FRAMEWORK FOR DIAGNOSTIC ASSESSMENT OF SCIENCE

Edited by
Benő Csapó • Gábor Szabó

NEMZETI TANKÖNYVKIADÓ

Framework for Diagnostic Assessment of Science

FRAMEWORK FOR DIAGNOSTIC ASSESSMENT OF SCIENCE

Edited by

Benő Csapó
Institute of Education, University of Szeged

Gábor Szabó
Department of Optics and Quantum Electronics, University of Szeged

Nemzeti Tankönyvkiadó
Budapest

Developing Diagnostic Assessment
Projekt ID: TÁMOP 3.1.9-08/1-2009-0001

National Development Agency
www.ujszechenyiterv.gov.hu
06 40 638 638

HUNGARY'S RENEWAL

The project is supported by the European Union
and co-financed by the European Social Fund.

Authors:
Philip Adey, Magdolna Adorjánné Farkas, Mária B. Németh, Benő Csapó,
Csaba Csíkos, Erzsébet Korom, Mariann Makádi, Lászlóné Nagy, Katalin Radnóti,
Ibolya Revákné Markóczi, Gábor Szabó, Zoltán Tóth, Éva Wagner

The chapters were reviewed by
Katalin Papp and Péter Tasnádi

ISBN 978-963-19-7289-4

Nemzeti Tankönyvkiadó Zrt.
a Sanoma company

www.ntk.hu • Customer service: info@ntk.hu • Telephone: 06-80-200-788

Responsible for publication: János Tamás Kiss chief executive officer
Storing number: 42 685 • Technical director: Etelka Vasvári Babicsné
Responsible Editor: Györgyi Pukli-Prokob • Technical editor: Nándor Dobó
Size: 27,35 (A/5) sheets • First edition, 2012

If you cannot measure it, you cannot improve it.

Lord Kelvin

Contents

Introduction (*Benő Csapó* and *Gábor Szabó*) 9

1 *Philip Adey* and *Benő Csapó:* Developing and Assessing Scientific Reasoning 17

2 *Mária B. Németh* and *Erzsébet Korom:* Science Literacy and the Application of Scientific Knowledge 55

3 *Erzsébet Korom* and *Gábor Szabó:* Disciplines and the Curricula in Science Education and Assessment 89

4 *Erzsébet Korom, Mária B. Németh, Lászlóné Nagy* and *Benő Csapó:* Diagnostic Assessment Frameworks for Science: Theoretical Background and Practical Issues 147

5 *Erzsébet Korom, Lászlóné Nagy, Mária B. Németh, Katalin Radnóti, Mariann Makádi, Adorjánné Magdolna Farkas, Revákné Ibolya Markóczi, Zoltán Tóth, Csaba Csíkos* and *Éva Wagner:* Detailed Framework for Diagnostic Assessment of Science 175

About the Contributors 307

Introduction

The motto chosen for the volumes discussing the frameworks of diagnostic measurements is a quotation by *Lord Kelvin* "If you cannot measure it, you cannot improve it". The truth of this dictum can be illustrated by an example taken from another domain of life, the practice of medicine. If we were unable to measure body temperature, it would be impossible to ascertain the effects of medication intended to reduce fever. We could, of course, estimate body temperature without measurement by touching the forehead, for instance, but the accuracy of the estimate may be influenced by several subjective factors. The work of a doctor is simply unimaginable today without a range of measuring tools aiding the diagnosis and the choice of the right therapy. Teachers, in contrast, are still obliged to rely mostly on methods of subjective estimation in their education programs, having no access to tools of measurement of either their students' level of development or the effects of intervention efforts or day-to-day teacher activities. The dilemma of measurement in education is also aptly summarised in a quotation, one by *Albert Einstein* this time: "*Not* everything that is *measurable* is *important*, and *not* everything that is *important* is *measurable*". Echoing that statement, the problem we need to face with respect to the diagnostic assessment of knowledge of science can be characterised as follows: The most important elements of knowledge are not always those that most readily lend themselves to measurement. It is understandable that the earliest efforts to measure knowledge of science focused on areas that were the easiest to measure, namely students' ability to reproduce the subject matter that had been presented to them the way it had been presented. The assessment of students' comprehension of the subject matter and their ability to apply that knowledge to new contexts is a more complicated task. We must progress even further if we wish to assess whether science education can meet the objective of developing students' mental abilities and scientific thinking.

Over the decades around the turn of the Millennium, a growing emphasis has been placed throughout the world on research and development programmes the integrated results of which may lead to a substantial improvement in public education if transferred into practice. The programme providing the framework for the present volume occupies the

intersection of three major development trends. First, various international surveys have given a great impetus to the development of educational assessment and testing. Second, recent research results in educational sciences and psychology have led to increasingly refined understanding of the concept of knowledge, which allows more precise definitions of what should be measured at different stages of development. Third, the availability of info-communication technologies allows measurements to be performed in the way and with the frequency required by public education.

The key to progress in an education system is the availability of efficient feedback mechanisms at the various levels of that system. Such mechanisms can be created through measurements providing objective data on various aspects of performance at each level of the system. These measurements allow us to ascertain whether the education program is successful in meeting its targets, or whether a given intervention strategy has achieved the desired results. At present, feedback mechanisms operate on three main levels in public education. Feedback is provided by international surveys, which have become regular events during the past decade. Hungary has been included in the major science education surveys (PISA, TIMSS). The data allow the performance of the Hungarian education system to be evaluated in the context of other countries' results and the comparison can be used to draw conclusions with regard to ways of improving system-wide features. The results of the recurrent cycles of the surveys also provide feedback on the effects of any interventions. The international assessment programmes are planned and implemented with the contribution of the top research and development centres in the world. The various solutions of measurement methodology developed in these centres are made use of in the preparation of national assessment systems.

Several countries, including Hungary, have introduced a system of annual assessment covering all students in selected grades of schooling. These surveys provide detailed feedback to individual schools on the performance of their own students. Based on an analysis of the results, schools may improve their internal processes and the efficiency of their activities. The results are also made public, which may act as an incentive to seek ways of improvement and development. The experiences of countries where a system of this sort has been in place for a relatively

long time show, however, that placing pressure on schools has the effect of improved efficiency only within certain limits. If the stakes associated with the evaluation are too high for either the teachers or the schools, various distortions may result. Further improvement in efficiency can only be achieved by devising methods and tools directly assisting the work of teachers. These include measurement tools that enable teachers to obtain a precise assessment of students' level of development in areas of key importance with respect to their further progress.

Traditional paper-and-pencil tests were, however, very costly and labour-intensive and were therefore unsuitable for performing sufficiently frequent assessments. The second important recent development is thus the explosive advancement of information and communication technologies, which offer novel solutions in every area of life, including educational measurement. Thanks to these technologies, tasks that used to be beyond solution have now become simple to implement in education also. One of these is educational assessment providing frequent diagnostic feedback. Computers were put in the service of education effectively as soon as the first large electronic computers appeared; educational computer software has been around for decades. The use of information technology in education was, however, often motivated by the technology itself, i.e., the reasoning was that now that these tools were available, it made sense to use them in education. Online diagnostic assessment approaches the question from the opposite direction: an appropriate technology is sought as a solution to the problem of implementing a task of key significance in education. From this perspective, info-communication technology is a tool that has no substitute in expanding the range of possibilities for educational assessment.

The third development, one which is closest to the concerns of this volume, is the cognitive revolution in psychology, which affected several areas towards the end of the last century and gave a new impetus to research efforts in connection with school learning and teaching. It has led to the emergence of new and more differentiated conceptions of knowledge allowing a more precise definition of the goals of public education and the development of scientifically established standards. This process has also opened the way to a more detailed characterisation of student development processes. The psychological approach penetrated early science education relatively soon. Piaget's classic works on cogni-

tive development used simple experiments of science to study child cognition, and later research on conceptual development and misconceptions also used cognitive processes related to science phenomena as their primary domain of inquiry.

Once the special significance of early childhood had been recognised, the focus of attention shifted to the first few years of schooling, especially to the encouragement of language development and reasoning skills. Several studies have provided evidence that the acquisition of basic skills is crucial for in-depth understanding of the subject matter taught at school, which is in turn essential for students to be able to apply their knowledge to new contexts rather than just reproduce exactly what they have been taught. If the required foundations are not constructed, serious difficulties will arise at later stages of learning: failures suffered during the first years of schooling will delimit students' attitudes towards education for the rest of their lives. The development of concepts related to science begins even before the start of formal education and the first years of school play a decisive role in steering conceptual development in the right direction. Early science education shapes children's thinking, their approach to the world and their attitudes towards empirical discovery.

The developmental processes discussed above have provided the basis of a project entitled "Developing Diagnostic Assessments" launched by the Centre for Research on Learning and Instruction at the University of Szeged. The project focuses on the development of detailed frameworks for diagnostic assessments in three major domains – reading, mathematics and science – in the first six grades of school. This involves the development of question banks containing several thousand questions and exercises, which will be accessible to students on the Internet through an online computer system. The system – the implementation of which is a lengthy process involving several hierarchically organised steps – will fulfil the function of providing frequent individual student-level feedback.

The diagnostic tests are designed to assess individual students' progress relative to various reference points. Similarly to system-wide surveys, the programme allows the population average to act as a standard of comparison: being able to compare an individual's performance to the performance of their peers can provide important information. In addition to this, certain developmental benchmarks and external reference

points should also be defined. The diagnostic tests should, however, go even beyond that: they should follow students' progress over time, i.e., compare performance at a given point in time with the results of previous measurements.

Diagnostic assessment can only be an efficient tool in student education if the measurement methods are based on scientifically based frameworks. Issues such as the target areas or dimensions of progress assessment, the desired direction of development, what constitutes progress in the various areas, and what constitutes advancement to the next step of development can only be decided on the basis of research evidence. Both the aim of diagnostic value and the fact that the focus is on early childhood call for a detailed specification of test contents, a well-rounded, scientifically based theoretical framework and the incorporation of considerations of developmental psychology, knowledge application standards and the discipline-specific characteristics of science education.

Frameworks define the object of measurement. Their development has been one of the most important tasks of the project. The results are presented in three uniformly structured volumes. The current volume discusses the frameworks of diagnostic assessment for science, and the two companion volumes summarise the conclusions for reading and mathematics. The development work for the three domains proceeded in parallel and the same broad theoretical framework and conceptual system were used for the development of the detailed contents of assessment for each of these domains. The three volumes therefore share not only their structure but also parts of their introduction and of one of the internal chapters. In accordance with international practice, the term *science* is used throughout the project as a general term referring to the domain of assessment.

The work presented in this volume draws on the experiences of several decades' research on educational assessment at the University of Szeged and on the achievements of the Research Group on the Development of Competencies, Hungarian Academy of Sciences with special reference (a) to the results of studies related to the structure and organisation of knowledge, educational evaluation, measurement theory, conceptual development, the development of reasoning skills, problem-solving and the assessment of school readiness; and (b) to the technologies developed for test item writing and test development. Our present work

on developing assessment frameworks has benefited a great deal from the results of several specific projects, including the Hungarian Educational Longitudinal Program.

The development of the frameworks of diagnostic assessments is, however, a complex task reaching beyond the experiences mentioned above. In order to achieve our goals, extensive international collaboration was required. Our work has therefore been carried out in cooperation with a large science community including experts in Hungary and abroad. The opening chapter of each volume has been prepared with the contribution of a leading researcher in the relevant field; thus our work rests upon the scientific foundations most widely valued in the international community. The details of the frameworks have been developed with the contribution of teachers and other professionals with practical experience in test construction.

The system of diagnostic assessments is based on a three-dimensional approach to knowledge, in line with the traditions characterising the entire history of organised education. The wish to educate the intellect, to cultivate thinking and general cognitive abilities has been around as long as organised education has. Modern public education also sets several goals applying to the students themselves as individuals. In order to achieve these goals, we must first of all be guided by evidence provided by the fields of inquiry concerned with the human being and the developing child, i.e., the results of studies in developmental psychology and the psychology of learning. In the context of sciences, the focus of this dimension is the development of scientific thinking.

Another area of educational goals is related to the usability of school knowledge. The dictum "*Non scholae sed vitae discimus.*" is perhaps more topical today than ever before, since our modern social environment is changing far too rapidly for public education to be able to keep pace with it. Past research has revealed that knowledge transfer is not an automatic process; students cannot automatically apply their knowledge to new contexts. For this reason, the assessment of applicable knowledge appears as an independent dimension in diagnostic assessments. This task requires a different approach to testing: we must define what is expected of students that will enable them to apply their knowledge in different school contexts and in contexts outside of the school. The third dimension concerns the selection of content knowledge accumulated by

science that public education should transmit. Not only because the above goals cannot be achieved without content knowledge but also because it is an important goal of its own right that students should become familiar with the knowledge generated by science and organised according to the internal values of science.

The above goals have been competing with each other over the past few decades with one or another coming into fashion at different times. For the purposes of the present project we assume that education integrates the three main goals in fulfilling its function but diagnostic assessments must differentiate between them. Diagnostic assessments must be able to show if there is insufficient progress in one or another of these dimensions.

The first three chapters of this volume discuss the theoretical background and research evidence pertinent to each of these three dimensions. In Chapter 1, Philip Adey and Benő Csapó discuss the role of science education in the development of thinking and the assessment goals related to this area. In Chapter 2, Mária B. Németh and Erzsébet Korom give an overview of theoretical issues related to scientific literacy and the application of scientific knowledge. Chapter 3 by Erzsébet Korom and Gábor Szabó summarises the content knowledge offered by science to the early stages of public education, especially for the purposes of the development of scientific thinking. Each chapter provides an extensive review of the literature and the included detailed bibliographies can assist future research. In Chapter 4, Erzsébet Korom, Mária B. Németh, Lászlóné Nagy and Benő Csapó discuss theoretical issues related to the development assessment frameworks, and outline a practical solution providing the foundations for diagnostic assessment programmes.

The second part of the volume contains the detailed frameworks for diagnostic assessment. The purpose of this section is to provide a basis for the development of measurement tools and test questions. Our diagnostic assessment program treats the first six grades of school as a continuous educational process. The results of assessment are therefore interpreted relative to scales spanning all six grades; students are placed along these scales according to their current level of development. The content specifications of assessment questions could also essentially form a single continuous unit. However, in an effort to allow greater transparency and to follow the traditions of educational standards, this process

has been divided into three stages, each of which covers approximately two years. For the three dimensions, therefore, a total of nine content blocks are described.

In their present state, the frameworks detailed in this volume should be seen as the first step in a long-term development process. They specify what is reasonable to measure and what the major dimensions of assessment are, given the present state of our knowledge. As the domains covered develop at a very rapid rate, however, the latest findings of science should be incorporated from time to time. The content specifications can be constantly updated on the basis of our experiences of item bank development and an analysis of the data provided by the diagnostic program in the future. Our theoretical models can also be revised through an evaluation of the test questions and an analysis of relationships emerging from the data. In a few years' time we will be in a position to look at the relationship between the various areas of early development and later performance allowing us to establish the predictive and diagnostic validity of test questions, which can be a further important source of information for the revision of theoretical frameworks.

Erzsébet Korom played a prominent role in the preparation of this volume. In addition to co-authoring four of the chapters, she also led the research team developing the detailed description of the contents of the assessment. Besides the authors of the chapters, several colleagues have contributed to the preparation of this volume, for which we are very grateful. Special thanks are also due to the team responsible for the organisation and management of the project: *Katalin Molnár*, *Judit Kléner* and *Diána Túri*. The development and final presentation of the content of the volume have benefited greatly from the comments of the reviewers of earlier versions. We would like to take this opportunity to thank *Katalin Papp* and *Péter Tasnádi* for their valuable criticism and suggestions.

Benő Csapó and *Gábor Szabó*