order to assemble and to deliver a test. The test is delivered to an identified subject through the test provider portal (6). The test resources are assembled automatically according to metadata associated to the set of resources that describe all the test components. Once the test has been passed by the subject, all results that have been collected during the execution are sent back to the result module (7) where they can be manipulated for further analysis. All modules are connected as a Peer-To-Peer network.

It should be noted that this architecture provides means both for internet-based testing and offline testing, as well as for all variants of local network solutions. A full description of TAO is beyond the scope of this framework and can be found in Martin et al. (2005) or Plichart et al. (2004).

Beyond the reengineering of the TAO platform, a further important step will be to integrate AgentSheets with TAO.

VIII. Conclusion

Based on results from basic research and first experiences with the assessment of problem-solving competencies in large-scale surveys (ALL, PISA) this framework provides a description of a more general framework for the assessment of problem-solving competencies. They can either be assessed as a completely separate dimension or by integrating another domain such as ICT literacy. Furthermore, the perspective of problem solving as an overarching dimension is outlined. It is proposed to implement the assessment of problem-solving competencies within a general architecture for technology-based assessment.

IX. References

Baethge, M./Achtenhagen, F./Arends, L./Babic, E./Baethge-Kinsky, V./Weber, S. (2005): Wie könnte eine internationale Vergleichsstudie zur beruflichen Bildung aussehen? Göttingen

Baxter, G. P./Glaser, R. (1997): An Approach to Analyzing the Cognitive Complexity of Science Performance Assessments. Los Angeles

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

Binkley, M. R./Sternberg, R./Jones, S./Nohara, D. (1999): An Overarching Framework for Understanding and Assessing Life Skills. Unpublished International Life Skills Survey (ILSS) Frameworks. Washington

Blech, Christine/Funke, Joachim (2005): Dynamis review: An overview about applications of the Dynamis approach in cognitive psychology.
URL (31.07.2006): http://www.die-bonn.de/esprid/dokumente/doc-2005/blech05_01.pdf

Brehmer, B./Dörner, D. (1993): Experiments with Computer-Simulated Microworlds: Escaping Both the Narrow Straits of the Laboratory and the Deep Blue Sea of the Field Study. In: Computers in Human Behavior, 9, pp. 171–184

Didi, H.-J./Fay, E./Kloft, C./Vogt, H. (1993): Einschätzung von Schlüsselqualifikationen aus psychologischer Perspektive. Gutachten im Auftrag des Bundesinstituts für Berufsbildung (BIBB). Bonn: IBF – Institut für Bildungsforschung

Dörner, D. (1987): On the Difficulties People have in Dealing with Complexity. In: Rasmussen, J./Duncan, K./Leplat, J. (eds.): New Technology and Human Error. Chichester, pp. 97–109

Frensch, P. A./Funke, J. (eds.) (1995): Complex Problem Solving: The European Perspective. Hillsdale, New Jersey

Funke, J. (1985): Steuerung dynamischer Systeme durch Aufbau und Anwendung subjektiver Kausalmodelle. In: Zeitschrift für Psychologie, 193, pp. 435–457

Funke, J. (1992): Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung. Berlin

Funke, J. (1993): Microworlds Based on Linear Equation Systems: A New Approach to Complex Problem Solving and Experimental Results. In: Strube, G./Wender, K.-F. (eds.): The Cognitive Psychology of Knowledge. Amsterdam, pp. 313–330

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

Funke, J./Müller, H. (1988): Eingreifen und Prognostizieren als Determinanten von Systemidentifikation und Systemsteuerung. In: Sprache & Kognition, 7, pp. 176–186

Gardner, P. H./Berry, D. C. (1995): The Effect of Different Forms of Advice on the Control of a Simulated Complex System. In: Applied Cognitive Psychology, 9, pp. 55–79

Görn, A./Vollmeyer, R./Rheinberg, F. (2001): Effects of Motivational Orientation and

Perceived Ability on Performance. In: Psychologische Beiträge, 43, pp. 153–171

Hensgen, A./Blum, F. (1998): Prüfen beruflicher Handlungsfähigkeit. Das KoPrA-Konzept. In: Zeitgemäß ausbilden – zeitgemäß prüfen. Theorie und Praxis handlungsorientierter Ausbildung und Prüfung im kaufmännischen Bereich. Bielefeld, pp. 37–53

Hensgen, A./Klieme, E. (1998): Wie bewähren sich die neuen Aufgaben zur Erfassung von Handlungskompetenz in den schriftlichen Zwischen- und Abschlussprüfungen für Bürokaufleute? In: Zeitgemäß ausbilden – zeitgemäß prüfen. Theorie und Praxis handlungsorientierter Ausbildung und Prüfung im kaufmännischen Bereich. Bielefeld, pp. 89–103

Hunt, E. (1994): Problem Solving. In: Sternberg, R. J. (ed.): Thinking and Problem Solving. San Diego, pp. 215–232

Hunt, E. (1995): Will we be Smart Enough? A Cognitive Analysis of the Coming Workforce. New York

Klieme, E./Leutner, D./Wirth, J. (eds.) (2005): Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie. Wiesbaden

Kluwe, R. H. (1988): Methoden der Psychologie zur Gewinnung von Daten über menschliches Wissen. In: Mandl, H./Spada, H. (eds.): Wissenspsychologie. München, pp. 359–385

Kunter, M./Stanat, P./Klieme, E. (2005): Die Rolle von individuellen Eingangsvoraussetzungen und Gruppenmerkmalen beim kooperativen Lösen eines Problems. In: Klieme, E./Leutner, D./Wirth, J. (eds.): Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie. Wiesbaden, pp. 99–115

Lennon, M./Kirsch, I./Davier, M. von/Wagner, M./Yamamoto, K. (2003): Feasibility Study for the PISA ICT Literacy Assessment. Educational Testing Services. Princeton

Martin, R./Latour, T./Burton, R./Busana, G./Vandenabeele, L. (2005): Covering Different Levels of Evaluation Needs by an Internet-Based Computer-Assisted Testing Framework for Collaborative Distributed Test Development and Delivery. URL (31.07.2006): www.tao.lu/downloads/publications/Edmedia2005b.pdf

Mayer, R. E. (1992): Thinking, Problem Solving, Cognition. 2nd edition. New York

Mayer R. E./Wittrock, M. C. (1996): Problem-Solving Transfer. In: Berliner, D. C./Calfee, R. C. (eds.): Handbook of Educational Psychology. New York, pp. 47–62

Moray, N./Lootsteen, P./Pajak, J. (1986): Acquisition of Process Control Skills. In: IEEE Transactions on Systems, Man, and Cybernetics, 16, pp. 497–504

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

Murray, T. S./Clermont, Y./Binkley, M. (2005): International Adult Literacy Survey.

Measuring Adult Literacy and Life Skills: New Frameworks for Assessment. Ottawa

OECD (2004a): Learning for Tomorrow's World: First Results from PISA 2003. Paris

OECD (2004b): Problem Solving for Tomorrow's World – First Measures of Cross Curricular Competencies from PISA 2003. Paris

Polya, G. (1945, 1980): Schule des Denkens. Vom Lösen mathematischer Probleme. 3. Auflage. Bern

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 International Conference on Advances in Intelligent Systems – Theory and Applications, Luxembourg]

Preußler, W. (1996): Zur Rolle expliziten und impliziten Wissens bei der Steuerung dynamischer Systeme. In: Zeitschrift für Experimentelle Psychologie, 43, pp. 399–434

Preußler, W. (1997): Effekte des Kontexts auf den Wissenserwerb bei der Steuerung eines dynamischen Systems. In: Sprache & Kognition, 16, pp. 48–59

Preußler, W. (1998): Strukturwissen als Voraussetzung für die Steuerung komplexer dynamischer Systeme. In: Zeitschrift für Experimentelle Psychologie, 45, pp. 218–240

Preußler, W. (2001): Problemlöseerfolg als Bedingung für den Erwerb von Eingriffswissen bei der Systemsteuerung. In: Zeitschrift für Experimentelle Psychologie, 48(3), pp. 214–225

Reeff, J. P. (ed.) (1999): New Assessment Tools for Cross-Curricular Competencies in the Domain of Problem Solving. URL (31.07.2006): http://www.ppsw.rug.nl/~peschar/TSE.pdf

Reeff, J.-P./Zabal, A./Klieme, E. (2005): ALL Problem Solving Framework. In: Murray, T. S./Clermont, Y./Binkley, M. (2005): International Adult Literacy Survey. Measuring Adult Literacy and Life Skills: New Frameworks for Assessment. Ottawa

Repenning, A.: Project Portfpolio.
URL (31.07.2006): www.cs.colorado.edu/~ralex/Portfolio.pdf

Rigas, G./Brehmer, B. (1999): Mental Processes in Intelligence Tests and Dynamic Decision Making Tasks. In: Juslin, P./Montgomery, H. (eds.): Judgement and Decision Making: Neo-Brunswikean and Process-Tracing Approaches. Hillsdale, New Jersey, pp. 45–65

Schmidt, F. L./Hunter, J. (2004): General Mental Ability in the World of Work: Occupational Attainment and Job Performance. Journal of Personality and Social Psychology, (86)1, pp. 162–173

Schoppek, W. (2002): Examples, Rules, and Strategies in the Control of Dynamic Systems. In: Cognitive Science Quarterly, 2, pp. 63–92

Schulz, C. A. (2003): Strategien zum komplexen Problemlösen. Eine experimentelle Analyse der Rolle von Analogien beim Wissenserwerb in komplexen, dynamischen Systemen.
URL (31.07.2006): www.ub.uni-heidelberg.de/archiv/3755

Smith, M. (1991): Toward a Unified Theory of Problem Solving. Hillsdale, New Jersey

Statistics Canada/OECD (2005): Learning a Living: First Results of the Adult Literacy and Life Skills Survey. Ottawa/Paris

Sternberg, R. J./Frensch, P. A. (eds.) (1991): Complex Problem Solving: Principles and Mechanisms. Hillsdale, New Jersey

Stevens, J. (1992): Applied Multivariate Statistics for the Social Sciences. 2nd edition. Hillsdale, New Jersey

Stigler, J. W./Gallimore, R./Hiebert, J. (2000): Using Video Surveys to Compare Classrooms and Teaching Across Cultures: Examples and Lessons from the TIMSS Video Studies. In: Educational Psychologist, 35(2), pp. 87–100

U.S. Department of Labor (1991): What work requires of schools: A SCANS report for America 2000. Washington (ERIC Document Reproduction Service No. ED 332 0-54)

Vollmeyer, R./Funke, J. (1999): Personen- und Aufgabenmerkmale beim komplexen Problemlösen. In: Psychologische Rundschau, 50, 213–219

Wirth, J. (2004): Selbstregulation von Lernprozessen. Münster

Wirth, J. (2005): Regulationsmaß der Wissensanwendung (Anhang B). In: Klieme, E./Leutner, D./Wirth, J. (eds.): Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie. Wiesbaden, p. 137