4

Diagnostic Assessment Frameworks for Science: Theoretical Background and Practical Issues

Erzsébet Korom
Institute of Education, University of Szeged

Mária B. Németh
Research Group on the Development of Competencies,
Hungarian Academy of Sciences

Lászlóné Nagy
Biology Methodology Group, University of Szeged

Benő Csapó
Institute of Education, University of Szeged

Introduction

The main function of this chapter is to create a link between the previous three theoretical chapters and the detailed content specifications appearing in the next chapter of this volume. We further provide a characterisation of the genre of frameworks and discuss the considerations justifying our choice of solutions.

Chapter 1 gave an overview of international research findings related to the development of scientific thinking and in general to the role of science in the improving thinking processes, approached mostly from the perspective of developmental psychology. Chapter 2 is similarly based on international research findings, but approaches the issue with the external goals of science education kept in mind. Chapter 3 moved on to the traditions and curricular features of public education in Hungary and a picture of the system emerged to which the diagnostic program would need to be tailored. All this information delineates the first problem to be solved: the achievements of the forefront of scientific research must be adapted to such an extent that they have the greatest educational effect both on individual students and on the public education system as a whole.

The diagnostic assessment system is developed in parallel for three main domains, each of which rests on the same set of principles.[1] The parallel treatment of reading, mathematics and science is justified by several principles of psychology and education as well as by considerations of education organisation. On the one hand, an appropriate level of reading comprehension is essential for learning both mathematics and science and on the other hand, mathematics and science enhance reading skills by offering texts that do not appear among literary styles. The logic of mathematics and that of language can mutually reinforce each other. Science is the best practice field for the application of relationships learnt in mathematics. Drawing attention to and making use of different types of relationship systems is especially important during the first stage of schooling, when students' intellectual development is very fast-paced and exceptionally sensitive to stimulating factors.

The parallel treatment of the three domains has the further advantage that they mutually fertilise one another, the ideas and formal solutions emerging in one can be used in the other two. The development of test questions, uniform measurement scales, data analysis methods and feedback systems also calls for the parallel treatment of the three domains and the sharing of certain principles. This parallel treatment also means, however, that certain compromises must be made: there is a limit to what extent the same principles can be adhered to in all three domains. In the interest of uniformity, the three-dimensional approach is preserved and uniformly applied, but the interpretation of each dimension takes the special features of individual assessment domains into account.

Another benefit of parallel treatment may be a complementarity effect. The three domains are discussed in a total of nine theoretical chapters. We made no effort to create parallel chapter outlines. This made it possible to give in-depth coverage to one issue in one domain and another issue in another domain. In the first chapter of the volume dedicated to the

1 This chapter also contains sections appearing in the corresponding chapters of the other two volumes.

domain of reading, for instance, special emphasis is given to issues in developmental psychology and neuroscience, which also offer important insights for mathematics and to some extent for science education. Certain reasoning skills are discussed in greater detail in the first chapter of the science volume, but the same skills are also important for mathematics education. The second chapters of the volumes focus on the issue of knowledge application and each of them draw general conclusions that equally apply to the other two domains. The third chapters examine practical questions related to the curriculum in their respective domains, but they share a commitment to the historical traditions and current principles of Hungarian public education. At the same time, the proposed choice and structuring of the contents of education also reflect the need to follow progressive international trends and to make use of the achievements of other countries.

In line with the above principles, we regard the nine theoretical chapters in combination as the theoretical foundation of the diagnostic assessment system. The background knowledge analysed in these chapters thus constitutes a common resource for each of the domains, without the need to detail the shared issues separately in the equivalent chapters of the different volumes.

The first section of the present chapter reviews the main factors taken into consideration during the development of the frameworks. First, the tools used for the specification of the goals of education and the contents of assessment are discussed and our solution to the problem of providing a detailed characterisation of the contents of diagnostic measurement is outlined. The next sections show how these principles are used in the development of the science frameworks.

Taxonomies, Standards and Frameworks

The development of frameworks of diagnostic assessment was assisted with a number of different resources. Our work followed an approach undertaking to offer a precise definition of educational targets and of the contents of assessment. First, we discuss various systems used around the world to characterise contents, which we then use as a standard of comparison in describing the method we developed.

Taxonomies

Efforts to define curricular goals in great detail first appeared in the 1950s. This was the time when as a combined result of various processes Bloom and his colleagues developed their taxonomic systems, which made a strong impact on education theoretic objectives for the next few decades. One of the triggers prompting the development of the taxonomies was a general dissatisfaction with the vague characterisation of curricular goals, and the other was the rise of the cybernetic approach to education. There appeared a need for controllability, which required feedback, which in turn presupposed the measurement of both intended targets and actual performance. By comparing targets with actual performance, weaknesses may be identified and interventions may be planned accordingly. During the same period, other processes led to a heavier emphasis on educational assessment and the expansion of testing also created a need for a more precise characterisation of the object of measurement.

Taxonomy is essentially a structured frame providing a system of ordering, organising and classifying a set of objects, in our case, the body of knowledge to be acquired. It is like a chest of drawers with a label on each drawer showing what should be placed in it. A taxonomy can also be interpreted as a data table with the headings indicating what can appear in its various rows and columns. Compared to the previous general characterisations of goals, planning based on such a formalised system constituted a major step forward, and prompted educators responsible for defining specific curricular objectives to think very carefully about what behavior could be expected as a result of learning.

The greatest impact was made by the first taxonomic system, one describing the cognitive domain (Bloom et al., 1956), which opened a new path for curriculum and assessment theory. This taxonomic system characterised expected student behavior in concrete, observable categories. The most obvious novelty was the system of six hierarchically organised frameworks, each of which was designed to apply uniformly to all areas of knowledge. Another significant improvement was the level of description that surpassed by far all previous efforts in detail, precision and specificity. As a further advantage, the same detailed description could be used to plan learning processes and to develop assessment tools. This is the origin of the name *taxonomies of objectives and assessments,* which refers to the two functions.

The Bloom taxonomies exerted a significant direct influence first in the United States, and later on this system provided the foundations for the first international IEA surveys (see also Chapter 2). The empirical surveys, however, did not corroborate every aspect of the hierarchy of knowledge proposed by the taxonomic system. Also, the behaviorist approach to psychology underlying the Bloom taxonomy lost its dominant position in the interpretation of educational processes and was replaced by other paradigms, most importantly by cognitive psychology. The original cognitive taxonomies thus became less and less popular in practice. The corresponding taxonomies for the affective and the psychomotor domains were constructed at a later stage and, although used in several areas, they did not make a wide-ranging impact similar to the cognitive taxonomy.

The taxonomies as organisational principles are 'blank systems', i.e. they do not specify content. References to specific contents only serve illustrative purposes in taxonomy handbooks. If, for instance, the six levels of Bloom's taxonomy – knowledge, comprehension, application, analysis, synthesis and evaluation – are applied to the educational goals in a specific area of chemistry, we need to specify what exactly must be remembered, understood, applied, etc. (see e.g., Kloppfer 1971).

The original taxonomies, their revisions or modernised versions gave rise and still continue to give rise to new systems and handbooks guiding the definition of objectives in a similar spirit (Anderson & Krathwohl, 2001; Marzano & Kendall, 2007). A common feature of these initiatives is that they carry on the tradition Bloom established and continue to treat the operationalisation of objectives and the decomposition of knowledge into empirically measurable basic elements as central issues. The methods emerging during the course of taxonomy development later became important methodological resources in the development of educational standards.

Standards in Education

The development of standards in education gained impetus in the 1990s. This process was especially spectacular in the English-speaking world, where previously there had been no normative documents regulating teaching content in public education. In some countries, for instance, – with some exaggeration – every school taught whatever was locally decided upon. Under these conditions, education policy had a very restricted margin of movement and there was little opportunity to improve the performance of the education system. This situation then gave rise to various processes leading to a centrally defined set of educational goals at some level (state or national).

Standards essentially represent standardised educational targets. In contrast with taxonomies – as systems, – standards always refer to specific education content. They are developed by specialist, professional teams, working groups composed of experts in a given field, and depending on the properties of the various fields, several methodological solutions may be used.

Although the development of standards takes the latest theoretical constructs and scientific achievements into account, there may be substantial differences between the science standards of different countries (see e.g., Waddington, Nentwig & Schanze, 2007). Standards are usually descriptive and define what a student should know in a given subject on completion of a given grade of school.

As the standards were developed, they were also put into practice both in assessment and in teaching processes, similarly to the earlier taxonomic systems. A series of handbooks were published discussing in great detail the methods of standard development and their applications. There are differences in emphasis, however, compared to the taxonomies. Standards have a direct effect first of all on the contents of education (see e.g., Ainsworth, 2003; Marzano & Haystead, 2008), and the question of assessment based on them is of secondary importance (e.g., O'Neill & Stansbury, 2000; Ainsworth & Viegut, 2006). Standards-based education essentially means that there *are* certain carefully specified, standardised education targets that students of a given age can be expected to attain.

The concept of standards and standards-based education is not entirely new to professionals working in the Hungarian or other strongly centralised education systems. In Hungary, before the 1990s, a single central curriculum specified all education content and a single textbook was published based on this curriculum. Every primary school student studied the same contents and in theory everyone had to achieve the same set of targets. The standardised subject curricula were polished through several decades of practical professional experience in some areas (mathematics, science), while other areas remained subject to the whims of political and ideological agenda. While the processes taking off in the 1990s were greatly influenced by the Anglo-American standards-based model, curriculum regulation could not avoid the pendulum effect and has swung to the other extreme: the current Hungarian National Curriculum contains only a minimum of central specifications. This process took a course contrary to what was taking place in other countries. As a comparison, it is worth noting that the volume discussing the American mathematics standards (National Council of Teachers of Mathematics, 2000) is alone longer than the entire first version of the Hungarian National Curriculum published in 1995. Since then the National Curriculum has become even shorter.

The appearance of standards and standards-based education is not, however, a simple matter of standardisation or centralisation but also introduces a professional and scientifically based method of organising education content. Standards constitute a new approach, which has become dominant even in countries that also had centrally developed curricula before. In Germany, for instance, where education content is already strongly regulated at the level of federal states, new research efforts have been initiated to develop new-style standards (Klieme et al. 2003). The most important defining feature of standards is that they are scientifically based. The development of standards and standards-based education has launched extensive research and development activities throughout the world.

Both the theoretical foundations of standards-based education and the contents and structure of individual specific standards were an important source of information in the development of frameworks for diagnostic assessments. The decision not to impose a uniform structural solution on the content specifications in reading, mathematics and science but, instead, respect the special features of the different content and assessment domains also reflects the traditions of learning standard development.

The frameworks developed here, however, differ from standards in that they do not define requirements or expectations. They share other features, however: the criteria of detailed, explicit and precise description and a strong scientific basis.

Frameworks

To mirror international practice, we use the term *frameworks* for the detailed specifications we have developed. The frameworks of assessments are similar to standards in that they contain a detailed, structured description of knowledge. They differ from standards, however, in that standards approach education from the perspective of outcomes. In contrast to traditional curricula, frameworks do not specify what should be taught or learnt. They also do not set attainable targets although they do convey implicitly what knowledge could or should be possessed at the highest possible level of achievement.

The most widely known examples of frameworks are the ones developed for international surveys. Self-evidently, in the case of assessment programs covering several countries, standards make little sense. These frameworks therefore characterise the knowledge that can be reasonably assessed. When defining contents, a number of different considerations may be observed. In the first waves of the IEA survey, for instance, the starting points of assessment contents were the curricula of participating countries, i.e., what was usually taught in a given domain.

The frameworks of the PISA surveys cover the three major domains of assessment and for each of these, characterise the applicable knowledge that fifteen year-old youths living in our modern society need to possess. In the development of these frameworks a dominant role is played by the typical contexts of application, and the focus is of course on the application of the knowledge of the given disciplines and school subjects.

A third approach to framework development is rooted in scientific research concerned with learning and knowledge, namely, in the achievements of developmental and cognitive psychology. These considerations also dominate in cross-curricular domains related to more than one (or just a few) school subjects. One example for this type of assessment is the fourth domain of the 2000 wave of the PISA survey, which focused on learning strategies and self-regulated learning. The frameworks of this domain were essentially shaped by psychological evidence provided by learning research (Artelt, Baumert, Julius-Mc-Elvany, & Peschar, 2003). The insights of psychology also help characterise learner attitudes, which have been an object of assessment in almost every international survey, and played an especially important role in the PISA science survey of 2006

(OECD, 2006). A further aspect of knowledge acquisition contributed by psychological research is the structure of problem-solving processes, which was a special domain of assessment in PISA 2003 (OECD, 2004), and the latest results of cognitive research provide the background for the assessment of dynamic problem-solving skills planned for PISA 2012.

The frameworks developed for diagnostic assessments (see Chapter 5) have drawn from the experiences of the frameworks of international surveys. They are similar to the PISA frameworks (e.g., OECD, 2006, 2009) in that they create the foundations for the assessment of the three major measurement domains of reading, mathematics and science. They differ, however, in that while PISA focuses on a single generation of students – 15 year olds – providing a cross-sectional view of student knowledge, our frameworks cover six school grades, assess younger students and place special emphasis on the issue of student progress over time.

Each set of the PISA frameworks is developed for a specific assessment cycle and although there is considerable overlap between individual assessment cycles, the frameworks are renewed for each. The PISA frameworks cover the entire assessment process from the defining of the assessment domains through to the characterisation of the organising principles of the domain, the specification of reporting scales and the interpretation of results. The frameworks we have developed cover selected sections of the assessment process: a definition of the assessment domains, a description of the organising principles and a detailed specification of contents. While the major dimensions of assessment and the contents of measurement scales are defined, performance scale levels and quantitative issues related to scales are not discussed. Given the longitudinal component of student development, the construction of scales requires further theoretical research and access to the empirical data.

Multidimensional Organisation of Assessment Contents

The dominant force shaping the educational innovations of the past decade has been the integrative approach. The competencies appearing in the focus of attention are themselves complex units of distinct knowledge components (and, according to some interpretations, also of affective components). Competency-based education, the project method, content-embedded skill development, content-integrated language teaching and various other innovative teaching and learning methods realise several different goals at the same time. The knowledge acquired through such integrative methods is presumably more readily transferable and can be applied in a broader range of contexts. Similar principles are likely to underlie summative outcome evaluations, and both the PISA surveys and the Hungarian competency surveys embrace this approach.

A different assessment approach is required, however, when we wish to forestall problems in learning and identify delays and deficiencies endangering future success. In order to be able to use assessment results as a tool in devising the necessary interventions, the tests we administer should provide more than global indicators of student knowledge. We need to find out more than just whether a student can solve a complex task. We need to discover the causes of any failures, whether the problem lies in deficiencies in the student's knowledge of basic concepts or in inadequacies in his or her reasoning skills, which are needed to organise knowledge into logical and coherent causal structures.

Since diagnostic assessment requires an enhanced characterisation of student knowledge, we adopt an analytic approach as opposed to the integrative approach dominating teaching activities. An assessment program intended to aid learning must, however, stay in tune with actual processes in education. In line with these criteria a technology of diagnostic and formative assessments is being developed drawing from the experiences of summative evaluations but also contributing several new elements of assessment methodology (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Leighton & Gierl, 2007).

The development of frameworks for diagnostic assessments can benefit a great deal from the experiences of previous work carried out in similar areas, especially from the assessment methods used with young children (Snow & Van Hemel, 2008) and the formative techniques developed for the initial stage of schooling (Clarke, 2001). For our purposes, the most important of these experiences is the need for a multifaceted, analytic approach and a special emphasis on psychological and developmental principles. Previous formative and diagnostic systems, however, relied on paper-based testing, which strongly constrained their possibilities. We replace this method by online computer-based testing, which allows more frequent and more detailed measurements. The frameworks must be accordingly tailored to this enhanced method of assessment.

The Aspects of the Organisation of the Content to be Assessed

The contents of assessments can be organised in terms of three major perspectives. This three-perspective arrangement creates a three-dimensional structure, which is schematised in Figure 4.1. In expounding the contents of measurements, however, the building blocks of this three-dimensional structure need to be arranged in a linear fashion. The components of the structure may be listed in various different ways depending on our first, second and third choice of dimension along which we wish to dissect it. In what follows, the structure is peeled open in the way best suited to the purposes of diagnostic assessment.

Our first perspective, the objectives of education, is a multidimensional system itself that encompasses the three major dimensions of our analysis: the psychological (cognitive), social (application) and disciplinary (school subject) objectives. It is these three dimensions for which development scales are constructed in each assessment domain (reading, mathematics and science) (see the next section for details).

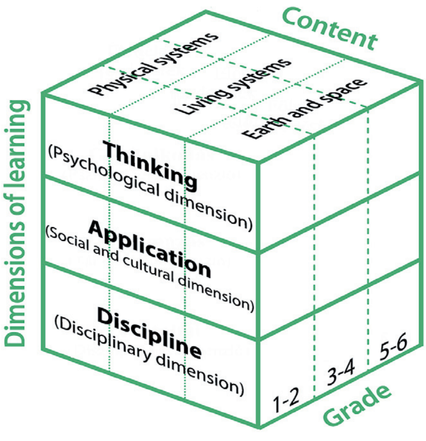

Figure 4.1
The multidimensional organisation of the content of assessments

Our second perspective is development. In this dimension, the six grades of school are divided into three blocks of two years each: Grades 1–2, 3–4 and 5–6. Since the period spanning the six grades is treated as a continuous development process, the above grouping is simply a technical solution to the problem of content disposition. In the absence of empirical evidence, the assignment of contents to different ages (grades) can in any case be no more than an approximation.

Finally, our third perspective is the question of contents available to a given domain of assessment. The content blocks thus broken up constitute the units of the detailed frameworks. With the various possible combinations of the different perspectives, increasing the number of values in any given dimension may easily lead to a combinatorial explosion. In order to avoid that, the number of assessment contents must be determined with caution. The combination of the three learning factors, three age groups and three main content categories of science creates a total of 27 blocks. Identifying further subcategories would substantially increase this figure.

Scales of Diagnostic Assessments, Psychological, Application and Disciplinary Dimensions

Drawing on our experiences of previous empirical studies, the model we have developed is structured along three dimensions corresponding to the three main objectives of education. These objectives have accompanied the history of education and also correspond to the main targets of modern educational performance assessment (Csapó, 2004, 2006, 2010).

The cultivation of the intellect and the development of thinking are objectives that refer to personal attributes rather than invoke external contents. In modern terminology this may be called a *psychological* dimension. As was mentioned in the previous section, this dimension also appeared in the PISA surveys. We have seen a number of assessment domains that interpreted the contents of measurement in terms of psychological evidence. In the case of science, the function of this dimension is to reveal whether science education improves thinking processes, general cognitive abilities or more narrowly defined scientific reasoning to the expected extent.

Another long-standing objective is that schooling should offer knowledge that can be used and applied in non-school contexts. This consideration is termed the *social* dimension and refers to the usability and applicability of knowledge. The concept of knowledge application is related to the notion of transfer of learning, which is defined as the application of knowledge acquired in a given context to a different context. There are degrees of transfer defined by the transfer distance.

The third major objective is that the school should ensure that students acquire the important elements of the knowledge accumulated by science and the arts. This goal is attained when students approach learning observing the principles and values of the given discipline or field of science. This is the *disciplinary* dimension. In recent years a number of educational initiatives have been launched in an effort to counterbalance the previous, one-sided disciplinary approach. Competency-based education and performance assessment focusing on the issue of application have somewhat overshadowed disciplinary considerations. However, for a course of studies to constitute – in terms of a given discipline of science – a coherent and consistent system, which can be reasonably understood, it is necessary to acquire those elements of knowledge that do not directly contribute to the development of thinking or application processes but are indispensable for the understanding of the essence of the discipline. That is, students must be familiar with the evidence supporting the validity of scientific claims and learn the precise definitions ensuring the logical connectedness of concepts in order to possess a system of knowledge that remains coherent in terms of the given scientific discipline.

The three-dimensional model ensures that the same contents (possibly with minor shifts in emphasis) can be used for test task specifications in all three dimensions. Let us illustrate this feature through the skill of organisation. At an elementary level, the operations subsumed under organisation skills, e.g., ordering, classification and grouping, appear during the childhood years. The objects in the world are grouped into categories and conceptual categories cannot be constructed without recognising similarities and differences between these objects or without deciding what attributes to use as a basis for categorisation. The various aspects of organisation skills are improved by classroom exercises and also by the structured presentation of scientific knowledge. The developmental level of organisation skills may be measured with the help of reasoning tasks based on simple content (e.g., classification of everyday objects, grouping of items of clothing according to the season of the year in which they are worn). The task of application may be embedded in an everyday situation such as the grouping of food items to plan a daily and weekly diet according to various criteria (e.g., composition and nutritional values). Finally, we can test whether students have acquired the principles used in biology to classify life forms, the basis of categorisation, the main groups of life forms, the names of these groups and ways of visualising the relationships between the groups and the hierarchy of life (e.g., tree diagrams or Venn diagrams). The last of these is a knowledge component that cannot be developed through exercises stimulating cognitive development but requires specific disciplinary knowledge.

The learning of science is closely connected to general intellectual development. Formal operations and thinking play a dominant role in every area of science and in several areas the applicability of knowledge also has a prominent place. For this reason, there may not be a sharp boundary between the three dimensions in all cases. Whether a certain task belongs to the dimension of thinking, application or disciplinary knowledge depends on the degree of association between the content it measures and disciplinary knowledge, the course syllabus or the context of classroom activities.

The Psychological Dimension of the Assessment in Science

The development of thinking skills and the assessment of their level of advancement as proposed in the detailed frameworks are discussed in the first section of the next chapter, where – in addition to the system of competencies shared with the domain of mathematics – examples are also provided for the assessment of domain-specific elements of scientific inquiry and research. The theoretical framework underlying the examples is presented in Chapter 1 of the volume, where the system of general thinking abilities and various issues in development and the fostering of development are discussed and the relationship between everyday thinking processes, general scientific thinking and the specific reasoning processes of the natural sciences is analysed.

Thinking in Sciences

Scientific thinking is often regarded as a specific mode of thinking. It is used as a cover term for all mental processes used when reasoning about some content of science (e.g., force in physics, solutions in chemistry or plants in biology), or when engaged in a typical scientific activity (e.g., designing and performing experiments) (Dumbar & Fugelsang, 2005). Scientific thinking encourages the development of general thinking skills and is at the same time a prerequisite to the successful acquisition of scientific disciplinary knowledge.

Scientific thinking cannot be reduced to familiarity with the methods of scientific discovery and their application. It also involves several general-purpose cognitive abilities that people apply in non-scientific domains such as induction, deduction, analogy, causal reasoning and problem-solving. Specific components of scientific thinking are linked with specific steps in scientific investigation (e.g., the formulation of questions, the recognition and clear definition of problems; the collection and evaluation of relevant data; the drawing of conclusions, an objective evaluation of results; and the communication of results). They involve the analysis of scientific information (e.g., the comprehension of scientific texts, evaluation of experiments and establishing connections between theories and facts). Further components of scientific thinking include knowledge related to the workings of science and to the evaluation of its impact (e.g., the explanation for the constant evolution of scientific knowledge; the recognition of the close relationship between the physical, the biological and the social world; the recognition of the utility and dangers of scientific achievements; evidence-based reasoning and decision-making), which leads to the dimension of knowledge application.

Development of Scientific Thinking

The intellectual development of children cannot be separated from the evolution of other components of their personality. Students' interests vary with their age: children of different ages think and act differently and have a different relationship to reality. Since there may be substantial individual variation in the pace of cognitive development, the different age-defined stages can have no rigid boundaries. For our frameworks, Grades 1 to 6 of schooling are treated as a single developmental process and, in the absence of empirical evidence, the developmental stages of thinking skills are not linked to the three age groups. However, for the interpretation of the development of thinking and for the analysis of thinking operations, we rely on the psychological attributes known from developmental psychology and make a distinction mainly between Grades 1–4 and Grades 5–6.

In terms of Piaget's stages of cognitive development, the age group covered by Grades 1–6 is essentially characterised by Concrete Operations but signs of the next stage, Formal Operations, may also appear in Grades 5–6. Students in Grades 1–4 are characterised by concrete operations related to their experiences: they can handle a limited number of variables; they can recognise and describe the relationship between the variables but cannot provide an explanation for it. In the Formal Operational stage children can handle problems involving several variables; they can predict and explain events. When characterizing an ecological system, for instance, a student in the Concrete Operational stage will be able to recognise and describe a simple food chain and identify the relationship between the members of the food chain. However, to be able to understand the dynamic balance of the ecosystem as a multivariate system and to understand that a change in the system may bring about further changes upsetting this balance, a higher level of thinking is needed (Adey, Shayer, & Yates, 1995).

The development of scientific thinking is closely related to the level of mathematical skills and to their applicability. The process of scientific inquiry and the operation of scientific research skills require, for instance, elementary counting skills, an ability to use the concept of proportionality, calculate percentages, convert units of measurement, display data, create and interpret graphs, and think in terms of probabilities and correlations.

The operations involved in scientific thinking may be developed from the start of formal education. During this period, a special role is played by direct experience and the observation of objects and phenomena but thinking operations may also be encouraged without performing experiments (e.g., by designing experiments and analysing the results of observations and experiments). As students get older and move forward in their school, the curriculum and the textbooks expect them to learn and apply increasingly difficult scientific methods with a growing number of content areas, while displaying an increasing level of independence (Nagy, 2006a, 2008, 2009).

Several methodological publications have pointed out that young children should be involved in doing science ('sciencing') rather than be taught ready-made scientific facts. The action-oriented and the inquiry-based approaches have also been adopted in science education for young children; with the help of activities and tasks, the children are encouraged to raise questions, search for answers, design experiments and collect data. The results of research on this method suggest, however, that only a few children can acquire the system of scientific knowledge based on simple discovery-based learning. A combination of directed discovery and explicit instruction is a more efficient method.

Chapter 5 discusses how to take into account in the assessment of scientific thinking the psychological attributes characterizing the stages of development of children in Grades 1–4 and 5–6 and the order of appearance of cognitive operations following from them. The operation of general thinking processes is characterised with reference to contents selected from the three science content areas. The development of the detailed content framework made use of the experiences of previous assessment programs in Hungary: with respect to general thinking abilities, the results of studies on inductive (Csapó, 2002), deductive (Vidákovich, 2002), analogical (Nagy, 2006b), combinatorial (Csapó, 1998) and correlational (Bán, 2002) reasoning and organisation skills (Nagy, 1990). The assessment of domain-specific processes is illustrated with examples from the areas of scientific inquiry, problem-solving, text comprehension, evidence analysis and decision-making.

The Application Dimension in the Frameworks

In the three-dimensional model of the contents of diagnostic assessments (Figure 4.1), application is the dimension reflecting social expectations related to learning, and focuses on the social utility of knowledge, its applicability to different contexts, the development of transfer of learning and the ability to create connections between science, technology, society and the environment. The social dimension carries approximately as much weight in the detailed frameworks as do the thinking and the disciplinary dimensions. It describes the standards along which it can be assessed whether at a given stage of development students possess scientific knowledge that can be applied in a way useful to their immediate or wider environment.

The theoretical foundations of the dimension of application are provided by the concept of scientific literacy representing the goals and principles of science education. Scientific literacy has several different definitions. While there are differences in emphasis, all of the interpretations invoke essentially the same social expectation. They construct a theoretical framework of applicable knowledge underlying individual decisions and supporting the interpretation and resolution of day-to-day problems.

Applicable Knowledge

Applicable knowledge may be defined as a complex system composed of content knowledge (factual knowledge) and operations (thinking skills) that remains functional in different contexts. Psychological studies (e.g., Butterworth, 1993; Clancey, 1992; Schneider, Healy, Ericsson, & Bourne, 1995; Tulving, 1979) reveal that learning is situational and the activation and application of knowledge are dependent on the relationship between the context of learning and the context of application. That is, application is not an automatic process; students must learn to transfer both contents and operations. During transfer, the similarities and differences between the two tasks or situations must be identified. The distance between the familiar and the novel task may be unequal in terms of contents versus operations. In addition to transfer distance, several attributes and forms of transfer are discussed in the literature (Molnár, 2006). The current detailed frameworks use the concepts of near and far transfer. Near transfer refers to cases where there is a high degree of similarity between the context of learning and the context of application. For instance, the knowledge acquired in the context of a given topic in a school subject may be applied in the context of a different topic of the same school subject or in a different school subject. Far transfer refers to an instance of application where there are substantial differences between the learning and the application situations, such as the application of school knowledge to complete tasks involving everyday situations and real-life problems (Figure 4.2). Transfer of learning and the application of knowledge are greatly influenced by the attributes of the task and the situation or context appearing in the task. For this reason, the context must be described before applicable knowledge can be evaluated.

The Context of Application

The interpretation of context varies considerably between the different disciplines of science (Butterworth, 1993; Goldman, 1995; Grondin, 2002; Roazzi & Bryant, 1993). For the purposes of the detailed frameworks, context is defined as the totality of objects (people, things and events), their properties and interrelationships, i.e., all the information characterising a situation that activates the relevant knowledge and determines the choice of solution to the task problem.

In the international standards and in the theoretical frameworks of the various surveys, context usually appears in the form of pairs of contrasting modifiers, such as ‘familiar versus unfamiliar/new;’ ‘in the classroom *versus* outside the classroom;’ or ‘scientific/academic versus real-life/realistic.’ The first program to provide a relatively detailed characterisation of context was the PISA survey (*OECD,* 2006). Our detailed frameworks essentially adopt the PISA system, where one test component focuses on the context (personal/social/global) and the other component focuses on the scientific contents and problems having social relevance (e.g., health, natural resources, risks) that are assessed in the various contexts. While these components are preserved in our frameworks, the program is extended to include the assessment of the application of knowledge not only in everyday situations but also in school contexts. Three types of school (classroom) context are distinguished: (1) a different topic within the same school subject, (2) a different science subject and (3) a non-science subject (see Figure 4.2). Non-school contexts cover everyday, real-life situations, which are grouped according to the PISA system into personal, social and global settings.

<table>
<tr><td rowspan="3">School</td><td colspan="2">Different topic in the same school subject</td></tr>
<tr><td colspan="2">Different science subject</td></tr>
<tr><td colspan="2">Non-science subject</td></tr>
<tr><td rowspan="3">Real-life</td><td>Authentic</td><td>Personal (self, family, peer groups)</td></tr>
<tr><td rowspan="2">Non-authentic</td><td>Social (community)</td></tr>
<tr><td>Global (life in the world)</td></tr>
</table>

Figure 4.2
The contexts of knowledge application

Real-life situations refer to phenomena, events, questions and problems that students of a given age are expected to be able to interpret and handle for various reasons, e.g., because they are elements of scientific literacy. Since for younger students (Grades 1 to 6), personal experiences play an important role both in learning and in application, and it is primarily the handling of problems in their immediate environment that constitutes relevant knowledge, real-life tasks are grouped into two categories depending on whether students may reasonably have a concrete experience of the situation represented by the task. A task may thus be classified as authentic or as non-authentic. The contexts of authentic tasks are related to situations taken from students' lives (e.g., travelling or sport) involving mostly their personal or occasionally their social environments: issues concerning their own selves, their families, their peer groups or their wider environment. Non-authentic tasks refer to day-to-day problems involving links between science, technology and society that are not directly relevant to children of the given age (e.g., global warming, alternative sources of energy). For Grades 1–6, the majority of social problems and the set of global issues, i.e., issues impacting on the human race in general, are non-authentic.

The Disciplinary Dimension of the Frameworks

Within the content dimension, science contents are organised in terms of two sets of factors: interdisciplinary and disciplinary considerations. With respect to interdisciplinary considerations, we place special emphasis – in agreement with the discussion of the disciplinary dimension in Chapter 3 – on the development of basic concepts, principles and relationship systems connecting individual disciplines. These constitute the foundations of scientific literacy and can be shaped and expanded not only in Grades 1–6 but throughout the period of science education. The science standards of other countries include several examples of specifying basic concepts and principles, and the Hungarian National Curriculum undertakes to follow this practice. The system we propose includes two basic concepts, matter and energy, and the relationships refer to the relationship between structure and properties, the nature of systems and interactions, the notions of constancy and change, the nature of scientific discovery and the relationship between science, society and technology.

The other approach to science contents follows disciplinary considerations. Based on the four disciplines of science, three content areas have been constructed: Non-Living Systems, Living Systems and Earth and Space Systems. The two disciplines of science concerned with the physical world, materials and their properties and states – chemistry and physics – are not treated separately but are contained within a single content area. Even though in Hungary science education is integrated combining the different disciplines into a single school subject in Grades 1–6 (Environmental Studies or Nature Studies), there are reasons to adopt the above division. The separation of the three content areas allows the various elements of disciplinary knowledge to be monitored in the different age groups, and the method provides an organised system showing the different topics, concepts, facts and relationships appearing within each discipline up to Grade 6. Another advantage of distinguishing these three content areas is that the system can be applied to the entire period of science education, including Grades 7–12, where science is taught divided into disciplinary subjects. The three content areas are in line with the system of categorisation used in the PISA surveys. The frameworks for the 2006 and 2009 waves use similar titles for the knowledge areas in the science domain: Physical Systems, Living Systems, and Earth and Space Systems. In addition to these three areas, the PISA surveys also include Technology Systems and topics related to scientific inquiry and scientific explanations (OECD, 2006, pp. 32−33; OECD, 2009, pp. 139−140). In our program, the latter three areas are positioned among interdisciplinary relationship systems.

For each of the three content areas (Physical Systems, Living Systems, Earth and Space), the knowledge components considered to be of special significance from the perspective of the disciplines of science are discussed in the third section of the Chapter 5. Our discussion of the knowledge, skills and competencies that can be taught and assessed in Grades 1–6 takes the research evidence related to students' thought processes and the development of their knowledge system, and notes variations in student knowledge across the different age groups into account. During the first stage of the study of science, students primarily rely on their own experiences, which is an exceptionally useful starting point but in several areas of science, everyday experiences cannot be directly linked to scientific knowledge; the path leading to understanding of science concepts stretches longer than that. Wherever possible, the relevant stages of conceptual development, their typical manifestations and diagnostic features are described. The description of knowledge development is illustrated with sample tasks that can be used in diagnostic assessments. As the disciplinary dimension takes the standpoint of science disciplines, the tasks appearing here assess the level of acquisition of science content knowledge in contexts familiar from classroom activities.

Physical Systems

This content area encompasses knowledge related to non-living systems in nature. Although the Hungarian National Curriculum places heavy emphasis on knowledge related to the physical world even during the foundational stage of science education, an analysis of the currently recommended framework curricula and the textbooks and practice books currently in circulation reveals that for Grades 1–6, contents providing the foundations of the study of physics and chemistry as science disciplines are considerably underrepresented compared to contents for other science disciplines. We consider the first years of schooling to be an exceptionally important preparatory period with respect to the discovery of the physical world and the acquisition of scientific knowledge and the scientific way of thinking. For this reason, the detailed frameworks – in line with the Hungarian National Curriculum and with curriculum and assessment standards in other countries – encourage the early development of the basic concepts of physics and chemistry, and place more emphasis on knowledge areas preparing the ground for the study of these disciplines (Properties of bodies and matter, Changes of matter, Interactions and Energy) than is currently typical of Hungarian schools.

Children are fascinated by the natural and social environment surrounding them, attempt to find explanations for natural phenomena and are curious to know how the technical tools they encounter every day work. The school plays an important part in helping children to organise the knowledge they have picked up in several different places. If the school fails to fulfill this function, the naive theories constructed by the children can lead to the emergence of misconceptions and to their entrenchment. It is a very important task of education to steer students from the very first years of schooling towards the knowledge and way of thinking that will later enable them to understand the role of science and technology in people's lives. The content framework of non-living systems also points out that the varied activities involved in the study of physics and chemistry develop thinking skills that will come useful in the study of other school subjects and will also be needed for later success in life.

Living Systems

The detailed content framework developed for the knowledge area of living systems describes what knowledge is expected of students in connection with living organisms while also referring to related knowledge in physics, chemistry and physical geography. The contents are fully compatible with the teaching principles defined in the National Curriculum and take into consideration the attributes of different age groups and the objective that the acquisition of the subject matter should help enhance students' cognitive abilities and increase their motivation to learn. The system of expected knowledge contents and the definition of knowledge areas (Criteria of life and the properties of living organisms, Single-celled organisms, Plants, Animals, Fungi, Humans, Populations and Environmental Protection) have been developed keeping the school leaving examination standards in biology in mind, thus allowing the system covering Grades 1–6 to be extended to cover the remaining grades of public education. An important feature of the system is that the detailed content framework emphasises the need to teach the methods of the science of biology (observation and experiments), to highlight the close relationship between biology, technology and society, and to describe concepts and relationships reaching across the various knowledge areas from different viewpoints.

Earth and Space

This content area fulfils a special function in the knowledge of science as it includes knowledge components that are closely related to other fields of knowledge (e.g., mathematics) and, due to their connections with social geography, act as a bridge between natural and social science.

The content framework has been developed with reference to the major logical dimensions of geographical and environmental contents. Geography being a science of space and time, the basic knowledge areas are orientation in space and time, the structures of and events in Earth's spheres (lithosphere, hydrosphere and atmosphere), the properties of regional space at different scales (home environment and Hungary, our planet and the Universe) and issues related to space (the relationship between the natural environment and society, the state of the environment). The content framework describes the contents of geography as environmental science in public education and the basics of the competencies required for the acquisition and application of these contents. The development of the framework relied to some extent on standards in other countries and to a larger extent on the results of Hungarian curriculum theoretical research, current educational documents (the National Curriculum and the school leaving examination standards) and recent trends in geography education theory. An important feature of the framework is that special attention is paid to the step-by-step development of skills and competencies related to the knowledge contents for the different age groups.

Summary and Future Objectives

The detailed frameworks of science are no more than the first step in the lengthy process of developing a diagnostic assessment system. Further work on the theoretical background and the detailed frameworks may be assisted by a number of different sources.

The limited time frame of development excluded the organisation of an external professional debate. Now that the frameworks are published in these volumes in both Hungarian and English, they become accessible to a broader academic and professional audience. The feedback we receive from this audience will be the main source of the first cycle of refinements.

A second, essentially constant source of improvements is the flow of new research evidence that can be incorporated in the system. Some areas develop at an especially rapid rate, such as the study of learning and cognitive development in early childhood. Several research projects are concerned with the analysis and operationalisation of knowledge, skills and competencies. Issues in formative and diagnostic assessment constitute a similarly dynamic research area. The results of these projects can be used to revise the theoretical background and to refine the detailed content specifications.

The most important source of improving the frameworks will be their use in practice. The diagnostic system will be constantly generating data, which may also be used to test and rethink the theoretical frameworks. The system offered here is based on the current state of our knowledge. The organisation of the contents and their assignment to different age groups rely not on facts but on what science views as a hypothesis. The measurement data will provide empirical evidence on *what students know* at a given age. This information and the results of further experiments will be needed to find an answer to the question of *how much further can students progress* if their learning environment is organised more efficiently.

An analysis of the relationships among the various tasks reveals correlations between the scales characterising development. In the short term, we can identify the tasks bearing on the nature of one or another scale and those affecting more than one dimension of assessment. The real benefit of the data, however, lies in the linked data points allowing the longitudinal analysis of the results of successive diagnostic assessments. In the long term, this makes it possible to determine the diagnostic power of the various tasks and to identify the content areas the results of which can predict later student performance.

References

Adey, P., Shayer, M., & Yates, C. (1995). Thinking Science: The curriculum materials of the CASE project (2nd ed.). London: Thomas Nelson and Sons Ltd.

Ainsworth, L. (2003). Power standards. Identifying the standards that matter the most. Englewood, CA: 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.

Artelt, C., Baumert, J., Julius-Mc-Elvany, N., & Peschar, J. (2003). Learners for life: Student approaches to learning. Results from PISA 2000. Paris: OECD.

Bán, S. (2002). Gondolkodás a bizonytalanról: valószínűségi és korrelatív gondolkodás. [Thinking about the uncertain: probabilistic and correlational thinking] In B. Csapó (Ed.), Az iskolai tudás (2nd ed., pp. 212–260). Budapest: Osiris Kiadó.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). Assessment for learning. Putting it into practice. Berkshire: Open University Press.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: the classification of educational goals. Handbook I: Cognitive Domain. New York: Longmans.

Butterworth, G. (1993). Context and cognition in models of cognitive growth. In P. Light, & G. Butterworth (Eds.), Context and cognition (pp. 1–13). Hillsdale, NJ: Erlbaum.

Clancey, W. J. (1992). Representations of knowing: In defense of cognitive apprenticeship. Journal of Artificial Intelligence in Education, *3*(2), 1391–168.

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. (1998). A kombinatív képesség struktúrája és fejlődése [The structure and development of combinatorial skills]. Budapest: Akadémiai Kiadó.

Csapó, B. (2002). Az új tudás képződésének eszközei: az induktív gondolkodás. [Tools of generating new knowledge: inductive reasoning] In B. Csapó (Ed.), Az iskolai tudás (2nd ed., pp. 261–290). Budapest: Osiris Kiadó.

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. (2008). A tanulás dimenziói és a tudás szerveződése. [Dimensions of learning the organisation of knowledge]. Educatio, *17*(2), 207–217.

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: Beltz.

Dumbar, K., & Fugelsang, J. (2005). Scientific Thinking and Reasoning. In K. J. Holyoak, & Morrison, R. G. (Eds.), The Cambridge Handbook of Thinking and Reasoning (pp. 705–725). New York: Cambridge University Press.

Goldman, A. (1995). A tudás oksági elmélete [A causal theory of knowledge]. Magyar Filozófiai Szemle, *39*(12), 231–248.

Grondin, J. (2002). Bevezetés a filozófiai hermeneutikába. [Introduction to philosophical hermeneutics]. Budapest: Osiris Kiadó.

Hartig, J., Klieme, E., & Rauch, D. (Eds.). (2008). Assessment of competencies in educational context. Göttingen: Hogrefe.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H. E., & Vollmer, H. J. (2003). Zur Entwicklung nationaler Bildungsstandards. Berlin: Bundesministerium für Bildung und Forschung.

Kloppfer, L. E. (1971). Evaluation of learning in Science. In B. S. Bloom, J. T. Hatings, & G. F. Madaus (Eds.), Handbook on formative and summative evaluation of student learning (pp. 559–641). New York: McGraw-Hill Book Company.

Leighton, J. P. & Gierl, M. J. (Eds.), (2007). Cognitive diagnostic assessment for education. Theory and applications. Cambridge: Cambridge University Press,

Marzano R. J., & Kendall, J. S. (2007). The new taxonomy of educational objectives. (2nd ed.). Thousand Oaks, CA: Corwin Press.

Marzano, R. J., & Haystead, M. W. (2008). Making standards useful in the classroom. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Molnár, Gy. (2006). Tudástranszfer és komplex problémamegoldás. [Transfer of learning and complex problem solving]. Budapest: Műszaki Kiadó.

Nagy, J. (1990). A rendszerezési képesség kialakulása. [The emergence of organisation skills]. Budapest: Akadémiai Kiadó.

Nagy, L. (2006a). A tanulásról és az értelmi fejlődésről alkotott elképzelések hasznosítása a természettudományok tanításában [Using theories of learning and intellectual development in science education]. A Biológia Tanítása, *14*(5), 15–26.

Nagy, L. (2006b). Az analógiás gondolkodás fejlesztése [Enhancing analogical reasoning]. Budapest: Műszaki Kiadó.

Nagy, L. (2008). A természet-megismerési kompetencia és fejlesztése a természettudományos tantárgyakban [Competency in the study of nature and its encouragement in school science]. A Biológia Tanítása, *16*(4), 3–7.

Nagy, L. (2009). Hogyan támogatják a környezetismeret-tankönyvek a tanulók kompetenciáinak, képességeinek fejlődését? [The role of Environmental Studies textbooks in the development of students' skills and competencies]. A Biológia Tanítása, *17*(5), 3–21.

National Council of Teachers of Mathematics (2000). Principles and standards for school mathematics. Reston, VA: National Council of Teachers of Mathematics.

O'Neill, K., & Stansbury, K. (2000). Developing a standards-based assessment system. San Francisco, CA: WestEd.

OECD (2004). Problem solving for tomorrow's world. First measures of cross-curricular competencies from PISA 2003. Paris: OECD.

OECD (2006). Assessing scientific, reading and mathematical literacy. A framework for PISA 2006. Paris: OECD.

OECD (2009). PISA 2009 Assessment Framework. Key competencies in reading, mathematics and science. Paris: OECD.

Roazzi, A., & Bryant, P. (1993). Social class, context and development. In P. Light & G. Butterworth (Eds.), Context and cognition (pp. 14–27). Hillsdale, NJ: Erlbaum.

Schneider, V. I., Healy, A. F., Ericsson, K. A., & Bourne, L. E. (1995). The effects of contextual interference on the acquisition and retention of logical. In A. F. Healy, & L. E. Bourne (Eds.), Learning and memory of knowledge and skills. Durability and specificity (pp. 95–131). London: Sage Publications.

Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). Early childhood assessment: Why, what and how. Washington DC: The National Academies Press.

Tulving, E. (1979). Relation between encoding specificity and levels of processing. In L. S. Cemark., & F. I. M. Craik (Eds.), Levels of processing in human memory (pp. 405–428). Hillsdale, NJ: Erlbaum.

Vidákovich, T. (2002). Tudományos és hétköznapi logika: a tanulók deduktív gondolkodása. [Scientific and everyday logic: students' deductive reasoning] In B. Csapó (Ed.), Az iskolai tudás (2nd ed., pp. 201–230). Budapest, Osiris Kiadó.

Waddington, D., Nentwig, P., & Schanze, S. (Eds). (2007). *Making it comparable. Standards in science education*. Münster: Waxmann.