

Chapter 3

Developing a Framework for Diagnostic Assessment of Early Science

Benő Csapó

Institute of Education, University of Szeged, Hungary

Abstract

This chapter stems from the research of the past two decades on conceptualizing and assessing the quality of knowledge, especially knowledge as an outcome of science education. The findings that students performed well on some test types, whereas results were poor in other kinds of assessments, initiated a series of research projects. The results of these studies indicated that students' knowledge fell into several segments which may be independent from each other. The attempts for theoretical explanations and generalizations led to the conclusion that at least three different kinds of knowledge exist that are relevant to consider; different theoretical traditions may be mobilized to understand and describe the different aspects of students' knowledge, and, considering them simultaneously may result in a better framework for assessment. After an elaboration of the co-existence of the different theoretical directions, we applied this approach to the development of assessment frameworks for a diagnostic assessment project. The project aims at developing an online assessment system for the first six grades of primary school for reading, mathematics and science. Thus, the approach I present in this paper for science works well for mathematics and with some limitations for reading as well. However, this approach focuses on the initial stage of schooling; therefore, not all of its aspects can be generalized to later phases of education. In this chapter I present the empirical predecessor and theoretical resources of our current work on framework development. In order to enhance the diagnostic power of the assessments, we use three different types of test tasks for assessing the development of students' scientific reasoning, their scientific literacy and their scientific expert knowledge. Thus, for developing the framework for the tests we tap three different sets of theoretical resources. As will be shown, framework development for diagnostic assessment should be an analytic activity where goals should be carefully differentiated; on the other hand, teaching science should adopt a synthetic approach where the goals are seamlessly integrated.

1 Preliminary Considerations

Let me start this chapter with two not only theoretical, but rather philosophical considerations. Both emphasize the need for changing the way we used to approach the definition of outcomes of science education in the second half of the past century. The change I suggest may result in innovative taxonomies on the one hand, and may also bring us back to some old traditions, on the other.

The first consideration is related to the purpose of science education. It seems that science education is in crisis in a number of countries, which I argue, cannot be overcome without reconsidering its mission in modern societies. In the introduction of his famous book, Postman wrote over a decade and a half ago that ‘... without a transcendent and honorable purpose schooling must reach its finish and, the sooner we are done with it, the better. With such a purpose, schooling becomes the central institution, through which the young may find reasons for continuing to educate themselves’ (Postman, 1995. x-xi). Let me specify the message of this citation to science education. Science education needs to find a new ‘transcendent and honorable’ purpose. With such a purpose we may have a chance to alter current trends and motivate young people to consider learning sciences as a mean for becoming more educated people.

The second consideration is related to the way we construct theoretical frameworks for defining outcomes of science education. There is still a strong temptation to look for a single theory which covers all aspects of learning and knowledge. If it fails to explain everything, it is thrown out and replaced with something else, based on entirely different, sometimes opposite assumptions to the previous one. One may recall a number of controversies, such as propositional knowledge or procedural skills, content or structure, Piaget or Vygotsky. I would argue for accepting several theoretical frameworks at the same time, acknowledging that none of them is able to describe all relevant aspects of the issue we are dealing with. To understand different features of the development of students’ knowledge, we may need different theoretical approaches. I propose that as the controversy between wave or corpuscular properties of light cannot be resolved in a single theory, we have to construct different models when we try to explain different properties of knowledge. Instead of choosing one from the seemingly competing goals of science education, we must try to integrate them.

Furthermore, I challenge Lewin’s claim that ‘there is nothing more practical than a good theory’ (1951, p. 169). There is something more useful: two good theories are more practical than one. In fact, I propose applying three good theories (or at least three different theoretical orientations) to set goals for science education and my aim is to show that this solution is more practical than using any one of them.

2 Research on Structure and Organization of Knowledge

2.1 Signals of Problems of Students' Science Knowledge

Traditionally, scientific research used to be a prestigious occupation in Hungary and scientist were highly regarded. This appraisal used to attract a number of bright young people to (natural) sciences in the past centuries until the late 1980s. This attraction resulted in not only renowned scientists, but also in talented and well prepared science teachers. Therefore, the results of the First International Mathematics and Science Study in 1972 were in line with expectations: Hungarian students ranked second in international survey, right after Japanese students. A decade later, in 1983, in the Second International Mathematics and Science Study, fourth graders exceeded all expectations as they ranked first, well ahead of their Japanese peers (Keeves, 1992). The first IEA assessments were based on Bloom's taxonomies, so the results of the application tasks could be separated from the rest of the tests, and more detailed analyses indicated that students performed relatively purely on the application items. Some other assessments in the late 1980s indicated that our students were good at the type of disciplinary knowledge IEA measured at the time. The explanation was quite obvious. The curriculum was demanding in terms of scientific content, the number of science lessons per week was relatively high, and science teachers were well selected and well trained.

This generally favorable picture was ruined by some findings of other types of assessments which indicated that students' general intellectual development was not as impressive as expected on the bases of the excellent science achievements and there were weak relationships between their reasoning skills and science achievements (Csapó, 1988). These findings questioned that the type of disciplinary science knowledge students possessed could be transferred easily to new contexts and situations. Some analyses of that time also indicated that the shortcomings of science knowledge were special cases of a more general problem. Hungarian schools focused on mastering expert-type knowledge in the particular disciplines but paid less attention on the application of knowledge outside the narrow disciplinary context. For example, a project comparing the Hungarian (*matura*) and Dutch school-leaving examinations pointed out that the Hungarian mathematics tests were mostly composed of pure, sterile mathematical tasks without referring to any practical contexts, whereas the Dutch mathematics tasks were more realistic and embedded in everyday contexts (Mátrai, 1997).

Based on these experiences, known within a smaller research community, but not among teachers or the general public, it was easily predictable that our students' results would not be as good as the early IEA assessments indicated, if their knowledge was assessed by different type of tests. This anticipation was confirmed by

the 1995 TIMSS results: Hungarian students scored somewhere at the end of the first third of the countries (Beaton, Mullis, Martin, Gonzalez, Kelly, & Smith, 1996; Beaton, Martin, Mullis, Gonzalez, Smith, & Kelly, 1996). The results of PISA 2000 were even more revealing: science achievements were not significantly different from the OECD means (OECD, 2001). Being in the middle in the OECD ranking was not an issue in itself, but falling down from the top to the middle was shocking. The consecutive TIMSS assessments have indicated a decline in science and mathematics performances in the past decade, whereas science and mathematics results in PISA studies have not changed at all; they have been stable within the margin of error. The picture has got even more complex, as PISA 2009 indicated a significant (14 points) improvement in reading. This improvement prompts the question why no change has occurred in PISA in mathematics and science. As good reading is a tool for all further learning, its improvement is expected to transmit to other fields of study (OECD, 2010). This complex pattern of changes and unchanged achievements requires a more sophisticated tool to find a consistent explanation.

A further observation is students' negative attitudes towards science subjects. By the end of the 1990s, chemistry and physics were the two least liked school subjects. Students' attitudes towards mathematics were somewhere in the middle and they liked biology very much (Csapó, 2000). Furthermore, many students achieving well in science dislike learning sciences itself. As these facts indicate, science-related affective issues are not less complex than the cognitive ones.

A national core curriculum was accepted in 1997, which dealt with science only in brief as one of ten content areas. The content description was further shortened in 2003 and changed in 2007. The current legal version accepts the concept of key competencies, defining science competence as one of them. Nevertheless, the detailed content description details the ten content domains defined in the first version and devotes ca. 30 pages to science of 1–12 grades. Thus, textbook publishers as well as schools enjoy freedom in determining the actual content of teaching.

A national assessment system was established in 2001 and has been systematically carried out since 2003. The Assessment of Basic Competencies tests every student at grade 4, 6, 8, and 10 in reading and mathematics, but not in science. Hungarian students attending one of the academic tracks of high school, take a school-leaving examination (*matura*, *Abitur*) at the end of grade twelve. History, mother tongue and literature, and a foreign language are among the compulsory subjects, but science is not. This arrangement underlies the fact that makes the saying true: what is not tested is not taught. Science lost its leading role among the school subjects it used to have some decades ago.

As a consequence of several interacting causes, science education is in a critical situation. Several indicators signal this problem. The number of science and technology graduates is much lower than the economy would absorb. The number of applicants to science teacher training faculties is the lowest ever, fewer than the available places. The major problems related to science education were identified several years ago, but so far little has been done to tackle them. Although the problems became visible in the later phase of schooling they probably are rooted in the early school years. A contribution to the solution may be a systematic diagnostic assessment from the very beginning of formal education, followed by systematic intervention. The work reported in this chapter belongs to the foundation phase of developing such a system.

2.2 Empirical Projects on the Development of Skills and Organization of Knowledge

Since the beginning of the 1990s, a series of research projects have been carried out to examine the development of specific or general cognitive skills and different aspects of students' knowledge. A general observation was that students achieved well if their knowledge was tested in the same context where they mastered it. They were not only able to reproduce the teaching material, but also to demonstrate a certain kind of understanding of it. They were able to connect pieces of knowledge, draw conclusions from facts if everything remained within the boundaries of the given scientific topic; these are indicators of a kind of understanding what Gardner (1992) calls disciplinary understanding. On the other hand, students performed very poor if they were expected to apply their knowledge outside of the given context or discipline. In general, contradictions were observed between the high science achievements of students on disciplinary type tests and the poorer results in assessments measuring the application of knowledge and skills (Csapó, 1988, 1994, 1997; Csapó & B. Németh, 1995).

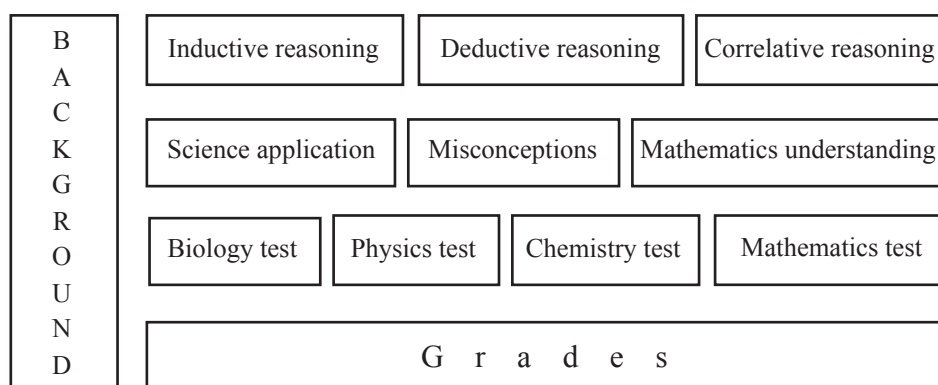


Figure 1: The System of Variables in the School Knowledge Project (Csapó, 1998a)

After several studies focusing only on a few variables, a systematic assessment project, called *School Knowledge Project* was designed that dealt with many relevant variables at the same time. Participants were 7th and 11th graders, and this way, development could be estimated from the cross-sectional design. The data collection was organized into a four-level model presented in Figure 1.

At the first level data characterizing students' school achievements were collected, such as grades given by the teachers. At the second level curriculum-based knowledge tests were used to measure how well students mastered what they were taught. These tests covered the teaching materials in four particular subjects: biology, physics, chemistry and mathematics.

At the third level three tests can be found. The science application test measured how students could apply their science knowledge outside of school, that is, in everyday situations. This test was constructed in a way that the tasks could be solved by utilizing the knowledge students were taught at school and what was measured by the tests at level two (B. Németh, 1998). As this test was based on authentic content placed in everyday situations, it was quite similar to the tests that PISA used some years later. The misconception test was built on some known, simple science misconceptions to examine if students used their scientific knowledge or their naive theories to interpret certain simple experiments or situations (Korom, 1997, 1998). The mathematics understanding test examined the depth of students' comprehension of some crucial mathematical concepts. It was composed of simple tasks which required deeper insight into certain mathematical issues, so they could not be solved mechanically by applying routine processes (Dobi, 1998).

The tests of the fourth level measured some general reasoning skills. The inductive reasoning test, used already in several other projects, contained tasks of verbal analogies, number analogies and number series (Csapó, 1997, 1998b). The deductive reasoning test covered the binary operations of propositional logic; the tasks mapped the truth functions of the logical operations (Vidákovich, 1998). The correlative reasoning tests examined students' reasoning about nondeterministic relationships (Bán, 1998).

The tests of the third and fourth level of the model were administered to both age groups in exactly the same form, so the difference between 7th and 11th graders could be directly assessed. Data collection took place in 1995, at the same time when the first TIMSS survey was administered.

The results of this project indicated again that students performed well at tests measuring their knowledge closely related to the curricular content, more precisely, close to the knowledge presented in the textbooks. On the other hand, they were unable to

apply their knowledge beyond the context given in the textbooks. The most shocking results were found when comparing the 7th and 11th graders. The four years between these two ages are the most intensive period of schooling: students master an enormous amount of knowledge in this sensitive phase of development, but their general thinking abilities develop at a slower pace. Although they are presented complex scientific theories, science education has little impact on their misconceptions. In some areas 7th graders even outperformed their 11th grader peers. For example, the older students were less likely to accept probabilistic relationships indicating that science education presents them mostly deterministic relationships and confirms a simple mechanistic world view, what was confirmed by the analysis of textbook content as well. In sum, results indicated that students learn a lot, but partly in vain, as their knowledge cannot be applied outside of the narrow disciplinary area, and learning sciences does not contribute to their intellectual development as much as it would be desirable (For a summary of the empirical findings see Csapó, 2007). Results of several other projects confirmed the findings that students' knowledge fell into at least two different parts. One is school knowledge they are able to mobilize in school contexts and their applicable everyday knowledge that they mobilize when solving practical problems (Molnár, 2001, 2006).

As for the affective aspect, an increasing negative attitude towards chemistry and physics was found here as well. What was even more problematic was that many students who performed well at disciplinary knowledge tests and general reasoning tests alike (thus potential candidates for science related studies and professions) formed negative attitudes toward science subjects (so, probably will not pursue scientific studies).

One general conclusion of the project was that this unfavorable pattern of achievements can be attributed to the methods of teaching. Science education is focusing on the disciplinary aspects of content; teaching materials are organized according to the values and logic of the corresponding scientific research areas. General principles of cognitive development are not taken into account in the process of organizing teaching materials and designing textbooks, and students' actual preliminary knowledge is not considered during the actual teaching. A deeper understanding of the material is often not possible at the actual developmental level; therefore, reasoning is rarely encouraged. Teachers often expect their students to reproduce the content of textbooks. Precise definitions and close scientific argumentation are valued and considered as high achievement honored with good grades. Other research confirmed this interpretation. As the PISA 2000 highlighted, Hungarian students considered the reproduction of teaching material as the main goal of learning, and memorization was one of the main learning strategies (OECD, 2003). Students prefer reproductive learning strategies in the earlier phase of schooling as well (B. Németh & Habók, 2006).

The methods of teaching are mostly rooted in the way science teachers are trained and developed during their career. They receive high-level scientific training in the particular discipline, but learn very little of developmental psychology and teaching methods. Their pre-service training in teaching practice is limited to a few weeks or a dozen of lessons they are expected to teach. Their professional identity is strongly connected to the particular disciplines; their professional organizations are also formed according to the scientific disciplines (e.g., separate societies for mathematics teachers, physics teachers).

There were two further conclusions of these empirical projects. (1) For changing the current situation, two other goals of education should be elevated to the same level of importance what today the disciplinary orientation enjoys. Cultivating students' general intellectual abilities and enhancing their capabilities to apply their knowledge should receive equal attention, scientific and technical support as the disciplinary approach receives today. (2) The three main goals may be supported by different theoretical frameworks. The consequences of these conclusions were elaborated in the framework of the most recent projects.

3 A Three-Dimensional Conception of Knowledge as Goal of Education

Recent tendencies in science education may be considered as integrative or synthetic. A number of recently preferred approaches to teaching sciences such as integrated science, project-based methods, problem-based learning, and inquiry-based science education emphasize global approaches, aiming at connecting pieces of knowledge in a way that goes beyond the traditional disciplinary views and boundaries. The rapid career of the concept of competence fits very well into this trend: competency is a proper counterpart on the outcome side of the teaching and learning processes that I have just listed on the methods side.

Competence is a useful umbrella conception; it helps to cope with the compartmentalization or fragmentation of students' knowledge. Competence-based teaching may contribute to solving the problems I have described in the previous section. The concept of competence may inform teaching and may be functional for constructing summative tests, such as the PISA tests, or the Hungarian national reading and mathematical competency assessment tests. On the other hand, such a general conception has limited value if we intend to construct 'high resolution' formative or diagnostic tests aiming at analyzing the fine nuances of students' knowledge. Therefore, one of the most promising directions in recent educational research aims at identifying and describing the most important competencies (key competencies) and

exploring their structure and function (see for example Schmid & Dinkelmann, 2007; Koeppen, Hartig, Klieme, & Leutner, 2008; Schott & Ghanbari, 2008; volume edited by Hartig, Klieme, & Rauch, 2008; and special issue of the *Zeitschrift für Pädagogik*, 2010).

If we aim at identifying different types of problems in students' learning, for example, insufficient mastery of the scientific concepts and domain specific skills, slowing or stagnating cognitive development or the difficulties in application of knowledge, we need different kinds of assessment instruments. These instruments should inform teachers about student's progress in several directions and should provide students with proper feedback on the strengths and weaknesses of their knowledge.

Inspired by the generalizable results of the empirical projects summarized in the previous section and by the needs of diagnostic assessments, I propose a three-dimensional model of the goals of school education. In this model I treat the goals separately that are related to (1) mastering the given curricular content, defined by the principles of the respective scientific disciplines; (2) the application of knowledge in relevant, authentic contexts, first of all in studying other subjects and in everyday life situations; and (3) the fostering students' general information processing capabilities, in short, developing their cognition. In this model, I consider each dimension of goals as equally important (for an elaboration of this model, see Csapó, 2004, 2010). Furthermore, I assume that they may be identified and defined separately at a conceptual-theoretical level, and different assessments scales may be constructed to measure them.

To pursue the different types of goals we need different underpinning theoretical resources. To understand the accumulation of disciplinary knowledge, a proper theoretical framework is based on the early approaches of cognitive science. This model distinguishes how novices and experts process information and represent knowledge. It explains the exceptional efficiency of expertise (expert knowledge): most of the time experts do not need to reason; they use their schemata which are ready for application. Expert knowledge is most efficient in familiar situations and cannot be easily applied in novel contexts. Thus, this model cannot explain a number of important characteristics of cognition and utilization of knowledge. One attempt to overcome the difficulties is to extend the 'novice-expert differences model' and to reformulate the conception of expertise. There are many successful applications of this approach or one of its derivatives in education as well. One of the most elaborated and most successful applications in education is the *How People Learn framework*, which has drawn from these theoretical resources as well (Bransford, Brown, & Cocking, 2000).

I agree with the critical remarks that the early information processing approach does not have a strong conception of development, and instead of sticking to it in every case, I propose to find a different theory for explaining the other aspects of knowledge.

Deanna Kuhn frames the following question in the title of the first chapter of her book: Why go to school? There may be several answers to this question; one is to master (a portion of the core) knowledge accumulated by sciences. This is the answer that relates to the disciplinary dimension of the goals of education. Nevertheless, she chooses another answer which then dominates her entire book: education for thinking (Kuhn, 2005).

The history of the goal of ‘cultivating the mind’ is as long as the history of education. However, scientifically established suggestions for doing it appeared only in the past century. This delay is quite understandable, if we consider that disciplinary knowledge is something that has an external representation (e.g., in textbooks). In contrast, constructs, such as intelligence, thinking, reasoning, creativity, and problem solving, do not have such a firm, visible external equivalent. When setting goals related to these constructs, we do not have external references; instead, we are dealing with the attributes of human mind. Therefore, we need scientific models of how the mind works to set goals for improving its working.

There are three main paradigms which accumulated relevant knowledge concerning this dimension of goals. The first one is intelligence research (psychometrics, individual differences approach) which by now has a century long history (see Carroll, 1993, for a synthesis of the findings of this approach, and for the reevaluation of the role of the concept intelligence can play in education: Adey, Csapó, Demetriou, Hautamäki, & Shayer, 2007). The second one is the work of Piaget and his school, especially useful for science education (Inhelder & Piaget, 1958). As they used science experiments for studying children’s cognitive development, there are many direct hints in their results of how science education can be utilized for the enhancement of reasoning. The third paradigm is the most recent one, cognitive educational neuroscience, which has some specific messages for teaching reading and mathematics, and a very general one is relevant from the point of view of science education: the human brain is more plastic than psychologists have assumed. Therefore, by stimulating teaching even the biological structure may be modified (OECD, 2007; Stern, Grabner, Schumacher, Neuper, & Saalbach, 2005).

Improving thinking emerged as a research field and later as a scientifically established goal in the second part of the last century. Those approaches which aimed at direct teaching of thinking resulted in fewer usable solutions for the educational practice. Other approaches which use the teaching material as means of facilitating cognitive development (embedding, infusion, enrichment, content based methods etc.)

practically integrate the two dimensions discussed before (Glaser, 1984; Bransford, Arbitman-Sith, Stein, & Vye, 1985). The content based approaches fertilized curriculum development (see the conception of thinking curriculum) as well as teaching methods (Csapó, 1990, 1992, 1999).

The third dimension of goals is also at least two millennia old; as Seneca's famous saying indicates, *Non scolae sed vitae discimus*. Schools are expected to teach something that can be utilized in the real life. This type of knowledge is best conceptualized in the assessment frameworks of PISA. The experts of PISA introduced and elaborated a generalized conception of *literacy*. Reading literacy, mathematical literacy and scientific literacy denote the type of applicable knowledge that a young person needs to live a successful life in a modern society. As this literacy conception is well-known, I have only one comment on it: it cannot (or should not) be taught directly, only in itself. As the concrete contexts and situations are constantly changing, a better strategy is to help students to master well understood transferable disciplinary knowledge and the reasoning skill necessary to process and transfer it in order to make it applicable in novel contexts. Content based methods of improving thinking may help to cope with 'inert' knowledge (a type of knowledge which cannot be applied) and facilitate transfer as well (Bereiter & Scardamalia, 1985).

In sum, I propose to distinguish different types of knowledge, in our case, three types, and use the best theoretical frames available for each one. The theories may be based on different assumptions, their explanatory power may be limited in themselves, but together they are more useful than relying solely on one of them.

4 Theoretical Foundations for a Diagnostic Assessment Framework

4.1 The Context of Framework Development

The context and the purpose of the framework development determine how this work is performed; therefore, I summarize some of the characteristics of the entire project in which the present effort is embedded. A research and development project is in progress at the University of Szeged to develop an Online Diagnostic Assessment System (ODAS) for the first six grades of primary school in reading, mathematics and science. The project aims to establish an item bank containing several thousands of items. This task requires the development of assessment frameworks that describe in detail what should be assessed and what should serve as the basis of item development.

In the first phase of the project, the items (600 per domain as a minimum) were prepared for paper-and-pencil assessment. These items were administered to large representative samples to determine their difficulty levels in 2010. The results of the paper-and-pencil testing will be used to correct and improve the existing items and to create new items to cover the difficulty scale proportionally. Those items that fit into the online system have been migrated to the computerized platform and parallel online assessments have been carried out. The results are used to study the media effect and to establish the validity of the online assessment (Csapó, Molnár, & R. Tóth, 2009). Another round of paper-based assessment will be carried out in 2011, and the results will be used to establish developmental scales. Three separate scales will be established for the three dimensions described earlier. These scales span from grade one to grade six. For reading and mathematics, one of the three scales, the literacy dimension dealing with the application of knowledge, will be connected to the already existing national assessment scales. The two systems have a common measurement point at the end of grade six, so there is an opportunity to measure reading literacy and mathematical literacy on the same scale from grade one to twelve. (Unfortunately, science is not measured in the national assessment program.)

The results of paper-and-pencil assessments will be used as a reference point for further online assessments. The paper-based items then can be used for the purposes of several assessments, but this line of item development will end, and the focus of the work will shift to technology-based assessment.

At present, not every item type which exists on paper can be reproduced in the computerized platform. On the other hand, new types of items can be created on the computer with no equivalents on paper. Therefore, parallel item development takes place for creating computerized items. The assessment frameworks play a crucial role in this process, because the framework–item correspondence will be used as one of the means to control the validity of the online assessment. As different cognitive processes are activated when solving items on paper and on a computer, it has been clear from the beginning of this work that it is not possible to create measurement instruments for both media which are equivalent in every respect. Furthermore, utilizing newer features of technology-based assessment requires new types of cognitive skills, and ICT familiarity may moderate the results in different ways. However, the intention is to map the same framework when creating items for the different media. Therefore, framework development has to take into account these requirements as well.

4.2 Frameworks for Early Diagnostic Assessment

The recent tendency to shift to standard-based education has inspired the creation of assessment standards in several countries. This movement has been generating a massive literature which includes the revision of Bloom's taxonomy (Anderson & Krathwohl, 2001) and the creation of new taxonomies (e.g., Marzano & Kendall, 2007). A number of principles, guidelines, handbooks and manuals have been published on these issues (e.g., O'Neill & Stansbury, 2000; Ainsworth, 2003).

Although there are differences between standards and frameworks, the literature on standards proved to be instrumental in framework development as well. The common bases of standards and frameworks are that both seek to answer 'What ...' questions: 'What to teach?' and 'What to assess?'. This way, similar methods and techniques can be used to describe and represent the content of teaching and assessment. On the other hand, frameworks in themselves do not set criteria; they do not define what levels should be reached by a certain age or school year. These criteria are set in the National Core Curriculum, and the ODAS is expected to measure students' progress so that it can be compared to these criteria. Furthermore, ODAS provides two other points of reference for each assessment. One is the norms calculated from the results of the actually assessed population (or its subgroups), and another one (as the assessment results are longitudinally connected) is the individual student's result on the previous assessment.

A segment of the standards and assessment literature deals with formative or diagnostic assessments explicitly (Nichols, Chipman, & Brennan, 1995; Clarke, 2001, 2005; Black, Harrison, Lee, Marshall, & Wiliam, 2003; Ainsworth & Viegut, 2006; Leighton & Gierl, 2007; Marzano & Haystead, 2008). Some other works focus on science standards (National Research Council, 1996; Waddington, Nentwig, & Schanze, 2007) or assess the outcomes of science education (Brown, 1996; Harlen, 2006).

From the point of view of the age range covered by the ODAS, research on early science education also resulted in findings and ideas that were influential in the framework development process (Nentwig & Schanze, 2006; Harlen & Qualter, 2009). Recent research on assessment of young students is also relevant for the project (e.g., Snow & Van Hemel, 2008).

The assessment frameworks of the ODAS project will be published in three volumes, one volume will be devoted for each assessment domain. The volumes have a common structure. Each one comprises three opening theoretical chapters summarizing the scientific foundations of the three dimensions. The second parts of the volumes elaborate the content of assessment in detail.

4.3 Scientific Reasoning

The aim of diagnostic assessment of reasoning within science is to monitor students' cognitive development, to make sure they possess the reasoning skills necessary for them to understand and master the science learning material in a meaningful way on the one hand, and to check if science education stimulates students' cognitive development as much as can be expected, on the other hand. Science education (with mathematics) has been considered as one of the most important means of developing students' reasoning skills (Minstrell, 1989). The content-based methods of enhancing cognition by applying science material for stimulating development provide rich resources for identifying reasoning processes which can be relevant in learning science and which can be developed through science education (Adey, 1999; Adey, Bliss, Head, & Shayer, 1989; Adey & Shayer, 1994; Shayer & Adey 1981).

Several simple and complex reasoning skills have been identified for diagnostic assessment. For example, thinking operations, such as control of variables, seriation, class inclusion, classification, multiple classification, combinatorial reasoning, proportional reasoning, and probabilistic reasoning are all essential for understanding science and early science education offers excellent opportunities for developing the appropriate skills (Lawson, Adi, & Karplus, 1979; Schröder, Bödeker, Edelstein, & Teo, 2000). Analogical reasoning, inductive reasoning, hypothetical-deductive reasoning etc. are also processes which are applied in learning sciences as well as more complex procedures, such as pattern-making through analysing wholes/parts and similarities/differences, making predictions and justifying conclusions, reasoning about cause and effect, generating ideas and possibilities, weighing up pros and cons, etc. The assessment frameworks identify and describe these thinking processes in an operationalized form (Adey & Csapó, 2011). The diagnostic tasks embed these cognitive processes in appropriate science content. Similar thinking processes were identified and assessed in mathematics.

4.4 Scientific Literacy

The aim of diagnostic assessment of scientific literacy is to monitor how students are able to apply their knowledge mastered in a school subject (or outside school) in other school subjects or in authentic real-life contexts.

The concept of scientific literacy is older than the PISA assessments (see e.g., Das & Ray, 1989; Klopfer, 1991; Roberts, 2007), but it became broadly known after the first PISA results were published. Although there is no straightforward way to identify the relevant contexts of application of scientific knowledge, a large number of pub-

lications have dealt with the issues of authentic contexts of acquiring and applying science knowledge (e.g., Nentwig & Waddington, 2005).

For the diagnostic assessment of science similar resources can be utilized to the ones serving as resources for PISA. However, for the purpose of ODAS, these conceptions have to be transferred to the early years (B. Németh & Korom, 2011). The diagnostic assessment tasks designed for the assessment of scientific literacy are based on the disciplinary learning material taught at school or on knowledge that can be expected to be learned outside of school, and use authentic contexts that are different from the ones where the knowledge was acquired.

4.5 Disciplinary Science Knowledge and Skills

The purpose of the diagnostic assessment of disciplinary science knowledge and skills is to monitor how well students master the disciplinary content. Research scientists are invited to define the relevant areas.

The framework organizes the content of assessment according to the logic and principles of the relevant disciplines and also takes into account the psychological developmental principles (see Shayer & Adey, 1981). As science education has been dominated by the principles of the respective disciplines, a lot of resources are available for selecting and organizing the content (Korom & Szabó, 2011).

The tasks used for diagnostic assessment present the content in scientific contexts and intend to measure the depth of disciplinary understanding. There is a special emphasis on concept development and conceptual change. The assessment intends to monitor not only how successful science education is in enhancing students' scientific theories and explanations but also how well schools cope with misconceptions.

5 Organizing the Content of the Science Framework

As described in the previous section, the diagnostic assessment frameworks are based on a three-dimensional conception of knowledge. The detailed description of the knowledge to be assessed has an even more complex structure, as further aspects have to be taken into account. This approach is depicted in Figure 2.

The assessment system is developed for the first six grades of the primary school, and the progression of students in science will finally be assessed in three devel-

omental scales. In the figure, these are represented as dimensions of learning. In accordance with the principles discussed earlier, these dimensions are the scientific thinking (which is informed mostly by the results of psychological studies), application of scientific knowledge (which can be identified as social needs and expectations concerning the quality of scientific knowledge), and scientific knowledge itself (derivates from the results of scientific disciplines).

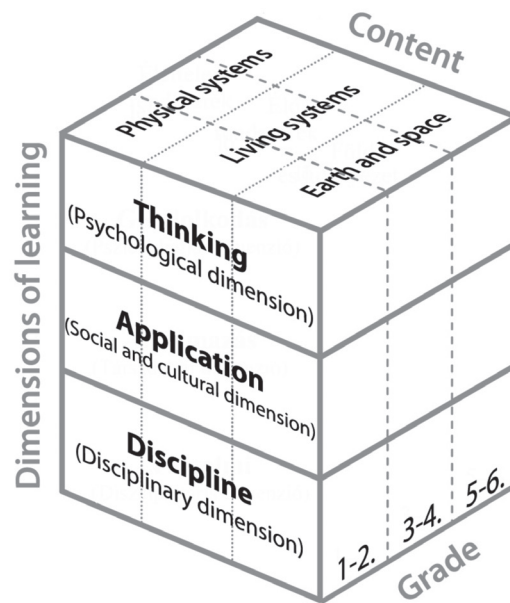


Figure 2: Organization of the Assessment Content (Source: Korom, B. Németh, Nagy, & Csapó, 2011)

The second aspect that has to be taken into account is the students' age. Although students' development will be measured on scales which overarch six grades, knowledge to be assessed could be arranged as a continuum. However, for practical reasons we followed the traditions of educational standards and curricula, and divided this continuum into three age groups (Grades 1–2, 3–4, and 5–6). Finally, the third aspect is the scientific content of the assessment where three areas are presented (Physical systems, Living systems, and Earth and space).

If the cube depicts the body of knowledge, the aspects discussed here identify 27 parts of it. If the frameworks are published in the form of a book, these blocks should be presented in a linear way. This means that the cube should be sliced up, and the blocks then should be systematically arranged. The way we arranged the materials is that first we separated the dimensions of learning, as they will constitute three differ-

ent developmental scales. Second, the three age groups are considered whereby the scientific content of the assessment is elaborated.

The same approach has been applied in each assessment domain, thus the reading, and mathematics frameworks have a similar structure. Only the content dimensions are different, where the actual content of the particular domain is presented.

6 Conclusions and Further Research

The first phase of developing assessment frameworks has been completed. Not only the conceptual foundations have been laid down, but also the detailed descriptions of the content of assessment have been accomplished. This phase of the work has been guided by theoretical models and has been carried out on the bases of results of previous research. The conceptual underpinnings and the detailed frameworks are based on the best knowledge available. On the other hand, there are rapid developments in the related areas and the results of current research should be continuously integrated into the assessment frameworks. There are some further steps in devising the online assessment system which also have some immediate conclusions for the conceptual frameworks.

Based on the frameworks, guided by the detailed description of the content to be assessed a large number of items have to be constructed. Item writing has already started at an early phase parallel with the framework development and a few hundreds of items have been created. The teams responsible for framework development and for creating test items have been working in close cooperation. The items were first designed for paper-based assessments, piloted in paper-and-pencil tests, then migrated to the electronic platform and piloted again in an online environment. These piloting processes revealed some difficulties of item construction that can then be reflected in the process of framework development as well. It turned out again, that although the goals of science education related to application of knowledge and development of thinking are often emphasized in educational documents and policy recommendations, they are less known among the practitioners. Most items proposed by the team of item writers were clearly related to the traditional pure disciplinary content. Creating items for the assessment of application skills was a bit more difficult, but the recent works in competency based education and the impact the PISA frameworks made this job easier.

The most difficult dimension proved to be that of assessing scientific reasoning, creating items for assessing science-related thinking skills using the content of science teaching materials. The conclusions of these experiences for the framework devel-

opment are that while items for the disciplinary dimension can be created more or less on routine bases, works in the other two dimensions require more attention and guidance. Especially, devising items in the reasoning dimension needs more support. As it is relatively novel and therefore more closely guided by the framework than by experience, more attention should be paid for the operationalization of the assessment content. The gap between research and practice should be bridged by extensive training of item writers.

Creating paper-based items and their migration to the electronic platform helped to reveal some characteristics of the knowledge to be assessed, for example, how the core issues remain unchanged while being represented in different ways. The first computer-based assessment of science (CBAS) in the PISA 2006 showed that there may be differences between the two media, and a study in the earlier phase of our current project has also indicated that some item types may prompt different cognitive processes if represented on paper or in computer (Csapó, Molnár, & R. Tóth, 2009). These findings call for further validity and media effect studies. More attention is needed when frameworks are mapped into electronic items, and the framework-item correspondence should be assured by careful analyses.

A further feedback for the revision of the framework is expected from the next phase of the field studies when the assessment scales will be established. Three separate science scales will be developed for the three dimensions discussed in this chapter. For this purpose, difficulty parameters of the items will be determined, and the item difficulty then may be compared with the position of the content measured by the item in the particular framework. The assumed hierarchy of the components of knowledge may also be verified on the bases of the empirical findings. These kinds of analyses can be carried out after a single assessment.

Additional analyses will be possible after several assessment cycles. As the data will be connected longitudinally, and students' development will be followed, the relationship between an earlier and a later assessment can be explored, even at the level of items. This way, the predictive power of the items can be studied, thus the diagnostic value of assessing certain components of knowledge can also be investigated. In general, detailed analyzes of the empirical data will serve rich sources for the revisions of the frameworks.

The ultimate reason of diagnostic assessment is to support teaching-learning processes. Its goals cannot be reached if teachers and students are not provided with training materials that can be used to fill the gaps indentified in students' knowledge. These materials can be practically delivered online by using the same network which is used for the assessment. As diagnostic assessment should focus on those components of knowledge that can be developed by the means available in school, training experi-

ments, planned for the next stage of the project will also inform framework development. In sum, the frameworks are planned to be a component of a renewed teaching and learning system, in which the development of the other components may have an effect on the frameworks themselves.

References

- Adey, P. (1999). Thinking science: Science as a gateway to general thinking ability. In J. H. M. Hamers, J. E. H. van Luit, & B. Csapó (Eds.), *Teaching and Learning Thinking Skills* (pp. 63-80). Lisse: Swets and Zeitlinger.
- Adey, P., Bliss, J., Head, J. & Shayer, M. (1989) (Eds.). *Adolescent Development and School Science*. New York: The Falmer Press.
- Adey, P., & Csapó, B. (2011). Developing and assessing scientific reasoning. In B. Csapó & G. Szabó (Eds.), *Framework for Diagnostic Assessment of Science*. Budapest: Nemzeti Tankönyvkiadó (in press).
- Adey, P., Csapó, B., Demetriou, A., Hautamäki, J., & Shayer, M. (2007). Can we be intelligent about intelligence? Why education needs the concept of plastic general ability. *Educational Research Review*, 2(2), 75-97.
- Adey, P., & Shayer, M. (1994). *Really Raising Standards: Cognitive Intervention and Academic Achievement*. London: Routledge.
- Ainsworth, L. (2003). *Power Standards. Identifying the Standards that Matter the Most*. Englewood, CO: Advanced Learning Press.
- Ainsworth, L. & Viegut, D. (2006). *Common Formative Assessments. How to Connect Standards-Based Instruction and Assessment*. Thousand Oaks, CA: Corwin Press.
- Anderson, L. W. & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching and Assessing*. New York: Longman.
- B. Németh, M. (1998). Iskolai és hasznosítható tudás. A természettudományos ismeretek alkalmazása [School knowledge and usable knowledge. Application of science knowledge]. In B. Csapó (Ed.), *Az iskolai tudás* [School knowledge] (pp. 115-138). Budapest: Osiris Kiadó.
- B. Németh, M. & Habók, A. (2006). A 13 és 17 éves magyar tanulók viszonya a tanuláshoz. [13- and 17-year-old students' views on learning] *Magyar Pedagógia*, 103(2), 83-105.
- B. Németh, M., & Korom, E. (2011). Science literacy and the application of scientific knowledge. In B. Csapó & G. Szabó (Eds.), *Framework for Diagnostic Assessment of Science*. Budapest: Nemzeti Tankönyvkiadó (in press).
- Bán, S. (1998). Gondolkodás a bizonytalanról: valószínűségi és korrelatív gondolkodás [Reasoning about the uncertain: Probabilistic and correlational reasoning]. In B. Csapó (Ed.), *Az iskolai tudás* [School knowledge] (pp. 221-250). Budapest: Osiris Kiadó.
- Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1996). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.

- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.
- Bereiter, C. & Scardamalia, M. (1985). Cognitive coping strategies and the problem of 'inert knowledge'. In S. F. Chipman, J. W. Segal & R. Glaser (Eds.), *Thinking and learning skills. Vol. 2. Research and Open Questions*. (pp. 65-80.). Hillsdale NJ: LEA.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning. Putting it Into Practice*. Berkshire: Open University Press.
- Bransford, J. D., Arbitman-Sith, R., Stein, B. S., & Vye, N. J. (1985). Improving thinking and learning skills: An analysis of three approaches. In J. W. Segal, S. F. Chipman, & R. Glaser (Eds.), *Thinking and Learning Skills* (pp. 133-206). Vol. 1. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How People Learn: Brain, Mind, Experience and School*. Washington, D. C.: National Academy Press.
- Brown, J. H. (1996). *Assessing Hands-on Science*. Thousand Oaks, CA: Corwin Press.
- Carroll, J. B. (1993). *Human Cognitive Abilities*. Cambridge UK: Cambridge University Press.
- Clarke, S. (2001). *Unlocking Formative Assessment. Practical Strategies for Enhancing Pupils Learning in Primary Classroom*. London: Hodder Arnold.
- Clarke, S. (2005). *Formative Assessment in Action. Weaving the Elements Together*. London: Hodder Murray.
- Csapó, B. (1988). *A kombinatív képesség strukturája és fejlődése* [Structure and development of combinative ability]. Budapest: Akadémiai Kiadó.
- Csapó, B. (1990). Integrating the development of the operational abilities of thinking and the transmission of knowledge. In H. Mandl, E. De Corte, N. Bennett, & H. F. Friedrich (Eds.), *Learning and Instruction. European Research in an International Context. Vol. 2.2. Analysis of Complex Skills and Complex Knowledge Domains* (pp. 85-94). Oxford: Pergamon Press.
- Csapó, B. (1992). Improving Operational Abilities in Children. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian Theories of Cognitive Development. Implications and Applications for Education* (pp. 144-159). London: Routledge.
- Csapó, B. (1994). Az induktív gondolkodás fejlődése. [The development of inductive reasoning] *Magyar Pedagógia*, 94(1-2), 53–80.
- Csapó, B. (1997). The development of inductive reasoning: Cross-sectional assessments in an educational context. *International Journal of Behavioral Development*, 20(4), 609-626.
- Csapó, B. (1998a). Az iskolai tudás vizsgálatának elméleti keretei és módszerei [Theoretical framework and methods of studying school knowledge]. In B. Csapó (Ed.), *Az iskolai tudás* [School knowledge] (pp.11-37). Budapest: Osiris Kiadó.
- Csapó, B. (1998b). Az új tudás képződésének eszközei: az induktív gondolkodás. [Tools for creating new knowledge: Inductive reasoning]. In B. Csapó (Ed.), *Az iskolai tudás* [School knowledge] (pp. 251-280). Budapest: Osiris Kiadó.

- Csapó, B. (1999). Improving thinking through the content of teaching. In J. H. M. Hamers, J. E. H. van Luit & B. Csapó (Eds.), *Teaching and Learning Thinking Skills* (pp. 37-62). Lisse: Swets and Zeitlinger.
- Csapó, B. (2000). A tantárgyakkal kapcsolatos attitűdök összefüggései [Students' attitudes towards school subjects]. *Magyar Pedagógia*, 100(3), 343-366.
- Csapó, B. (2004). Knowledge and competencies. In J. Letschert (Ed.), *The integrated person. How curriculum development relates to new competencies* (pp. 35-49). Enschede: CIDREE.
- Csapó, B. (2007). Research into learning to learn through the assessment of quality and organization of learning outcomes. *The Curriculum Journal*, 18(2) 195-210.
- Csapó, B. (2010). Goals of learning and the organization of knowledge. In E. Klieme, D. Leutner, & M. Kenk (Eds.). *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. 56. Beiheft der Zeitschrift für Pädagogik* (pp. 12-27). Weinheim u.a.: Beltz.
- Csapó, B. & B. Németh, M. (1995). A természettudományos ismeretek alkalmazása: mit tudnak tanulóink az általános és a középiskola végén? [The application of science knowledge: What do our students know at the end of primary and secondary schools?] *Új Pedagógiai Szemle*, 8, 3-11.
- Csapó, B., Molnár, G. & R. Tóth, K. (2009). Comparing paper-and-pencil and online assessment of reasoning skills. A pilot study for introducing electronic testing in large-scale assessment in Hungary. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment. New Approaches to Skills Assessment and Implications for Large-Scale Testing* (pp. 113-118). Luxemburg: Office for Official Publications of the European Communities.
- Das, R. R. & Ray, B. (1989). *Teaching Home Science*. New Delhi: Sterling Publishers.
- Dobi, J. (1998). Megtanult és megértett matematikatudás [Mathematics knowledge mastered and understood]. In B. Csapó (Ed.), *Az iskolai műveltség* (pp. 169-190). Budapest: Osiris Kiadó.
- Gardner, H. (1992). *The Unschooled Mind. How Children Think and How Schools Should Teach*. New York: Basic Books.
- Glaser, R. (1984). Education and thinking. The role of knowledge. *American Psychologist*, 39(2), 93-104.
- Harlen, W. (2006). *Teaching, Learning and Assessing Science 5-12*. London: Sage.
- Harlen, W. & Qualter, W. (2009). *The Teaching of Science in Primary Schools*. London: Routledge.
- Hartig, J., Klieme, E., & Rauch, D. (Eds.) (2008). *Assessment of Competencies in Educational Context*. Göttingen: Hogrefe.
- Inhelder, B., & Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. New York: Basic Books.
- Keeves, J. P. (1992). *The IEA Study of Science III. Changes in Science Education and Achievement: 1970 to 1984*. Oxford: Pergamon Press.

- Klopfer, L. E. (1991). Scientific literacy. In A. Lewy (Ed.), *The International Encyclopedia of Curriculum*. (pp. 947-948). Oxford: Pergamon Press.
- Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie*, 216(2), 61-73.
- Korom, E. (1997). Naiv elméletek és tévképzetek megjelenése a természettudományos fogalmak tanulása során [Naive theories and misconceptions in learning science]. *Magyar Pedagógia*, 97(1), 19-41.
- Korom, E. (1998). Az iskolai tudás és a hétköznapi tapasztalat ellentmondásai: természettudományos tévképzetek [Contradictions of school knowledge and everyday experience: science misconceptions]. In B. Csapó (Ed.), *Az iskolai tudás* [School knowledge] (pp. 139-167). Budapest: Osiris Kiadó.
- Korom, E. & Szabó, G. (2011). Scientific and curriculum aspects of teaching and assessing science. In B. Csapó & G. Szabó (Eds.), *Framework for Diagnostic Assessment of Science*. Budapest: Nemzeti Tankönyvkiadó (in press).
- Korom, E., B. Németh, M., Nagy, L., & Csapó, B. (2011). Diagnostic assessment frameworks for science: Theoretical background and practical issues. In B. Csapó & G. Szabó (Eds.), *Framework for Diagnostic Assessment of Science*. Budapest: Nemzeti Tankönyvkiadó (in press).
- Kuhn, D. (2005). *Education for Thinking*. Cambridge, MA: Harvard University Press.
- Lawson, A. E., Adi, H. & Karplus, R. (1979). Development of correlational reasoning in secondary schools: Do biology courses make a difference? *American Biology Teacher*, 41(7), 420-425
- Leighton, J. P. & Gierl, M. J. (2007) (Eds.). *Cognitive Diagnostic Assessment for Education. Theory and Applications*. Cambridge: Cambridge University Press.
- Lewin, K. (1951). *Field Theory in Social Science. Selected Theoretical Papers*. D. Cartwright (Ed.). New York: Harper & Row.
- Marzano, R. J. & Haystead, M. W. (2008). *Making Standards Useful in the Classroom*. Alexandria: Association for Supervision and Curriculum Development.
- Marzano, R. J. & Kendall, J. S. (2007). *The New Taxonomy of Educational Objectives*. 2nd ed. Thousand Oaks, CA: Coewin Press.
- Minstrell, J. A. (1989). Teaching science for understanding. In L. B. Resnick, & L. E. Klopfer (Eds.), *Toward the Thinking Curriculum: Current Cognitive Research* (pp. 129-149). Alexandria: Association for Supervision and Curriculum Development.
- Mátrai, Z. (1997) (Ed.). *Középiskolai tantárgyi feladatbankok I. Biológia – Matematika – Angol nyelv* [Subject matter item banks for high schools I. Biology – Mathematics – English]. Budapest: Országos Közoktatási Intézet.
- Molnár, Gy. (2001). Az életszerű feladat-helyzetekben történő problémamegoldás vizsgálata [Examination of problem solving in real life situations]. *Magyar Pedagógia*, 101(3), 347-373.
- Molnár, Gy. (2006). *Tudástranszfer és komplex problémamegoldás* [Knowledge transfer and complex problem solving]. Budapest: Műszaki Kiadó.
- National Research Council (1996). *National Science Education Standards*. Washington DC: National Academy Press.

- Nentwig, P. & Schanze, S. (2006). *Es ist nie zu früh! Naturwissenschaftliche Bildung in jungen Jahren* [It is never too early! Science education in the early years]. Münster: Waxmann.
- Nentwig, P. & Waddington, D. (2005). *Making it Relevant. Context Based Learning of Science*. Münster: Waxmann.
- Nichols, P. D., Chipman, S. F., & Brennan, R. L. (1995). *Cognitively Diagnostic Assessment*. Hillsdale, N. J.: Lawrence Erlbaum Associates Publishers.
- O'Neill, K. & Stansbury, K. (2000). *Developing a Standards-Based Assessment System*. San Francisco: WestEd.
- OECD (2000). *Measuring Student Knowledge and Skills. The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy*. Paris: OECD.
- OECD (2001). *Knowledge and Skills for Life. First Results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.
- OECD (2003). *Learners for Life. Student Approaches to Learning. Results from PISA 2000*. Paris: OECD.
- OECD (2007). *Understanding the Brain. The Birth of a Learning Science*. Paris: OECD.
- OECD (2010). *PISA 2009 Results, Vol. I.: What Students Know and Can Do. Student Performance in Reading, Mathematics and Science*. Paris: OECD.
- Postman, N. (1996). *The End of Education*. New York: Vintage Books.
- Roberts, D. A. (2007). Scientific literacy/Science literacy. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 729-780). London: Routledge.
- Roth, W. M. (1995). *Authentic School Science*. Dordrecht: Kluwer Academic Publishers.
- Schmid, C. & Dinkelmann, I. (2007). *Bildungsstandards im nichtfachlichen Bereich: Probleme und Perspektiven am Beispiel der Lern- und Denkstrategien. Forschungsbericht*, Zürich: Pädagogische Hochschule Zürich, Departement Forschung und Entwicklung.
- Schott, F. & Ghanbari, S., A. (2008). *Kompetenzdiagnostik, Kompetenzmodelle, Kompetenzorientierter Unterricht. Zur Theorie und Praxis überprüfbarer Bildungsstandards*. Münster: Waxmann
- Schröder, E., Bödeker, K., Edelstein, W., & Teo, T. (2000). *Proportional, Combinatorial, and Correlational Reasoning. A Manual Including Measurement Procedures and Descriptive Analyses*. Study "Individual Development and Social Structure". Data Handbooks Part 4. Berlin: Max Planck Institute for Human Development.
- Shayer, M. & Adey, P. (1981). *Towards a Science of Science Teaching. Cognitive Development and Curriculum Demand*. London: Heinemann Educational Books.
- Snow, C. E., & Van Hemel, S. (2008). *Early Childhood Assessment. Why, What and How*. Washington, DC: National Academies Press.
- Stern, E., Grabner, R., Schumacher, R., Neuper, C., & Saalbach, H. (2005). Educational research and neurosciences – expectations, evidence and research prospects. Bonn/Berlin: Bundesministerium für Bildung und Forschung.
- Waddington, D., Nentwig, P., & Schanze, S. (2007) (Eds.). *Making it Comparable. Standards in Science Education*. Münster: Waxmann.

Vidákovich, T. (1998). Tudományos és hétköznapi logika: a tanulók deduktív gondolkodása [Scientific and everyday logic: Students' deductive reasoning]. In B. Csapó (Ed.), *Az iskolai tudás* [School knowledge] (pp. 191-220). Budapest: Osiris Kiadó.

Acknowledgement

This chapter is based on the results of studies of the *Research Group on the Development of Competencies, Hungarian Academy of Sciences*.