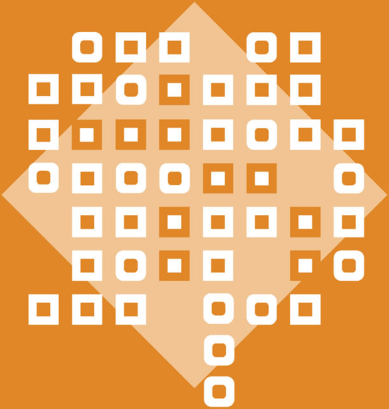
Sascha Bernholt, Knut Neumann,
Peter Nentwig (Eds.)

Making it tangible

Learning outcomes in science education



WAXMANN

Making it tangible

Sascha Bernholt, Knut Neumann,
Peter Nentwig (Eds.)

Making it tangible

Learning outcomes in science education

Waxmann 2012
Münster / New York / München / Berlin

Bibliographic information published by die Deutsche Nationalbibliothek
Die Deutsche Nationalbibliothek lists this publication in the
Deutsche Nationalbibliografie; detailed bibliographic data
are available in the internet at http://dnb.d-nb.de.

ISBN 978-3-8309-2644-3

www.waxmann.com
info@waxmann.com

Cover Design: Verena Hane, Kiel
Print: Hubert & Co., Göttingen
Printed on age-resistant paper, acid-free as per ISO 9706

FSC
www.fsc.org
MIX
Papier aus verantwortungsvollen Quellen
FSC® C016439

Printed in Germany

Preface

The Leibniz Institute for Science and Mathematics Education (IPN) at the University of Kiel in Germany and the University of York Science Education Group (UYSEG) in the UK are major centres of research and development in science education. Both have many years' experience of research projects on different aspects of science education practice and policy, and of the development and evaluation of innovative approaches to science teaching and learning. In recent years, one focus of the work at both centres has been the drive to improve scientific literacy at all levels of school education, in particular through the development of context-based and context-led science courses.

As a means of exploring specific issues in some depth, the two Centres inaugurated a series of small conferences on selected themes. The idea was that each would involve a maximum of 30 participants, selected staff drawn from the two centres and invited experts on the chosen topic from a range of countries. In contrast to international conferences with hundreds of participants and dozens of parallel sessions, this setting allowed extensive and in-depth discussions and the opportunity to contrast in detail the different approaches and experiences contributed by leading international experts. This book presents some of the outcomes of the fifth of these conferences. The topic was learning outcomes in science education, how they might be defined and assessed – and how this might stimulate improved outcomes. This is a topic which is widely discussed and debated, as a common feature of today's educational reforms is the increasing focus on educational outcomes. However, despite this international trend, considerable differences can be found in the details.

The chapters of this book are based on the papers presented at the conference – revised in the light of the discussions. The expertise of the participants, their experience and the exchange of ideas and controversies during the symposiums are reflected in the quality of each individual contribution. The composition of the papers in this book therefore provides on the one hand an overview of different national approaches and developments concerning the topic of learning outcomes in science education; on the other hand, it provides a well-grounded compendium of challenges and open questions that need to be addressed by the international educational research community in the future, which makes this book of particular interest to the field.

December 2011

Olaf Köller (IPN, Kiel)
Robin Millar (UYSEG, York)

Acknowledgements

We thank Professor Dr. Ilka Parchmann for contributing inspiring ideas and for her support in the process of planning and preparing this, the fifth IPN-UYSEG Conference in Science Education.

We are grateful to the staff of the Strandhotel Strande near Kiel who made our stay most effective as their aid allowed us to fully focus on the topic of the meeting. We moreover thank the student assistants who helped in organizing the symposium, planning arrivals and departures, organizing taxis or who just stepped into the breach whenever their help was needed.

We are much obliged to our patient and helpful authors. They wrote their papers prior to the symposium and read all contributions as preparation for the meeting. Subsequent to the symposium, we asked them for revisions in order to gain coherence and to reflect some of the major issues that emerged from the discussions. This was an iterative process and the authors were unfailingly cooperative.

We thank the German Research Foundation (DFG) for substantial funding without which it would not have been possible for the meeting to take place. In equal measure, we thank the Leibniz Institute for Science and Mathematics Education (IPN) for the financial and organisational support of the meeting.

Finally, we wish our readers to note that any mistakes that they find are our responsibility.

December 2011

Sascha Bernholt
Knut Neumann
Peter Nentwig

Contents

Part A Setting the Scene

Part B The Big Picture

Part C The Devil in the Details

Part D Perspectives from Different Countries

Part E A Coda

Part A

Setting the Scene

Chapter 1

Making it Tangible –
Specifying Learning Outcomes in Science Education

***Sascha Bernholt*[1]*, Knut Neumann*[2] *& Peter Nentwig*[1]**
IPN · Leibniz Institute for Science and Mathematics Education, Kiel, Germany
[1]Department of Chemistry Education, [2]Department of Physics Education

Abstract

Today, many countries have established educational standards. In addition, most of the recent educational reforms across the world are shifting away from traditional content-driven curricula. Instead, these reforms aim to provide clear, specific descriptions of the skills and knowledge that teachers are supposed to teach and students are expected to learn. The specification, however, what exactly students are expected to know and be able to do, varies widely between countries. In Germany for example, so called *competences* that students are expected to have acquired at a given stage in their school career (end of primary and end of lower secondary school, respectively) are specified. In other countries, learning outcomes are defined sometimes with a different notion, sometimes with just different terms. In March 2011, an international symposium brought renowned researchers together from countries with different notions of the nature and quality of learning outcomes for a conceptual clarification and better understanding among the international science education community.

1 Introduction

Triggered by international large-scale assessments, such as TIMSS (Beaton et al., 1996), PISA (OECD, 2001, 2004, 2007), or PIRLS (Mullis, Martin, Kennedy, & Foy, 2007), discussions about the achievement and effectiveness of the national school system have emerged in many countries. Several countries found themselves far behind their expectations with regard to the results of these international comparative studies. In response to these often disappointing findings, a debate about the aims of science education arose in many countries whether the current curricula are sufficiently prospective to provide a fruitful foundation for a successful and future-oriented science education in schools.

For a long time “the content of the science curriculum has largely been framed by scientists who see school science as a preparation for entry into university rather than as an education for everyone. No other curriculum subject serves such a strong dual mandate” (Osborne & Dillon, 2008, p. 21). However, with regard to the future needs of the students, the question pertained to what young adults need to learn in formal education in order to play a constructive role as a citizen in society (Tenorth, 2006). In the case of science education, the question arose regarding what is important to know, to value, and to be able to do in situations involving science and technology in a modern society (cf. deBoer, 2000).

In recent years and especially induced by the results of PISA, the idea of scientific literacy, albeit in different notions, received wide acceptance as the overall aim of science education (DeBoer, 2000; Gräber & Bolte, 1997; Gräber, Nentwig, Koballa, & Evans, 2002; Roberts, 2007). Although there is no general consensus about the exact meaning of or what constitutes scientific literacy, it is considered a functional educational concept which can provide both a basis for lifelong learning and a preparation for life in a modern society (e.g., Bybee, 1997; Millar & Osborne, 1998; Roberts, 2007).

Following the adoption of scientific literacy as the overarching aim of science education, a growing number of countries introduced science education standards in order to ensure that students in fact obtain scientific literacy (cf. Nentwig & Schanze, 2007). These standards intend to provide precise descriptions of the skills and knowledge that teachers are supposed to teach and students are expected to learn in order to become scientifically literate. Along with the introduction of standards assessment systems were developed to benchmark students’ achievements in their struggle to obtain scientific literacy (Ravitch, 1995; Resnick & Resnick, 1983). The combination of clearly formulated learning outcomes and the assessment of students’ achievement is a concept of educational governance and aims to enable a data-driven steering mechanism for the educational system towards a higher quality of teaching (Altrichter, Brüsemeister, & Wissinger, 2007; Amos, 2010).

In summary, the common feature of today’s educational reforms is the increasing focus on educational outcomes. While some countries (e.g., the United States) traditionally maintained science education standards, others (e.g., Germany or France) relied on a different educational tradition. Due to students’ mediocre science performance in large-scale assessments, attention shifts towards educational outcomes. The number of countries building on science education standards is growing. Countries that built on science education standards before initiated revisions of standards and assessment systems (e.g., Australia) (cf. Waddington, Nentwig, & Schanze, 2007). However, despite scientific literacy being the overarching aim of science education and the movement towards science education standards, the learning outcomes defined in these standards differ considerably amongst countries.

2 International Perspectives on Conceptualizing Educational Standards

2.1 Standards in Science Education

The country with the longest tradition in the use of science education standards probably is the United States with a history of standards that can be traced back to the 19th century – including a history of different conceptions, paradigms, debates, and controversies. In recent history the debate about standards in the United States was stirred up by the report “A Nation at Risk” (National Commission on Excellence in Education, 1983). This report evoked fundamental reforms of the American educational system. Since then, standards, standard-based curricula and subsequent assessments were considered a central shaping agent for what can be expected from students (Resnick, 1985). Standards for science education were published in the early 1990s by the American Association for the Advancement of Science [AAAS] (1993) and the National Research Council [NRC] (1996). Several other English-speaking countries developed science education standards around the same time, among them Canada (Council of Ministers of Education of Canada [CMEC], 1997) and the UK (Department for Education and Science [DES], 1989). About a decade later, Australia also followed the movement towards educational standards (Hafner, 2007).

European countries traditionally have not maintained educational standards (cf. Waddington, Nentwig, & Schanze, 2007). In response to the results of international large scale assessments, however, several countries introduced educational standards, among them Austria (Weiglhofer, 2007), Germany (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005a, 2005b, 2005c), Luxembourg (Ministère de l’Éducation nationale et de la Formation professionnelle, 2009), and Switzerland (Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008).

Some Asian countries such as Japan and Korea maintain national science curricula (cf. Han, 1995; Ogawa, 1998), others such as China or Singapore build on something more closely related to what can be considered national science content standards (Liu & Ruiz, 2008). Interestingly, whereas countries/regions such as Korea or Taiwan have maintained cycles of curriculum revision ever since the introduction of national curricula (Chiu, 2007; Han, 1995), China and Singapore have only recently initiated major revisions of these standards (Curriculum Planning & Development Division [CPDD], 2000; Wei & Thomas, 2005; cf. Liu & Ruiz, 2008).

2.2 Different Concepts, Different Standards

Although science education standards in most countries are based on scientific literacy as the overarching aim of science education, in detail the standards differ considerably. In the **United States**, the Educate America Act (1994) dictates national and state standards on content, student performance, and opportunities to learn. Content standards refer to what is to be taught and learned, performance standards define degrees of mastery or levels of attainment, and opportunities-to-learn standards specify the availability of resources provided for learning (Ravitch, 1995). However, the national science education standards only contain content standards, teaching standards, and assessment standards (NRC, 2006). On state level, although all states have implemented content standards, many have not yet developed performance or opportunity-to-learn standards (Liu & Ruiz, 2008).

Currently, the Common Core State Standards Initiative is a state-led effort of 48 states, two territories and the District of Columbia to provide a clear and consistent framework of what students are expected to learn in English Language Arts and Mathematics. The standards are to provide appropriate benchmarks for all students, regardless of where they live. The draft standards were opened for public comment and are presently waiting to be adopted and implemented by the states.

Wilson (2006) lists several features of high-quality content standards including proficiency levels and performance expectations. Despite his demand for content standards to include proficiency levels, only few current content standards explicitly do so (cf. NRC, 2006). Instead, most content standards lack coherence and focus (Valverde & Schmidt, 2000). In a recent report, Duschl, Schweingruber, and Shouse (2007) criticize that in the current U.S. science content standards "little attention is given to how students' understanding of a topic can be supported and enhanced from grade to grade. As a result topics receive repeated shallow coverage with little consistency, which provides a fragile foundation for further knowledge growth" (p. 217). The report concludes that science standards need to be reorganized around core ideas of the respective domain and that learning progressions need to be developed which describe how students progress in developing a deeper understanding of the core ideas when proceeding from grade to grade (Duschl et al., 2007).

The idea of learning progressions became increasingly popular in the science subjects (Duncan & Hmelo-Silver, 2009). Learning progressions describe students' progression in understanding scientific phenomena (Steedle & Shavelson, 2009), i.e. "successively more sophisticated ways of reasoning within a content domain that follow one another as students learn" (Smith et al., 2006, p. 1). Learning progressions are not only about knowledge, but also about abilities and skills which allow successful solving of real-life problems (Schwarz et al., 2009; Smith, Wiser, Anderson, & Krajcik, 2006; Songer, Kelcey, & Gotwals, 2009).

As in the US in the early 1990s, a similar movement towards standards also took place around the same time in other countries such as **Canada** (CMEC, 1997) and the **UK** (DES, 1989). Based on a common framework of science learning outcomes (CMEC, 1997), the Canadian provinces have developed individual science education standards or curricula respectively. For example, the Ontario Science Curriculum (Ministry of Education, 2007, 2008a, 2008b) describes the knowledge and skills students are expected to acquire. Students' knowledge is organized around fundamental concepts, so-called big ideas. Students are expected to develop an increasing understanding of these concepts whilst proceeding from grade 1 to grade 12. Regarding scientific inquiry and experimentation skills, the Ontario Science Curriculum describes four categories of skills such as "Initiating" and "Planning" to operationalize the continuum of scientific inquiry and experimentation skills. The standards also define four performance levels which describe different levels of proficiency to be used for assessment.

Australia released national science education standards which were based on a first formulation of national goals for science education in 1989 and were revised in 1999 (cf. Hafner, 2007). Adapted to these goals, a measurement framework was developed (Ministerial Council on Education, Employment, Training and Youth Affairs [MCEETYA], 2003) that was heavily influenced by the framework used in PISA (e.g., OECD, 2001). Within this framework, a scientific literacy progress-map describes six proficiency levels with respect to three domains of scientific literacy (Hafner, 2007). On the level of the individual Australian states, science education standards take a similar approach. The New South Wales science education standards define content and outcomes in syllabi for K-6, 7-10, as well as for post-secondary education. The content is described in terms of what students learn about and what students learn to do while the outcomes describe the knowledge, the skills, and the understanding that are expected to be obtained by students (Hafner, 2007).

Asian countries like **China** and **Singapore** have traditionally maintained national science content standards (Liu & Ruiz, 2008). However, both countries have recently initiated major revisions of their national standards. Within the scope of these revisions, the idea of scientific literacy was integrated as the central purpose of science education (Curriculum Planning & Development Division [CPDD], 2000; Wei & Thomas, 2006). In most Asian countries/regions such as **Japan**, **Korea**, and **Taiwan**, the idea of scientific literacy worked its way into the formerly very content-oriented documents during the regular process of science curriculum revision (cf. Chiu, 2007; Han, 1995; Ogawa, 2001). In Taiwan for example, the recently revised science education curriculum includes so called guidelines of science education. Based on general goals for science education, these guidelines embrace eight core components of scientific literacy, so-called core competences, and five science content areas in which students are expected to develop the core competences (cf. Chiu, 2007). However, although the concept of competence obtains a central role in the PISA framework and made its way into revised versions of science education standards and curricula in Asian countries/regions, the concept of competence did not become as central as it did in Germany.

In the German-speaking countries, the concept of competence became widely used to specify the educational goals, to guide the designing of learning environments, and to direct the development of assessments (Ditton, 2002; Slavin, 2002). Turning away from previously used terms like knowledge, skills, ability, or qualification, the term competence is used to acknowledge the changed requirements of everyday- and working-life (Klieme & Leutner, 2006). This change was intended to be reflected in the formulation of the educational standards the development of which was highly influenced by the report of Klieme et al. (2003). With regard to educational settings, competences are mainly interpreted (at least in the German speaking countries) as clusters of cognitive prerequisites that must be available for an individual to perform well in a particular content area (Weinert, 1999). In summary, competences require long-term, cumulative learning opportunities, a broad experience, and a deep understanding of the topic (Weinert, 2001).

The concept of ‘competence’ is not only used in scope of research projects but was also chosen as the constitutive concept for educational standards in **Germany** (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005a, 2005b, 2005c). The standards specify competences students are expected to have achieved by the end of secondary education. Because competences are considered domain-specific, standards were defined for each of the three science subjects (chemistry, physics, and biology). In an effort to show that science goes beyond factual knowledge, the standards for these subjects distinguish between four areas of competence (content knowledge, epistemological and methodological knowledge, communication, and socio-scientific reasoning), each of which embraces a list of specified competences. In order to grasp different degrees of achievement, three levels of cognitive complexity are proposed for all four areas as a rather rough measure: reproduction, application, and transfer. This conceptualization is intended to facilitate the measurement of learning outcomes as well as of students’ progress across time of schooling.

To describe the structure, graduation, and development of competences, the formation and empirical examination of competence models have become a major research area in Germany (cf. Klieme, Leutner, & Kenk, 2010). Different domain-specific models have been published with the aim to describe competence in science subjects (e.g., Labudde & Adamina, 2008; Neumann, Kauertz, Lau, Notarp, & Fischer, 2007; Schecker & Parchmann, 2006). Additionally, national and international large-scale studies, e.g., PISA, use competence models as assessment frameworks (e.g., OECD, 2006).

Austria and **Switzerland** basically followed the German approach. Their standards describe competences students are expected to have obtained at certain points of time during their school career. In accord with the development in Germany, the same definition of competence is used and competence models are developed the structures of which even bear resemblance to the models discussed in the context of the German education system (cf. Labudde, Metzger, & Gut, 2009; Weiglhofer, 2007).

In the **UK**, the National Science Curriculum was first introduced in 1989 and has been revised several times since (Department for Education, 1995; Department for Education and Employment, 1999; DES, 1989). Science education and the science education curriculum in the UK are divided into several key stages. The original curriculum provided extensive lists of science content students should know, understand, or be able to do (DES, 1989). Each statement was assigned to one of ten levels of proficiency which were intended to span schooling grades K-12. In the recent revision, stages one to three embrace less extensive lists of what is to be taught (instead of to be learnt) and a series of eight performance levels which more generally describe what is expected from students on that particular level. As the National Science Curriculum considers the listed contents as to be taught, external examinations and school inspections are an integral part of the education system, to ensure that the achieved curriculum actually corresponds to the intended curriculum (cf. Millar, 2007). These external examinations fall back on the detailed definitions of performance levels that are provided in the curriculum material.

With regard to other European countries, the intended learning outcomes in science still demonstrates a wide variety. The **Danish** science curriculum in the Folkeskole (primary and lower secondary level school) for example describes core knowledge and skills for students to learn as well as guidelines for teachers to develop their teaching so students can reach these goals. Beyond formative assessment, there is no particular culture of standard-based assessments (cf. Dolin, 2007). **Finland** has a long tradition of decentralized assessments based on content-oriented curricula (as used to be the case in Germany). This led to disparities in how the goals specified in the curricula were assessed and consequentially to discrepancies in performance records between schools. Now, educational standards describe what students should know and be able to do at certain stages of schooling (cf. Lavonen, 2007).

3 Common Trends, Shared Issues

So far it seems that a consensus exists about the overarching aim of science education. Scientific literacy has developed into a common base of many national science education standards. As a consequence, by all standards students are expected to become scientifically literate in terms of being able to master everyday situations related to scientific phenomena. However, the exact understanding of the components of scientific literacy differs from country to country. In some countries, the components of scientific literacy are considered to be knowledge and skills; in other countries they are seen as competences (which embrace combinations of knowledge and skills); and then in some countries scientific literacy is understood to correspond to an understanding of scientific concepts and the ability to apply scientific inquiry skills.

Accordingly, there is no consensus with regard to clear, specific descriptions of the skills and knowledge that students are expected to learn. Nevertheless, most educational reforms currently tend to focus on educational outcomes as well as the development (or alignment) of assessment systems. In correspondence with the different national interpretations of scientific literacy, the focus of these assessments differs between countries. In some cases, however, the different perspectives and emphases are not as diverse as they seem at first sight. Albeit using a different notion, the recent development in the US is very similar to the movement towards models of competence in Germany. For instance, the idea of learning progressions in the US and Wilsons' (2006) conception of high-quality content standards including proficiency levels and performance expectations are similar to the German idea of competence development (cf. Schecker & Parchmann, 2006). The levels of understanding in a learning progression may be considered as levels of competency (Reiser, Krajcik, Moje, & Marx, 2003). However, only few researchers have established a connection (e.g., Liu, 2009) as the term 'competence' in the US is strongly limited to the vocational field (Melton, 1994). Similarly, the Australian national educational standards very much correspond to the German educational standards which were also strongly influenced by the PISA framework (Neumann, Kauertz, & Fischer, 2010). Similar to competence levels in Germany, the proficiency levels in the Australian national science education standards make frequent use of the idea of complexity (MCEETYA, 2002).

However, the assessment of learning prerequisites and outcomes is theoretically and methodically challenging. The measurement of learning outcomes must exhibit the differentiated internal structure of the knowledge base, i.e. the components as well as the level of proficiency. Furthermore, measurement should reveal changes and transformations in the process of learning and development (Klieme & Leutner, 2006). For this purpose, three major tasks can be identified: (1) the development of cognitive models of learning that can serve as the theoretical basis for assessment, (2) research on new measurement models and their applicability, (3) research on assessment designs. “Much hard work remains to focus psychometric model building on the critical features of models of cognition and learning and on observations that reveal meaningful cognitive processes in a particular domain” (Pellegrino, Chudowsky, & Glaser, 2001, p. 6). Without doubt, considerable expertise exists in the area of assessing students’ knowledge or theorizing about educational goals; but the question of how to measure students’ mastering of standards still remains largely unanswered (Fensham, 2009). Therefore, teachers in schools and science education reformers need additional detailed information about what students individually have already mastered and what their next level of development could be (Bernholt & Parchmann, 2011; Neumann, Fischer, & Kauertz, 2010). The use of theoretical constructs for the characterization of students’ learning outcomes intends to align curricula, instruction, and assessment. The ambition to align different facets of the school system is a central feature of current developments in most educational reforms throughout the world.

With regard to the parallels and differences between the developments in many countries across the world, a discussion about the different national notions of what constitutes scientific literacy and how this superordinate goal can be specified and linked to student performance seems necessary. The comparison of different concepts of students’ learning outcomes might also reveal misunderstandings due to different national traditions as well as to difficulties in the exact definition of concepts like ‘knowledge’, ‘skills’, and ‘understanding’ – especially with regard to translation issues. Additionally, a clarification of the central national conceptions could provide a constructive basis for further developments and scientific exchange.

4 The Symposium

Motivated by the controversy concerning the definition of learning outcomes, an international symposium in Kiel, Germany, brought renowned researchers together from countries/regions with different notions of the nature and quality of learning outcomes in March 2011. This is the fifth in a series of international symposia organized by the Leibniz Institute for Science and Mathematics Education at the University of Kiel (IPN) and the Science Education Group at the University of York (UYSEG) (cf. Bennett et al., 2005; Gräber, Nentwig, Koballa, & Evans, 2002; Nentwig & Waddington, 2005; Waddington, Nentwig, & Schanze, 2007). Twenty-seven science educators, psychologists, and educationalists were invited from 13 countries/regions. The participants produced papers about the situation in their countries/regions or about specific research programmes. These papers were made available well in advance of the meeting to all participants and served as basis for the detailed discussions at the symposium, which were held alternately in small groups and in the plenum. Some of the participants had been invited to the roles of chairs and discussants for particular sub-sections of the discussion. Their engagement and creativity in structuring their themes led to a variety of activities and proved to be crucial for the success of the event. After the meeting, the contributors were asked to revise their papers in the light of the discussions, and these revised versions became the core of this book.

The general structure of the book attempts to take advantage of the diversity and individuality of the 19 articles that were submitted by grouping them in three sections. Each of the individual contributions to this first section provides a broad overview about different approaches, challenges, and pitfalls on the road to the clarification of meaningful and fruitful learning outcomes. Additionally, the role and impact of different systemic factors and stakeholders are highlighted. This parts emphasizes how complex the endeavor of defining, assessing, and promoting of learning outcomes in science education is. The second set of papers provides deep insights into different, although comparable approaches which aim to frame, to assess, and to promote learning and learning outcomes in science education. Smaller projects are presented as well as broad, coordinated national programs. In addition, general and specific reproaches are included to reveal gaps and blind spots within current projects. The third set of papers reflects this ambition of striving for individual solutions in several countries. From different national perspectives, these papers outline the individual historical development, the deficits and problems that led to the current reforms, and finally what these reforms look like. Despite common trends, these national reports picture a diversity of school systems and educational traditions. These papers outline the problems and challenges in defining learning outcomes in science education and indicate that there will be probably no single solution that fits the needs of every country.

In summary, this book intends to provide an overview about different conceptions and different notions of the nature and quality of learning outcomes, about difficulties and challenges concerning research, school practice, and teacher education, as well as about different national perspectives and experiences. It intends to make tangible what in the literature sometimes remains blurred.

References

Altrichter, H., Brüsemeister, T., & Wissinger, J. (2007). *Educational Governance*. Wiesbaden: VS Verlag.

American Association for the Advancement of Science (AAAS) (1993). *Benchmarks for Scientific Literacy*. New York: Oxford University Press.

Amos, K. (2010). *International Educational Governance*. Bingeley, UK: Emerald Group Publishing.

Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1996). *Science Achievement in the Middle School Years: IEA's third International Mathematics and Science Study* (TIMSS). Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Bennett, J., Holman, J., Millar, R., & Waddington, D. (2005). *Making a Difference. Evaluation as a Tool for Improving Science Education*. Münster: Waxmann.

Bernholt, S. & Parchmann, I. (2011). Assessing the complexity of students' knowledge in chemistry. *Chemistry Education Research and Practice, 12*, 167-173.

Bybee, R. W. (1997). Toward an understanding of scientific literacy. In W. Gräber & C. Bolte (Eds.), *Scientific Literacy* (pp. 37-68). Kiel: IPN.

Chiu, M.-H. (2007). Standards for science education in Taiwan. In D. Waddington, P. Nentwig, & S. Schanze (Eds.), *Making it Comparable – Standards in Science Education* (pp. 303-347). Münster: Waxmann.

Council of Ministers of Education of Canada. (1997). *Common Framework of Science Learning Outcomes*. Toronto: Council of Ministers of Education.

Curriculum Planning & Development Division [CPDD] (2000). *Science Syllabuses: GCE "O" & "N" level*. Singapore: CPDD.

DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching*, *37*(6), 582-601.

Department for Education. (1995). *Science in the National Curriculum*. London: Her Majesty's Stationary Office (HMSO).

Department for Education and Employment. (1999). *The national curriculum for England*. London: Her Majesty's Stationary Office (HMSO).

Department for Education and Science. (1989). *Science in the National Curriculum*. London: Her Majesty's Stationary Office (HMSO).

Ditton, H. (2002). Evaluation und Qualitätssicherung [Evaluation and Quality Control]. In R. Tippelt (Ed.), *Handbuch Bildungsforschung* (pp. 775-790). Opladen: Leske + Budrich.

Dolin, J. (2007). Science education standards and science assessment in Denmark. In D. Waddington, P. Nentwig, & S. Schanze (Eds.), *Making it Comparable – Standards in Science Education* (pp. 71-82). Münster: Waxmann.

Duncan, R. G. & Hmelo-Silver, C. E. (2009). Learning Progressions: Aligning Curriculum, Instruction, and Assessment. *Journal of Research in Science Teaching*, *46*(6), 606-609.

Duschl, R. A., Schweinguber, H. A., & Shouse, A. W. (Eds.) (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8*. Washington, DC: The National Academies Press.

Fensham, P. J. (2009). Real world contexts in PISA science: Implications for context-based science education. *Journal of Research in Science Teaching*, *46*(8), 884-896.

Gräber, W. & Bolte, C. (1997). *Scientific Literacy – An International Symposium*. Kiel: IPN.

Gräber, W., Nentwig, P., Koballa, T., & Evans, R. (2002). *Scientific Literacy: Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung* [Scientific Literacy: the contribution of the sciences to general education]. Opladen: Leske & Budrich.

Hafner, R. (2007). Standards in science education in Australia. In D. Waddington, P. Nentwig, & S. Schanze (Eds.), *Making it Comparable – Standards in Science Education* (pp. 23-59). Münster: Waxmann.

Han, J.-H. (1995). The Quest for National Standards in Science Education in Korea. St*udies in Science Education*, *26*(1), 59 - 71.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., et al. (2003). *Zur Entwicklung nationaler Bildungsstandards* [Regarding the development of National Education Standards]. Berlin: Bundesministerium für Bildung und Forschung.

Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG [Models of Competencies for Assessment of Individual Learning Outcomes and the Evaluation of Educational Processes. Description of a new DFG priority research program]. *Zeitschrift für Pädagogik*, *52*(6), 876-903.

Labudde, P. & Adamina, M. (2008). HarmoS Naturwissenschaften: Impulse für den naturwissenschaftlichen Unterricht von morgen [HarmoS Science: impetuses for the science classes of tomorrow]. *Beiträge zur Lehrerbildung*, *26*(3), 351-360.

Labudde, P., Metzger, S., & Gut, C. (2009). Bildungsstandards: Validierung des Schweizer Kompetenzmodells [Education standards: validation of the Swiss competence models]. In D. Höttecke (Ed.), *Chemie- und Physikdidaktik für die Lehramtsausbildung* (pp. 307-309). Münster: LIT Verlag.

Lavonen, J. (2007). National science education standards and assessment in Finland. In D. Waddington, P. Nentwig & S. Schanze (Eds.), *Making it Comparable – Standards in Science Education* (pp. 101-126). Münster: Waxmann.

Liu, X. & Ruiz, M. (2008). Using Data Mining to Predict K–12 Students' Performance on Large-Scale Assessment Items Related to Energy. *Journal for Research on Science Teaching*, *45*(5), 554-573.

Melton, R. F. (2004). Competence in perspective. *Educational Research, 36*(3), 285-294.

Millar, R. (2007). How standards in science education are set and monitored in the English education system. In D. Waddington, P. Nentwig & S. Schanze (Eds.), *Making it comparable – Standards in science education* (pp. 83-100). Münster: Waxmann.

Millar, R. & Osborne, J. (1998). *Beyond 2000: Science Education for the Future*. London: King's College London, School of Education.

Ministère de l’Éducation nationale et de la Formation professionnelle (2009). *Naturwissenschaften – Kompetenzorientierte Bildungsstandards – Niveau 8e* [Science: competence oriented education standards]. Luxembourg: Le Gouvernement du Grand-Duché de Luxembourg.

Ministerial Council on Education, Employment, Training and Youth Affairs [MCEETYA] (2002). *National Year 6 Primary Science Sample Assessment: Assessment Domain and Progress Map.* MCEETYA.

Ministerial Council on Education, Employment, Training and Youth Affairs [MCEETYA] (2003). *A Measurement Framework for National Key Performance Measures.* MCEETYA.

Ministry of Education. (2007). *The Ontario Curriculum, Grades 1–8: Science and Technology*. Toronto: Ontario Ministry of Education.

Ministry of Education. (2008a). *The Ontario Curriculum, Grades 9–10: Science*. Toronto: Ontario Ministry of Education.

Ministry of Education. (2008b). *The Ontario Curriculum, Grades 11–12: Science*. Toronto: Ontario Ministry of Education.

Mullis, I. V., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 International Report: IEA’s Progress in International Reading Literacy Study in Primary School in 40 Countries*. Boston, MA: TIMSS & PIRLS International Study Center.

National Commission on Excellence in Education (1983). *A Nation at Risk: The Imperative for Educational Reform*. Washington, DC: U.S. Government Printing Office.

Nentwig, P. & Schanze, S. (2007). Making it Comparable – Standards in Science Education. In D. Waddington, P. Nentwig & S. Schanze (Eds.), *Making it Comparable – Standards in Science Education* (pp. 11-19). Münster: Waxmann.

Nentwig, P. & Waddington, D. (2005). *Making it relevant. Context based learning of science*. Münster: Waxmann.

Neumann, K., Fischer, H. E., & Kauertz, A. (2010). From PISA to standards: The impact of large-scale assessments on science education in Germany. *International Journal of Science and Mathematics Education*, *8*(3), 545-563.

Neumann, K., Kauertz, A., Lau, A., Notarp, H., & Fischer, H. E. (2007). Die Modellierung physikalischer Kompetenz und ihrer Entwicklung [Modelling structure and development of students‘ physics competence]. *Zeitschrift für Didaktik der Naturwissenschaften*, *13*, 103-123.

OECD (2001). *Knowledge and Skills for Life – First Results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.

OECD (2004). *Learning for Tomorrow’s World. First Results from PISA 2003*. Paris: OECD.

OECD (2006). *Assessing Scientific, Reading and Mathematical Literacy – A Framework for PISA 2006*. Paris: OECD Publishing.

OECD (2007). *PISA 2006: Science Competence for Tomorrow’s world* (Vol. 1). Paris: OECD.

Ogawa, M. (1998). A cultural history of science education in Japan: an epic description. In W. W. Cobern (Ed.), *Socio-Cultural Perspectives on Science Education: An International Dialogue* (pp. 139-161). Dordrecht: Kluwer Academic Publishers.

Ogawa, M. (2001). Reform Japanese style: Voyage into an unknown and chaotic future. *Science Education*, *85*(5), 586-606.

Osborne, J. & Dillon, J. (2008). *Science education in Europe: Critical Reflections*. Technical report, King's College London, London.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington: National Academy Press.

Ravitch, D. (1995). *National Standards in American Education: A Citizen's Guide*. Washington, DC: Brookings Institution.

Reiser, B., Krajcik, J., Moje, E., & Marx, R. (2003). *Design strategies for developing science instructional materials*. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching.

Resnick, D. P. & Resnick, L. B. (1983). Improving Educational Standards in American Schools. *Phi Delta Kappan*, *65*, 178-180.

Resnick, L. B. (1985). Cognition and instruction: Recent theories of human competence and how it is acquired. In B. L. Hammonds (Ed.), *Psychology and Learning: The Master Lecture Series* (Vol. 4, pp. 123-186). Washington, DC: American Psychological Association.

Roberts, D. A. (2007). Scientific literacy / science literacy. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 729-780). Mahwah, NJ: Lawrence Erlbaum Associates.

Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz [Modeling scientific competence]. *Zeitschrift für Didaktik der Naturwissenschaften*, *12*, 45-66.

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., et al. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632-654.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005a). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. München: Luchterhand.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005b). *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss*. München: Luchterhand.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005c). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss*. München: Luchterhand.

Slavin, R. E. (2002). Evidence-Based Education Policies: Transforming Educational Practice and Research. *Educational Researcher*, *31*(7), 15-21.

Smith, C., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications for children's learning for assessment: A proposed learning progression for matter and the atomic molecular theory. *Measurement*, *14*(1&2), 1-98.

Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, *46*(6), 610-631.

Steedle, J. T. & Shavelson, R. J. (2009). Supporting Valid Interpretations of Learning Progression Level Diagnoses. *Journal of Research in Science Teaching*, *46*(6), 699-715.

Tenorth, H.-E. (2006). Erziehung zur Persönlichkeit: Worauf zielt Bildung? [Education towards personality]. In H.-E. Tenorth, M. Hüther, & M. Heimbach-Steins (Eds.), *Erziehung und Bildung heute* (pp. 7-24). Berlin: Walter-Raymond-Stiftung der BDA, Kleine Reihe, H. 77.

Valverde, G. A. & Schmidt, W. H. (2000). Greater expectations: learning from other nations in the quest for 'world-class standards' in US school mathematics and science. *Journal of Curriculum Studies*, *32*(5), 651-687.

Waddington, D., Nentwig, P., & Schanze, S. (2007). *Making it Comparable – Standards in Science Education*. Münster: Waxmann.

Wei, B. & Thomas, G. P. (2006). An examination of the change of the junior secondary school chemistry curriculum in the P. R. Chine: In the view of scientific literacy. *Research in Science Education*, *36*(4), 403-418.

Weiglhofer, H. (2007). Austria at the beginning of the way to standards in science. In D. Waddington, P. Nentwig, & S. Schanze (Eds.), *Making it Comparable – Standards in Science Education* (pp. 61-70). Münster: Waxmann.

Weinert, F. E. (1999). *Concepts of Competence*. Contribution within the OECD-Project Definition and Selection of Competencies: Theoretical and Conceptual Foundations (DeSeCo). Neuchâtel: Bundesamt für Statistik (Ch).

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and Selecting Key Competencies* (pp. 45-65). Seattle, WA: Hogrefe & Huber.

Wilson, M. & Bertenthal, M. (Eds.) (2004). Systems for state science assessment. Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.Wissenschaftliches Konsortium HarmoS Naturwissenschaften+ (2008). *Harmos Naturwissenschaften+, Kompetenzmodell und Vorschläge für Bildungsstandards* [HarmoS Science+: competence model and proposals for educati ards]. Wissenschaftlicher Abschlussbericht. Ostermundingen: suterprint.

Part B

The Big Picture