



Diagnostic Assessment Frameworks for Reading: Theoretical Foundations and Practical Issues

Benő Csapó

Institute of Education, University of Szeged

Krisztián Józsa

Institute of Education, University of Szeged

János Steklács

Institute of Human Studies, University of Kecskemét

Ágnes Hódi

Center for Research on Learning and Instruction, University of Szeged

Csaba Csíkos

Institute of Education, University of Szeged

Introduction

This chapter serves the purpose of creating a link between the previous three theoretical chapters and the detailed content descriptions given in the next part of this volume. We further provide a characterisation of the genre of frameworks and discuss the factors justifying our choice of solutions.

Chapter 1 gave an overview of international research findings related to the cultivation of reading skills and to the role of literacy in general, approached mostly from the perspective of developmental and cognitive psychology. Chapter 2 looked at reading acquisition from a social and cultural point of view with the focus, once again, on international re-

search evidence. The traditions and curricular features of Hungarian public education appeared in Chapter 3, and a picture of the practices emerged to which the diagnostic programme would need to be tailored. All this information delineates the first problem to be solved: The achievements of the leading edge of scientific research must be adapted in such a way that they have the greatest educational effect both on students as individuals 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 science 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 and comprehension 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. Taking into account and making use of different types of relationship networks are 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 items, measurement scales, data analysis methods and feedback systems also calls for the parallel treatment of the three domains and the sharing of certain principles. However, this parallel treatment also means, that certain compromises must be made: There is a limit to the extent to which 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 into account the special features of individual assessment domains.

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

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 to another issue in another domain. In the first chapter of the volume dedicated to the 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. Reasoning skills are discussed in greatest detail in the science volume but the same skills are also fostered in mathematics instruction. The second chapters of the volumes focus on the issue of knowledge application and each of them draws general conclusions that equally apply to the other two domains. The third chapters examine practical curricular issues sharing a commitment to the historical traditions and current principles of Hungarian public education. At the same time, the discussion of contents and disciplinary questions also reflects the need to follow progressive international trends and to make use of international achievements. In line with the above principles, we regard the nine theoretical chapters in combination as the theoretical foundations of the diagnostic assessment system. The background knowledge analysed in the theoretical chapters thus constitutes a common resource for each of the domains, without needing to detail the shared issues separately in the parallel chapters of the different volumes.

The first section of the present chapter reviews the main factors taken into consideration in developing the frameworks. First, the system used for the specification of the detailed frameworks is described and the subsequent sections show how these principles are used in the development of the reading framework.

The Development of the Frameworks of Diagnostic Assessment

Taxonomies, Standards and Frameworks

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 defining educational 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, while the other driving force 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 growing 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; or we can interpret a taxonomy as a data table with the headings indicating what can appear in its various rows and columns. Compared to previous general characterisations of goals, planning based on such a formalised system constituted a major step forward, and encouraged its developers to give careful thought to the descriptions contained in it.

The greatest impact was made by the first taxonomic system, one describing the cognitive domain (Bloom et al., 1956), which opened up a new path for curriculum and assessment theory. This taxonomic system characterised expected student behaviour in concrete, observable categories. The most obvious novelty was the system of six hierarchically or-

ganised 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. The empirical surveys, however, did not corroborate every aspect of the hierarchy of knowledge proposed by the taxonomic system. Also, the behaviourist 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 domain were constructed at a later stage and 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. The six levels of Bloom’s taxonomy, for instance, are knowledge, comprehension, application, analysis, synthesis and evaluation. When these were used to describe knowledge of a given domain, such as chemistry, they specified what exactly had to be learnt, understood, applied, etc.

The original taxonomies, their revisions or modernised versions gave rise, and 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 despite the decreasing influence of behaviourism, they maintain the tradition Bloom established, the operationalisation of objectives and the decomposition of knowledge into empirically measurable basic elements. The methods emerging during the course of taxonomy development later became important methodological resources in the development of educational standards.

Standards

The development of learning standards gained new impetus in several countries 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, whether state or national.

Learning standards essentially represent standardised educational targets. In contrast with taxonomies, as systems, standards always refer to specific instruction content. They are usually developed by special professional teams and may rely on several different methodological solutions depending on the properties of the various fields. Standards are often developed (or commissioned) by education authorities and tend to be descriptive, defining what a student should know in a given subject on completion of a given grade of school.

As the standards were being developed, they were also put into practice both in assessment and in instruction processes, similarly to the taxonomic systems. A multitude of handbooks were published discussing in great detail the methods of standards development and their applications. There are differences in emphasis, however, compared to the taxonomies. Standards have a direct effect first of all on the content 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 content 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. The processes taking off in the 1990s were greatly influenced by the former Anglo-American standards-based models, but curriculum regulation could not avoid the pendulum effect and has swung to the other extreme: The current Hungarian National Core 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 defining the American mathematics standards (National Council of Teachers of Mathematics, 2000) is alone longer than the entire first version of the Hungarian National Core Curriculum published in 1997. The National Core Curriculum has become even shorter since.

The appearance of learning 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. The development of standards embracing the new approach has become the dominant trend even in countries that had standardised curricula before (e.g., in Germany, see Klieme et al., 2003). The most important defining feature of standards is that they are scientifically based. The development of learning standards and standards-based education has launched extensive research and development activities throughout the world.

Frameworks

Both the theoretical foundations of standards-based education and the contents and structure of specific individual standards were important sources of information in the development of our frameworks of diagnostic assessment. The decision not to impose a uniform structural solution on the content specifications in reading, mathematics and science but, instead, to respect the special features of the different content and assessment domains also reflects the traditions of standard development. The frameworks developed here, however, differ from learning standards in that they do not define requirements or targets. They share other features though: the criteria of detailed, explicit and precise description and a firm scientific basis.

To mirror international practice, we use the term ‘framework’ for the detailed specifications we have developed. The frameworks of assessment are similar to standards in that they contain a detailed, structured description of knowledge. They differ from standards, on the other hand, 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 the content descriptions 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 programmes covering several countries, learning standards or targets make little sense. These frameworks therefore characterise the knowledge that can be reasonably assessed. When defining content, a number of different considerations may be observed. In the first waves of the IEA survey, for instance, the starting points of assessment content were the curricula of participating countries, i.e., what was usually taught in a given domain. The frameworks of the PISA surveys characterise for each main assessment domain the applicable knowledge that fifteen year-olds living in our modern society need to possess. In the development of these frameworks a dominant role is played by the needs of modern societies and typical contexts of application, and the focus is of course on the application of the knowledge of given disciplines or 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 of 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 considerations and the results of 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). Psychological studies have also mapped the

structure of problem solving processes, which was a special domain of assessment in PISA 2003 (OECD, 2004).

The frameworks developed for diagnostic assessments 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 focus on three major assessment domains creating the foundations for the assessment 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, apply to 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. Although there is considerable overlap between individual assessment cycles, the content descriptions 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 considerations of student progress, the construction of scales requires further theoretical research and access to the empirical data.

The Dimensions of the Diagnostic Assessment of Reading

The dominant force shaping the educational innovations of the past decade has been the integrative approach. The competencies receiving attention are themselves complex units of various knowledge components (and, according to some interpretations, also of affective components). Competency-based education, the project method, problem-based learning, content-embedded skill development, content-integrated language teaching and various other innovative teaching and learning methods

realise several different goals at the same time. They both transmit knowledge of a given subject or subjects and also foster the development of various general-purpose skills and competencies. 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 needed, however, when we wish to forestall problems in learning, to identify delays and deficiencies endangering future success and to support direct learning processes. 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 the reasoning skills needed to organise knowledge into logical and coherent causal structures.

To be able to fulfil the above requirements, student knowledge must be described in great detail for diagnostic assessments. We therefore adopt an analytic approach as opposed to the integrative approach dominating teaching activities. An assessment programme intended to aid learning must, however, stay in tune with the various aspects of learning and knowledge application. In line with these criteria, a technology of diagnostic and formative assessment is being developed drawing on the experiences of large-scale summative evaluations but also contributing several new elements of assessment methodology (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Leighton & Gierl, 2007).

The development of the frameworks of diagnostic assessment can draw on the experiences of previous projects, especially of the assessment methods used with young children (Snow & Van Hemel, 2008) and the formative techniques developed for the initial stage of schooling (Clarke, 2001). The most important of these experiences are 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 and pencil testing, which heavily constrained

their possibilities. We replace this method with online computer-based testing, which allows more frequent and more detailed measurements and new task formats. That is, assessments can be of a higher resolution than before and the frameworks must be tailored to this new method.

Scales of Diagnostic Assessment, the Psychological, Applicational 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 which have accompanied the history of schooling (Csapó, 2004, 2006, 2010). The development of the frameworks of diagnostic assessment also moves along these three dimensions. This three-dimensional approach can be directly applied to the domains of mathematics and science and in a somewhat extended sense also to the domain of reading. Having a testing system developed with a uniform approach and according to a uniform interpretation framework for the domains of reading, mathematics and science alike facilitates the implementation of assessments and the utilization of the feedback information transmitted to the user. Psychological factors have priority in the cultivation of foundational skills and competencies, the goals and contexts of the use of knowledge need to be considered in the application of acquired knowledge, while teaching itself may start with the content knowledge to be transmitted, which requires tasks to be organised in curricula.

The cultivation of the intellect and the fostering of thinking skills are objectives that refer to personal attributes rather than invoking external contents. In modern terminology this may be called the psychological dimension. As was mentioned in the previous section, this dimension also appears in the PISA surveys, where problem-solving skills are assessed with this consideration in mind. We have also seen a number of assessment domains that interpret the contents of measurement in terms of psychological evidence. In the domain of reading, the function of this dimension of diagnostic assessment is to reveal whether the cognitive skills needed for literacy are sufficiently developed.

Another long-standing objective is that schooling should offer knowledge that can be used and applied outside the classroom. This consid-

eration is termed the social dimension and refers to the external 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 in a different context. There are degrees of transfer defined by the transfer distance. In the domain of reading, the question to ask is whether the development of reading skills can adequately assist students in comprehending texts in other school subjects, in solving reading problems in real everyday life and in extracting and interpreting the information content of texts appearing in different formats.

The third major objective is the acquisition of 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. In the disciplinary dimension of our assessment programme, the acquisition of the subject matter is tested directly in a familiar school context and according to the principles of the given discipline. 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 basics 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. Since the instruction of reading and its place as a school subject differ from mathematics and science in that reading does not have direct disciplinary content similar to that of the other two domains, this dimension is interpreted in a slightly different way from its interpretation for mathematics and science.

The theoretical background to the three dimensions of assessment in the domain of reading is summarised in the first three chapters of the current volume. As was also apparent in the theoretical discussion, the

acquisition of reading is closely related to general cognitive development, and the three dimensions are not always separated by sharp boundaries, which is also indicated by the lesser or greater overlap between the chapters. Since text comprehension plays a decisive role in every other domain of learning, the level of reading skills is closely related to performance in other areas of knowledge. The assignment of a given task or of the comprehension of a given text to one or another of the psychological, applicational or disciplinary dimensions partly depends on the specific comprehension skill the task is designed to assess.

The Organisation of the Contents of Assessment

The contents of assessment are organised in terms of the three major perspectives previously discussed. Within this arrangement, a further level of classification is used taking students' ages and level of development into account. This system is schematised in Figure 4.1. The six grades of school are divided into three blocks of two years each, Grades 1-2, 3-4 and 5-6, in line with the usual grouping of curricular and learning standards. However, 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. Given the large differences between individual students, the assignment of contents to different ages (grades) can in any case be no more than an approximation. A more precise grouping of test items into age groups can only be achieved on the basis of solid empirical evidence.

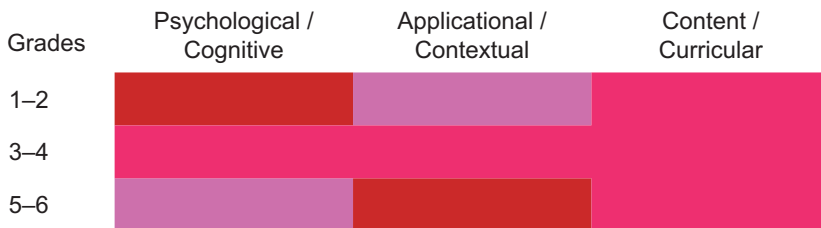


Figure 4.1
The dimensions of diagnostic assessment
and the focal points of instruction

For the diagnostic assessment of reading, the three dimensions have varying significance across the different age groups, as shown by the different shades of colour in the figure. At the first stage of schooling, the fostering of skills and the construction of firm foundations have special significance, which gives special emphasis to tasks incorporating psychological factors. A greater number of test items and more extensive coverage allow more frequent assessments. Application and the comprehension of different types of text are, in contrast, less important at this stage, and the test items discussed here are intended to test students who learnt to read earlier on, possibly before starting school. Tasks assessing the specific curricular targets of reading instruction have equal significance for all three age groups.

For the domains of mathematics and science, we can identify the content constituting the foundations of the given discipline which students need to acquire at school. In this respect reading occupies a special place among school subjects because it does not have core content that should be acquired. Nevertheless, we can identify a range of texts for every culture that has been shown to be suitable for learning reading techniques and can be a tool of reading instruction for a relatively long period of time. The assessment of the curricular dimension focuses on these familiar texts, which tend to be used in reading instruction allowing students to practise reading techniques.

A further explanation for the special role of reading assessment is that reading is a form of linguistic communication. As such, from Grade 5 onwards it is no longer a school subject in the sense that the other two domains are. In terms of disciplines or school subject structure, reading is replaced by Hungarian language and literature. The fostering of reading skills, however, cannot be equated with or mapped onto the school subjects of language and literature, which focus on the transmission of written cultural knowledge. Reading and text comprehension form the foundations of every school subject and are indispensable for social adjustment and success.

Since we are dealing with a linguistic activity or skill, the different levels of reading and text comprehension can be simply interpreted in terms of the structural levels of language. We can thus assess the phonological level through letter to sound mapping, the morphological level through syllabification, the lexical level through word recognition rou-

tines, the syntactic level through sentence reading and the textual level through text comprehension. Approaching the question from a different perspective, the syntactic, semantic and pragmatic levels of language may be interpreted in terms of the tripartite division into psychological, social and disciplinary dimensions. In this case, the knowledge dimension is closer to the three linguistic or communication dimensions applied to the domain of reading.

If we now want to find the analogues of the psychological, social and disciplinary dimensions defined for mathematics and science in the domain of reading, these three dimensions can be interpreted as follows.

The psychological dimension comprises the skills constituting the cognitive preconditions of reading acquisition: Phonological and phonemic awareness, speech sound (phoneme) processing, letter, word and sentence reading and inference which is needed for sentence and text comprehension. The latter appears among the cognitive operations discussed in the social dimension at a different level: Information retrieval, interpretation and critical evaluation, i.e., thinking processes underlying the comprehension of written texts. In brief, we place the components of reading as a mental structure in the psychological dimension.

Those aspects of knowledge that are pertinent to application can be classed with the social dimension. For reading, these include familiarity with the functions of various text formats and text types and their uses, and the texts and reading operations needed for further studies, everyday life and day-to-day coping. An indispensable precondition of success in these areas is the acquisition of habits and routines related to reading. The social dimension of reading has been thoroughly researched by the PISA surveys. This background knowledge provided a solid starting point for the development of our detailed frameworks.

The disciplinary (curricular) dimension encompasses the components of reading directly related to school subjects, covering the topics of reading in the first four grades and those of Hungarian language and literature in higher grades. Knowledge related to written text, a grammar topic, is included here. Another topic is familiarity with reading strategies that can be successfully fostered and assessed in the classroom. Further examples in this dimension are oral reading and its assessment, the use of reading skills to learn poems, rhymes, stories, fables and other literary works, indeed, any text that is worth reading because it has inherent sig-

nificance. The definition and organisation of texts of this type are, however, beyond the scope of the detailed framework described in the present volume. In addition to the above examples, texts read by students in other classes (e.g., history or mathematics) for study purposes can also be included here.

Keeping all these considerations in mind, we find that in the domain of reading, the skills dominating the development of the detailed framework are the reading skills belonging to the psychological dimension. The description of the framework accordingly starts with this dimension. Needless to say, the roles of the other two dimensions are also beyond doubt. We should remember, however, that the three dimensions are often difficult to isolate from one another in a given reading task, word, sentence or text in the description of the framework. In most cases, the three dimensions have equally important roles in reading or in the process of comprehension.

The Phases of the Acquisition of Reading Skills

Preparation, Sound, Letter, Word

Preparing children for reading has recently received an increasing amount of attention. In Hungary, the view emphasising the role and complexity of this process rose to the forefront in the nineties (Gósy, 1990, 1999). The recognition of the importance of pre-reading skills owes a great deal to psychology, special education theory and speech-language pathology and has by now become widely accepted throughout the Hungarian education system (Csépe, 2005; Nagy, 2002).

The length of the preparation phase and the nature of activities best suited to this goal may vary greatly across individuals. It is therefore crucial to establish differences between students and to have them to work in groups in the first grade. The areas of development during the preparation phase include thinking skills and linguistic abilities including linguistic awareness and knowledge related to language and reading. Other areas of special significance here are general cognitive processes such as attention, hearing and vision.

Since the Hungarian language has a shallow orthography, the identification and distinguishing of individual speech sounds have a fundamental role in learning to read and write. Speech sound perception can be successfully cultivated even during the pre-school period (Fazekasné, 2000; Józsa & Zentai, 2007). For children who have suboptimal speech sound perception skills when they enter school, practice activities with sounds and speech sounds have a significant role in the reading preparation phase in Grade 1.

The successful recognition and identification of speech sounds is a prerequisite to the acquisition of letters (Nagy, 2006a). A number of different methods of teaching letters are used around the world and in Hungary, and the choice depends to a large extent on the properties of the given language and on its writing system. The timing of the introduction of letters and the order in which they are taught are determined by the method of reading instruction used. An especially wide range of methods is in use in Hungary relative to other countries. Most teaching materials and methods take into account the principle of homogeneity or Ranschburg inhibition effect, i.e., letters similar in shape are preferably not taught simultaneously. Confident letter recognition is next in line among the preconditions of good reading skills.

The same observation can be made of the availability of a sufficient number of word recognition routines. The acquisition of a sufficient number of word recognition routines is also required for fluent reading (Nagy, 2006a). In addition to the number of these routines, another crucial question is which words belong to this set for a particular child, i.e., which words he or she can recognise as a whole. For good performance in reading comprehension, there should be the greatest possible overlap between this set of words and the vocabulary used in schoolbooks or, at a later stage, the vocabulary used in everyday texts.

The framework introduced in this volume covers the assessment of both letter reading skills and word reading skills. A further ability discussed is syllabification. Syllabification has a long history in Hungarian reading instruction; the debate over its usefulness and importance was closed at the end of the past century, when its place in reading instruction was validated (Adamikné, 1993, 2001).

Sentence Comprehension

Another fundamental precondition of meaningful reading is the confident comprehension of sentences, which encompasses not only general linguistic factors but also knowledge of the rules of writing. The comprehension of a sentence also involves the identification of sentence modality (declarative, interrogative, etc. modes), which signals the intentions of the speaker in a written text. With respect to content, we have a more intricate picture since sentence-closing punctuation marks convey substantially less information than do intonation and other suprasegmental features of spoken language. For this reason, the recognition of the discrepancy between the speaker's intentions and the information conveyed by punctuation marks is a crucial component of text comprehension.

Research on the mechanisms and nature of sentence processing mostly focuses on spoken language; this is primarily an area of psycholinguistics (Pléh, 1998). Since sentence comprehension is analysed here in the context of diagnostic assessment, factors such as the identification, confident recognition and use of letters, words and various non-verbal elements (punctuation marks, text arrangement and typographical features) all need to be integrated.

Sentence comprehension is difficult to define in terms of the levels of language as it encompasses a large area both in speech and in writing ranging from one-word sentences, simple subject-predicate structures (i.e., sentences consisting of only a subject and a predicate) and expanded, modified versions of these structures at one end to long complex-compound sentences with multiple clauses that are similar to a simple text both in terms of formal and content properties at the other. A reasonable option is to analyse the microstructure of texts, namely paragraphs and linguistic units shorter than paragraphs, at the level of sentences. Our framework therefore classes short dialogues and brief texts containing just a couple of sentences with sentence comprehension.

Oral and Silent Reading, Fluency

With respect to oral and silent reading we must note that both are indispensable components of the cultivation of reading skills and each has its

special role in the process. It is difficult to give guidelines as to their relative weight in reading instruction as this is dependent on the children's age and the characteristics of their development.

Although the detailed framework described in the next chapter is dominated by tasks related to silent reading, we must not forget the significance of expressive oral reading. Oral reading is primarily used in the classroom to foster comprehension skills and at the same time to encourage expressive and meaningful reading. It is the acoustic channel of reading aloud that allows teachers to test and evaluate the strengths and weaknesses of children's reading techniques and their reading rates. Oral and silent reading skills tend to have similar structures and to be closely interconnected within an individual. In some cases, however, one or the other of these reading skills may be less developed than the other. Our framework treats the assessment of oral reading as having the closest relationship with the comprehension of sentences. The justification for this principle is that the combined use of expression, interpretation and vocal variation first appears at the level of sentences. Although the tasks refer to texts, students should be asked to read the individual sentences making up these texts in an expressive manner.

The relative weights of oral and silent reading change substantially over the first six grades of formal education. Initially, both classroom activities and everyday reading situations are dominated by oral reading, and at later stages silent reading takes precedence. By the time adulthood is reached, the significance of oral reading over silent reading is reduced to a minimum. During the years of schooling, however, it is useful to test students' oral reading occasionally in all grades. Enhanced attention to this skill is also warranted by empirical data indicating an increase in error rates in oral reading over the past decades (Adamikné, 2000; Molnár, 1993).

Reading Strategies, Texts, Text Comprehension

There are two areas showing salient deficiencies when the Hungarian system of reading instruction and classroom practices are compared to education systems producing outstanding outcomes as measured by international surveys (e.g., Finland and Norway). The first problem is that

the teaching of reading strategies is still not used widely enough in Hungarian schools. The second difference is that the schoolbooks used in the Hungarian system offer a less heterogeneous set of text types or genres and lack the varied text structures typical of everyday life. We may contend that devoting attention to these two factors could bring about substantial positive changes in Hungarian reading instruction and the cultivation of reading skills.

The acquisition of reading strategies is accompanied by enhanced motivation, text comprehension and metacognitive knowledge related to reading and to an individual's own reading processes. A number of methods have been developed for the instruction of these strategies. One of these, the transactional instruction of reading comprehension strategies, has offered several useful experiences. The method emerged in the early nineties and aptly illustrates the theoretical approach characteristic of the reading instruction efforts of the past two decades (Pressley, El-Dinary, Gaskins, Shuder, Gergman, Almasi, & Brown, 1992). Our detailed framework incorporates reading strategies from the first to the sixth grade. For higher grades, we recommend activities which involve practising and integrating the acquired strategies. The potential of strategy-based reading instruction has been demonstrated by a successful Hungarian teaching experiment involving fourth grade students (Csíkos & Steklács, 2010).

The potential for improvement regarding the issue of text variety mentioned above is closely related to the learning of strategies. Specifically, varied reading strategies must be used to understand varied text types. Lower primary school textbooks contained almost exclusively narrative and educational texts up until the first decade of the new Millennium. More recent books include more texts displaying a modern approach and a larger variety of genres and styles. The task remains, however, to choose reading texts matching the structures of texts students are likely to encounter during and after their school years and to use for their studies and to be successful in their everyday lives. This structural change would enable the Hungarian education system to achieve better performance in international surveys assessing system-wide text comprehension (PISA and PIRLS). The development of the detailed framework described in the next chapter was also guided by these considerations.

Frameworks of Reading Assessment: International Practices

The assessment of reading skills, including reading as a tool, is a central issue in most international system-wide surveys. International comparative surveys are designed to obtain summative evaluations to be able to compare education systems and are not suitable for individual diagnostic assessment.

In recent decades there has been a substantial shift in the definition of reading skills in international surveys (Csíkos, 2006; D. Molnár, Molnár & Józsa, 2012). The selection of texts and tasks included in the assessment is strongly dependent on the definition and interpretation of reading skills. The PISA 2009 framework defined reading literacy as the ability to understand, use and reflect on written texts and to engage with written texts in order to achieve one's goals, to develop one's knowledge and potential and to participate in society. It also mentions the need to engage with reading, which refers to the motivation to read and encompasses a number of affective and behavioural characteristics. The PISA assessment framework therefore comprises the set of reading operations and in addition focuses on the motivation to read and on metacognitive processes and strategies related to reading (PISA 2009 Assessment Framework, 2009).

Since PISA is not concerned with the assessment of basic-level reading techniques, the following comprehension skills were tested in 2009: Retrieving information, forming a broad understanding, integrating, reflecting on and evaluating a text. It follows from the PISA definition of reading that these mental operations should be assessed on a variety of written text types. In terms of medium, printed and electronic texts are distinguished, and the latter type is grouped into authored and message-based texts. Authored texts are static, i.e., the reader cannot modify the text appearing on the screen. Message-based electronic materials, in contrast, can be modified, expanded or edited by the reader. In terms of text format, PISA distinguishes continuous versus non-continuous texts, and there are tasks based on mixed formats or on a set of different texts (OECD, 2009).

PIRLS (Progress in International Reading Literacy Study) also views reading as a means to an end. It defines reading as the ability to construct meaning from a variety of texts. In addition to comprehension, the study

also assesses reading behaviours and attitudes supporting lifelong reading. Similarly to the PISA surveys, PIRLS 2011 measures a variety of reading processes: The retrieval of information explicitly stated in the text, the making of straightforward inferences, interpretation and integration, and the examination and evaluation of the content and formal properties of the text (Mullis, Martin, Kennedy, Trong, & Sainsbury, 2009).

The PIRLS surveys originally targeted fourth grade students but in 2011 their coverage was extended. The PrePIRLS framework was introduced, which preserves the goals and principles of the original fourth grade assessment but comprises easier tasks to test students whose basic reading skills have not reached an advanced level (PrePIRLS Information Sheet, 2011). The test items are easier in the sense that they are based on shorter texts, a smaller vocabulary and simpler syntax and the focus is on lower level reading processes. PrePIRLS assessments, like PIRLS assessments, are concerned with reading comprehension and lack tasks testing decoding or word reading.

The NAEP (National Assessment of Educational Progress) framework was created to assess schools in the United States. Reading is one of its main domains. NAEP defines assessment criteria for the fourth, eighth and twelfth grades. The framework in use between 1992 and 2007 covered three main reading processes (Reading Framework for the 2007 National Assessment of Educational Progress, 2006). These are (1) forming a general understanding of the text and developing an interpretation (e.g., Describe what the text is about.); (2) making connections between the text and the reader's background knowledge (e.g., What can the frog eat other than what is listed in the text?); (3) examining content and structure (e.g., Compare the two texts.). In 2009 there was a major shift in the conception of reading assessment: The measurement of vocabulary was added to the skills under assessment, the former three text types (literary, informational and functional) were replaced by only two types (literary and informational), and the PISA 2009 scheme of testing cognitive processes was adopted. The NAEP framework makes measurements relative to benchmarks. Three achievement levels are defined, each of which is linked to a set of standards. NAEP views reading as a complex process in which various cognitive skills interact. This assessment method cannot, however, be used for individual diagnostic purposes because it does not separate the various factors influencing performance.

CBAL (Cognitively Based Assessment of, for and as Learning) is an innovative research-based approach to reading assessment, which undertakes to develop methods of measuring reading skills providing both formative and summative assessments. Both reading ability itself and the cognitive skills involved in reading are assessed. The framework relies on a competency model, which defines three components of reading competency: A prerequisite reading skill, which is the acquisition of the technique of reading, a mental model building skill, which refers to the construction of meaning, and the skill of application, which is the ability to make use of the text. The model includes three categories of reading strategies as its core component: Pre-reading strategies, model building strategies and strategies for going beyond the text (O'Reilly & Sheehan, 2009).

In contrast to the above assessment frameworks focusing on text comprehension, The Abecedarian Reading Assessment targets the initial phase of reading acquisition, and, consequently, basic-level reading skills. The system is suitable for diagnostic assessment. Six basic skills are distinguished: letter recognition, phonological and phonemic awareness, the alphabetic principle, vocabulary and decoding. Various tasks are provided for the assessment of the different components, e.g., finding synonyms and antonyms to test vocabulary size, letter reading to test letter recognition, etc. The six components are assessed through six subtests. Students' performance in the tests can be used to make recommendations on further learning. It is proposed that children should be able to complete the first three subtests with no more than two errors before the start of the first grade and the remaining subtests should be completed with at most two mistakes before the student enters the second grade (Wren & Watts, 2002).

In the United States the Goals 2000: Educate America (1994) and the No Child Left Behind (in effect since 2001) Acts have required every state to develop an unambiguous and transparent framework defining the knowledge targets students must achieve by the end of their public education. The states' learning standards therefore define targets for the 12 grades of schooling focusing on two domains: English language and mathematics. The standards of two states of the US are described in some detail below as an illustration.

The Nevada English Language Arts Standards (2007) describe achievement indicators starting with the kindergarten year. The indicators de-

scribe the content standards related to individual reading skills and specify minimum expectations. The standards are organised by school grade but the framework treats the period from the kindergarten year to Grade 4, the period of Grades 5-8 and the period of Grades 9-12 as single developmental units. The standards are defined for three major areas: (1) word analysis, which involves prerequisite skills and basic-level reading skills; (2) reading strategies and (3) the comprehension of different types of text (literary and non-literary). The framework is intended for teachers, parents and students alike, and provides guidelines for the development of assessment principles and test items.

The Wisconsin Knowledge and Concepts Examinations (WKCE) Reading Assessment Framework is a system of standards relying on educational content targets and it is designed primarily for assessment. The framework uses a criterion-referenced assessment system and describes the principles and standards related to the development of reading skills from Grade 3 to Grade 10. It uses three types of text: literary texts, informational texts and everyday texts. The latter category includes texts such as application forms, product labels and schedules. The reading skills assessed are word reading/word comprehension, text comprehension and text analysis and evaluation (Wisconsin Student Assessment Framework Criterion-Referenced Test Framework, 2005).

The international frameworks described above cover the three dimensions defined in our framework in different contexts. The synthesis of their results and their adaptation to the diagnostic assessment of reading and text comprehension therefore requires further work. Nevertheless, the international models are a useful starting point for the development of the Hungarian framework. The process of adapting the international principles must take into account the characteristics of the Hungarian language and culture.

Reference Points for Assessment Scales

Traditionally, two methods of rating are used in performance assessment: norm-referenced assessment and criterion-referenced assessment. In norm-referenced assessment students' level of achievement is rated relative to the mean score of the population sample and the focus is on

performance differences. The norm-referenced approach is used by the PISA programme and the Hungarian competency assessments, for instance (Balázsi, Ostorics, Szalay, & Szepesi, 2010; D. Molnár, Molnár & Józsa, 2011; Vári, 1999, 2003). In the PISA system, the evaluation of students based on their relative performance, i.e., the norm-referenced method, is supplemented by an achievement standard scale: The scores are standardised such that the mean performance of the OECD countries is set at 500 points with a standard deviation of 100 points characterising the scale of differences between the students (OECD, 2000). These scores are used to define the boundaries of good and less good reading comprehension skills. The same method of standardisation with a mean of 500 and a standard deviation of 100 is in the Hungarian national assessment of reading competency (Balázsi, Rábainé, Szabó & Szepesi, 2005). In these systems, therefore, a good reader is one who displays substantially better reading performance than his or her peers, i.e., who performs better than average.

Criterion-referenced assessment compares test results to an external target and each student's performance is evaluated relative to that target (Csapó, 1987; Nagy, 2010). When skills and abilities are assessed, criterion-referenced testing can be used only if all components of the given skill are known. Criterion-referenced tests are well suited to diagnostic purposes because they reveal the level reached by a student in the acquisition of a particular skill.

Some components of reading competency may be evaluated through criterion-referenced testing. The Hungarian DIFER diagnostic assessment package, for instance, is a system which relies on this method (Nagy, Józsa, Vidákovich & Fazekasné, 2004a, 2004b). The package is widely used by Hungarian kindergartens and primary schools. DIFER comprises criterion-referenced diagnostic tests for the measurement of seven foundational skills. One of these is speech sound processing, which plays an essential role in reading acquisition. The test covers the system of Hungarian speech sounds and the results help to reveal which sounds cause difficulties thus hampering the development of reading.

Knowledge of the letters of the alphabet is another critical precondition of reading. Letter recognition (the recognition of lower and upper-case printed and handwritten letters) can also be measured through criterion-referenced tests. Criterion-referenced diagnostic assessments allow

teachers to find out whether a student has learnt every letter covered in class.

For the assessment of word reading, Hungarian criterion-referenced tests were developed by Nagy (2006b), which test the knowledge of the 5000 most frequent Hungarian words. This set of words is characterised as the critical vocabulary of an optimal level of word reading competency. The words are arranged into 20 interchangeable tests, each of which is divided into four subtests: Headwords, morphologically complex words, synonyms and word meanings. The analysis of the tests uses two empirical indicators: Vocabulary and fluency. Vocabulary is measured by the percentage of correct answers, with a performance of 90% or more satisfying the achievement target of the criterion-referenced test. The fluency indicator measures reading rate. The target level is determined with reference to teachers' performance, i.e., students' reading rates are compared to the average reading rate of teachers. The measurements relying on these criterion-referenced tests reveal whether reading these words is an automated routine process for a student.

Criterion-referenced tests are also found among tests linked to reading standards in other countries. One example is the *Wisconsin Assessment Framework for Reading* described above, which defines content and assessment programmes for Grades 3-8 and Grade 10. The framework uses criterion-referenced tests, which define five proficiency levels. Reading comprehension is a complex skill and several cognitive abilities contribute to its development. For this reason, there may be considerable performance differences between children even within a school grade. The complexity of the skill calls for the use of both norm-referenced and criterion-referenced tests in the assessment of the components of reading competency.

Summary and Further Tasks

The reading framework constitutes a starting point for the development of the diagnostic assessment system. The content of measurement as it is described in this volume is not considered to be complete; further work is needed for the construction of a model that maintains lasting validity but can at the same time be straightforwardly updated at any time. There

are several sources that can contribute to the enhancement of the theoretical background and the assessment framework.

The limited time frame of development did not allow the organisation of an external professional debate. Now that the content frameworks are published in these volumes in both Hungarian and English, they will 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 content 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 students can progress if their learning environment is organised more effectively.

An analysis of the relationships among the various tasks reveals correlations between the scales characterising development. In the short term, we can identify the test items bearing on the nature of one scale or another 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 test items and to identify the areas the results of which can predict later student performance.

References

- Adamikné Jászó, A. (1993). Nyelvi tudatosság – olvasástanítás – helyesírás [Linguistic awareness – reading instruction – spelling]. *Magyar Nyelvőr*, 117(3), 320–331.
- Adamikné Jászó, A. (2000). Változott-e húsz év alatt a főiskolások kiejtése és olvasása [Has there been any change in the articulation and reading of college students over twenty years]? In Gósy Mária (Ed), *Beszédkutatás 2000* [Speech research] (pp. 124–131). Budapest: MTA Nyelvtudományi Intézet.
- Adamikné Jászó, A. (2001). *Anyanyelvi nevelés az ábécétől az érettségiig* [Language education from the alphabet to the school leaving exams]. Budapest: Trezor Kiadó.
- Adamikné Jászó, A. (2006). *Az olvasás múltja és jelene* [The past and present of reading]. Budapest: Trezor Kiadó.
- 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*. Paris: OECD.
- Balázsi I., Ostorics, L., Szalay, B., & Szepesi, I. (2010). *PISA 2009. Összefoglaló jelentés. Szövegértés tíz év távlatában* [Summary report. Text comprehension after ten years]. Budapest: Oktatási Hivatal.
- Balázsi I., Rábainé Szabó, A., Szabó, V., & Szepesi, I. (2005). A 2004-es Országos kompetenciamérés eredményei [The results of the of 2004 National Assessment of Competencies]. *Új Pedagógiai Szemle*, 55(12), 3–21.
- 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: Longman.
- 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. (1987). A kritériumorientált értékelés [Criterion-referenced assessment]. *Magyar Pedagógia*, 87(3), 247–266.
- 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 and the organisation of knowledge]. *Educatio*, 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.

- Csépe, V. (2005). *Kognitív fejlődés-neuropszichológia* [The neuropsychology of cognitive development]. Budapest: Gondolat Könyvkiadó.
- Csíkos, C., & Steklács, J. (2010). Metacognition-based reading intervention programs among 4th grade Hungarian students. In A. Efkiades, & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (pp. 345–366). New York: Springer.
- Csíkos, C. (2006). Nemzetközi rendszerszintű felmérések tanulságai az olvasástanítás számára [The experiences of international system-wide surveys for reading instruction]. In K. Józsa (Ed.), *Az olvasási képesség fejlődése és fejlesztése* [The learning and teaching of reading skills] (pp. 175–186). Budapest: Dinasztia Tankönyvkiadó.
- D. Molnár, É., Molnár, E. K., & Józsa, K. (2012). Az olvasásvizsgálatok eredményei [The results of reading assessments]. In B. Csapó (Ed.), *Mérlegen a magyar iskola* [Hungarian schools in an international context] (pp.17–81). Budapest: Nemzeti Tankönyvkiadó.
- Fazekasné Fenyvesi, M. (2000). A beszédhanghallás kritériumorientált fejlesztése [The criterion-referenced cultivation of speech sound perception]. *Új Pedagógiai Szemle*, 50(7–8), 279–284.
- Fazekasné Fenyvesi M. (2006). *A beszédhanghallás fejlesztés 4–8 éves életkorban* [The cultivation of speech sound perception at age 4-8]. Szeged: Mozaik Kiadó.
- Gósy, M. (1990). Modell az olvasástanításhoz [A model for reading instruction]. *Fejlesztő Pedagógia*, 2, 28–32.
- Gósy Mária (1999). *Pszicholingvisztika* [Psycholinguistics]. Corvina Kiadó, Budapest.
- Hartig, J., Klieme, E., & Rauch, D. (Eds.). (2008). *Assessment of competencies in educational context*. Göttingen: Hogrefe.
- Józsa, K. (Ed.). (2006). *Az olvasási képesség fejlődése és fejlesztése* [The learning and teaching of reading skills]. Budapest: Dinasztia Tankönyvkiadó.
- Józsa, K., & Steklács, J. (2009). Az olvasástanítás kutatásának aktuális kérdései [Current issues in reading instruction research]. *Magyar Pedagógia*, 109(4), 365–397.
- Józsa, K., & Zentai, G. (2007). Óvodások kritériumorientált fejlesztése DIFER Programcsomaggal [Criterion-referenced assessment of pre-school children using the DIFER package]. In J. Nagy (Ed.), *Kompetencia alapú kritériumorientált pedagógia* [Competency based criterion-referenced education] (pp. 299–311). Szeged: Mozaik Kiadó.
- 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.
- 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: Association for Supervision and Curriculum Development.
- Molnár, I. (1993). Tévesztés és megakadás kötött szöveg felolvasásában [Errors and hesitation in oral reading]. In M. Gósy, & P. Siptár (Eds.), *Beszédkutatás* [Speech research] 1993 (pp. 175–184). Budapest: MTA Nyelvtudományi Intézet.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L., & Sainsbury, M. (2009). *PILRS 2011 Assessment Framework*. Boston College: TIMSS and PIRLS International Study Center.

- Nagy, J. (2002). *XXI. Század és nevelés* [The 21st century and education]. Budapest: Osiris Kiadó.
- Nagy, J. (2006a). Olvasástanítás: a megoldás stratégiai kérdései [Reading instruction: strategic issues]. In K. Józsa (Ed.), *Az olvasási képesség fejlődése és fejlesztése* [The learning and teaching of reading skills] (pp. 17–42). Budapest: Dinasztia Tankönyvkiadó.
- Nagy, J. (2006b). A szóolvasó készség fejlődésének kritériumorientált diagnosztikus feltérképezése [A criterion-referenced approach to the diagnostic assessment of word reading]. In K. Józsa (Ed.), *Az olvasási képesség fejlődése és fejlesztése* [The learning and teaching of reading skills] (pp. 91–106). Budapest: Dinasztia Tankönyvkiadó.
- Nagy, J. (2010). *Új pedagógiai kultúra* [New education culture]. Szeged: Mozaik Kiadó.
- Nagy, J., Józsa, K., Vidákovich, T., & Fazekasné Fenyvesi, M. (2004a). *DIFER Programcsomag: Diagnosztikus fejlődésvizsgáló és kritériumorientált fejlesztő rendszer 4–8 évesek számára* [The DIFER package: a system of criterion-oriented testing and diagnostic assessment for 4–8 year-olds]. Szeged: Mozaik Kiadó.
- Nagy, J., Józsa, K., Vidákovich, T., & Fazekasné Fenyvesi, M. (2004b). *Az elemi alapkészségek fejlődése 4–8 éves életkorban* [The development of foundational skills at age 4–8]. Szeged: Mozaik Kiadó.
- Nevada English Language Arts Standards (2007). *Integrating content and process*. Nevada Department of Education.
- 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: WestEd.
- O'Reilly, T., & Sheehan, K. M. (2009). *Cognitively based assessment of, for and as learning: A 21st century approach for assessing reading competency. Research memorandum*. Princeton: Educational Testing Service.
- OECD (2000). *Measuring student knowledge and skills. The PISA 2000 Assessment of reading, mathematical and scientific literacy. Education and Skills*. Paris: OECD.
- OECD (2004). *Problem solving for tomorrow's world. First measures of cross-curricular competencies from PISA 2003*. Paris: OECD.
- OECD (2006). *Assessing scientific, mathematical and reading literacy. A framework for PISA 2006*. Paris: OECD.
- OECD (2009). *PISA 2009 assessment framework. Key competencies in reading, mathematics and science*. Paris: OECD.
- Pléh, C. (1998). *A mondatmegértés a magyar nyelvben* [Sentence processing in Hungarian]. Budapest: Osiris Kiadó.
- PrePIRLS Information Sheet (2011). TIMSS and PIRLS International Study Center, Boston College. <http://pirls.bc.edu/pirls2011/downloads/prePIRLS.pdf>
- Pressley, M., El-Dinary, P. B., Gaskins, I., Shuder, T., Gergman, J., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *Elementary School Journal*, 92, 511–552.
- National Assessment Governing Board (2006). *Reading Framework for the 2007 National Assessment of Educational Progress*. Washington: U.S. Department of Education.

- National Assessment Governing Board (2008). *Reading Framework for the 2009 National Assessment of Educational Progress*. Washington: U.S. Department of Education.
- Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early childhood assessment*. Washington DC: The National Academies Press.
- Vári, P. (Ed.). (1999). *Monitor '97. A tanulók tudásának mérése* [Monitor '97. The assessment of student knowledge]. Budapest: Országos Közoktatási Intézet.
- Vári, P. (Ed.). (2003). *PISA-vizsgálat 2000* [PISA survey 2000]. Budapest: Műszaki Könyvkiadó.
- Office of Educational Accountability (2005). *Wisconsin Student Assessment Framework. Criterion-Referenced Test Framework. Assessment Framework for Reading*. Wisconsin Department of Public Instruction. http://www.dpi.wisconsin.gov/oea/pdf/read_framework.pdf
- Wren, S., & Watts, J. (2002). *The Abecedarian Reading Assessment*. <http://www.balance-dreading.com/assessment/abecedarian.pdf>