New Assessment Tools for Cross-Curricular Competencies in the Domain of Problem Solving

Final report of project ERB-SOE2-CT98-2042
Funded under the Targeted Socio-Economic Research (TSER) Programme - Directorate General XII Science, Research and Development/ Directorate F
EUROPEAN Commission

Project coordinator:
Dr. Jean-Paul Reeff
Ministère de l´Education Nationale, de la Formation Professionnelle et des Sports
29, rue Aldringen
L-2926 Luxembourg

Partners:

Institut für Bildungsforschung (IBF), Germany
Max Planck Institute for Human Development, Germany
Center for School Development, Austria
University of Groningen, Netherlands
University of Heidelberg, Germany

1 Executive Summary

The NATCCC-PS network has been created as a small and short-term co-operation network of European researchers to improve the visibility and to increase the impact of European research in the field of Problem Solving, with special emphasis given to large scale international comparative studies. The main objectives of the network were

- to connect European policy demands with European experts and experience in the field of Problem Solving;
- to bring together a range of scientific experts working in the field of Problem Solving to develop new concepts and tools, and to create synergies;
- to produce a research and development plan directed to educational indicators in the field of Problem Solving.

While weak on the first objective, the network proved to be extremely successful on the two remaining goals. The networks' activities had significant impact both on the OECD PISA study and the Adult Literacy and Life Skills Survey (ALL). The activities went far beyond the originally defined work plan and generated both new research projects and applied research transfer contracts.

2 Individual network members

who actively participated in the preparation of this report

Judith Ebach (IBF)
Birgit Eschmann (MENFPS)
Joachim Funke (Univ. Heidelberg)
Eckhard Klieme (MPI)
Iris Langer (MENFPS)
Jules Peschar (Univ. Groningen)
Jean-Paul Reeff (MENFPS)
Astrid Schorn-Buchner (MENFPS)
Erich Svecnik (ZFS)
Hendrik Vogt (IBF)
Marieke van der Waal (Univ. Groningen)
Sietske Waslander (Univ. Groningen)
Joachim Wirth (MPI)

3 Background and objectives of the project

One of the biggest challenges to educational research and policy is providing relevant information regarding the education system's outcome at different levels. Student outcome indicators are of special interest as obviously this information is a primary criterion for different activities such as teaching, assessment, quality improvement programs, evaluation studies, and steering (as expressed by the French word "pilotage") the educational systems. In the last few decades a major effort was exerted at the international level (e.g. IEA, OECD, EU) to develop student outcome indicators for comparative purposes. The most recent of these enterprises, the Third International Mathematics and Science Study (TIMSS) conducted by IEA between 1994 and 1996, exemplifies the progress that has been made.

In the next decade, a major source of information about education will be provided by a recently launched OECD-study in 26 countries (comprising all EU-countries). While the main focus of the three three-year surveys will be on Reading Literacy, Mathematics and Science, all three domains will be included in all of the surveys. The effort required to carry out this study is substantial. It requires well-tuned and well co-ordinated actions within and between countries. This is especially true for European countries, where the need for co-ordinated actions and intra-European comparative studies has been stressed in all major EU-programs during the last several years.

Substantial scientific improvements have been made in the field of student outcome indicators over the years. The focus of these studies is, however, limited to the traditional assessment domains of Reading, Mathematics and Science. Due to rapid technological and social changes there is a growing demand for competencies beyond these traditional domains, such as problem-solving, communication skills, and learning to learn. There is thus a widely acknowledged need for broader indicators and the ability to monitor these competencies that lie outside traditional domains.

To date, the most successful and innovative initiative to develop additional educational indicators was conducted by the OECD. A small component on cross-curricular competencies has been included in the first cycle of the PISA study, results will be available in autumn 2001. The DeSeCo (Definition and Selection of Competencies) initiative, studies the theoretical foundations of Cross-Curricular Competencies and key qualifications and will propose relevant competencies to be assessed in the futures. The CCC-initiative will have a major impact on advancements made in the field of student outcomes in the years ahead. For education policy and education science alike, it is therefore highly important that countries of the European Union ensure the contributions of their scientific expertise and ensure incorporation of a European perspective into these new developments.

Beyond the initiatives mentioned above a mid-term project, to the year 2003, in the domain of Problem Solving has been launched by the Network A of the OECD INES programme. The goal of this project is to develop instruments to be included in the PISA survey cycle of the year 2003. For the moment these activities are co-ordinated by Dr. Eugene Owen (NCES, USA), chair of the OECD network A.

At the moment of the decision, no suitable instruments were available in the field of Problem Solving. There was a clear need for input from experts in the field. The activities in this domain had been initiated by the United States of America but European countries had expressed a need for more European input. The main challenges in developing new assessment tools for Cross-Curricular Competencies in the domain of Problem Solving were therefore twofold:

(1) To connect policy needs with expertise and experience. This requires close co-operation between policy-makers and scientific experts.
(2) To bring together scientific experts from different disciplines who are working within different conceptual frameworks. By finding commonalties and enhancing synergy, new concepts and new assessment tools may be developed.

Consistent with the political and scientific demands described above, the main objectives of the Thematic Network were

- To establish co-operation between European policy-makers and a range of scientific experts and
- To work towards a research and development plan to accomplish the demanding task of instrument development in the area of Problem Solving that is suitable for large-scale and international comparative research for indicator purposes.

High quality scientific work in the field of Problem Solving is available within Europe. The following objectives were proposed to the end of ensuring this European input for OECD activities:

(A) Connect European policy demands with European experts and experience in the field of Problem Solving;
(B) Bring together a range of scientific experts working in the field of Problem Solving to develop new concepts and tools and to create synergies;
(C) Produce a research and development plan directed to educational indicators in the field of Problem Solving.

4 Network activities

Bearing in mind the above mentioned objectives, network results can be roughly assessed as follows:

1) The network has not been successful on objective A (co-operation between decision-makers and researchers), at least not to the extent we had expected to be.

2) The network has been extremely successful on objectives B and C.

In the following chapters we will describe the work done in the network.

In order to reach the above mentioned goals, the network proposal list several work packages, related to the three objectives A-C:

A) Connect European policy demands with European experts and experience
- WP 1 Overview and clarification
- WP 2 EU policy issues

B) Bring together a range of scientific experts
- WP 3 Problem Solving in general education
- WP 4 Problem Solving in vocational education and training
- WP 5 Alternatives
- WP 6 Technological issues

C) Produce a research and development plan
- WP 7 Co-operation & extension
- WP 8 Documentation & exploitation
- WP 9 Project definition
- WP 10 Network management

Connect European policy demands with European experts and experience in the field of Problem Solving (objective A)

The work in this field has been driven by two different work streams, the first one focussing on a literature based, interdisciplinary approach aiming at a conceptual clarification. This part of the work strongly emphasises a societal point of view and can be seen as a major point of reference for the rest of the network's work. The chapter indirectly grasps essential political views, mainly by referring to policy oriented work within OECD. While the work of WP1 can be regarded as fully successful, the work for WP2 was far less successful.

The original idea behind this work package had been to get a clearer idea of what policy makers mean when speaking about Problem Solving in an educational context. Reading the minutes of quite a few high level political meetings dealing with Cross Curricular Competencies, it repeatedly turns out that Problem Solving is considered to be one of the most important competencies identified by decision makers at minister level. The WP work plan asked for a two step procedure:

a) review of documents;
b) a quick survey among decision makers.

It turned out very quickly that the document review was not a very useful exercise. While the topic was mentioned on a regular basis, the minutes and documents do not point to any operational definitions. "Problem Solving" is used as a very broad label, pointing more at a rather blurred political ambition than any concrete educational framework. This situation had been expected to a certain extent, and this expectation had been one of the reasons to envisage a quick survey.

The small questionnaire we developed (Annex 1) focuses on important characteristics of Problem Solving and on a short assessment of its importance when compared to other CCCs. To be very clear: We did not want to involve decision makers in a lengthy scientific process, but rather get some rough indications to better tune the subsequent research process.

The "result" is rather amazing: most of the minister counsellors we contacted, refused (!) to submit the questionnaire to their minister. The chairman of one important international organisations network did the same. The argument seemed to be that this type of questions is not admissible at a high political level. Very obviously, the top-level advisors of the minister protect their ministers against a situation where more concrete questions could be asked. One of these counsellors explicitely said, that ministers "SHOULD" remain vague and the subsequent "definitions of what is meant" is "the job of the researcher and education experts".

As a result we had to stop this top down approach.

This situation seems to be quite common, as is revealed e.g. by a text published on the European Commissions web page on "Indicators and Benchmarks of Quality of School Education" (http://europa.eu.int/comm/education/indic/indic1en.html). Among other attainment indicators the page describes indicators on ICT:

"....
No recent data in this rapidly changing area exist. ... *SITES will provide more descriptive information regarding the status of ICT in schools. These texts could be used to describe the area and raise central policy questions.* (bold & italic by the author of this paper)
..."

This passage obviously describes a very different approach as compared to the one taken by our network. While we expected the decision-makers to raise central policy questions a priori, this passage expects the already rather elaborated SITES instrument to precede the raising of central political questions.

In line with this approach we changed the strategy and subsequently fed the networks results to the appropriate policy networks. For instance did we have extensive presentations at the Helsinki Conference on "Learning-to Learn as part of CCC", organised by the "European Network of Policy Makers for the Evaluation of Education Systems" (22-24 March 1999).

Networking of scientific experts (objective B)
Research and development plan (objective C)

The basic idea of this 12 month network had been to establish a platform of and for European researchers in the field of assessment and large scale comparative studies (objective B) and while running some preparatory research activities, to prepare a research and development plan. This plan should generate a European project to stronger impact on international studies like PISA (objective C). The very early success of the network speeded up this process substantially. This is the reason why objectives B and C cannot be clearly separated in this paper.

A closer look at the background description and/or the original proposal to the European Commission clearly shows, that we were mainly focussing on the international aspects of the PISA study when speaking about large-scale comparative studies. Surprisingly, the only fact of submitting the proposal to the European Commission yielded a request from the project management of the International Life Skills Survey (ILSS) on whether the planned network could contribute to the development of a Problem Solving instrument for this survey. ILSS, today known as ALL (Adult Literacy and Life Skills Survey), is an international comparative study on adults (16-65 years old) in the domains of literacy, numeracy and Problem Solving. The study is co-directed by the US National Center for Education Statistics and Statistics Canada. As the first contacts were quite promising the network directed its work both to the trends in PISA and in ILSS/ALL. Moreover, after the EC funding was secured, it turned out that the network was an extremely good catalyst for a broader German subgroup working on the German national PISA option on Problem Solving, which was directed by Eckhard Klieme from MPIB, one of our network partners.

These new opportunities had several consequences:

1) The timelines for ALL and PISA being very different, the speed of work had to be dramatically increased

2) While systematically working on the basis for future contributions to PISA, important choices had to be made for ALL

3) The project we had envisaged to propose at the end of the network lifetime was already started while the network was still working. Actually one of our network partners (IBF) was contracted by Statistics Canada to develop both a framework and instruments for Problem Solving within ALL. The Berlin, Luxembourg and Heidelberg partners were heavily involved in this work. Thus the original aim of having a stronger European impact on large-scale international studies already became true. All network partners are fundamentally convinced that without the funding of the NATCCC-PS network this task could not have been successfully completed.

Despite of this more complex and more complicated situation, the original work-plan has been fully respected. Work packages 3,4,5 were dealt with in parallel, the results will be described in the following chapters. While WP3 took a more general approach and was strongly directed to large scale comparative assessment, the work in WP4 focussed on aspects closer to school, and above all to vocational education and training issues. The work not only shows how the concept of problem solving is implemented in "real" curricula, but also how it is related to other relevant concepts in vocational education. The work on WP5 illustrates some alternative concepts that go beyond of what can be implemented in today's large-scale surveys. By the very nature of the concepts described, this work package is closely related to technology issues, some of them being described in the paper. Some of the concepts developed have already been implemented within the German national PISA option. Technology issues going even beyond these alternative concepts have been analysed with different technology partners. It became clear that dealing with advanced technology like e.g. interactive video, virtual reality or complex simulations was far beyond the reach of our network activities. Instead of writing a purely descriptive document, we submitted a proposal for "Accompanying Measures " on "Skills Measurements and Technology Based Assessment". The proposal has been rejected, the reviewers considered the proposal to be "a fully fledged project proposal" and suggested to resubmit as a *project* proposal within an extended partnership.

Beyond this content oriented work, the network has been actively dealing with network extension issues. After careful analysis of the situation, the network members decided in a very early phase not to aim at a formal extension of the network, for two major reasons:

1) The one-year run time of the network activities proved to be an obstacle. Negotiations at different levels to reach formal commitments would haven take too much time without yielding useful results before closing network activities.

2) The success of the network in a very early phase of work (actually already before the work started officially!) gave the network members plenty of opportunities to become involved in different networks and working groups. Further extension of networking activities would have endangered the major goals of the network.

Thus the global strategy was to fully exploit the opportunities provided through co-operation with other working groups and organisations, or within sub-networks created as a result of the NATCCC networks activities.

Major co-operations can be identified:

1) Several network members (Berlin, Heidelberg, Luxembourg, Groningen) became strongly involved in the OECD activities on assessment of cross-curricular competencies. The networks work had a strong impact on the overall strategy in this field of OECD INES Network A. The work will substantially impact the instrument development for the PISA study.

2) Within Germany the network facilitated the creation of a German working group who prepared the German national option for the PISA study. Under the lead of the Berlin Partner and with very close co-operation of the Heidelberg partner, this group prepared a set of highly innovative instruments.

3) The whole network became strongly involved in the development of a framework and of instruments for Problem Solving for the International Life Skills Survey (ILSS, today known under the acronym ALL). As a result, the Bonn partner got a contract from Statistics Canada to develop the Problem Solving instruments to be used in the survey.

4) As a side-effect of the work on Problem Solving for ILSS, a major research contract on translation and “cognitive equivalence” issues has been given to the Psychological Institute of the University of Trier (PD Dr. Axel Buchner) by the US National Center for Education Statistics.

5 Main Results

In the following four chapters we will describe the main results of the networks work. Chapter one will give an overview of Problem Solving as a cross-curricular competence and already reflects issues concerning general education. Chapter two focuses on aspects of vocational training. Chapter three briefly outlines new concepts and tools. Finally, chapter four discusses the assessment aspects in depth, with special emphasis being given to large-scale comparative studies. Parts of the following chapters have been/will be published as separate papers.

5.1 Problem Solving as a Cross-Curricular Competence

5.1.1 Read instruction

This chapter is about problem solving as a cross-curricular competence and addresses the following goals:

Review of concepts of problem solving, which shows a relation with cross-curricular competencies
Description of instruments to measure problem solving and their methodological problems.

We have worked out these goals in the following way:

In the first part of the paper we describe and explain the concept ‘cross-curricular competencies’ and its relation with problem solving. We view problem solving as a cross-curricular competence. In short, we argue that problem solving is a necessary competence a person needs in order to survive in society. We emphasise that this societal point of view is the leading starting point across the whole paper. Our minor starting point is education. We argue the school has the task to prepare individuals for their societal life and therefore, should learn individuals how to solve problems besides other competencies.

In part two, we give a review of theoretical conceptualisations of problem solving. This paragraph may be a bit too elaborate for readers who know the field. However, for a good understanding of problem solving and a critical review of instruments to measure problem solving with regard to problem solving as a cross-curricular competence, we thought it essential for the argument of the paper.

In part 3 three we give an overview of instruments to measure problem solving, which have potential for measuring problem solving as a cross curricular competence. We will discuss the advantages and disadvantages of each instrument.

We conclude this chapter with a short summary and discussion about what should be done in this field of research in the near future.

5.1.2 Method

The main method for the elaboration of the results presented in the following paragraph has been research literature analysis. For the first part a mainly societal point of view was taken and the analysis was triggered by a sociology rationale. Part two follows a cognitive-psychological rationale and describes different aspects of problem solving and related theories. The focus of part three is on measurement. The results of an extensive literature analysis fed a discursive process to prepare the ground for first ideas on a measurement framework. In selected cases literature analysis was supplemented by a (re-)analysis of empirical data stemming from studies of some of the network partners.

5.1.3 Introduction into Problem Solving as a Cross Curricular Competence

5.1.3.1 Cross Curricular Competencies

Cross Curricular Competencies (CCC) can be described as competencies an individual needs in order to play a constructive role as a citizen in the society and to live a valuable individual and social life (Bongers, Kleine, Waslander & Peschar, 1996). Although the concept CCC is relatively new, already many shed their light on it. In general, all seem agree on a not explicated vague content, but everyone makes it's own accents (Salganik, Rychen, Moser & Konstant, 1998). This has led to many diverse descriptions and caused confusion in the field.

In this paper we ally with the opinion of Waslander (1999), who argues that cross-curricular competencies can be understood as an umbrella-concept, which has the function of a search guide, instead of a concept with an explicitly specified meaning. The concept has to be filled in by taking the society as a starting point. This means that the content of CCC may change in the future, due to a changing context. This thought represents the societal point of view and is the main criterion for evaluating conceptual frameworks of problem solving, problem solving definitions and instruments to measure problem solving.

The idea of CCC has been introduced by Trier (1991), though under a different label (Non-curricula bounded socio-cultural knowledge and skills = NOBS) in the context of Network A of the OECD. Trier argued that schools prepare students also in other dimensions than the traditional ones, to facilitate students with competencies. These competencies should make students able to participate and to survive in society when they finish school. The dimensions concern a variety of skills, knowledge and attitudes which are developed in school, even though this is not always explicated in the formal curriculum. Trier connects here society with education.

It is important to acknowledge that CCC can be approached from these different points of view. Therefore, cross-curricular competence is actually a misleading term, because it stresses the educational point of view and neglects the societal point of view. Again, we want to emphasise that the society is the main starting point for conceptualising CCC. The educational point of view comes in second instance.

The interest of researchers and politicians to distinguish cross-curricular competencies has grown due to several developments in society, which started decades ago but which impact is strongly felt nowadays. While the diversity of cultures, values and preferences increases within countries, social cohesion seems to decrease (OECD, 1997). Traditional institutions like the family and local communities seem to produce less cohesion than in the past. Individuals are getting more connected with each other by bureaucratic instead of personal ties. Such a decline of integration puts pressure on the society, which may go along with negative consequences like social isolation, delinquency, discrimination etc. (Holmes, 1992). Since the industrial revolution, society is growing increasingly complex. In the last decades developments in the information technology follow each other rapidly. This has changed and still changes the nature of work and daily life drastically (OECD, 1997).

One may safely conclude that individuals have to deal with a multicultural, complex and continuously changing environment nowadays. To cope with such an environment individuals need certain competencies (knowledge, skills and attitudes) which are useful in different dimensions of their lives.

Individuals need competencies for getting, keeping and executing a job sufficiently. Due to the continuous developments of technology, jobs can change easily and suddenly. To cope with these changes individuals should be able to solve problems they are confronted with on the work floor, be flexible and communicative and be prepared for continuing learning (OECD, 1997);
Individuals face a growing complex social world. On the level of neighbourhood, town and region individuals are confronted with different cultures and ethnicity's and their corresponding values. Also developments beyond the region and or even the nation have an impact on the social world, f.e. the integration of the European Union. For ordering this world and to participate in and contribute (in the sense of integration and social cohesion) to it individuals need competencies like self-confidence, critical thinking, tolerance etc. (Nickerson, 1994; OECD, 1997).
Most modern societies have a democratic system for governing the country. Such a system can function only if citizens know their democratic rights and duties and know how to participate actively in the local and national democratic institutions. Democratic behaviour implies citizens have democratic attitudes and relevant competencies, which make them able to participate in the democratic institutions (Van der Ploeg, 1995).
The three dimensions above come together in the last dimension. Individuals need competencies to manage their own personal life and have the ability to adapt to changing situations (Nickerson, 1994). There is a strong argument that certain competencies are related to a person mental health and its social well being (Fend, 1995; OECD, 1997).

It seems obvious that individuals need certain competencies besides the traditional cognitive competencies like reading, math and science in order to survive in society (OECD, 1997). A general consensus is felt that the school (should) play(s) an important role in the transfer of these other competencies on students (Calvert & MacBeath, 1995). However a strong dissent exists about which concrete competencies these would be (Waslander, 1998).

When we look at the concept 'cross-curricular competencies' from an educational point of view, instead of the society, we can distinguish four general characteristics of CCC.

CCC are

independent of situations and content
learnt within and “between” several subjects (between in the meaning of ‘reading between the lines’)
are necessary in situations which demand complex and extensive achievements (original sentence: bei der Bewältigung komplexer, ganzheitlicher Anforderungen von Bedeutung sind)
can be transferred to new tasks, which are not explicated in the formal curriculum
(Klieme, 1998).

Both points of view meet each other in a first effort of the OECD (1997) to distinguish and measure cross-curricular competencies. The four cross-curricular competencies measured are:
Civics
Problem Solving
Communication
Self-Perception / Self-Concept

This paper will go into further detail about problem solving. The feasibility study of the OECD (1997) to measure CCC showed that an instrument to measure problem solving in a large survey is not available yet. The instrument used in the feasibility study had several difficulties and is not good enough yet for a large survey research. In the year 2003 the OECD wants to include problem solving in PISA (a large survey in many countries to measure output indicators of national educational systems) and therefore it needs a sufficient instrument. To meet this goal the existing instrument should be improved, a search should be done for other available instruments or in the extreme case, a totally new instrument should be developed. This paper intends to contribute to this discussion, by describing theories on problem solving and instruments or possibilities for developing an instrument to measure problem solving as a CCC.

5.1.3.2 Problem Solving as a Cross-Curricular Competence

What is problem solving? At this stage, we will answer this question more or less intuitively following from the notion that an individual has to solve problems to get along in the society in all its facets.
All the time individuals meet simple and complex problems in their daily life i.e. the labour market, the multicultural society, the democracy and their personal social life. A problem arises when the individual experiences difficulties with solving a task i.e. a task becomes a problem by interaction with an individual. When individuals face a task they experience as a problem, they have to invent activities to solve the problem, but the exact activities are often unknown. In other words, for individuals the solution of a problem is not always obvious, but has to be created. Most daily life problems can be solved in different ways, which might result in different possible outcomes. In the case of different possible outcome it is often hard to say if one solution or outcome is better than the other one.

Why would problem solving be a cross-curricular competence?

It is striking how much the arguments to stress the importance of teaching problem solving and thinking resemble the arguments which support the importance of cross-curricular competencies (Nickerson, 1994). Problem solving and thinking would be a necessary competence to cope with a continuously changing work situation, to preserve a democratic way of life, to cope with a growing complex social world and to be equipped to manage ones own life and to adjust to changing realities. The last argument is the most compelling for Nickerson. He argues that the competence to deal with daily life problems effectively, could be a major determinant of the level of happiness a person achieves (Nickerson, 1994: p.414).

The paragraph above supports our opinion that problem solving is as a cross-curricular competence. From the educational point of view we can basically draw the same conclusion, because the characteristics of CCC by Klieme (1998) correspond with the feature of problem solving. Problem solving competence can be used for very diverse problems in very different and sometimes complex situations. A problem may demand a solution where the individual has to combine and integrate knowledge and skills from different subjects. Due to this interdisciplinary character of problem solving, individuals implicitly learn to solve problems within and between subjects and explicitly with specific problem solving tasks.

In sum, from the societal as well as from the educational point of view problem solving can be seen as a CCC. The scientific research on problem solving however, is mostly done in the field of psychology. Before we are going into further detail about the measurement of problem solving as a CCC, we will first explore several psychological theories on problem solving in the next part of this chapter.

5.1.4 Theoretical approaches towards Problem Solving

In this part we will give an overview of theories on thinking and problem solving. The overview aims at examining whether theoretical approaches provide instruments to measure problem solving or give leads for developing instruments. Across the whole part two, we will also pay attention to how the theoretical approaches as well as the possible instruments ally to the CCC concept.
In the second section of this part, the most important historical theories on thinking and problem solving are described shortly. In the third section, we concentrate on the information processing approach, an approach, which is mostly adhered to nowadays. The fourth and fifth section is about two approaches towards problem solving, which comes closer to the idea of problem solving as a CCC. First, we start with a definition of problem solving.

5.1.4.1 Definition of Problem Solving

Many definitions of problem solving exist. Nevertheless, today, most psychologists agree that problem solving is … *a cognitive process to transform a given situation to a desired situation. However, the way to accomplish the change is not direct and obvious to the problem solver* (Mayer, 1992).
The definition above is a proper one for understanding the theoretical approaches on problem solving described in this chapter. However, the definition is not a definition of problem solving as a CCC. Therefore, the definition differs from the description of problem solving as a CCC that we wrote in section 1.2, but it does not contradict that description. In the first section we will work out a definition of problem solving as a CCC in more detail.

5.1.4.2 Traditional theories on Problem Solving

There are two main traditional views to look at thinking and problem solving: Associationism and Gestalt theory (Mayer, 1992; Dominowski & Bourne Jr., 1994). Both views concentrate on the problem solving process and on the transfer effects after solving a specific problem.

Associationism (Trial and Error)
According to the associationist view, thinking can be described as the trial and error application of already existing response tendencies. In this view a person, confronted with a problem, will try to solve a problem on past experiences. When the first solution is not successful, he will try another one, just as long as the problem is solved. In more theoretical words associationists assume that for any problem situation S, there are associations or links to many possible responses R1, R2, R3 and so on. The three elements in the associationist theory are (1) the stimulus (problem solving situation), (2) the responses (problem solving behaviour) and the associations between a particular stimulus and a particular response. The links are presumed to be in the problem solver's head. Responses vary in strength, with some associations being very strong and some being very weak. Responses for any given situation may be put into a hierarchy in order of their strength. The response on top of the hierarchy will be executed first to solve the problem (Mayer, 1992).

The response hierarchy can change in time. When a response has been practised in past situations it is more likely to be performed when that situation is presented again, in other words the response will move in the direction of the top of the hierarchy (*law of exercise)*. A response will also increase in strength and go up in the hierarchy when it has helped to solve a past problem (*law of effects*). The process of trial and error can take place in the mind of a person. This is called *covert thinking* (Mayer, 1992).

Associationists also thought about the question why a transfer of problem solving competence takes place. Is this because task A and B have identical elements (specific transfer) or because learning task A provides a person a general cognitive skill that also somehow helped in learning task B (general transfer). Empirical research gives support to the first explanation. A person can solve a problem faster, when the problem is similar to a past problem.

Gestalt theory (Insight)
Gestalt-psychologists put a strong emphasis on restructuring of a problem as part of the problem solving process. Instead of trying solutions (as the associationists argue), a person starts the problem solving process with distinguishing and structuring the components of the problem. While structuring the problem the solution of the problem will come suddenly as a 'flash of insight' (also called 'aha-Erlebnis) (Mayer, 1992).

Gestalt-psychologists distinguish two kinds of thinking. One, based on creating a new solution to a problem, is called *productive thinking*, because a new organisation is produced. The other, based on applying past solutions to a problem, is called *reproductive thinking*, because old habits and behaviours are reproduced. Productive thinking is associated with 'insight' and reproductive thinking with 'trial and error'. What makes this distinction interesting is that according to the Gestalt theory productive thinking directs at understanding. Several researches support the opinion that learning by understanding holds longer, than learning by simply memorising. In addition, problems, which are slightly different from the problem learned, are solved better by students who learned by understanding (Mayer, 1992).
Another interesting point in the Gestalt theory is that Gestalt-psychologists view problem solving as a process, which can be divided in several stages. Polya (1965) distinguishes four steps:

Understanding the problem (gathering information about the problem)
Devising a plan (searching for a solution by using past experiences; this is the stage of flashes of insight)
Carrying out the plan
Looking back (checking the result and asking whether the solution is applicable for other problems)
These four steps are often taken as starting point for describing the problem solving process.

Summarising, one can see that in the past, psychologists, who studied thinking and problem solving, were mostly interested in how exactly the mental activity in the human brain looks like, when a human being is thinking and solving problems. They assumed that the activity to solve a simple problem could be generalised to complex problems. That is why they directed their research on simple problems in a laboratory setting. They paid little attention to the assessment of problem solving and so instruments to measure problem solving can not be derived from these theories. The importance of these theories lies in the fact that present theoretical frameworks on problem solving build further on the insights of associationism and Gestalt theory.

5.1.4.3 Information processing approach: a present conceptual framework on Problem Solving

Today, most psychologists view problem solving as an information processing model (Mayer, 1992; Hunt, 1994; Nickerson, 1994). In this approach problem solving is viewed as '*a sequence of mental processes or operations performed on information in a subject's memory*' (Mayer, 1992, p. 172).

The main idea of the information processing approach is that computer programs can be used as models for human thought (Hunt, 1994). Computer simulations of problem solving strategies have shown to be successful for well-defined problems, more explicitly for problems with a well-defined initial state, a well-defined goal state and a well-defined set of operators or moves (Mayer, 1992).

According to Simon (1978, 1979) three components are important in the information processing approach: (1) the problem solver or 'information processing system' (2) the problem and (3) the problem representation or 'problem space' (Mayer, 1992). Especially the construct problem space is interesting.

The problem space refers to the problem solvers' internal representation of (a) the initial state, (b) the goal state, (c) the intermediate problem states and (d) operators or moves that a person can make to come from one state to the other state (Mayer, 1992). The problem space is a set of all possible sequences of operators for solving a problem that the problem solver is aware of. This means that the problem space of one person is likely to be different from the problem space of another person and could contain errors. By 'drawing' the problem space the problem solver can get a notion how to solve the problem. Solving a problem in this approach can be viewed as finding the correct path or route through the problem space.

For building a problem space a problem solver needs information about the problem, and previous knowledge and experiences it preserves in its memory. Therefore, it uses the immediate memory system and the long-term memory. The immediate memory system contains active information and can be divided into a working memory and short-term memory (Mayer, 1992). In the immediate memory, the problem space or so-called mental model is constructed. The long-term memory contains episodic memory for prior experiences, procedural information about how to react to situations, and semantic memory about how the world is organised (Hunt, 1994). The long-term memory is used as a sort of database, from where the immediate memory can get it's information. The complexity of a mental model is determined by the capacities of the immediate memory of the problem solver (Hunt, 1994). When the immediate memory cannot handle the complexity, the mental problem will be simpler than the actual problem situation really is. This may have negative consequences for effectively solving problems.

How then, is it possible that most individuals seem to cope with their daily life problems, which often are more complex than their immediate memory can handle?
According to Hunt (1994) individuals solve their daily-life problems by using schemata. A schema is a bundle of knowledge and experiences structured around an abstract subject. A schema serves to select and organise incoming information into an integrated meaningful framework (Mayer, 1992). Hunt argues that although schemata often are redundant and contradictory it helps people to solve their daily-life problems, because it relieves the limited immediate memory.

If Hunt is right, we do not have many leads for developing an instrument to measure problem solving, because the construct 'schema' still is very vague, lacks any empirical support and therefore, is difficult to measure. However, if we leave out the schema(ta) theory for the moment and concentrate on the information processing approach we can detect several leads for development of an instrument. The information processing approach incorporates the idea of problem solving as a process, consisting of several stages. Especially this idea of a process divided in stages serves as a starting point. A possible instrument to measure problem solving could measure the following things.

The end state of the problem solving process: has the person reached the goal state or how far is he away from the goal state? By taking the end state as measure moment we are confronted with the complication that a problem may not have an absolute correct end state. This may lead to difficulties for giving scores.

The problem solving process: Polya (1957) divided the problem solving process in four stages: (1) understanding the problem; (2) devising a plan; (3) carrying out the plan; (4) looking back. A more detailed division is also possible as Hayes (1981) shows: (1) finding the problem; (2) representing the problem; (3) planning the solution; (4) carrying out the plan; (5) evaluating the solution; (6) consolidating gains. Other conceptualisations of the problem process exist, but the main idea of a problem solving process stays the same. Regardless of which conceptualisation one takes, the explicated stages can serve as measurement moments. Possible questions to measure are: do persons plan their problem solving activities properly, i.e. do they follow the steps, which are distinguished in the theory? An important stage in the problem solving process is the representation of the problem. This representation of the problem can also function as a measurement moment, by judging and giving scores on a (written) representation. Also planning is an important component for solving a problem effectively and might be scored in some way.

Strategies for solving a problem: During the problem solving process, people can use several different strategies (Mayer, 1992; Nickerson, 1994). The chosen strategy has influence on how effective a person can solve the problem. Therefore, it seems relevant to measure if persons choose a right problem solving strategy. It can been seen as a competence when a person knows when to use a certain problem solving strategy given a specific problem. Such competence may be measured by multiple-choice questions afterwards.

Because, researchers on problem solving are mostly interested in defining and analysing the activity of the human brain during problem solving, they mainly have concentrated their research on the process of problem solving and in second instance on problem solving strategies. Their most interesting result in regard to the subject of this paper is that successful problem solving depends more on the domain specific knowledge rather than the general problem solving competence (Mayer, 1992). The consequence of this result is that most American recent research directs on problem solving in a specific domain instead of general problem solving and that the question about a general problem solving theory has been postponed. The ‘domain-specificity’ of an instrument is an impediment for measuring problem solving as a CCC. Problem solving as a CCC directs at daily life problems, which can occur in many different settings. Instruments, which concentrate on a specific domain, put limitations on the possible different daily-life problem situations. Besides this domain-specific approach, another mostly European based approach exists, which aims at a theory on and measurement of general problem solving competence. In the section 2.5, we will describe this approach in more detail.

A second disadvantage is that many researchers, who use information processing approach concentrates on problems, which have a well-defined initial and well-defined goal state. This is not applicable on problem solving as a CCC, because of its concentration on daily life problems. Daily life problems are usually ill-defined and may have more possible goal states. The concept ‘creativity’ seems to be more in the direction of problem solving as a CCC. We will address section 2.4 to this subject.

5.1.4.4 Creativity

Creative thinking can be characterised as divergent thinking (Mayer, 1992). A person, who thinks divergent does not direct it’s thoughts to one answer, but is spreading its thoughts to many possible directions. Problems, which demand creative (divergent) thinking can be solved in many ways and can have more than one correct answer. In other words, according to Mayer, creative thinking is a cognitive activity that results in one or more novel solutions for a problem. Creative thinking goes along with critical thinking, which means that a person can critically evaluate its novel solutions and pick the appropriate one for executing (Mayer, 1992). Creative thinking and critical thinking both are useful in the problem solving process. Sometimes, these concepts are replaced with problem solving and the other way around. Creative thinking, critical thinking and problem solving seem to be interchangeable.
Creative thinking as described here comes very close to the idea of problem solving as a CCC. Therefore, it is justifiable to have a closer look at this concept.

A huge amount of literature exists about creativity and it goes beyond the goal of this paper to give a more or less complete overview of this literature. What we describe here is a short overview of the confluence approach. This approach is supported by several researchers and clearly described by Lubard (1994).

Lubard (1994) sets creativity in a broader context than Mayer does. According to Lubard creativity not only can be applied in problem situation but also in a creating process, like that of artists, designers or composers. Several elements are involved in the creative process. Besides the cognitive elements like intellectual abilities, knowledge and thinking styles, personality, motivation, and socio-cultural environment are sources for creativity as well.
A person, who has intellectual abilities, is able (1) to recognise a problem, (2) to define / represent the problem, (3) to choose the right problem solving strategy and (4) to evaluate the outcome of the problem solving process. These steps in the problem solving process are familiar for us. We also saw them in the information processing approach towards problem solving (see section 2.3). Besides these so-called high-level abilities, a creative person needs insight abilities and divergent thinking skills. These three components form the intellectual abilities.

The second component, which leads to creativity, is knowledge. Knowledge is necessary during the whole process. Lubard (1994) describes six reasons why knowledge is necessary for creativity. All reasons are compelling but in summary knowledge is the known information about the problem, the problem's content and the problem's context. This information should make it possible to think in directions of new ideas. On the other hand, knowledge can also block the creative thinking process. For example, a person successfully solved a problem in the past. Now the person has to solve a new problem, which seems similar as the first problem. The person tries to solve the problem with the same solution, but this time the solution is not successful. The person has to think about a new solution, but it is likely that the person gets stuck in this new thinking process, because it comes up with the old solution all the time. So, knowledge can make the person less flexible in its thinking process (Lubard, 1994; Mayer, 1992). This so-called blockbusting was already recognised by Gestalt-Psychologists (Mayer, 1992).

The third component 'thinking style' is the link between the cognitive elements and personality traits. Two thinking styles can be distinguished. A person has a sensing style, when it relies on the available information and focuses on the realities of a situation. Persons who use the intuitive style rely on their hunches, feelings and internal sources of knowledge. Studies show that the intuitive style leads to more ideas and solutions that are new, than the sensing style does.

Personality is the fourth component. Personality traits may facilitate the effective use of the cognitive components and help to turn ideas into real results. During the problem solving process there are moments when the problem solver is uncertain about the solution or senses that the solution is not right and that he has to think about other solutions. To handle the uncertainty and obstacles a person needs tolerance of ambiguity and perseverance. Openness to new experiences and a willingness to take risks are characteristics, which stimulate the person to think about and try new alternative solutions. What binds these four personality traits is self-esteem. When the person doubts it's ideas, but also when it has to try new ideas a belief in oneself and one's capabilities is essential.

Close to these personality traits lies the fifth component: motivation. Motivation is the drive to continue with the problem solving process until the problem is solved, because the completion of the task is satisfying for the person one way or the other.The final component of creativity is environmental context. Some environments are stimulating for forming new ideas, others put barriers in the thinking process.

Much more can be said about the six components of creativity. What is essential is that results of empirical studies seem to give support to all the components separately. How the components mix together during the problem solving process and which component influences the process the most, are still unanswered questions, although Lubard suggests that a minimal of knowledge about the problem is essential.

When it comes down to measure creativity or creative thinking, again problems arise. Empirical studies on creativity concentrate on creativity in a productive sense, like painting pictures, composing music etc. or on how creative thinking can be learned. Especially the measurement instruments in the first category are of no use for realising our goal, because they are not applicable on problem solving in daily life situations.

The approach on creative problem solving as a learnable skill is less theoretically based than the confluence approach (Mayer, 1992). The empirical research in this field addresses the question whether creative thinking can be learned and improved. The results of this research point out two things: (1) Specific knowledge and general problem solving strategies are necessary to solve a problem in a specific domain. (2) Problem solving skills are often applicable only within the domain for which they have been learned and there is little proof that these skills can be transferred to other domains (Mayer, 1992).

The theoretical confluence approach is promising and useful for our conceptualisation of problem solving as a CCC, especially concerning the addition of non-cognitive elements of problem solving competence. However, the existing instruments to measure creative thinking are disappointing. The instruments seem to have the same disadvantages we already expressed in section 2.3. Most existing creativity tests are domain specific or do not measure creative problem solving in daily life situations (look for an overview Mayer, 1992 and Lubard, 1994).

5.1.4.5 Complex Problem Solving

Complex Problem Solving (CPS) is a new branch in psychological research on problem solving and decision making. Having started about 1975 with the seminal work by Dietrich Dörner (currently Bamberg University) and his crew, this European rooted research is now an international recognised and established approach for analysing problem solving processes under uncertainty and complexity. The CPS-research can be embedded in the information processing approach (Frensch & Funke, 1995). This section sketches the main ideas, illustrates the historical background and explains main features of CPS.

The main idea of CPS consists in confronting human subjects with complex and dynamic computer simulations of selected domains of reality, which they have to identify and control without much prior training. This situation overcomes the deficiencies of classical problem solving and intelligence research by introducing ill-defined problem situations, which at the same time can be controlled and manipulated by the experimenter.

The background of the emergence of CPS research in the mid-70s can be found in a deep disappointment with the low predictive power of traditional IQ tests for problem solving in everyday situations. One of the reasons for this flop of IQ tests was the academic nature of the tasks. Thus, the alternative approach to IQ measurement offered by CPS research was based on the construction of computer simulated scenarios and their presentation to naive persons as a new tool for analysing complex problem solving under controlled conditions in the lab and with more task requirements than traditional IQ testing. Due to this enrichment of the problem space, results from CPS research seem to achieve much higher validity with respect to everyday problem solving than is achieved by traditional diagnostic tools.

Features of complex problems and the related requirements on the side of the problem solver are (in line with the description given by Frensch & Funke, 1995):
complexity and connectivity between a huge number of variables: complexity requires information reduction, connectivity requires anticipation of side effects;
dynamic aspects of problem situations: they require prediction of future developments (planning) as well as long-term control of decision effects;
intransparency and opaqueness of situation: this feature requires systematic collection of information;
polytely (=multiple goals): polytely requires decisions about goals and goal balancing.

All these features are inherent in many situations known from politics, management, or biology, to mention only a few of the areas in which CPS research becomes important.

CPS has several obvious advantages. With complex problems, it is possible to achieve real-life situations which go far beyond the puzzles of traditional problem solving or intelligence testing. Instead of presenting only well-defined problems from artificial settings, CPS research started with the appeal of introducing real-world problems similar to those, which are faced by every human being throughout its life. Also, the emphasis on problem solving in specific domains is increased by this approach.

Like the confluence approach towards creativity, CPS also overcomes the limitation of classical problem solving which restricts the tasks to purely cognitive demands. CPS requires not only cognitive but also emotional competencies: on the cognitive side, identification and control of dynamic systems is required; but, on the emotional side, additionally the control of emotions during problem solving is required. This integration of cognitive and emotional aspects of problem solving is an important step beyond classical problem solving and leads to a more integrated view of the problem solver and his coping.

5.1.4.6 Conclusion

Overlooking the theory described in the forgoing sections, we can see a certain development in the conceptualisation and research on problem solving. The recent conceptual frameworks have been built on former theories and have - besides differences – also many similarities. After reviewing different frameworks we may conclude that the concept 'divergent thinking and the complex problem solving approach are the most promising to operationalize problem solving within the CCC-concept.
Overlooking all the theoretical frameworks described in this chapter, we can detect several leads for developing an instrument to measure problem solving as a CCC as well. Firstly, the idea of a problem solving process divided in stages can serve as a starting point in the development of an instrument. Secondly, the instrument should measure cognitive and non-cognitive elements of problem solving competence. The theory on creativity and CPS both stress the importance of emotional, personal and social abilities during the problem solving process and including these elements in the instrument as well, will make the instrument more valid.
Nevertheless, in a certain way, the available literature on thinking and problem solving is also disappointing. The main problem we feel after studying the literature is that there is little experience with measuring problem solving for assessment. Most literature stays very vague in their descriptions on how problem solving is actually measured and what procedures are followed to give a person a score on problem solving, because the main interest lies in the activity of the human brain during the problem solving process. Therefore, we cannot simply derive from theory an adequate well-developed instrument to measure problem solving as a CCC in a large survey. Measuring problem solving in a large survey demands a concentration on the end state of the problem solving process and in second instance on the process itself. In this sense, there is a sort of gap between the continuing development of theory on thinking and problem solving and the empirical experience to measure problem solving for assessment purpose. In the following part we will describe several instruments to measure problem solving, which have been developed recently or will be developed for assessment purpose.

5.1.5 Available instruments to measure problem solving competence

5.1.5.1 Revising the definition of Problem Solving as a CCC

Thinking about an instrument to measure problem solving implies we know 'exactly' what we want to measure. In section 1.2 we gave a temporary description of problem solving, which in summary was as follows:

Problem solving is a set of activities a problem solver has to invent in order to overcome a problem. The problem is situated in a daily-life situation and unknown. The problem may have more than one unknown possible solution and more than one unknown possible outcome.

After studying the literature on problem solving we are able to explicate the features of a problem and problem solving, which we refer to when we talk about problem solving as a CCC.

A problem arises when there is a difference between the initial state and the desired goal state and the solution to diminish this difference is unknown to the individual. A problem, in regard to problem solving as a CCC is a daily-life problem, which means that the problem is usually not well-defined, often complex, intransparent, changes dynamically during problem solving, and can be solved by one or more solutions. The desired goal state may have more than one possible feature and therefore, is not always well-defined either.
Problem solving is a cognitive and emotional process with several steps in order to overcome obstacles between a given state and a desired goal state. During the problem solving process cognitive, emotional, personal and social abilities and knowledge are involved.

This definition is almost the same as the definition of Frensch & Funke (1995):

'CPS occurs to overcome barriers between a given state and a desired goal state by means of behavioral and/or cognitive, multistep activities. The given state, goal state and barriers between given state and goal state are complex, change dynamically during problem solving and are intransparent. The exact properties of the given state, goal state and barriers are unknown to the solver at the outset. CPS implies the efficient interaction between a solver and the situational requirements of the task, and involves a solver's cognitive, emotional, personal and social abilitities and knowledge.'

Both definitions can be embedded in the information processing approach.

In the definition we mention intransparancy as characteristic of a problem. A problem is intransparent when the necessary information for solving the problem is not completely available. This is the case when just some variables can be detected from direct observation or when so many variables are involved in the problem that it is only possible to take a few relevant variables into account during the problem solving process (Funke, 1991). The extent of intransparancy of a problem depends on the ability of the problem solver. A specific problem can be more intransparent for one person than for the other.

Another characteristic, which needs some clarification is the unknown solutions and goal state of the problem solving process. The phrases about the solutions and goal state imply that three hypothetical outcomes of the problem solving process exist: (1) one solution – one goal state; (2) more than one solution – one goal state and (3) more than one solution – more than one goal state. In all three cases the solution(s) are unknown to the individual when it starts the problem solving process. In most daily life problem situations the individual does not perfectly know the feature of the goal state either. This makes a daily life problem often difficult.

In our opinion, problem solving competence as a CCC has to be measured by using a task in a test situation, which is experienced by the individual as a problem and which has the characteristics, mentioned in the definition.
We acknowledge that a problem in a test situation, which meets the characteristics perfectly, is an ideal and hypothetical situation. It is not feasible to develop a task for measuring problem solving in a large survey, which is equally intransparent, unknown etc for all respondents. It is also not feasible to develop a task for measuring problem solving in a large survey, because of practical constraints of large-scale assessment. However, we have the opinion that one should strive for an instrument which comes as close as possible to the ideal hypothetical situation.

Below we describe several existing instruments to measure problem solving. We will evaluate them on validity and reliability, by taking the concept of CCC and the theoretical views on problem solving into account.

5.1.5.2 The "Project approach"

The Institute for Educational Research (IBF= Institut für Bildungsforschung) has a long experience in test development. The institute developed a paper and pencil test for adolescents of 13 -14 years old to measure problem solving competence. The problem to be solved is relatively unknown, daily life problem and domain independent, e.g. laying out a school garden or editing a school journal. These problems fit in the daily context of pupils of 12-14 years old and therefore, may be appealing for them. Although the problem is relatively unknown, the methods to be applied during solution are supposed to be well known (IBF, 1998).

The stages of the problem solving process have been taken as a starting point. The test is structured around the following stages:

Define goals
Analyse the situation
Plan the solution
Executive the plan
Evaluate the result

Each step then is operationalized by several items. These are mostly multiple-choice items, but other formats like writing numbers or letters, ordering items and correspondence items are possible as well. The main characteristic is that all the items have a closed format (IBF, 1998; Klieme, 1998). The test aims to measure the ability to identify, sort, combine and judge complex information given in the several items. Moreover, the constructors of the test did not intend to measure the strategic ability to complete problem solving steps. The results of this test are promising. This has led to involvement of the IBF in the International Life Skill Study (ILSS) for which IBF is working out new projects.

The main advantage of the project-approach-instrument is that it measures traits of problem solving competence without significant overlap of other concepts like achievement and intelligence, although a general ability factor explains a great amount of the test results. The relatively simple administration of the scoring on the items, due to the closed item format is another advantage. However, the instrument has serious shortcomings as well. In the first place, the test is structured so much, that we may question our selves, whether the test really measures the problem solving competence as a CCC. The theory on problem solving stresses the importance of the activity of organising the problem solving process by the problem solver. In this test, the test developer has already done a big part of this job and not many activities are left for the pupils. Metaphorically speaking, the pupils are prompted the sequence of the stages and more or less pointed in the direction of the solution. It is likely, this instrument measures only a part of problem solving competence, which corresponds with the intention of the test developers (see the paragraph above). Moreover, the definition of problem solving on which this instrument is based, does not perfectly correspond with the definition, we described in 3.1. The test is not intransparent, because the information needed is offered in the booklet. Also, the test is not dynamic, because the problem cannot change during the problem solving process.

5.1.5.3 *“Plan a trip”*

In the feasibility study of the OECD (1997) problem solving competence was measured by a test called ‘Plan a trip’. ‘Plan a trip’ is a paper and pencil test and consists of one integrated task. The test intended to measured communicative skills as well.

In ‘Plan a trip’ students are asked to organise a trip for a youth club. The students get information about several alternative activities, transport, costs and the composition of the group. With a first small set of multiple choice question about the task, understanding of the task is measured. The main questions are open questions. The first core question asks students to make a list of consideration they would take into account before they start the actual task. The second core question invites students to write a report. Both questions correspond with stages of the problem solving process. The questions imply that students make a problem representation, by reading and ordering the information and set out the problem situation, plan the solution and execute and evaluate the plan. It is difficult to say, whether students already run through these stages, when they make a list of consideration or whether they are still in the middle of the process, when they are writing the report. The test concludes with multiple choice questions about motivation, perception of time and difficulty. The administration of the multiple choice questions speaks for itself. For the administration of the two core questions a general and holistic coding schema was used.

In ‘Plan a trip’ students have to organise their problem solving process by themselves. This is an important advantage of the instrument. The problem definition, on which the test is based implicitly, has similarities with the problem definition of problem solving as a CCC. The problem addresses to a daily-life situation and has more than one possible solution and outcome. However the problem does not change dynamically during the problem solving process and is not intransparent, because all the relevant information is given in a booklet.

The feasibility study showed serious problems with the administration of the test. Although judgement criteria were set, the scoring of the test remained subjective, differed between countries and was very time-consuming. The test itself appeared to be difficult. Many students, especially the students in the lower tracks could not complete the task or did not even start with the core question, because they did not understand the task or, the task did not appeal to them. The consequence of this was a lot of missing cases.
We may conclude that in principal the instrument is promising, but in this stage not applicable. The instrument needs more development work, especially in regard to the difficulty of the task and the administration.

5.1.5.4 *Questionnaire Life Skills Study (O’Neill, 1998)*

The University of California / CRESST is developing an instrument on problem solving for the Life Skill Study. The instrument is in a developing phase and is not empirically tested yet. The instrument is meant for persons from 16 to 65 years old and is especially aiming at measuring problem solving competence needed in the workforce. The test is a paper and pencil test and consists of one individual task. The test measures three components of problem solving:

Content Understanding	
Problem Solving strategies:	- domain dependent
	- domain independent
Metacognition:	- planning
	- self monitoring

According to CRESST the best way to measure problem solving is a thinking aloud protocol. This way of measuring is not feasible mainly due to the high costs. In order to stay close to this way of measuring CRESST proposes to measure the content understanding (similar to domain knowledge) by asking persons to draw a concept map. To measure problem solving strategies persons are asked to write an essay in which they explain how they solved the problem of the task. Both measurements have an open item format and need to be judge by professional coders afterwards. Metacognition will be measured by questions with a closed item format

It is difficult to give an opinion about an instrument, which does not exist yet. A positive element of this instrument is the open item format, which lets persons free to plan the problem solving process and to choose problem solving strategies by their own. However, the definition of problem solving stays unclear. It seems that this question is left open by purpose, so the characteristics of the problem can be manipulated for a specific situation. In principal the problem can be manipulated in such a way, that it corresponds with the problem definition we use (daily- life, complex, intransparent, dynamic, polytely). We do not know how the CRESST will concretise the instrument, but we have the feeling that it develops into a domain specific test, due to the purpose of the International Life Skill Study. Only metacognition will be measured with domain independent items. In Germany this approach will be worked out and tested by Eckhard Klieme of the Max Planck Institute Berlin in the frame of the national part of PISA study of the OECD.

The CRESST instrument has several similarities with ‘Plan a trip’ especially in regard to the administration of the test. Therefore, it is likely that the administration of this test will suffer the same problems as the administration of ‘Plan a trip’.

Again, we have to conclude that this instrument is not applicable at this stage for measuring problem solving as a CCC. Firstly, only the shape of the instrument is available now. The actual task and the judgement schemes still have to be developed and it is uncertain if the actual task (problem) will fit in our framework. Secondly, the instruments aims at a target group differing in age from 16 to 65 years old. This is in contradiction with the target group of the OECD, which has set the age on 16 years old (actually 18 years old is the most preferable). Thirdly, the administration of the test will probably be problematic. Results of the national PISA study in Germany will show if these conclusions will hold.

5.1.5.5 Complex Problem Solving

Measuring CPS needs computer presentation because the dynamic aspects of a situation consisting of many variables cannot easily be reduced to paper-pencil situation. Instead of being interested in single-step decisions, research on CPS concentrates on a series of inter-related decisions showing side-effects, time-delayed effects, etc. The necessity to use computers leads on the one side to an increased effort during assessment but also - what seems to be important - to an increased motivation and acceptance for pupils (or human subjects in general). Very important: use of computers would allow for adaptive testing (branched testing) allowing adapting item difficulty to person ability and, thus, producing better measurements. This would solve the problem to develop a totally new task, when the original task showed to be too difficult for a target group, as was the case with the ‘Plan a trip’ task

In concreto Funke (1998) suggested to construct a set of finite state automata which represent digital black-boxes on varying levels of difficulty and confront students with them. A finite state automate is a system where you can switch between different states by means of a simple button press. Examples of such a system are a ticket machine and a software application. Funke argues that this sort of systems have the feature of CPS (complex, intransparent, dynamic). The task of a test with finite state automata is to find out how the system works (identification) and produce certain states of the system (control). Identification and control could be measured by measuring how long a student needs to identify the system and to produce certain states and how correct and complete the identification and the produced states are.

The state of this ‘instrument’ is more or less the same as the CRESST instrument. We do not know which concrete problems will / can be put in the black-box; how the exact administration must be organised and how feasible this approach is in regard to costs. A lot depends on how fast future developments in computer technology and computer use will go. Probably already within the next 10 to 15 years it is feasible to assess problem solving with computers on a large scale.

Although many uncertainties exist about this new way of measurement, it is theoretically an appealing ‘instrument’ in the light of CCC. The instrument corresponds with the definition of problem solving as a CCC. The concrete problem can vary in difficulty, which overcomes the disadvantage of ‘Plan a trip’.

According to Funke (1998) the ‘instrument’ taps important areas of future competence. In real life these systems are not restricted to technical ones. Besides the identification and control of social systems, learning and understanding the grammar behind these interaction, are competencies with an utmost future aspect.

5.1.6 Conclusions

It seems to us some specific points appear from the theory and the empirical instruments. Firstly, there seems to be a general consensus about the conceptual framework of problem solving: the information processing approach. Most researchers seem to agree upon viewing problem solving as a <u>cognitive process</u> with <u>several stages</u>, for which a person uses it’s <u>working, short term and long term memory</u>.

Secondly, elaborations of these elements differ enormously among researchers. Researchers vary in definition of problem and problem solving at least on two points. The first point is the debate whether problem solving competence is a domain specific competence or general competence. The second point is whether a problem for measuring problem solving should be simple or complex. In figure 1, we mention the most important characteristics of a simple problem versus a complex problem

Figure 1

SIMPLE PROBLEM	COMPLEX PROBLEM
- static	- dynamic
- transparent	- intransparent
- few variables, simple connected	- many variables, difficult connected
- well defined problem state and goal state	- ill defined problem state and goal state
- one correct solution/outcome	-more than one correct solution/outcome possible

Since we took the society as starting point, we view problem solving as a competence needed in daily life problem situations. A daily life problem can occur in many different settings and to solve such a problem an individual needs a problem solving competence which can be used in all these different settings. Therefore we argue for an instrument which measures a general problem solving competence . Subsequently, it is likely that a daily life problem will usually correspond with the characteristics of a complex problem. The problem in the final instrument should meet these characteristics as much as possible.

Thirdly, a good instrument to measure problem solving as a CCC does not exist yet, because most instruments do not meet the definition of problem solving as a CCC. In general, we have two sorts of instruments to measure problem solving. Most instruments to measure problem solving are based on a paper and pencil test. Very few instruments measure problem solving by using computers. This is logical when we realise that assessment with computers on a large scale is not feasible yet. This might change in the future.

The existing paper-and-pencil instruments differ in their proportion of closed and open question. We described three paper and pencil tests in chapter three. One test contained only questions with a closed item format. The other two tests contained both, but the core questions had an open item format.

A serious disadvantage of the closed item format in a problem solving test is that the test loses the planning activities and directs the student in the right direction, as is the case in the "project approach" by IBF. This disadvantage touches the heart of the instrument and can only be tackled by reviewing the structure and content of the instrument itself. A possible alternative might be to restructure the test in such a way that the problem solver gets more space to plan the process by itself. How this exactly should take shape is a task for the near future. However, when we assume that schools will have a number of computers for educational purposes, one might think of a paper and pencil task in a menu structured programme on the computer. The programme should let students free to follow their own problem solving route, e.g. by making for each stage an icon a student can click on. Behind the icon the student finds tasks and/or question. Information gathering could be done through links with the Internet. We admit that this is just a preliminarily idea, but it might be worth thinking about it.

From plan a trip we know that the main disadvantage of questions with an open item format is that the written text has to be judge by (professional) coders. Although judgement schemes are used, the results are very vulnerable for subjectivity and cross-national bias. More specific judgement schemes, training sessions of the coders and computing inter-reliability may encounter these difficulties. However, this will not make the administration less time-consuming.

Because the 'black-box' instrument is a relatively new idea and is not properly tested, yet, we cannot evaluate the instrument as we did the paper-and-pencil instrument. Nevertheless, the idea of a black-box seems very promising, mostly because the test can be adjusted to the definition of problem solving as a CCC. Therefore, it deserves an endeavour to develop the instrument further and test it on the students of 16 years old.

This chapter began with an enumeration of characteristics of problem and problem solving, which should ideally be included in an instrument to measure problem solving as a CCC. In the ideal situation, we want to measure a <u>general</u> problem solving competence, by using a daily life problem, which is mostly <u>complex</u>. This general competence for a great part consists of the competence of divergent thinking.

As we argued in section 3.1 it is most likely that the ideal situation is not feasible and that choices have to be made which characteristics of a problem are most essential and should at least be maintained in the instrument and which are less worth keeping.

5.1.7 References

Bongers Y., J. Kleine, S. Waslander & J.L. Peschar (1996) *Mag het ietsje meer zijn? Het ontwikkelen van onderwijs-indicatoren voor Cross-Curriculaire Competenties; een vooronderzoek in Nederland*. Groningen: University of Groningen, Department of Sociology.
Calvert, Carol & John MacBeath (1995). The Findings. In: OECD. *Public Expectations of the Final Stage of Compulsory Education.* (p.65 – p.103). Paris.

Dominowski Roger L. & Bourne, Jr. Lyle E. (1994). History of Research on Thinking and Problem Solving. In: Sternberg, Robert J. *Thinking and problem solving*. (p.1 – p.33). California: Academic Press, INC.

Fend, Helmut (1995). Personality Theories and Developmental Processes: Their Implications for Indicators of the Quality of Schooling. In: OECD. *Measuring What Students Learn.* (p.125 – p.141). Paris.

Frensch, Peter A. & Funke, Joachim (1995). Definitions, traditions and a framework. In: Frensch, Peter A. & Funke, Joachim (Eds.). *Complex problem solving: The European Perspective*. (p.3 – p.25) Hillsdale, NJ: Lawrence Erlbaum Associates.

Funke, Joachim (1991). Solving Complex Problems: Exploration and Control of Complex Systems. In: Sternberg, Robert, J. & Frensch, Peter A (Eds). *Complex Problem Solving. Principles and Mechanisms.* (p.185 – p.222). Hillsdale, New Jersey: Lawrence Erlbaum Aossiciates, Publishers.

Funke, Joachim (1998). *Learning and problem solving in dynamic task environments.* Presentation on the Network A meeting, OECD in San Fransisco. Heidelberg: University of Heidelberg

Holmes, Mark (1992). *Educational Policy for the pluralist democracy: the common school, choice and diversity.* Washington D.C., London: The Falmer Press.

Hunt, Earl (1994). Problem Solving. In: Sternberg, Robert J. *Thinking and problem solving*. (p.215 – p.232). California: Academic Press, INC.

IBF (Institut für Bildungsforschung) (1998) Assessment of problem solving. The application of the problem solving 'project approach' of IBF in the framework of ILSS. (unpublished paper) Bonn: IBF

Klieme, Eckard (1998). "Cross-curricular Competence" als Gegenstand vergleichender Bildungsstudien. Unpublished paper. Berlin: Max Planck Institut fur Bildungsforschung Berlin.

Klieme, Eckard (1998). Problem solving as a cross curricular competence. Presentation on the Network A meeting, OECD in San Fransisco. Berlin: Max Planck Institut fur Bildungsforschung Berlin.

Lubart, Todd I. (1994). Creativity. In: Sternberg, Robert J. *Thinking and problem solving*. (p.289 – p.332). California: Academic Press, INC.

Mayer, Richard E. (1992). *Thinking, problem solving, cognition*. New York: W.H. Freeman and Company.

Nickerson, Raymond S. (1994) The teaching of thinking and problem solving. IN*:* Sternberg, Robert J. *Thinking and problem solving*. (p.409 – p.449). California: Academic Press, INC.

OECD (1997) *Prepared for Life? How to measure Cross-Curricular Competences*. Paris: OECD.

O'Neill, Harry (199?). *Framework for problem solving assessment. The international life skill survey.* Draft paper. Los Angeles: University of Southern California/CRESST

Ploeg, van der, Pieter A. (1995). *Opvoeding en politiek in de overleg-democratie: democratische verdeling en normering van pedagogische autoriteit*. Utrecht, University of Utrecht.

Polya, G.(1957). *How to solve it.* Garden City, NY: Doubleday / Anchor.
Simon, H.A. (1978). Information processing theory of human problem solving. In: W.K.Estes (Ed) *Handbook of learning and cognitive processes*. Hillsdale, NJ: Erlbaum

Simon, H.A. (1979). *Models of thought.* New Haven, CT: Yale University Press.

Salganik, Laura, Dominique Simone Rychen, Urs Moser & John Konstant (1998). *Projects on Competencies in the OECD Context: Analysis of Theoretical and Conceptual Foundations*. (Draft) Presented on the Network A meeting of OECD in Brussels, October, 1998.

Trier, Uri P. (1991). *Non-curriculum bound outcomes*. Proposal presented at the Network A meeting of the OECD-INES Project on educational indicators, Paris.

Waslander, Sietske (1999). *Koopmanschap en Burgerschap. Marktwerking in het onderwijs*. Assen: van Gorcum.

5.2 Problem Solving in Vocational Training and Education

5.2.1 Point of departure

In order to prepare vocational trainees for the demands of their professions, demands which are changing at an ever increasing fast pace, the key pedagogical idea of "conveying comprehensive professional competence in action taking" is gaining acceptance throughout Europe. Professional competence in action taking, defined as the ability to cope with vocational assignments, by means of planning, executing and checking them independently, has become a professed educational aim of many curricula. To reach this goal, general interdisciplinary abilities are cultivated in addition to the comprehensive vocational qualification. These abilities, so-called "key qualifications", are intended to put future vocational graduates in a position to react flexibly to new demands. There is wide agreement that particular key qualifications such as communication skills, the ability to co-operate and problem-solving abilities are of central importance. Yet the understanding of a term such as "problem-solving ability" has not yet been the focus of discussion. The TSER network made an initiative capable of investigating precisely this aspect possible. Thus, the issue of the concept underlying problem-solving ability and its status within vocational training was given priority over the question of "assessment."

5.2.2 Method

What concept of problem-solving ability is reflected in the regulations governing vocational training, and how is this ability to be promoted within the context of vocational training? Within the TSER Network, these questions were investigated in an international-model-curriculum-analysis. At a meeting in Bonn on the 29th and 30th of October, 1998, the participating partners decided to examine the following training courses: from the commercial sphere the "office administration" course; and from among the technical trades, courses in electro-technology. The subject of the analyses was the respective course curriculum valid at the time.

In Luxembourg the curricula for training in the electro-technical fields of "energy technology" and "communications electronics" were included in the study. The German contribution comprised courses in "electro-mechanics," "electrical engineering," "electrical installation," "electrical system installation" and "information and telecommunications (IT) system electronics," as there is no direct equivalence of vocations between the countries involved. Whereas in Luxembourg training in these areas is purely scholastic[1], the courses offered in the above-mentioned fields in Germany provide for the participation in both schools and businesses. For this reason, both the federal-German-framework curriculum governing in-school training as well as the regulations for in-company training were investigated. Within the in-school training, a distinction was also made between the respective federal framework curriculum and the corresponding curricula of the states. In view of these criteria, only an exemplary study of the selected vocations was possible. This was carried out for the federal German states of Bavaria, North Rhine-Westphalia, Saxony-Anhalt and Schleswig-Holstein. The training regulations for in-company training, on the other hand, apply nation-wide, so that there are no state-specific differences in the shaping of the curricula.

The goal of this analyses was to take an inventory of vocational training curricula and to determine whether "problem-solving ability" plays a role and, if so, which one. The study was outlined by the establishment of the following key questions:

[1] Here, the scholastic training integrates theory and practice. In addition to theoretical training, the trainees receive practice-oriented training carried out in school workshops.

What are the overriding goals pursued by the curricular concept?
Is the term “problem solving” mentioned?
What is the context in which the term “problem solving” is mentioned in each case?
What measures for the promotion of the “problem-solving ability” are cited?
Which references are made to the assessment of this ability?

Using these questions, it was possible to analyse the vocational training curricula in Luxembourg and Germany. However, Austria turned out not be suited for this analysis.

5.2.3 Results

The presentation of the results is structured according to the key questions listed above.

5.2.3.1 What are the overriding goals pursued by the curricular concept?

The curricula were first investigated with regard to their overriding goals, central ideas and principles. This made it possible to view the status and the context of “problem solving” in vocational training from the perspective of the underlying educational objective. It thus became evident that professional action-taking competence / professional action-taking ability is unanimously defined in the preambles as independent and responsible thought and action. The German and Luxembourg curricula agree that, in order to attain this goal, action-oriented methods are particularly suitable. An example may be found in the provisional curricular guidelines for vocational schools in the state of Bavaria, there it is stated: “…vocational schools must base their instruction on an educational theory emphasising action orientation and geared to its specific teaching objectives ...” (p. 1).

In the North Rhine-Westphalia curriculum for training in office administration, competence in action taking is defined as the ability to cope with professional problems and tasks. “In the vocational school, the didactic point of reference for the development of competence in action taking is the capability of coping with tasks and problems…” (p. 12). Furthermore, in Schleswig-Holstein training in office administration is linked with the goal of “… comprehending and learning company internal tasks and problems, developing solution proposals and justifying possible solutions …” (p. I).

Some curricula expressly point out that action orientation in training can be achieved not only through the taking of concrete action, but also through purely cognitive operations, such as the intellectual comprehension of an action (e.g., federal German framework curriculum: electrical systems installation, 1997; preliminary curricular guidelines for the Bavarian vocational school: electrical systems installation, 1998; trial curriculum: IT system electronics, 1997; curriculum of North Rhine-Westphalia: office administration, 1994). In this context, an important role is played by problem-oriented learning situations, which in many cases are directly linked to the term “action orientation.” In learning situations, action taking comprises a sequence of action phases, derived in part from the analysis of practice-oriented problems.

In addition to the ability to take action, professional flexibility is to be promoted. This would enable the employee to better cope with the continually changing demands of work and society.

Particularly the Luxembourg curricula that was included in the study, point out that the ability to think in terms of interconnections, i.e., networked thought, is the necessary prerequisite for responsible professional action. In North Rhine-Westphalia, the ability to think in terms of models and interconnections is regarded to be one of the fundamental keynotes of vocational training in several of the technical training courses[2].

5.2.3.2 Is the term “problem solving” mentioned?

The term “problem solving”[3] is found throughout the investigated office administration curricula. The acquisition of problem-solving ability is regarded both as a general learning goal (curriculum of North Rhine-Westphalia), and as a specific learning goal in various subjects or areas, such as information / data processing (curricula of Bavaria, Saxony-Anhalt), organisation (curricula of Bavaria, North Rhine-Westphalia, Saxony-Anhalt, Schleswig-Holstein) and general economics (curriculum of Bavaria). In the German training regulations for in-company training as an office administrator, however, the term is not mentioned at all.

In the Luxembourg curricula for both electro-technical training courses, terms such as “problem solving thought,” “independent problem solving” and “the ability to solve problems” are mentioned. In the corresponding German curricula, the term “problem-solving ability” is not used with as much regularity. The term appears only twice in connection with electro-mechanics (curricula of Bavaria and North Rhine-Westphalia) and once each in the curricula for electrical installation (curriculum of North Rhine-Westphalia), electrical engineering (curriculum of North Rhine-Westphalia), electrical system installation (federal framework curriculum) and IT system electronics (federal framework curriculum).

5.2.3.3 What is the context in which the term “problem solving” is mentioned?

Neither of the various curricula supply a definition or an even more detailed description of “problem-solving ability,” which could provide insight into the concept underlying the term in each case. References to the concept of problem solving as a competence, which should be promoted during training can, however, be inferred from the respective context in which the term is used. It was thus determined that, throughout the geographical areas and the fields of training comprised by the study, problem solving is not only linked with specific subject matter but also referred to as a “comprehensive” or “key” qualification.

Regardless of subject area, the ability to solve problems is often cited in connection with the formulation of goals within the respective training. On the basis of professional knowledge and abilities: vocational tasks and problems are to be solved; the results to be judged, in a manner which is goal-oriented, appropriate to the material and to the situation in question; and correct with regard to the demands of the particular profession; methodical; systematic; and independent.

In connection with “competence in action taking,” “problem solving” is expressly understood as a part of the cognitive structure of learning processes that contribute to the development of “independent professional action.” Professional action-taking competence is defined as: “…the ability and willingness of the person to act in professional situations in a manner appropriate to the material and the field of specialisation, with careful personal deliberation and social responsibility." This includes solving current problems independently and in a goal-oriented manner on the basis of an acquired plan of action, assessing the solutions thus arrived at, and further developing one’s repertoire of plans of action (e.g. office administration curriculum of North Rhine-Westphalia, pp. 11f.). Independent planning,

---

[2] Guidelines and curricula for the electrotechnical trades: electrical engineering — electromechanics — electrical installation

[3] In addition to the term “problem solving,” the terms “solving of problems,” “problem-solving process,” “problem-solving learning,“ “problem-solving thought” and “problem-solving strategies” were taken into consideration.

execution and checking / independent problem solving / independent mastering of tasks is mentioned in all of the training regulations for in-company training and in several scholastic curricula.

A goal of training is the ability to apply the acquired knowledge and skills to similar and new problems, and to solve problems in various ways and with the employment of various aids. This goal creates a connection between problem-solving ability and the capacity for (knowledge) transfer. On the basis of familiar situations, the trainees should be able to adapt themselves to changed demands and to apply what they have learned to the treatment of new problems arising in practice. This goal is to be attained by conveying exemplary solution strategies or forming analogies, and by placing specific circumstances into a larger context (e.g. federal German framework curriculum: electrical engineering, p. 17; electro-mechanics, p. 64; curriculum of North Rhine-Westphalia: office administration, pp. 12f.; Saxony-Anhalt, p. 84; curriculum of North Rhine-Westphalia: electro-mechanics, pp. 25f., etc).

In the training curricula for electrical system installation and IT system electronics, problem solving is regarded as an aspect of "field-specific competence." Field-specific competence is defined here as "…the willingness and ability to solve tasks and problems; judge the results on the basis of professional knowledge and in a manner which is goal-oriented, appropriate, methodical and independent." (IT system electronics, p. 64; electrical system installation, p. 14). Problem solving is further connoted by "procedural competence." Here, procedural competence is defined as the application and further development of working procedures / solution strategies for coping with tasks and problems.

Particularly in the Luxembourg curricula, problem-solving thought — which is described as a comprehensive qualification — is linked throughout the selected vocations with competence in "thinking in terms of systems" and competence in "networked thought." Within the framework of information acquisition and information processing, thinking in terms of systems is regarded to be an ability particularly worthy of promotion. Thinking in terms of models and systems is also referred to as a requirement for problem solving within an electro-technical context in technical trades. This connection is made in the Luxembourg curricula as well as in the curricula for training in the technical trades in North Rhine-Westphalia.

Within the framework of the curriculum analyses undertaken in this study, comparable contextual incorporations of the term "problem-solving ability" and, in this connection, repeated goal attributes were found. Yet no concrete definition of the term was supplied. One interesting exception can, however, be cited here: in the office administration curriculum of Bavaria (curriculum of Bavaria, p. 52), in connection with the learning goal designated as "the ability to systematise work routines" (organisation of routines), a "systematics of problem solving" was explicitly defined according to the following stages:

Problem definition
Goal formulation
Task definition
Identification, analysis and criticism of the actual situation
Development of nominal suggestions
Decision
Execution
Checking
Feedback

This "systematics of problem solving" comprises the stages of a complete action and thus emphasises the process character of the problem-solving strategy under consideration here. In the federal-German-framework curriculum for training in office administration, within the context of the subject "special economics," problem solving is analogously described as a process comprising the phases “problem definition, goal definition and solution possibilities, e.g., advice, brainstorming, decision, execution and checking.”

5.2.3.4 What measures for the development of “problem-solving ability” are cited?

The question as to the measures through which “problem-solving ability” is to be supported and cultivated in vocational training was a further focus of this investigation. The instructional measures cited in direct connection with problem solving / problem treatment include:
Case studies
Learning assignments analogous to the working situation or examples of vocational practice
Project work or project tasks

In all of the geographical areas and vocational fields studied, these methods adhere to the principles of action-oriented instruction, according to which trainees cope with and resolve problems and tasks typical of their vocational fields. This is a means of promoting the trainees’ general ability to solve problems / develop one’s own solution strategies systematically and independently as well as of developing their approach to problem solving in a specific vocational context, such as in work with control technologies.

As pointed out by the office administration curriculum of North Rhine-Westphalia, vocationally relevant, action-oriented situations of this kind should be created by bringing together interrelated subject matter, working methods and field-specific procedural and problem-solving strategies from a range of vocational subjects. The recommendations for integral learning stated in the corresponding Luxembourg curriculum point in the same direction.

5.2.3.5 What references are made to the assessment of problem-solving ability?

The required achievements related to the solving of problems can be classified as follows:

Linking information taken from several vocational fields or learning areas
Within the framework of testing, field-specific problems are to be analysed, assessed and suitable solutions presented by means of linking informational, technical and mathematical elements. (training regulations: electro-mechanics, p. 51; electrical engineering, p. 6; electrical installation, p. 27; electrical system installation, p. 5)

Appropriate, target-group-specific presentation of solutions
By means of a presentation, including a field-specific discussion, the examinee should show that he/she is able to: describe problems and solution concepts specifically related to his/her area of specialisation in a manner appropriate to the respective target group, indicate his/her knowledge of the background of the project work and justify his/her manner of proceeding in the project. (training regulations: IT system electronics, p. 7)

Problem-related assignments are to be resolved by the trainees in a goal-oriented and independent manner, solution processes and solutions to be presented and assessed in an appropriate way. (curriculum of North Rhine-Westphalia: office administration, p. 15)

Classroom tests should not only require the reproduction of the material learned but also provide trainees with the opportunity of developing connections between thoughts and presenting ideas. (curriculum of Saxony-Anhalt: office administration, p. 84)

Goal-oriented employment of methods
Interdisciplinary tests aim towards the solution of a specialised, field-oriented problem. The application of methods for the goal-oriented handling of the assignment is an essential aspect

for the determination and assessment of achievement. (curriculum of North Rhine-Westphalia: electrical installation, p. 15; electrical engineering, p. 14; electromechanics, p. 15)

Transfer of knowledge
On the basis of work carried out in class, trainees are to be given the opportunity — within the framework of a field-related assignment — “to undertake a transfer to more comprehensive problems; they are to deal more extensively with the problems and find new ways of solving them, and they should be able to cope with such assignments independently.” (curriculum of North Rhine-Westphalia: office administration, pp. 71f.)

In written tests, the trainees should apply the knowledge and skills they have acquired to similar and new problems. (curriculum of Saxony-Anhalt: office administration, p. 84)

Further references
“Substantiated alternative solution possibilities are to be taken into appropriate consideration in the assessment of achievement” (curriculum of North Rhine-Westphalia: office administration, p. 108). We mention this reference here because it implies that even solution possibilities, which have not been previously supplied, are accepted if they have been well thought out. This gives the trainees scope for creative thinking.

In the Luxembourg curricula investigated, there were no references to means of checking learning success in connection with problem solving.

5.2.4 Conclusion

The training goal of “professional competence in action taking” is to be attained through practical instruction oriented towards actions typical of the respective vocation. Here action orientation, the dominating principle of the curricula investigated, is consistently and closely associated with problem-oriented instruction. The demand for action orientation is evidently made with a view not (only) to routine vocational procedure / standard situations, but also to new demands and problems. This view is particularly apparent in connection with the knowledge transfer requirement. The trainees are to be able to transfer action schemata they have learned to new material, and to demonstrate the professional flexibility necessary for coping with the constantly changing demands of work and society. This close association between action orientation and problem solving is expressly stated in the office administration curriculum of North Rhine-Westphalia: “They (the learning processes) are to be structured in such a way that active, action-oriented, problem-solving learning is made possible for the trainees.” (p. 68)

No definition of the term “problem solving” could be found within the framework of the curriculum analyses under consideration here. For this reason, references to implicit conceptions of the term were sought, for example, by examining the respective context in which the term “problem-solving” is cited, as well as the attributes with which the term is more closely described.

Problem solving is referred to as a part of the cognitive structure of learning processes.
The problem-solving systematics follows the successive phases of an action, thus revealing its process character.
Competence in action taking is defined here as the ability “to solve the respective problems independently and in a goal-oriented manner on the basis of acquired action schemata, to assess the solutions found, and to further develop one’s repertoire of action schemata.” (e.g. curriculum of North Rhine-Westphalia: office administration, p. 11) Within the context of this definition, problem solving is described as a somewhat less creative learning process. This interpretation is furthermore supported by the description of problem solving as a systematic

and structured process, this description was repeatedly found in the curricula. On the other hand, the reference to the further development of one's repertoire of action schemata and the demand made on the trainees to develop their own solution means, also contain the aspect of innovation.

The question also arose as to whether problem solving should be cultivated as a competence implicitly, or expressly raised to the status of a learning topic, for the teaching of which strategies should be conveyed on a more abstract level. A passage in the federal-German-framework curriculum for the vocation IT system electronics seems to point to the latter solution by referring to the learning contents as "methods and tools for the development of a solution concept." (framework curriculum: IT system electronics, p. 68). The above mentioned "systematics of problem solving" in the office-administration-training course (curriculum of Bavaria), which identifies the problem-solving process as a learning subject, also points in this direction.

In summary, it can be observed that in numerous curricula problem solving represents an important learning goal, both in the specifically field-related and in the comprehensive sense. Nevertheless, no definition of the term can be found. Only references to the underlying conception can be gleaned from the respective context in which the term is used. Although the development of problem-solving ability is required in many training courses, only rarely are recommendations made for teaching this ability, and where they do appear they are very general. References to the determination of the problem-solving ability in examinations are made only in a few of the investigated German curricula. They describe formal requirements such as the appropriate target-group-specific presentation of solutions; or the goal-oriented employment of methods in the solution-finding process; as well as requirements with regard to content, such as: the transference of knowledge or the linking of information from several fields / learning areas.

5.2.5 References

Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst (1998). Vorläufige Lehrplanrichtlinien für die Berufsschule Fachklassen Elektroanlagen-monteur/ Elektroanlagenmonteurin. München: Staatsinstitut für Schulpädagogik und Bildungsforschung.

Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst (1995). Lehrpläne für die Berufsschule Fachklassen Elektromechaniker/Elektromechanikerin. München: Staatsinstitut für Schulpädagogik und Bildungsforschung.

Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst (1992). Lehrpläne für die Berufsschule Fachklassen Bürokaufmann/Bürokauffrau. München: Staats-institut für Schulpädagogik und Bildungsforschung.

Bundesminister der Justiz (Hrsg.). (1991). Bekanntmachung der Verordnung über die Berufsausbildung zum Bürokaufmann/zur Bürokauffrau nebst Rahmenlehrplan. Bundesanzeiger, 43, 37-72.

Bundesministerium der Justiz (Hrsg.). (1998). Bekanntmachung der Verordnung über die Berufsausbildung im Bereich der Informations- und Telekommunikationstechnik nebst Rahmenlehrplänen. Bundesanzeiger, 50, 1-113.

Bundesministerium der Justiz (Hrsg.). (1997). Bekanntmachung der Verordnung über die Berufsausbildung zum Elektroanlagenmonteur/zur Elektroanlagenmonteurin nebst Rahmenlehrplan. Bundesanzeiger, 49, 1-24.

Bundesministerium der Justiz (Hrsg.). (1988). Bekanntmachungen der Verordnungen über die Berufsausbildung zum Elektromaschinenbauer/zur Elektromaschinenbauerin; zum Elektroinstallateur/zur Elektroinstallateurin; zum Elektromechaniker/zur Elektromechanikerin nebst Rahmenlehrplänen. Bundesanzeiger, 40, 3-71.

Die Ministerin für Bildung, Wissenschaft, Jugend und Kultur des Landes Schleswig-Holstein (1991). Lehrplan Bürokaufmann/Bürokauffrau Berufsschule. Kiel.

Enseignement secondaire technique: Horaires et Programmes 1998/99. Volume VI: Métiers de l'électricité. 2. Régime de la formation de technicien. Ministère de l'Education Nationale et de la Formation Professionnelle, Grand Duche de Luxembourg.

Enseignement secondaire technique: Horaires et Programmes 1998/99. Volume IV.2: Formations administratives et commerciales. Régime de la formation du technicien. Ministère de l'Education Nationale et de la Formation Professionnelle, Grand Duche de Luxembourg.

Enseignement secondaire technique: Horaires et Programmes 1996/97. Volume VI: Formations des métiers de l'électricité. 2. Régime de la formation du technicien.

Ministère de l'Education Nationale et de la Formation Professionnelle, Grand Duche de Luxembourg.

Enseignement secondaire technique: Horaires et Programmes 1997/98. Volume IV.2: Formations administratives et commerciales. 2. Régime de la formation de technicien. Ministère de l'Education Nationale et de la Formation Professionnelle, Grand Duche de Luxembourg.

Kultusministerium des Landes Nordrhein-Westfalen (Hrsg.). (1994). Richtlinien und Lehrpläne für die Berufsschule in Nordrhein-Westfalen Bürokaufmann/Bürokauffrau. Die Schule in Nordrhein-Westfalen, Heft 4182. Düsseldorf: Verlagsgesellschaft Ritterbach.

Kultusministerium des Landes Nordrhein-Westfalen (Hrsg.). (1991). Richtlinien und Lehrpläne industrielle und handwerkliche Elektroberufe Grundbildung industrielle und handwerkliche Elektroberufe. Die Schule in Nordrhein-Westfalen, Heft 4173/4174-01. Düsseldorf: Verlagsgesellschaft Ritterbach.

Kultusministerium des Landes Nordrhein-Westfalen (Hrsg.). (1991). Richtlinien und Lehrpläne handwerkliche Elektroberufe Elektroinstallateur/Elektroinstallateurin. Die Schule in Nordrhein-Westfalen, Heft 4173-12. Düsseldorf: Verlagsgesellschaft Ritterbach.

Kultusministerium des Landes Nordrhein-Westfalen (Hrsg.). (1991). Richtlinien und Lehrpläne handwerkliche Elektroberufe Elektromaschinenbauer/Elektromaschinen-bauerin. Die Schule in Nordrhein-Westfalen, Heft 4173-11. Düsseldorf: Verlagsgesellschaft Ritterbach.

Kultusministerium des Landes Nordrhein-Westfalen (Hrsg.). (1991). Richtlinien und Lehrpläne handwerkliche Elektroberufe Elektromechaniker/Elektromechanikerin. Die Schule in Nordrhein-Westfalen, Heft 4173-14. Düsseldorf: Verlagsgesellschaft Ritterbach.

Kultusministerium Sachsen-Anhalt (1996). Vorläufige Rahmenrichtlinien Berufsschule berufsbezogener Lernbereich Bürokaufmann/Bürokauffrau. Magdeburg.

Ministerium für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig-Holstein (1997). Lehrplan für die Berufsschule IT-Berufe. Kiel.

Ministerium für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig-Holstein (1991). Lehrplan für die Berufsschule Elektroinstallateur/Elektroinstallateurin. Kiel.

Ministerium für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig-Holstein (1991). Lehrplan für die Berufsschule Elektromaschinenbauer/ Elektromaschinen-bauerin. Kiel.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (Hrsg.). (1997). Lehrplan zur Erprobung Informations- und Telekommunikationssystem-Elektronikerin/ Informations- und Telekommunikationssystem-Elektroniker. Schule in NRW, Heft 4275. Düsseldorf.

Ministerium für Schule und Weiterbildung, Wissenschaft und Forschung des Landes Nordrhein-Westfalen (Hrsg.). (1998). Sekundarstufe II Berufsschule. Lehrplan Elektroanlagenmonteurin/Elektroanlagenmonteur. Schule in NRW, Heft 4268. Düsseldorf: Verlagsgesellschaft Ritterbach.

5.3 Alternative Concepts and Tools

5.3.1 Introduction

This chapter is about alternative approaches (concepts as well as tools) to the assessment of problem solving. These approaches should possibly range from purely conceptual ideas to test blueprints or to first prototypes, implemented on computer or in paper-and-pencil format. A few examples of potential alternative approaches, which do not restrict themselves to mathematics and science skills, had been mentioned in the proposal:

Everyday planning and Problem Solving: Identify relevant and irrelevant sources of information, produce a proper sequence of timely events and look for constraints which are imposed on certain daily errands. Control emotions in case of disaster.

Script-Monitoring: Watch a video-clip showing a certain type of interaction between people and identify acceptable and unacceptable behaviour. Observe how people solve social problems and produce better solutions.

Identification and control of a complex but unknown finite state automaton: identify the correct interaction grammar and produce a certain state. This grammar could be one behind a technical device (tamagotchi approach) or behind a social interaction.

As it turned out during the project time, only the first and third mentioned issue could be further elaborated. For the second issue, one doctoral student at Heidelberg gave up her work on the video approach because of the high costs for video equipment and video production. Thus, this line of assessment had to be postponed.

5.3.2 Method

Based on previous basic research some promising concepts have been reviewed for potential use in measuring problem solving in a large-scale survey. Beyond literature analysis and conceptual work, fast prototyping was used during the first development steps. Based on evidence from small scale testing, more extensive pre-piloting of some instruments was used to prepare the use of new tools in the German PISA pilot study. Finally some experimental prototyping was used to overcome traditional measurement constraints. Because of the very *network* character of the whole work, this last effort was rather limited.

5.3.3 New assessment tools

The two really successful branches of development within this work package have been the Project Approach and the Finite State Automaton Approach. Both of them will be illustrated in more detail now.

5.3.3.1 The Project Approach[4]

Broad consensus exists within the assessment and research communities on the importance of problem solving as a cognitive ability. However, due to the enormous range of tasks and behaviours implied by the term, there is little agreement on the exact definition of problem solving and how it should be measured. Further, there is also an issue of whether or not problem solving is distinguishable enough from general intelligence/cognition to allow for measurement of a distinct competency.

[4] This section uses parts of the report given by Ebach, Hensgen, Klieme, and Sidiropoulou (1999).

In attempting to meet these challenges, the Problem-Solving Framework of ILSS takes a "project approach" to assessing problem solving that focuses on the cognitive competencies needed for solving complex tasks. These tasks are embedded in a "project" which is itself constructed in accordance with the individual steps of the problem-solving process, including (1) defining the goal, (2) analysing the situation, (3) planning the solution, (4) executing the plan, and (5) evaluating the result. Recognising the situational nature of problem solving, these tasks, or projects, are placed in readily identifiable contexts. The situation, though new, is meant to be conceivable for the participant such that the participant solves tasks, which might occur in a similar situation in real life.

"Projects" contain a description of the problem situation, the participant's role in the situation, the participant's task, and a listing of the steps in the work to which later items refer. The items are designed to address the five areas of problem solving outlined above. Multiple-item formats may be used in this approach, including multiple-choice items, locating and identifying requested data, ordering items, correspondence items, and free-response items.

While the "project approach" to problem solving is not designed to yield sub-scale scores, previous experiences with the measurement approach have demonstrated three levels of proficiency for score interpretation: identifying information, ordering and evaluating information, and analyzing dependencies.

There are some limitations of the project approach because some aspects cannot be measured within this approach:

The "dynamic" aspects of task regulation (continuous processing of incoming information, coping with processes that cannot be influenced directly, coping with feedback, and handling of critical incidents) has to be addressed by computer simulated tasks (CPS; see Finite State Automata).

The motivational, affective, and self-regulatory aspects of task regulation might be addressed implicitly by problem-solving tasks or explicitly by questionnaire methods.

Problem-solving behaviour within this test will depend on general, context-specific, domain-specific and situation-specific processes. Nevertheless, even though not yet proven, it is highly reasonable to assume that a general (latent) problem-solving competency can be identified.

Nevertheless, even though not yet proven, it is highly reasonable to assume that a general (latent) competency for analytical reasoning (as an essential part of problem solving) can be identified within this approach.

5.3.3.2 Finite State Automaton Approach[5]

Within the "Program for Student Assessment" (PISA), the TSER consortium was successful with respect to the implementation of the "Finite State Approach" into the German national pilot study (about 600 German pupils in the age of 15, from about 5 different school types).

Based on the formalism of finite state automata, Buchner and Funke (1993) have proposed to construct new instruments for assessing complex problem solving. This type of problems can be easily implemented in hand-held palm-tops. Such a device could simulate a complex dynamic machine which has to be identified and controlled by a subject (because of the

[5] Because the FSA approach is part of the German National PISA project, we are not allowed to go into detail too much due to publication restrictions posed on us by the contractor to keep ongoing assessments fair.

similarity between this task and a well-known toy product we call this the „Tamagotchi" approach). The degree of correct identification and the closeness of a shown control activity to a given goal value offers a new approach to the assessment of technical problem solving.

The following figure shows the interface of such an automaton to the subjects. Their task is to find out how this space shuttle works and later on to control this shuttle and reach certain goal states as quickly as possible.

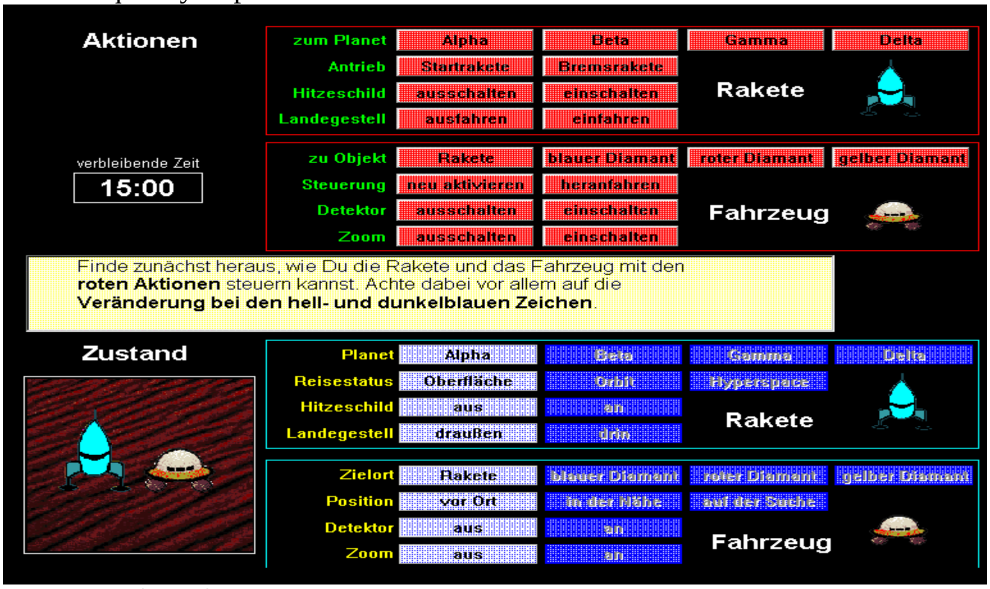

Figure. Interface of the Space Automaton (German PISA pilot study)

In the PISA German national pilot study from summer 1999, a prototype was developed and tested in schools. Based on data from more than 200 students, the psychometric qualities of the new instrument turned out to be acceptable as well as the data on validity do.

5.3.4 Next steps

Further work in the area of new assessment tools is strongly needed. With an increasing need for computer-based instruments (due to the need for process tracing and because of ease of data collection), much more effort has to put into these issues. Also, the whole area of technology-based assessment (TBA) waits for creative ideas. One of the strongest arguments for innovative steps in assessment comes from the use of new technologies. Mainly, there are four reasons to go into this direction:

Technology based assessment allows the creation of complex, dynamic, realistic, interactive stimuli in the assessment situation. This is a mayor improvement to paper-pencil-based assessments, which are much more restricted in what they can present to the subjects. Think of all environments, which require interaction between certain devices and the subject: only through proper use of the adequate technology such situations can be simulated within the assessment situation. Therefore, validity of measurement can be increased dramatically.

Process tracing. A major disadvantage of current assessment instruments is their result-orientation. Result-orientation is a problem insofar as in many areas the process quality is at least equally important as is the result quality. Technology-based assessment allows for process-tracing in a way which has not been available up to now. This is also one of the reasons that new psychometric models are needed and therefore co-operation with basic researchers from the area of measurement and test theory is needed urgently.

Adaptive testing. In standard assessments, items have to be selected which cover the full range of difficulty within the domain because one has to be prepared for measuring subjects with a high level of performance as well as for those with a low level. Consequently, subjects with high abilities see the simple items as boring, where subjects at the lower end can solve only a small subset of items and are frustrated by many items which are above of their ability. Technology-based assessments offers with the feature of adaptive testing the possibility to stay with the item selection near to the subjects‘ performance level, therefore proving subjects with items which are neither boring nor frustrating.

Cost effectiveness in the long run: This argument might be astonishing because at the first look a huge financial investment is necessary in order to establish a technology-based assessment environment. But besides of this hardware costs, there is the tremendous advantage of automatic data collection, automatic data analyses, and even automatic reports. Also, changes within such an environment, if it is once established, can become implemented in very short time cycles.

Besides these four reasons, there might be additional factors in favour of a technology-based assessment approach, for example, the increased acceptance of this type of measurement by the testees, compared to the traditional paper-pencil-tests. Subjects as well as „consumers“ of the assessment data (i.e., policy makers, managers) accept those new measurement approaches because of the assumed higher validity.

When talking about TBA, we think of at least four major areas of „technologies“.

Simulations. As mentioned before, simulations are an important step to increase validity of the assessment situation though not every simulation is also a good one with respect to assessment purposes (Funke, 1998).

Video. Because of the high context sensitivity of our human cognitive system and because of the high importance of the visual information channel, video-based assessments are very promising (Schuler et al., 1993).

PC networks. There is growing evidence that people at the work place don’t work by themselves but have to co-operate with other members within the organization as well as with people from outside the company. To assess these competencies the use of PC networks in an assessment situation seems to be very promising to realize collaborative problem solving tasks. Specially, with respect to command and control decision making (see Omodei et al., in press), such tasks would allow not only to assess an individuals contribution to a group task, but also to analyze the effectiveness of an individual with respect to different control structures (hierarchic, distributed, etc.).

Virtual reality. This technology is – due to the necessary effort and the still high costs – the least explored one with respect to assessment. Up to now, we don’t know any VR-based procedure for assessment, most of the scenarios currently in use are for learning and training purposes (Schank, 1997). This technology seems to be the most interesting and most promising because of its flexibility on the one side and because of its close approximation to reality on the other side.

5.3.5 Conclusion

The work done within this work package has been from our point of view very successful. We have put two new instruments from within the European TSER consortium into discussion for two big international large-scale studies (OECD-PISA and ALL). These steps have to be followed to consolidate the work done. Also, at the same time, further steps towards innovation are strongly needed. The intelligent use of TBA methods seems to be promising and leads far ahead compared to current paper-pencil-based assessment methods.

5.3.6 References

Ebach, Judith, Hensgen, Anne, Klieme, Eckhard & Sidiropoulou, Eftychia (1999). *Analytical reasoning in real worls contexts*. Framework presented to ILSS.

Frensch, Peter A. & Funke, Joachim (1995). Definitions, traditions and a framework. In: Frensch, Peter A. & Funke, Joachim (Eds.), *Complex problem solving: The European Perspective* (pp.3 – 25). Hillsdale, NJ: Lawrence Erlbaum Associates.

Funke, Joachim (1998). Computer-based testing and training with scenarios from complex problem-solving research: Advantages and disadvantages. *International Journal of Selection and Assessment*, *6*, 90-96.

Funke, Joachim (1998). *Learning and problem solving in dynamic task environments.* Presentation on the Network A meeting, OECD in San Fransisco. Heidelberg: University of Heidelberg

Hunt, Earl (1994). Problem Solving. In: Sternberg, Robert J. (Ed.), *Thinking and problem solving* (pp. 215 –232). California: Academic Press, INC.

Klieme, Eckard (1998*). "Cross-curricular Competence" als Gegenstand vergleichender Bildungsstudien*. Unpublished paper. Berlin: Max Planck Institut fur Bildungsforschung Berlin.

Klieme, Eckard (1998). *Problem solving as a cross curricular competence*. Presentation on the Network A meeting, OECD in San Fransisco. Berlin: Max Planck Institut fur Bildungsforschung Berlin.

Mayer, Richard E. (1992). *Thinking, problem solving, cognition*. New York: W.H. Freeman and Company.

O'Neill, Harry (1998). *Framework for problem solving assessment. The international life skill survey. Draft paper*. Los Angeles: University of Southern California/CRESST

OECD (1997) Prepared for Life? How to measure Cross-Curricular Competences. Paris: OECD.

Omodei, Mary, Wearing, Alex & McLennen, Jim (in press). Relative efficacy of an open versus a restricted communication structure for command and control decision making: An experimental study. In C. McCann & R. Pigeau (Eds.), *The human in command*. New York: Plenum.

Schank, Roger (1997). *Virtual learning. A revolutionary approach to building a highly skilled workforce*. New York: McGraw-Hill.

Schuler, Heinz, Diemand, A., & Moser, Klaus (1993). Filmszenen. Entwicklung und Validierung eines neuen eignungsdiagnostischen Verfahrens. *Zeitschrift für Arbeits- und Organisationspsychologie, 37,* 3-9.

5.4 Assessment of Cross-Curricular Problem-Solving Competencies in Large-Scale Surveys.

Problem solving is highly popular in the human ressources literature as well as in educational policy, pedagogics and even in curriculum development. It is ranked as an important key qualification by labour market experts (see Binkley, Sternberg, Jones, & Nohara, 1999) as well as in the literature on vocational training and education (Didi, Fay, Kloft, & Vogt, 1993). It was classified as an important outcome of schooling by OECD experts (OECD, 1997), and it is stated quite often as a high-level aim in curricula (see, e.g., Svecnik, 1999). Likewise, problem solving is highly relevant for a lot of – if not all – school subjects, e.g. for mathematics, since "solving problems is an activity at the heart of doing mathematics" (De Corte, Greer, & Verschaffel, 1996).

But what kind of relation exists between solving problems of a certain kind, e.g. mathematical problems, and problem solving as a generalized capability? Although problem solving seems to be crucial for academic, professional and everyday performance, it is quite unclear what exactly it means to define problem solving as a general competence or life skill, and how it might be assessed in large scale assessments, especially in contrast to domain-specific competencies like mathematics or science literacy. The present chapter tries to review the state of the art with regard to these issues.

Since there is a considerable body of research on problem solving in general cognitive psychology as well as in differential psychology, the first and main part of the paper defines problem solving in psychological terms and discusses related questions such as the domain-specificity of problem solving and its relation to intelligence. The second part discusses the notion of problem solving as a cross-curricular or non-curriculum-bound competence and outlines a framework for problem solving assessment within educational evaluation and systems monitoring. The third and last part gives an overview of existing instruments, including a report on the German national enhancement to the Program of international student assessment (PISA) which piloted and validated a broad set of problem solving instruments.

5.4.1 Method

Drawing upon the work described in the previous chapters, a more refined literature analysis has been done. Furthermore, both literature analysis and conceptual work followed strictly an assessment rationale. Different elements from previous work have been integrated with regard to the scientific validity of the underpinning concepts and the constraints to be found in large-scale international surveys. As one of the ambitions of the network had been to lay the ground for a problem solving assessment in PISA, particular attention has been given to the specific needs of problem solving assessment in an educational setting. Beyond literature analysis and conceptual and development work, instrument piloting and data (re-)analysis have been core methods of the work described in this chapter.

5.4.2 The concept of Problem Solving

5.4.2.1 Definition and psychological modeling

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*.

This definition is based on the understanding of *problem solving* as it is established in psychology (Hunt, 1994; Mayer, 1992; Mayer & Wittrock, 1996; Smith, 1991). The definition is relatively broad, but makes it clear that problem solving is to be understood as a cognitive process. This is in contrast to the everyday understanding of the term or to the clinical psychological concept in which "problem solving" is associated with the resolution of social and emotional conflicts. We are concerned with cognitive-analytical problem solving only. Here, too, the social context plays a role – when problems have to be approached interactively and resolved co-operatively, for example – and motivational factors such as interest in the topic and task-orientation influence the problem-solving process. However, the quality of problem solving is primarily determined by the comprehension of the problem situation, the thinking processes used to approach the problem, and the appropriateness of the solution.

Our definition encompasses a wide range of problem-solving processes:

The *scope* of a problem can range from working on limited parts of a task to planning and executing of actions and even to extensive projects.
The *context* can reflect different domains (school subjects, topic areas, areas of experience), which may be of a theoretical or a practical nature, related to academic situations or to the real world. In these domains, problems can be more or less authentic. The 'real world' context may be related to school life, out-of-school activities of the students, games, adult life or professional situations.
The problem can have an open or a closed format, a well-defined or an ill-defined goal, transparent (explicitly named) or intransparent constraints, and involve a few isolated elements or numerous interconnected ones. Altogether, these different features determine the *complexity* of the problem.

Following Pólya (1945, 1980), the process of problem-solving has frequently been described in terms of the following stages:
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, and
5. evaluate the result.

Such a sequential model may be of use for training purposes or for the implementation of problem-solving heuristics in practical life. However the cognitive processes, which are activated in the course of problem solving are certainly more diverse and more complex, and they are most probably organised in a non-linear manner. Among these processes, the following components may be identified:

Searching for information, structuring and integrating it into a mental representation of the problem („situation model“),
reasoning, based on the situation model,
planning of actions and other solution steps,
executing and evaluating solution steps,
continuously processing external information and feedback.

Baxter and Glaser (1997) present a similar list of “cognitive activities”, also labeled “general components of competence in problem solving”: problem representation, solution strategies, self-monitoring, and explanations.

Psychological models of these processes and the mental structures (representations) on which they operate have changed over the history of psychology, each being tailored to certain kinds of problems that were focused by the respective research paradigm. In the early years of cognitive psychology, for example, “insight” was seen as a major mechanism. This notion was appropriate in limited, but ill-defined problem situations, where a sudden restructuring or reinterpretation could solve the problem. Newell and Simon (1972), in their seminal book “Human problem solving” which served as a framework for numerous studies in information processing psychology and artificial intelligence, described problem solving as a process of search in a "problem space" consisting of states (including given state and target state) and operators. This model was appropriate for the study of well-defined, "puzzle"-type problems. While Newell and Simon believed they had discovered rather universal mechanisms, research on scientific reasoning and expertise later proved that problem solving heavily depends on the use of domain-specific knowledge which was described in terms of rule systems, schemata, mental models or "mental tools” (see, e.g., Chi, Glaser, & Farr, 1988; Weinert & Kluwe, 1987). At the same time, it became clear that metacognition (defined as the *process* of planning, monitoring, evaluation and regulation of ongoing cognition as well as the *knowledge* and *beliefs* about cognitive functioning) is of crucial importance to process and product of problem solving activities (Brown, 1987; Flavell, 1976).

In order to find out how well students can solve particular types of problems, it is not necessary to identify mental structures or process components in detail. Assessment frameworks need not meet the sophistication of cognitive-psychological models. However, even a purely functional approach to problem solving assessment has to take into account some important results of psychological research, associated with the key terms „general intelligence”, „complex problem solving“, and „domain specificity“. In the following, we will give a short review of these findings and discuss consequences for the design of problem solving assessments.

5.4.2.2 Problem Solving, reasoning and related constructs

Problem solving as it was defined above is quite similar to some other constructs in modern psychology. Among them are critical thinking (Ennis, 1996; Norris, 1989) which mainly means judging the credibility of arguments, and naturalistic decision making (Zsambok & Klein, 1997), defined as the use of knowledge and expertise to act under complex and uncertain conditions. Each of these constructs describes some kind of intellectual activity, based on reasoning and the application of knowledge. Therefore, they are closely linked to the construct of intelligence which by many modern psychologists is understood as a generalized capability to acquire, integrate, and apply new knowledge. Intelligence in turn is linked to more basic features of the human information processing system such as working memory capacity or mental speed (Neisser, 1996).

In the tradition of psychometric research, the core of general intelligence is named *reasoning* (Carroll, 1993) or *information processing capacity* (Süß, 1999).. It is operationalized by tests using mathematical word problems, number series (e.g. 1, 2, 4, 7, 11, ...?), and analogical reasoning, in particular by figural analogies like "/ is to \ as # is to ... ?". All these may be subsumed under the broad concept of *problem solving* as it was defined above – with the exception of rare cases in which highly trained persons solve such tasks using special algorithms. Thus, whatever indicator for *problem-solving competence* we use, it will be correlated to psychometric measures of *reasoning ability*. How strong this correlation is and hence the extent to which *problem-solving competence* can actually be distinguished from *reasoning*, is an open question in cognitive-psychological research. Even with respect to complex, dynamic, computer-based problem-solving tasks (Frensch & Funke, 1995); see section 1.3 below) several studies suggest that inter-individual performance differences can be fully explained by reasoning ability and basic features of the human information processing system (Süß, 1999).

Another perspective to this controversy is taken in recent publications from Differential Psychology, namely by Robert Sternberg and his associates (see, e.g., Sternberg & Kaufman, 1998). Sternberg supports a very broad concept of *intelligence*, basically equating it with problem-solving abilities. He identifies three subcomponents of intelligence: a) analytical abilities such as "identifying the existence of a problem, defining the nature of a problem, setting up a strategy for solving the problem, and monitoring one's solution process", b) creative abilities "required to generate problem-solving options", and c) practical abilities needed to apply problem-solving strategies to real-life tasks. Sternberg assumes that practical intelligence is clearly discernible from intelligence assessed by means of the classic psychometric measures (IQ). However, a method for measuring practical aspects of intelligence and problem-solving abilities independently has yet to be devised. In computer-based complex problem solving, the practical and the analytical presumably mix. The procedures Sternberg himself proposes to measure practical intelligence cannot be regarded as performance tests: He presents respondents with descriptions of real-life or job-related problem situations and asks them to evaluate different response alternatives. If the evaluations of the respondent correspond to those of a reference group ("experts" in an occupational field or representatively selected control groups for real-life problems), the respondent is said to have "tacit knowledge", which Sternberg sees as the core of practical intelligence (Sternberg & Wagner, 1986).

The assessment of the third aspect in Sternberg's triarchic concept of intelligence — creativity — appears to be just as difficult, going by the present state of research. As problem solving emphasises new situations, which cannot be routinely dealt with, it always requires a certain degree of creativity. Attempts to measure creativity independently (as originality, flexibility and fluency of ideas; see (Krampen, 1993) or to assess it as a distinctive feature of problem-solving performance (Mumford, Supinski, Baughman, Costanza, & Threlfall, 1997), however, have not yet yielded convincing results.

In accordance with this research, we recommend that a framework for problem solving assessment should cover practical as well as analytical components. Some of them will be very close to psychometric reasoning, while others – especially strategy indicators from computer-based complex problem solving as discussed in section 1.3 - may prove to make unique contributions to the cognitive profile of students. Work done in related areas such as critical thinking might be helpful when instruments have to be developed or adapted for use in large scale assessment.

5.4.2.3 Addressing the complexity and dynamics of problem solving

In recent years, psychological research on problem solving has turned to increasingly complex, authentic problems with a broader scope (Sternberg & Frensch, 1991). It is no longer concerned with well-defined "puzzles" (in the extreme case, reasoning tasks as used in psychometric tests of human intelligence) which can be solved by the application of suitable operations. Instead, it addresses the thinking of experts in scientific and professional domains (Reimann & Schult, 1996; Zsambok & Klein, 1997), planning and problem solving in real-life contexts (Funke & Fritz, 1995; Jeck, 1997; Lave, 1988) , and the understanding and control of complex ecological, economic and technical systems (Dörner, Kreuzig, Reither, & Stäudel, 1983; Frensch & Funke, 1995).

Computer simulation has proven to be an important tool for investigating complex problem-solving performance. In interaction with the computer, the problem solver explores the simulated system, generates and tests (more or less systematically) hypotheses about relationships and regularities, acquires knowledge, and is finally able to govern the system by purposeful intervention. Systems used include realistic simulations of highly interconnected ecological or economic systems (Dörner, Kreuzig, Reither, & Stäudel, 1983); for school contexts see (Leutner, 1992), systematically constructed, discrete, smaller-scale systems ("finite state automata"; Buchner & Funke, 1993) and virtual experimental environments. From the perspective of educational psychology, such procedures can be understood as environments for discovery learning (Boshuizen, van der Vleuten, Schmidt, & Machiels-Bongaerts, 1997; Leutner, 1992). From the perspective of problem-solving research, these instruments provide a new quality of problems, distinguished by high levels of complexity and, in particular, by a dynamic character. These dynamic tasks have three advantages over static paper-and-pencil tasks:

The *demands* of the tasks are enhanced by an *active search and continuous processing of external information and feedback*. In solving written problem-solving tasks, it is also possible to apply, evaluate and — if necessary — modify processing strategies. Interaction with the computer, however, makes such a course of action inevitable.
In this medium, the *problem situation* can be made much more *authentic* than in a written test.
Not only the results, but also the course of the problem-solving process can be recorded and assessed, i.e., the type, frequency, length and sequence of interventions made by the subjects. This provides *process-based indicators of problem-solving strategies*.

These three advantages are mutually independent, and demonstrate the benefits of using computers in the assessment of problem-solving performance. There are, however, serious theoretical and methodological problems when it comes to the measurement of "strategies". The definition of such measures, their reliability, the extent to which they are comparable across different simulated systems, and the impact of motivational factors are research questions which are not yet fully answered. The advantages as well as the disadvantages mentioned above also apply to so-called performance assessment, where students engage in hands-on activities or experiments (Linn, 1994). In fact, computer-based discovery tasks may be understood as a variant of performance assessment.

Mumford et al. (1998) suggested that the degree to which problem solving is focused on salient and eventually contradictory features of the problem situation may be an important indicator of strategic behaviour. However, their attempt to operationalize this kind of strategy in paper-and-pencil-tasks resulted in rather low correlations between strategy indicators and external criteria. Veenman and his associates (1997) proposed a set of indicators for *metacognitive skillfulness*, namely orientation, systematical orderliness, evaluation and elaboration, based on extensive qualitative analyses of thinking-aloud-protocols. For these indicators, Veenman reports rather strong predictive power with regard to academic success.

Most other research on metacognition has focused on reading or it has been based upon self reported learning strategies (for an adaptation to large scale assessment, see Baumert et al, 1998). Problem solving assessment, especially the evaluation of processes and strategies, overlaps with these paradigms, but it is based on behavioral data, i.e. on problem solving performance rather then on questionnaires. Our national framework for problem solving assessment (Klieme, Funke, Leutner, Reimann, & Wirth, in press) aims at identifying strategy indicators by automatic analysis of the testee's operations within computer-simulated scenarios (see section 3).

We conclude that a framework for problem solving assessment should – in addition to more traditional multiple-choice or extended-answer formats - incorporate complex, dynamic tasks which require continuous processing of information in more authentic settings and which may allow for the assessment of strategies rather then just outcomes of problem solving behaviour. Portable computers with simulation software, which under present conditions would be accessible at least for some sub-population within large-scale assessments, would serve this purpose. Strategy indicators may be in part derived from research on metacognition.

Another way to enhance problem solving tasks in terms of complexity and dynamics is to implement them in small group settings where participants contribute different knowledge and different goals. A common solution has to be reached by negotiating goals and strategies, taking into account all the information available. This approach has been widely used in job placement and selection (*assessment centers*), and it may benefit from research recently done in the area of *distributed cooperative problem solving.* Scoring would be based on solutions generated by groups of students (not on individual performance), and it would be interpreted as a measure of complex problem solving outcome which implicitly also measures strategic qualities and some aspects of social competence.

5.4.2.4 Domain-Specificity of Problem-Solving Performance

One of the most important insights of recent research in cognitive psychology is that demanding problems cannot be solved without a fund of knowledge in the domain in question. The concept of a *problem space* through which a General Problem Solver moves by means of domain-independent search strategies (Newell & Simon, 1972) turned out to be too simple to describe the understanding of problem situations and the process of finding a solution. Efforts to identify a general, domain-independent competence for steering dynamic systems ("operative intelligence") within the framework of complex problem solving research were also unsuccessful; performance with such systems can only partially be transferred to other systems (Funke, 1991).

Problem solving is dependent on knowledge of concepts and facts (declarative knowledge) and knowledge of rules and strategies (procedural knowledge) in the relevant subject domain. Empirical support for the importance of knowledge is provided by comparison of the problem-solving strategies used by experts and novices (Gruber & Mandl, 1996; Reimann & Schult, 1996; Schunn & Anderson, 1999), by analysis of the factors affecting the development of school performance (Helmke & Weinert, 1997), by laboratory experiments on learning in the natural sciences (Glaser, Schauble, Raghavan, & Zeitz, 1992) and by studies of real-world problem-solving performance – for example, studies of Brasilian street vendors who use complex rules, detached from those learned in school maths lessons, to do demanding price calculations (Lave, 1988).

The amount of relevant previous knowledge available could also account for the relation between intelligence and problem-solving performance, as shown in the work of Raaheim (1988) and Leutner (1999). People with no relevant previous knowledge at all cannot explore the problem situation and plan a solution in a systematic manner, but have to rely on trial and error. Those who are already very familiar with the task are able to deal with it as a matter of routine. General intellectual ability, as measured by reasoning tasks, plays no role in either of these cases. When problem solvers are moderately familiar with the task, however, reasoning strategies can be successfully implemented, thus resulting in clear correlations between intelligence and problem-solving performance.

These findings are crucial for the assessment of problem-solving competencies. They show that the idea of universal problem-solving competence is not tenable. Instead, a framework for problem solving assessment should aim at defining a profile of competencies in particular (subject or real-world) domains. The selection of these domains and the balance between high specificity on the one hand, proximity to general intelligence in the sense of psychometric reasoning on the other hand, is of great importance. Problem-solving tasks must be carefully directed at the target population to ensure that they can neither be solved as a matter of routine, nor place excessive demands on students, thus triggering trial-and-error behaviour. The students' relevant previous knowledge should, if possible, be empirically controlled for.

But if the idea of a universal problem solving ability has to be abandoned, in what sense can we speak of problem solving as a cross-curricular competence? This issue will be addressed in the next section.

5.4.3 The assessment of Problem Solving across and cross domains in PISA

When OECD policy makers reached out to define an international program to assess the outcome of schooling, the guiding question they had in mind was the following: “What do young adults at the end of education need in terms of skills to be able to play a constructive role as citizens in society?” (Trier & Peschar, 1995). Thus, they crossed the boundaries of school curricula as well as the limitations of classical models of human ability. They neither restricted educational assessment to knowledge and skills within a few school subjects nor referred to psychological theories of general intellectual abilities. Instead, they took a functional view, asking whether young adults are prepared to cope with the affordances and challenges of their future life. This type of disposition for mastering unforseen demands and tasks has been called *life skills* (Binkley et al., 1999), *non curriculum-bound outcomes* or *cross curricular competencies* (CCC) (OECD, 1997; Trier & Peschar, 1995).

In using the term *competencies* rather then *abilities*, this effort fits into a scholary discussion in psychology launched by McClelland (1973; see also Barrett & Depinet, 1991), Bandura (1990), and others and has recently been reviewed by Weinert (1999). These authors claim that the measurement of general, domain-independent skills or traits is of little use when performance in real life situations is to be understood or optimized. As Bandura (1990) put it, “there is a marked difference between possessing knowledge and skill and being able to use them well under diverse circumstances, many of which contain ambiguous, unpredictable and stressful elements”. Being able to cope with a certain range of situations is called a *competence*. Weinert (1999) suggested that a similar, purely functional approach should be used in large-scale assessments of educational outcome. Competencies should be defined by the range of situations and tasks which have to be mastered, and assessment might be done by confronting the student with a sample of such (eventually simulated) situations. This kind of assessment should be of grater practical use because it goes beyond compartmentalized and inert knowledge.

In fact, the functional understanding of competencies became central to the whole Program for international student assessment (PISA) as it has been implemented by OECD experts since 1998. For example, the PISA framework defines *Mathematical literacy* as “an individual’s ability, in dealing with the world, to identify, to understand, to engage in and to make well-founded judgements about the role that mathematics plays, as needed for that individual’s current and future life as a constructive, concerned, and reflective citizen” (OECD, 1999). Likewise, *reading literacy* and *science literacy* are related to everyday applications and authentic tasks. Obviously, such competencies heavily rely on non-routine, goal-oriented, high-level cognitive processes, and thus they incorporate problem solving in the broad sense of our definition. This becomes even clearer when we look at the sub-components or “aspects” of the literacy dimensions described within the frameworks. For *mathematical literacy*, for example, the framework distinguishes between three *competency classes*: Class 1 items require “reproduction, definitions, and computations”, while class 2 items require “connections and integration for problem solving”, and class 3 items address “mathematical thinking, generalisation and insight”. Roughly, the distinction between class1 and class2/3 may be understood as a distinction between non-problem (routine or reproductive) and problem-type tasks. Within the framework for science literacy, among the so called *science processes*, “identifying evidence needed in a scientific investigation” and “drawing or evaluating conclusions” most probably address some sort of domain-specific problem solving competence.

The assessment of problem solving competencies in PISA and similar programs has to include an explicit approach to problem solving within the literacy domains (Domain-specific problem solving competencies, DSC). In general, the test items of each domain will be anchored on a latent scale that fits to the Rasch model so that each student’s performance can as well be anchored on that scale. What would be the position of problem-type tasks within these latent scales? Theoretically, for each of the domains three different structures are possible:

DSC-1) There is one latent dimension, and the different types of tasks (the so-called aspects, competency classes or processes) form a hierarchy within the dimension. This would imply that problem-type tasks consistently have higher difficulty parameters than routine or reproductive items.
DSC-2) Problem-type tasks vary in difficulty as routine-type tasks do, and they form independent sub-dimensions of the literacy domain. This would imply that, e.g., class 2 math items would constitute a unique Rasch scale, distinctive from class1 items, which may be interpreted as a "mathematical problem solving" scale.
DSC-3) Problems and non-problems both vary in difficulty accross a wide range, but they do not spread out independent competency dimensions. Those students who can do routine and reproductive tasks in most cases are also good problem and vice versa. This would mean that problem solving competence cannot be distinguished from routine or reproductive competencies within the domains – neither as a higher level of proficiency (as in DSC-1) nor as a distinctive sub-dimension (as in DSC-2).

The PISA frameworks mainly argue against alternative DSC-1, but they leave the ultimate answer to empirical investigation. Psychological findings about the dependency of problem solving on domain-specific knowledge (see section 1.4) would instead support DSC-1, and in fact our own work on proficiency scales in TIMMS (Klieme & Köller, in press) as well as analyses of national PISA field trial data are in accordance with this hypothesis.

Assuming that an approach to identify domain-specific problem solving *dimensions* (DSC-2) or *levels* (DSC-1) within the three literacy domains has been developed, we may try to discriminate it from problem solving as a *cross-curricular competence* (CCC). Here again, several alternatives seem to exist:

CCC-1) Cross-curricular problem-solving is made up of all problem solving indicators from the three domains, combined into *one latent dimension*. Thus, the term "cross-curricular competence" is understood in the rather narrow sense of "*measured across curricular domains*".
CCC-2) Problem solving as a CCC is understood as problem solving in contexts which clearly go beyond (or "cross") the boundaries of the three literacy domains into what could be regarded as *transfer domains*. These domains may include school subjects other then mathematics or science, integrative domains/subjects (as, e.g., the combination of science, math, and geography when the student has to work in a simulated environment), extra-curricular activities (such as editing a school journal – which is part of school life, but not part of a particular domain/subject), or games.
CCC-3) Problem solving as a CCC is assessed within certain settings which are not used for the literacy domains. These settings may be distinguished by their complexity: On one hand, CCC indicators may include more complex, computer-based or small-group settings, while on the other hand they may include more constrained, reasoning-type tasks. Thus, cross-curricular problem solving would be characterised by particular *performance expectations* such as the continuous processing of external information and feedback on one hand and analogical or combinatorial reasoning on the other hand.

Based on psychological research as discussed in Section 1.4 of this paper, the expectation that problem solving indicators from different domains will constitute a general "problem solving competence" is highly unreasonable. If a indeed a broad range of problem solving tasks could be scaled on a single latent dimension, this dimension most probably would be identical to general intelligence or reasoning. Thus, alternative CCC-1 is ruled out, and we will use context (CCC-2) and setting (CCC-3) to specify cross-curricular problem solving.

To sum up our discussion on domain-specific and cross-curricular problem solving competencies, we can conclude:

The framework for the assessment of problem solving competencies should define a profile of competencies rather than a unique competence.
Domain-specific problem solving competencies related to the "traditional" literacy domains will be part of the profile. Whether they can be characterised as proficiency levels or sub-dimensions of the respective literacy scales is open to investigation.
Cross-curricular problem solving competencies should be assessed by tasks which differ from literacy measures in terms of context (focusing various transfer domains) and setting (focusing complex, dynamic environments as well as reasoning tasks).
These cross-curricular tests again will spread up a profile of competencies rather then a single dimension. Most probably, it will be possible to discriminate reasoning-like problem solving from strategic behaviour (including meta-cognitive regulation) in dynamic problem situations.
In order to discriminate cognitive problem solving competencies from motivational factors and from effects of domain specific knowledge, both types of prerequisites should be assessed independently and controlled for in statistical analyses.

For an example of how such an assessment program may look like see section 3.2 below. We conclude the general discussion by asking what kind of insights can be reached by implementing such a framework within large-scale assessments on an international level.

Student performance can be described and evaluated in domains and settings which go beyond classical areas of cognitive functioning into transfer domains and into new kinds of settings such as exploration and regulation of complex, dynamic environments. This would be appropriate for the *life skill* orientation of programs like PISA.

The structure of cognitive capabilities, including reasoning, literacy domains and cross-curricular competencies, will be better understood. The kind of relations which can be empirically identified might even differ between educational systems (countries). For example, low correlations between competency domains and general reasoning can be expected when the system is successful in fostering domain-specific competencies for all students, independent of their general intellectual level. Low correlations between curriculum-bound and cross-curricular competencies would be expected if the system is *not* successful in supporting links and transfer between domains.

The impact of personal prerequisites (like domain-specific knowledge and motivation), social background and school-related variables on competencies of different scope, context and complexity can by studied. Thus, the analytical power of the program may be enhanced. The possibilities and limits of fostering problem solving may be demonstrated and educational recommendations may be derived. Again, these insights will be amplified by cross-cultural comparison.

5.4.4 Indicators of problem solving competencies in large scale assessments

5.4.4.1 A short overview of Problem Solving assessment instruments

According to our broad definition, problem solving can occur in any domain, and so there is of course abundant experience with problem solving tests in educational research and practice. Most stimulating may be tests that use innovative formats like the "Clinical reasoning test" (Boshuizen et al., 1997), based on case studies in patient management, the "Overall-test" of complex, authentic decision making in business education (Segers 1997), or the "What if – Test" which addresses intuitive knowledge – apart from declarative academic knowledge – attained in exploring simulations of science phenomena (Swaak & de Jong, 1996). For science, Baxter and Glaser (1997) provide a systematic approach to performance assessment tasks, allowing for an analysis of cognitive complexity and problem solving demands. Within the domain of mathematics, there is a long tradition of problem-oriented thinking and learning (Hiebert et al., 1996; Schoenfeld, 1992) and related assessment strategies (Charles, Lester & O'Daffer, 1987; see Klieme, 1989, for an integrated discussion from an educational, cognitive-psychological and measurement perspective). For example, Collis, Romberg, and Jurdak (1986) developed a "Mathematical problem solving test" which used so-called "superitems", each composed of a sequence of questions which address subsequent levels of cognitive complexity. Since the seminal work by Bloom and colleagues (Bloom, Hastings, & Madaus, 1971), there have been various attempts at differentiating task complexity levels, a more recent example being the SOLO taxonomy (Collis et al. 1986). It is interesting to note that the former literature on taxonomy of learning objectives and related tasks did not contain any category like "problem solving", since Bloom and his colleagues conceptualised problem solving as an integration of all the levels they proposed (reproduction, understanding, application, and so on). Later work like the test rubrics developed for TIMSS (Robitaille & Garden, 1996) and PISA (OECD, 1999), in specifying *performance expectations*, includes some kind of problem solving category, leaving the question, whether this may constitute a separate dimension or level, open (see section 2 above).

Recently, several attempts have been made to implement measures of cross-curricular problem solving in large-scale assessments. A general test of cross-curricular competence, developed by Meijer and Elshout-Mohr (1999), is based on "critical thinking" inventories, but it seems to be rather heterogeneous. Trier and Peschar, working for OECD-Network A (OECD, 1997), addressed problem solving as one of the important cross-curricular competencies. Their "item", which at the same time served as an operationalization of skills in written communication, is an essay-like planning task, based on "in-basket" documents. The testee had to plan a trip for a youth club. This task turned out to be too difficult for the target population, and rather weak levels of objectivity were reached in scoring the answers.

However, the idea of planning tasks as typical instances of cross-curricular problem solving was independently invented by other research groups. In Germany, Funke and Fritz (1995) devised several experimental variants of planning tests, while Klieme, Ebach et al. (in press) developed a multiple choice test of problem solving competence for a large scale assessment program in one of the German federal states. A "project" such as arranging a party or planning a trip (sic!) is decomposed into action steps (clarifying goals – gathering information – planning – making decisions and executing the plan – evaluating the result) which are addressed by a sequence of items. A typical item within this *project approach* requires subjects to judge the consistency of goals, to analyze maps, schedules and other documents, to reason about the order of activities, to diagnose possible errors in the execution of actions and so on. While the tasks make heavy use of seemingly authentic material, answering the questions (most often in closed formats) is based on analytical and combinatorial reasoning. Structural models proved that performance in this tests is not confounded with mathematical and language competencies, i.e. it adds a specific branch to the students' profile of competencies. The *project approach* has now also become part of the International Life Skills Survey (ILSS; see Binkley et al., 1999). Ebach et al. (1999) developed several "projects" for this large-scale assessment and succeeded in establishing Rasch scores. Analysis of critical item elements showed that the number and connectivity of information elements as well as the cognitive complexity of the reasoning account for item difficulty. Thus, the project approach is an example of a reasoning-like, analytical, though contextualized test of cross-curricular problem solving.

Also within the ILSS context, pilot versions of Sternberg's "practical cognition" measure (see section 1.2 above) and a problem solving test developed by Baker (1998), O'Neil (1999) and their colleagues at the US Center for Research on Evaluation, Standards and Student Testing (CRESST) have been tried out. The latter is based on a framework which defines domain-dependent strategies, metacognition, content understanding, and motivation as components of problem solving. (Note that only the first two of these are understood as aspects of problem solving competence in the present paper, while the third and fourth are understood as *prerequisites* which have to be assessed independently.) To assess strategies, these authors confronted subjects with information on a technical device (tire pump) or a similar biological system which was described as malfunctioning and asked them to think about trouble shooting actions. To assess content understanding; subjects had to explain how the the device works by drawing a *knowledge map*. Several field trials showed that in principle the instrument was feasible, although its difficulty (with less then 25 % of the adults being able to solve the trouble shooting questions) and reliability were not fully convincing.

From the preceding overview, it becomes clear that the existing approaches to large-scale assessment of cross-curricular problem solving do by no means address the full range of tasks, contexts and (more or less dynamic) settings which are covered by the general framework outlined in section 2. In fact, the only CCC approaches which so far have been shown to be feasible for large scale implementation and to be theoretically sound are the project approach and the CRESST problem solving test ("tire pump task"). Therefore, the national expert group for problem solving assessment in PISA/Germany[6] started a major research and development enterprise, adapting instruments from various paradigms of experimental psychology and integrating them in a large validation study. The intention was to use as much input from basic cognitive research on problem solving as possible for the development and validation of new instruments.

5.4.4.2 Instruments used for cross-curricular Problem Solving assessment in PISA/Germany

As outlined in section 2, cross-curricular problem solving competencies should be assessed by tasks which differ from literacy measures in terms of

context, focusing transfer domains which are more or less distant from school subjects, and
setting, focusing complex, dynamic environments as well as reasoning tasks.

We also wanted to vary the paradigm of psychological problem-solving research from which the task has been taken.

[6] The author is indebted to Professors Joachim Funke, Detlev Leutner, Peter Reimann and Peter Frensch for serving on the expert goup, providing instruments, and contributing to the theoretical framework. The design, implementation and analysis of the study was indeed a joint activity of this group, supported by Joachim Wirth, Thomas Schmitt, Stefan Wagener and other scientific staff at the Max Planck Institute.

As illustrated in Fig.1, the following variants of *context* were implemented in the study:

- Fictitious, game-playing contexts (e.g. a space scenario) which are nonetheless meaningful and authentic for today's students;
- Real-world contexts (repairing a simple technical instrument; planning);
- School-, but not curriculum-related contexts (planning and reasoning about extra-curricular activities);
- Cross-curricular or integrated contexts, which relate to topics dealt with in school, but cross the borders of individual subjects (e.g. a simulated ecosystem which requires not only geographical, but economic, arithmetical and real-world knowledge);
- A virtual laboratory in which the fundamental principles of hypothesis generation and hypothesis testing are used.

The following settings implemented variants of *complexity and dynamics:*

- completely transparent, well-defined tasks with a closed format, mainly calling for combinatorial reasoning (see below; tests 4 and 8);
- less well-defined tasks with an open format, in which conceptual relations have to be understood, analogies drawn, or approaches to a practical problem outlined (tests 6 and 7);
- computer-assisted learning environments which can be explored freely, and provide t he student with continual feedback on the effects of his or her interventions.

The programs used in PISA (tests 1, 2 and 3) have differing levels of transparency and connectivity.

Co-operative problem solving (test 5): Here, students first work on individual subsections of a problem and must then agree on a common solution in which all relevant aspects are integrated. Here, the problem-solving process gains complexity and dynamics from the group discussions in which the differing perspectives are introduced.

With respect to the psychological foundation of the instruments, the following *paradigms* are used:

- Problem solving as search in a problem space with well-defined operators (test 8);
- Problem solving as combinatory reasoning (test 4);
- Analogical problem solving by transferring solutions to novel problems (test 7);
- Problem solving as drawing conclusions in mental models (test 6);
- Problem solving as inductive generating and testing of rules (test 1);
- Complex problem solving as exploration and steering of highly interconnected systems which change dynamically during the problem-solving process (test 2);
- Complex problem solving as exploration and steering of well-structured, discrete systems (test 3);
- Co-operative problem solving (test 5).

Variants a) to d) are presented in the form of written tests, e) to g) are computer-assisted.

A total of eight different problem-solving tasks have been used in the study, each lasting approximately 45 minutes (with the exception of tests 5 and 8, which were both administered in one 45-minute lesson).

Test 1: Virtual Laboratory (Authors: Reimann & Schmitt)
The student has to identify certain regularities in a series of simulated experiments. He or she specifies the experimental conditions and is played a video of the experiment. The sequence of the experiments and their results are recorded in a table. On the basis of this information, the problem solver formulates hypotheses which can then be either confirmed or rejected. The experimental environment thus consists of three components: a virtual laboratory (shown in the videos), a table of results and a hypothesis window.
Evaluation is based on whether the problem solver generates hypotheses central to the subject domain and on how systematically he or she proceeds.

Test 2: Ecological simulation (Authors: Leutner & Schrettenbrunner)
Topics from various subject domains are addressed in this planning game which consists of a computer simulation of a small farm. Moreover, the system is relatively complex (extensive, interconnected, intransparent, changing dynamically during the problem-solving process) and is thus able to reflect the research tradition of "complex problem solving". Experts in the field have judged that the system represents reality rather well.
Here, assessment is based on how efficiently the system is explored, how much knowledge is acquired, and how well the subject manages his or her "concern" in a concluding test.

Test 3: Space Game (Authors: Funke & Wagener)
Not only dynamic systems with quantitative variables (such as the ecological simulation described above), but also discrete systems have been studied in the domain of "complex problem solving" research. These consist of a series of states which can be altered by certain interventions (represented as the pushing of "buttons"). Real-world examples are ticket machines and electronic appliances (so-called "finite state automata"). The variant developed for PISA is embedded in a space scenario.
Here, too, evaluation is based on how efficiently the system is explored, the knowledge acquired, and the performance in control tasks.

Test 4: Project Tasks (Authors: Klieme, Ebach et al.)
A "project" such as the organisation of a class party is broken down into several steps, for each of which individual tasks are set (see section 3.1 above). The PISA test includes two such "projects" with differing contexts.

Test 5: Co-operative Planning (Author: Klieme)
A new variant of the project task in which three students work on a project together was developed for use in PISA. The three students are given different introductory information and are each responsible for different goals. Each student first works out a partial solution for his or her part of the problem; a common solution then has to be arrived at in a group discussion. The result is interpreted as a combination of problem-solving competence and co-operative abilities.

Test 6: Technical Problem Solving (Authors: Baker, O'Neil et al.; Adaptation: Wirth)
See description in section 3.1. The "knowledge mapping" instruction as well as the scoring rubric where revised for our study, and two additional questions were added.

Test 7: Analogical Problem Solving (Author: Wirth)
Several short planning and ordering tasks are set, each embedded in an real-world context. For each target task a (more or less) analogous base task with a worked-out solution is provided. The problem solver has to recognise the relationship between the target task and the base task, and transfer the solution strategy. The success of this *analogical transfer* (Vosniadou & Ortony, 1989) is evaluated.

Test 8: Transformation Problems (Author: Leutner)
This is a rather short, puzzle-like test. A fictitious biological object can be altered by means of certain operations ("genetic transformations"). In a series of tasks, the given state and the goal state are named, and the problem solver is asked to find the shortest possible sequence of operations leading from the given state to the goal.

5.4.4.3 Selected results of the validation study

The national enhancement to the PISA field trial allowed for extra testing time within a sub-sample of about 650 15- year-old students from different states, representing the full range of school types, intellectual and socio-economic background. A balanced matrix design was used where each student worked on some of the problem solving instruments. To administer the computer-based instrument, a Max Planck Institute team visited approximately 20 schools in three German *Länder* with a set of 30 laptops.

In addition to the problem solving tests, students went through the PISA reading, mathematics and science booklets, worked on a psychometric reasoning test (figural and verbal analogies) and answered – in addition to the PISA questionnaires – questions regarding task-specific knowledge, motivation, and metacognitive regulation. Thus, performance on the various problem solving tests could be correlated to other CCC indicators, to the literacy scores, and to general reasoning scores, allowing for sophisticated analyses of the structure of competencies. Effects of motivation and knowledge could be condidered, as could be a whole set of socio-cultural, individual and educational background variables.

In the following we very shortly report on two aspects that are of greatest concern for the evaluation of the instruments:
Feasibility: Are the instruments appropriate for large-scale assessments in the target group? We calculate the difficulty level (which on an average should be close to 50 % of the maximum score) and the index of reliability (Cronbach's alpha should be above .70), and we look at eventually biasing effects such as advantage of one gender.
Convergent and discriminant validity: In general, the indicators of cross-curricular problem-solving competence should show high intercorrelations with one other, but somewhat lower correlations with reading, math and science scores. Also, if we differentiate between types of literacy items (problem solving vs. routine tasks) as well as types of CCC indicators (reasoning-like vs. dynamic), appropriate structures should show up.

Instrument	Difficulty (Mean proportion correct)		Reliability (Cronbach's alpha)		
Transformation problems	.32			.88	
Analogical Transfer	.33			.79	
Technical problem solving	.36			.63	
Project tasks	.68			.74	
Cooperative planning	.60			.71	
Space game (Knowledge acquired)	.53		-	.82	-
Virtual Lab (Knowledge acquired)	.50	-	-	.78	-
Ecological simulation (Knowledge acquired)	.62		-	.63	-

Table 1: Difficulty and reliability of problem solving instruments

As can be judged from table 1, most of the tasks, especially the knowledge tests for computer-based learning and the co-operative planning task, do well with respect to difficulty level and reliability. The project tasks are rather easy for the target population, and the knowledge test for the ecological simulation is not reliable enough. (A posteriori inspection showed that it contains several questions which may be answered by domain-specific knowledge alone, without having explored the system.) Both tests can be easily revised. Technical problem solving has some shortcomings, replicating the findings reported by Baker (1998), despite our attempts to make the knowledge mapping task easier and the scoring more reliable. The two newly created paper-and-pencil-tasks (transformation problems and analogical transfer) are strongly reliable, but rather difficult for the majority of students.

Task-specific motivation had no effect at all – with one significant exception: Performance in the virtual laboratory depended upon pre-task-motivation. From observation of student behaviour we concluded that this program, as it had been implemented, was too demanding for lower-track students. However, both of the other computer-based tests were highly self-motivating. Most students – independent of their background – enjoyed working with the programs, so that motivational differences had no impact.

There were some gender effects, favouring males, for technical problem solving ($r=.16$, $p<.05$) and for three out of eight computer-based problem solving indicators (r between .28 and .35, $p<.001$). When we controlled for computer experience (which is higher in male students), all of these effects are reduced to values below .30 and thus – by applying standard rules from psychological research – may be qualified as "weak". The more computer experience the students reported, the more efficiently they controlled the space game and the ecological simulation ($r = .36$ and $r = .28$, resp., $p<.001$). However, performance in space game control tasks is the only one indicator (out of eight) which shows "medium size" effects of computer experience. Thus, in accordance with our expectations, both gender and computer-experience are relevant for computer-based problem solving processes, but in no sense do they determine the test results.

Some first insights into the validity of problem solving indicators are illustrated in Figure 2. This diagram is based on an analysis of the correlational pattern between seven competencies, each of them being measured by two or three measures in our design. These competence factors are math, science and reading literacy, reasoning (measured by figural and verbal analogies), and three components of cross-curricular problem solving: (1) "PS-paper", which is based on paper-and-pencil-tests (technical problem solving, analogical transfer, and project tasks), (2) "PS-Know", integrating the tests which assess knowledge aquired in exploring the three simulations, and (3) "PS-strat", which is based on indicators of efficient and systematic processing. This model, which differentiates *between* cross-curricular problem solving, reasoning and literacy domains, and which also differentiates three components *within* cross-curricular problem solving, fits the data better then models without such a differentiation. Thus, separating several domain-specific and cross-curricular competencies seems to be reasonable, as proposed in section 2[7].

The diagram also shows that reasoning is in deed the very core of the structure of competencies, and our paper-and-pencil problem-solving tests are quite close to that center. In search for a more unique type of problem solving competence, knowledge and strategy indicators based on computer-based instruments should be used.

[7] It should be noted that this analysis is a preliminary one, since not all of the test components have been integrated so far. In particular, the collaborative planning task can not be included here, since it does not provide individual performance indicators.

5.4.5 Conclusions

The validation study in PISA/Germany demonstrated the feasibility of both paper-pencil- and computer-based instruments for problem solving assessment in transfer domains. These instruments address competencies which are clearly distinct from literacy domains. As expected, the following propositions have been supported: (1) There is no way to define a general, unique “problem solving competence”. Rather, problem solving assessment produces a profile of competencies, varying in terms of context and setting (complexity). (2) Problem solving indicators, especially those based upon paper-and-pencil-tasks, are strongly correlated with reasoning, which in fact is the core of problem solving. (3) Strategy indicators derived from students’ behaviour in simulated environments do provide specific additional information.

The national enhancement in PISA/Germany will include cross-curricular problem solving also for the main study in year 2000. It has been decided to use a written test (project tasks), a co-operative problem situation, and – for a sub-sample of about 800 students - two computer based tests (space game and ecological simulation). Based on this work and similar work done in other assessment projects, it seems reasonable to set up an international framework which integrates both domain-specific and cross-curricular problem solving. On the international level, new challenges will arise, such as the question of culture-fair testing. However, theoretical concepts as well as assessment techniques are ready for new developments.

Setting (ordered by complexity)				
Exploration and decision making in dynamic environments	Virtual Lab	Ecological simulation	Cooperative planning	Space game
			Technical problem solving	
Understanding and solving of open-ended problems				Analogical problem solving
			Project tasks	Transformation problems
Reasoning on closed, transparent problems				
	Integration of Curriculum-Related domains		Extra-curricular, but school-related	Out-of-school life Games and puzzles

Context (ordered by distance from curricular domains)

Fig. 1: Variation of task characteristics in the problem solving validation study

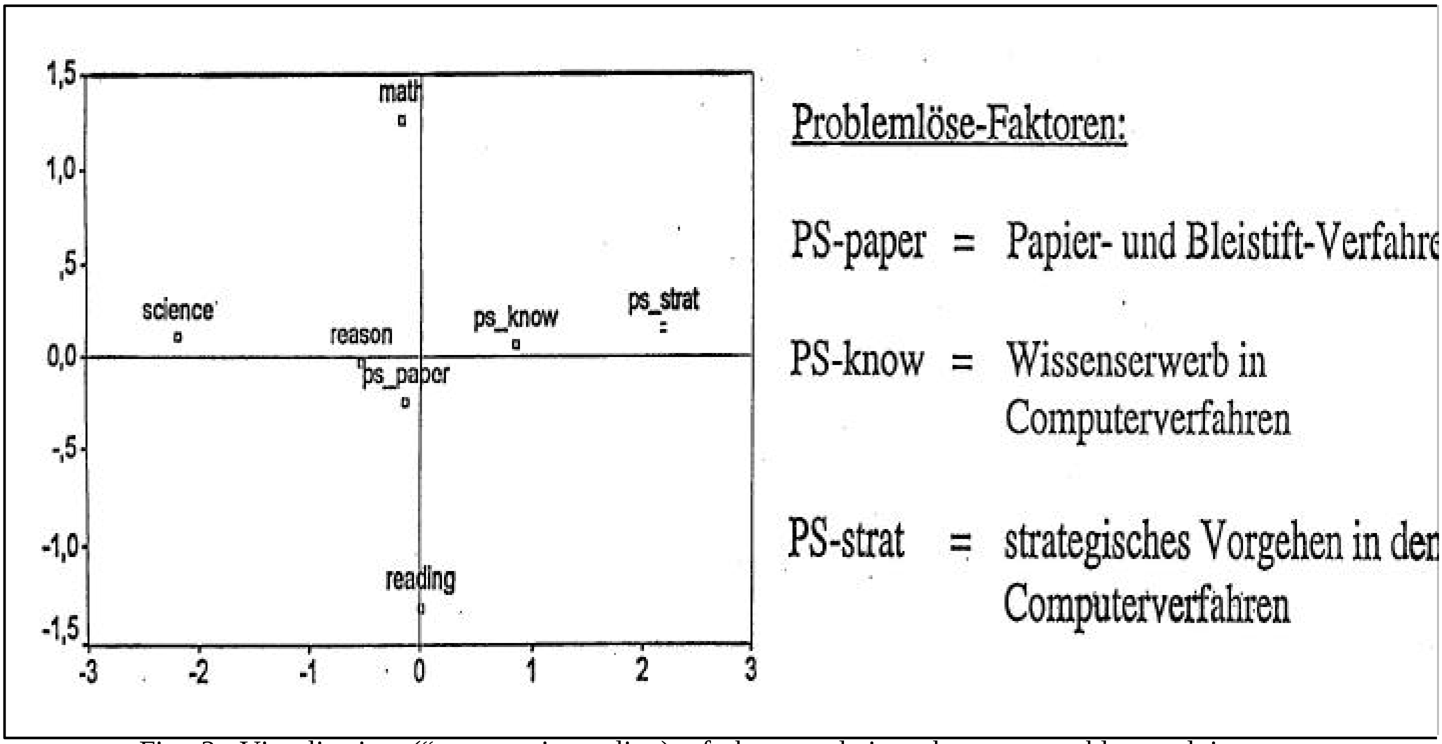

Fig. 2: Visualization (“non-metric scaling) of the correlations between problem solving factors, literacy domains, and reasoning. The closer two factors are in this diagram, the higher they are correlated in our data.

5.4.6 References

Baker, E. L. (1998). *Final report for validation of teamwork skills questionnaire using computer-based teamwork simulations* (85300-97-0056). Ottawa, Canada: National center for Research on Evaluation, Standards, and Student Testing (CRESST).

Bandura, A. (1990). Conclusion: Reflections on nonability determinants of competence. In R. Sternberg & J. Kolligian Jr. (Eds.), *Competence considered* (pp. 315-362). New Haven, London: Yale University Press.

Barrett, G. V., & Depinet, R. L. (1991). A reconsideration for testing for competence rather than for intelligence. *American Psychologist, 46*, 1012-1024.

Baxter, G. P., & Glaser, R. (1997). *An approach to analyzing the cognitive complexitiy of science performance assessments* (CSE Technical Report 452). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

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.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw Hill.

Boshuizen, H. P. A., van der Vleuten, C. P. M., Schmidt, H. G., & Machiels-Bongaerts, M. (1997). Measuring knowledge and clinical reasoning skills in a problem-based curriculum. *Medical Education, 31*, 115-121.

Brown, A. L. (1987). Metacognition executive control, self-regulation and other more mysterious mecanism. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, Motivation and Understanding* . Hillsdale, NJ.: LEA.

Buchner, A., & Funke, J. (1993). Finite State Automata: Dynamic Task Environments in Problem Solving Research. *Quaterly Journal of Experimental Psychology, 46A*, 83-118.

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. New York: Cambridge University Press.

Charles, R., Lester, F., & O´Daffer, P. (1987). *How to evaluate progress in problem solving*. (Vol. 6): National Council of Teachers of Mathematics.

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hilldale, NJ.: Erlbaum.

Collis, K. F., Romberg, T. A., & Jurdak, M. E. (1986). A technique for assessing mathematical problem-solving ability. *Journal for Research in Mathematics Education, 17*(3), 206-221.

De Corte, E., Greer, B., & Verschaffel, L. (1996). Mathematics Teaching and Learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* . New York: Simon & Schuster Macmillan.

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., Kreuzig, H., Reither, F., & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität.* Bern: Huber.

Ebach, J., Hensgen, A., Klieme, E., & Sidiropoulou, E. (1999). Analytical reasoning in real world contexts (pp. 1-15). Luxembourg: International Life Skills Survey (ILSS).

Ennis, R. H. (1996). *Critical Thinking.* Upper Saddle River, NJ 07458: Prentice Hall.

Flavell, J. H. (1976). Metacognitive aspects of Problem solving. In T. n. o. intelligence (Ed.), *Hillsdale, NJ.* . Erlbaum.

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

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. Sternberg & P. A. Frensch (Eds.), *Complex Problem Solving: Principles and Mechanisms* (pp. 185-222). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Funke, J., & Fritz, A. (Eds.). (1995). *Neue Konzepte und Instrumente zur Planungsdiagnostik.* Bonn: Deutscher Psychologen Verlag.

Glaser, R., Schauble, L., Raghavan, K., & Zeitz, C. (1992). Scientific Reasoning Across Different Domains. In E. de Corte, M. C. Linn, H. Mandl, & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving.* (Vol. 84, pp. 345-371).

Gruber, H., & Mandl, H. (1996). Das Entstehen von Expertise. In J. Hoffmann & W. Kintsch (Eds.), *Lernen* (Vol. 7, pp. 583-615). Göttingen, Bern, Toronto, Seattle: Hogrefe.

Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Ed.), *Psychologie des Unterrichts und der Schule* (Vol. 3, pp. 71-176). Göttingen: Hogrefe.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., & Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: the case of mathematics. *Educational Researcher, 25*(4), 12-21.

Hunt, E. (1994). Problem solving. In R. J. Sternberg (Ed.), *Thinking and Problem Solving* (pp. 215-232). San Diego: Academic Press.

Jeck, S. (1997). *Planen und Lösen von Alltagsproblemen*. Lengerich: Pabst.

Klieme, E. (1989). *Mathematisches Problemlösen als Testleistung*. Frankfurt/M: Lang.

Klieme, E., Ebach, J., Didi, H.-J., Hensgen, A., Heilmann, K., & Meisters, K.-H. (in press). Problemlösetest für die 7. Jahrgangsstufe. In R. Lehmann & I. Steinbach (Eds.), Hamburger Schulleistungstest für sechste und siebente Klassen.Göttingen: Hogrefe.

Klieme, E., Funke, J., Leutner, D., Reimann, P., & Wirth, J. ((in press)). Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Ergebnisse aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik.*

Klieme, E., & Köller, O. ((in press)). Mathematisch-naturwissenschaftliche Grundbildung: Erfassung und Skalierung von Kompetenzen. In J. e. a. Baumert (Ed.), *TIMSS - Mathematisch-naturwissenschaftliche Bildung am Ende der Sekundarstufe II* . Opladen: Leske & Budrich.

Krampen, G. (1993). Diagnostik der Kreativität. In G. Trost, K. Ingenkamp, & R. S. Jäger (Eds.), *Tests und Trends - Jahrbuch der Pädagogischen Diagnostik, Bd. 10* (pp. 11-39). Weinheim: Beltz.

Lave, J. (1988). *Cognition in practice*. Boston, MA.

Leutner, D. (1992). *Adaptive Lehrsysteme*. Weinheim: Beltz/PVU.

Leutner, D. (1999). *Discovery learning, intelligence, and problem solving ability.* Unpublished Paper presented at the EARLI Conference. Gothenburg, September 1999, Erfurt Educational University, Erfurt.

Linn, R. L. (1994). Performance Assessment: Policy Promises and Technical Measurement Standards. *Educational Researcher, 24*(9), 4-13.

Mayer, R. E. (1992). *Thinking, problem solving, cognition*. (2 ed.). New York: W. H. Freeman and Company.

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Simon & Schuster Macmillan.

McClelland, D. C. (1973). Testing for competence rather than for "intelligence". *American Psychologist, 28,* 1-14.

Meijer, J., & Elshout-Mohr, M. (1999). *An instrument for the assessment of cross curricular skills.* Unpublished Paper presented at the EARLI conference, Göteborg, Schweden, 1999.

Mumford, M. D., Baughman, W. A., Supinski, E. P., & Anderson, L. E. (1998). A construct approach to skill assessment: Procedures for assessing complex cognitive skills, *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 75-112).

Mumford, M. D., Supinski, E. P., Baughman, W. A., Costanza, D. P., & Threlfall, K. V. (1997). Process-based measures of creative problem-solving skills: V. overall prediction. *Creativity Research Journal, 10*(1), 73-85.

Neisser, U., Boodoo, G., Bouchard Jr., T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and Unknowns. *American Psychologist, 51,* 77-101.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs: Prentice-Hall.

Norris, S. P. (1989). Evaluating critical thinking ability. *History and social science teacher, 21*(3), 135-146.

O´Neil, H. F. J., Chung, K. W. K. G., & Herl, H. E. (1999). *Computer-based collaborative knowledge mapping to measure team processes and team outcomes.* , University of Southern California
University of California, Los Angeles.

OECD. (1997). *Prepared for life? How to measure cross-curricular competencies*. Paris: OECD.

OECD. (1999). *Measuring student knowledge and skills. A new framework for assessment*. Paris: OECD.

O'Neil, H. (1999). *A theoretical basis for assessment of problem solving.* Unpublished Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada, April 19, 1999, University of Southern California.

Pólya, G. (1945, 1980). *How to solve it. Deutsche Übersetzung: Schule des Denkens. Vom Lösen mathematischer Probleme*. (Vol. 3). Bern: Francke.

Raaheim, K. (1988). Intelligence and task novelty. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 73-97).

Krampen, G. (1993). Diagnostik der Kreativität. In G. Trost, K. Ingenkamp, & R. S. Jäger (Eds.), *Tests und Trends - Jahrbuch der Pädagogischen Diagnostik, Bd. 10* (pp. 11-39). Weinheim: Beltz.

Reimann, P., & Schult, T. J. (1996). Turning Examples Into Cases: Acquiring Knowledge Structures for Analogical Problem Solving. *Educational Psychologist, 3*(2), 123-132.

Robitaille, D. F., & Garden, R. A. (Eds.). (1996). *Research qustions & study design*. Vancouver, Canada: Pacific Educational Press.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, Metacognition, and Sense Making in Mathematics. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 334-370).

Schunn, C. D., & Anderson, J. R. (1999). The generality/specifity of expertise in scientific reasoning. *Cognitive Science, 23*(3), 337-370.

Segers, M. S. R. (1997). An alternative for assessing problem-solving skills: the overall test. *Studies in Educational Evaluation*, 23(4), 373-398.

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

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

Sternberg, R. J., & Kaufman, J. C. (1998). Human Abilities. *Annual Review of Psychology, 49*, 479-502.

Süß, H.-M. (1999). Intelligenz und komplexes Problemlösen: Perspektiven für eine Kooperation zwischen differentiell-psychometrischer und kognitionspsychologischer Forschung. *Psychologische Rundschau, 50*(4), 220-228.

Svecnik, E. (1999). *Stellenwert der Problemlösefähigkeit im österreichischen Schulwesen.* Unpublished Bericht des EU-Netzwerk "New Assessment Tools for Cross-Curricular Competencies in the Domain of Problem Solving", Graz.

Swaak, J., & de Jong, T. (1996). Measuring intuitive knowledge in science: the develoment of the what-if test. *Studies in Educational Evaluation, 22*(4), 341-362.

Trier, U. P., & Peschar, J. (1995). Cross-Curricular Competencies: Rationale and Strategy for Developing a New Indicator. In OECD (Ed.), *Measuring what Students Learn* (pp. 99-109). Paris: OECD.

Veenman, M. V. J., Elshout, J.J. & Meijer, J. (1997). The generality vs. domain-specificity of metacognitive skills in novice learning across domains. Learning and Instruction, 7, 187-209.

Vosniadou, S., & Ortony, A. (1989). Similarity and analogical reasoning: a synthesis. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 1-18). Cambridge: Cambridge University Press.

Weinert, F. E. (1999). *Konzepte der Kompetenz.* Unpublished Unveröffentlichtes Gutachten zum OECD Projekt "Definition and Selection of Competencies: Theoretical and Conceptual Foundations (DeSeCo)., München.

Weinert, F. E., & Kluwe, R. H. (Eds.). (1987). *Metacognition, Motivation and Understanding.* Hillsdale, NJ.: LEA.

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

6 Conclusions and policy implications

Based on the results described in chapter four, several conclusions can be drawn:

1) Assessment of more complex competencies like Problem Solving is feasible outside a pure laboratory environment.

The review presented in chapter 5.1 outlined different approaches for measuring problem Solving, many of them stemming from basic research paradigms. The educational and societal view taken in this chapter already shows on how to further develop these approaches towards more a more applied research paradigm. The discussion on problem solving and vocational training from chapter 5.2 definitely shows how the combination of the work from very different (scientific) communities can foster the development of new approaches. The project approach described throughout the different chapters has important roots in an effort to develop a meaningful assessment in vocational education.

2) Assessment of Problem Solving skills in a large-scale assessment paradigm is possible.

In the very beginning of the networks activities there were quite a few voices arguing against the feasibility of such a project. The main argument was that the many constraints to be found in international, large-scale comparative studies would make a measurement of problem solving skills either impossible or too expensive. The most optimistic assumptions expected an instrument covering only a very narrow aspect of Problem Solving. The work described in chapter 5.3 and 5.4, and above all the empirical results available today, show that meaningful aspects of problem solving are caught with the instruments described in this paper. Moreover, the initial work and the success have encouraged further development, both in the field of Problem Solving and other rather complex domains like e.g., social competencies, ICT literacy, etc.

3) Technology based assessment is a *must* for future work

Both the development work and the empirical studies have clearly shown that the use of technology raises new opportunities in the assessment of skills and competencies. The dynamic aspects of problem solving cannot be grasped without technological support. While the usefulness of technology from a content point of view seems to be beyond any doubt, there are still remaining concerns because of the increased financial efforts required by such a technology-based assessment. Beyond the arguments against this position described in chapter 5.3, it must be seen that the increasing impact of (comparative) assessment on decision making in education *requires* assessment tools for the *relevant* skills and competencies. Otherwise the ease of administration of certain traditional assessment types triggers the learning and teaching process! This is exactly what is happening in the today's US education, certainly an example not to be imitated.

4) Further research and development work is needed

Instruments available today cover well main aspects of problem solving, above all analytical problem solving. The chapter on alternative concepts and tools, as well as the reflections on complex problem solving show one major new direction to follow. Based on the available theoretical and empirical work instruments to be used in large-scale survey have to be developed and tested. Beyond the complex problem aspect, research on co-operative problem solving has to be fostered. The situation here is different as compared to complex problem solving. Substantially more research efforts have to be invested before being able to successfully develop instruments for large-scale surveys. Nevertheless, the research process and the development process should continue in parallel in order to reach a maximal efficiency.

5) Cultural fairness of the assessment should be systematically addressed

Cultural fairness issues might turn out to be even more important in the field of problem solving than in traditional assessment areas like reading, numeracy or science. More research is needed to fully understand the potential impact of cultural differences on the problem solving process. The development of new assessment instruments should systematically include precautionary measures to avoid culturally biased tests. This will become even more important as soon as results from large-scale comparative studies will be used for international benchmarking.

6) Clustering of European competence

Although substantial competence is available in Europe in the field of both the development of assessment instruments for large-scale surveys and the design of such surveys, it is quite obvious that this competence is dispersed and most difficult to bundle within a short time. Given the time constraints usually to be found in the development process in large-scale surveys this turns out to be a major competitive deficit as compared to the United States. The clustering of European competence should be fostered systematically, above all in the essential field of (statistical) methods.

7 Annex

7.1 Problem Solving Competency of Students: Ad-hoc Questionnaire for Education Policy Makers

1) What does the word "problem" mean to you? How would you define the word "problem"?

2) Which situational factors and behavioral features would lead you to conclude that a school-leaver has problem solving competency?

3) In your opinion, what skills and abilities does a school-leaver need in order to be able to solve problems that arise in everyday life and at work? In other words: What components are important for problem solving competency?

4) Do you believe problem solving competency should be an objective of school education? If yes, why?

5) How would you rate the importance of problem solving competency in relation to other cross-curricular competencies (CCC)?
For example an OECD study identified four important CCCs:

	Your rating *:
A) Civil, political and economical competencies	____
B) Problem solving competency	____
C) Self-perception (self-concept, self-esteem)	____
D) Communication competency	____

* Please rate the CCCs with regard to their relative importance by assigning a "1" to the most important, a "2" to the second, a "3" to the third, and a "4" to the least important one.

6) In your opinion, what role should problem solving competency play in comparison to subject-oriented aims of school education?

7) Is problem solving competency explicitely mentioned as an educational objective in the curricula of your country? Do programs to support it exist?

Your country:_____________________

Your year of birth: 19_____ Your sex (m/f): _____

***** Thanks for your assistance. *****