Deborah Corrigan · Richard Gunstone
Alister Jones *Editors*

Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy

Springer

Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy

Deborah Corrigan • Richard Gunstone
Alister Jones
Editors

Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy

Springer

Editors
Deborah Corrigan
Centre for Science, Mathematics
and Technology Education
Faculty of Education
Monash University
Clayton, Victoria
Australia

Richard Gunstone
Centre for Science, Mathematics
and Technology Education
Faculty of Education
Monash University
Clayton, Victoria
Australia

Alister Jones
Faculty of Education
University of Waikato
Hamilton, New Zealand

ISBN 978-94-007-6667-9 ISBN 978-94-007-6668-6 (eBook)
DOI 10.1007/978-94-007-6668-6
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013941142

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book is the third in a series, the first two of which resulted from the collaboration between Monash University and King's College London through the Monash-King's College International Centre for the Study of Science and Mathematics Curriculum. This book introduces a third partner, the University of Waikato, an institution which already has strong links with both Monash University and King's College London in the areas of science and technology education.

The first book in the series, *The Re-emergence of Values in Science Education* (D. Corrigan, J. Dillon & R. Gunstone [Eds.], Rotterdam: Sense Publishers, 2007), considered the state of science education in the twenty-first century through the lens of values. The book presented the 'big picture' of what science education might be like if values once again became central in science education. However, overwhelmingly the experiences of those who teach science have taken place in an environment which has seen the de-emphasising of values in school science. So there is a disparity between the evolutionary process that science is undertaking and that undertaken by science education (and school science in particular). In the second book in the series, *The Professional Knowledge Base of Science Teaching* (D. Corrigan, J. Dillon & R. Gunstone [Eds.], Dordrecht: Springer, 2011), the focus was on exploring what expert science education knowledge and practices may look like in the emerging 'bigger picture' of the re-emergence of values. The exploration of the knowledge bases considered necessary for science teaching and how evidence of high-quality science teaching may reflect the knowledge bases of the modern-day professional science teacher are the central themes throughout the book.

This third book, *Valuing Assessment in Science Education: Pedagogy, Curriculum and Policy*, focuses on examining different aspects of generating understanding about what science is learnt. Some overview of the 'big picture' is again provided through explorations of policy issues associated with science education and the assessment of science knowledge of worth. There is also attention given to what the science curriculum may look like if more consideration is given to assessment practices during the processes of curriculum development (rather than the common approach of considering assessment only after the curriculum development is seen to be completed).

The consideration of values, particularly creativity, provides an example of how the relationship between curriculum and assessment can be explored from different perspectives. A range of types of assessment is considered with particular insights provided into how these assessment types are enacted in a range of contexts and educational sectors. Particular examples of pedagogical practices in different classroom contexts are provided in later chapters of the book.

We used the same approach to the creation of this book as we did with the two previous books. In order to attempt both the creation of a cohesive contribution to the literature and having authors able to assert their own voices without restrictive briefs from us as editors, we again organised a workshop involving the authors and ourselves to enable a more interactive and formative writing process. Authors completed the first drafts of their chapters in time to distribute these to all workshop participants before we met. The workshop then involved discussions of individual chapters and feedback to authors and considerations of the overall structure and cohesion of the volume. Authors then rewrote their chapters in the light of these forms of feedback. As with the previous books, the workshop was scheduled around the European Science Education Research Association (ESERA) conference at the Monash University Centre in Prato, Italy, rather than in the same city as ESERA.

This procedure had been previously successfully used in the production of other books we had been involved in, such as *The Content of Science: A Constructivist Approach to its Teaching and Learning* (P. Fensham, R. Gunstone & R. White [Eds.], London: Falmer, 2000) and *Improving Science Education: The Contributions of Research* (R. Millar, J. Leach & J. Osborne [Eds.], Milton Keynes: Open University, 1994). We believe this process significantly improves the quality of the final product and provides an opportunity for what is, sadly, a very rare form of professional development—considered formative and collaborative (and totally open) discussions of one's work by one's peers.

We gratefully acknowledge the funding of the workshop through contributions to the Monash-King's College International Centre for the Study of Science and Mathematics Curriculum from the Monash University Research Fund and from King's College London. We also acknowledge the funding from the Faculty of Education, University of Waikato.

A special thank you must go to Dr. Cathy Buntting for her excellent work in assisting the editors in both the academic work and administrative roles in producing this edited volume.

Clayton, Australia
May 2012

Deborah Corrigan
Richard Gunstone
Alister Jones

Contents

1 Valuing Assessment in Science Education: An Introductory Framework 1
Deborah Corrigan, Cathy Buntting, Alister Jones, and Richard Gunstone

2 International Assessments of Science Learning: Their Positive and Negative Contributions to Science Education 11
Peter J. Fensham

3 International, National and Classroom Assessment: Potent Factors in Shaping What Counts in School Science 33
Alister Jones and Cathy Buntting

4 Improving Science Education: Why Assessment Matters 55
Robin Millar

5 Towards an Authentically Assessed Science Curriculum 69
Peter J. Fensham and Léonie J. Rennie

6 Aligning Science Curriculum Renewal Efforts and Assessment Practices 101
Wilmad Kuiper, Elvira Folmer, and Wout Ottevanger

7 Content to be Assessed Across the History of the National Assessment of Educational Progress 119
Audrey B. Champagne

8 The Influence of Assessment on Moderating Science Teachers' Beliefs and Practices 153
Deborah Corrigan and Rebecca Cooper

9 Issues in Teaching for and Assessment of Creativity in Mathematics and Science 169
Mike Askew

10 Science Teachers' Understanding of the College Entrance Examination in a Climate of National Curriculum Reform in China 183
Hongming Ma

11 Pedagogy in Theory and in Practice: Formative and Summative Assessments in Classrooms and in Systems 207
Paul Black

12 An Assessment *Perezhivanie*: Building an Assessment Pedagogy *for, with* and *of* Early Childhood Science Learning 231
Marilyn Fleer and Gloria Quiñones

13 Classroom Assessment: Making Space for Diversity 249
Bronwen Cowie

14 Formative Assessment as a Cultural Practice: The Use of Written Formative Assessment in Samoan Science Classrooms 267
Desmond Lee Hang and Beverley Bell

15 The Disparity Between Achievement and Engagement in Students' Science Learning: A Case of East-Asian Regions 285
Jinwoong Song

16 Embedding Assessment Within Primary School Science: A Case Study 307
Angela Fitzgerald and Richard Gunstone

17 Formative Assessment in Learning to Teach Science 325
Pernilla Nilsson and John Loughran

18 Changing Assessment Practices in Science Classrooms 347
Chris Harrison

19 Assessment: Where to Next? 359
Deborah Corrigan, Cathy Buntting, Richard Gunstone, and Alister Jones

Author Biographies 365

Index 371

Chapter 1
Valuing Assessment in Science Education: An Introductory Framework

Deborah Corrigan, Cathy Buntting, Alister Jones, and Richard Gunstone

Valuing assessment in science education appears—on the surface—to be a self-evident tenet within science education and science education research. However, while assessment might be valued, *what* is valued in science assessment and how we demonstrate that it is valued are not so obvious. It is also interesting to consider whether assessment in science differs from assessment in other school subjects. What are the links, for example, between science assessment and science as a curriculum area? What are the links between science assessment and the pedagogies employed in a science classroom? How does policy influence what is valued in terms of science assessment, and how does this impact on science teaching and learning programmes? Of course, how science assessment is valued is perceived differently from different perspectives. In order to define the parameters of this volume, we decided that it is the perspectives of pedagogy, curriculum and policy, and the complexities between these that we wanted to articulate.

Pedagogy as a concept can be used in widely varied ways. In this instance, it is not merely the action of teaching (which itself can be easily misinterpreted as the transmission of information), but rather it is about the relationship between teaching and learning and how together they lead, through meaningful practice, to enhanced knowledge and understanding (Loughran 2006). Our interpretation of pedagogy is therefore inclusive of ways in which assessment practices are embedded into everyday classroom experiences.

D. Corrigan (✉) • R. Gunstone
Centre for Science, Mathematics and Technology Education, Faculty of Education, Monash University, Building 6, Wellington Road, Clayton, Victoria 3800, Australia
e-mail: debbie.corrigan@monash.edu; richard.gunstone@monash.edu

C. Buntting • A. Jones
Faculty of Education, University of Waikato, Gate 1 Knighton Road, Hamilton 3240, New Zealand
e-mail: BUNTTING@waikato.ac.nz; a.jones@waikato.ac.nz

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_1,

In the development of this volume, there was much debate about the title of the book, the discussions highlighting how complex the issue of assessment in science education actually is. Suggestions such as 'The many faces of assessment' or 'Assessment in science: Valuing multiple perspectives' attempted to capture the many facets of science assessment explored in this volume. Ultimately, we decided on the chosen title—*Valuing assessment in science education*—in order to reflect our position as multiple authors and contributors that assessment in science does indeed need to be, and is, valued.

Assessment, with its close links with learning, is a fundamental issue in science education: in science curriculum development and implementation, in the classroom teaching and learning of science, and in research. Exploring the relationships between assessment and learning has led to a great deal of research focussed on formative and summative assessments of science learning. While this is a helpful lens through which to view assessment of learning in science, a broader perspective considers assessment *as* learning, assessment *of* learning, and assessment *for* learning:

- Assessment *as* learning focuses on the role that assessment can play in monitoring learning as it occurs, how this role can be used to enhance the quality of learning, and how this monitoring of learning as it occurs can then be used to plan further learning experiences and goals. Assessment as learning seeks to help students take responsibility for their own learning and so build metacognition in the learner.
- Assessment *of* learning involves the monitoring of learning that has occurred. It most often takes the form of the common approaches to 'formative' and 'summative' assessment.
- Assessment *for* learning is a practice deeply embedded in pedagogy and occurs when teachers use inferences based on evidence they have gathered about student progress to inform their teaching. By necessity, assessment for learning is frequent, can be formal or informal (such as through quality questioning, anecdotal notes, written comments), shapes planning for learning, and provides clear and timely feedback for students as their learning progresses. Thus, it is often formative in nature. It also frequently involves students explicitly being made aware of what they are expected to learn. For this reason it may contribute to the development of metacognition in learners. Considerations of progression in learning are perhaps particularly important in assessment for learning as they enable teachers to make judgements about the best ways to support students in their ongoing learning.

These definitions are not unproblematic, nor are they uncontested. For example, assessment for learning is sometimes interpreted simply as continuous summative assessment. Formative assessment is seen by some as being distinct from pedagogy. Summative assessment is often perceived as being directly related only to external assessments, particularly high-stakes assessments. Summative assessment can also be used for formative purposes to provide feedback to enhance learning. Our view is that assessments as, of, and for learning cannot be seen in isolation, but rather that they overlap and interact in ways that are integral to effective science pedagogy. Our

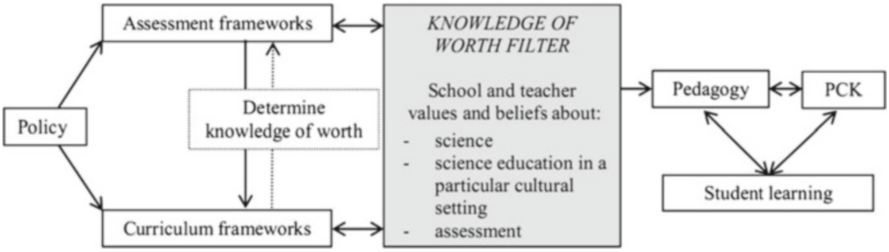


Fig. 1.1 Interactions between assessment, policy, curriculum, and pedagogy

concern through this volume is therefore on the interplay between science assessment practices (including what the teachers says and does in the classroom) and how this impacts on student learning.

In this volume, a range of authors explore assessment philosophies and practices and possibilities from different sociocultural contexts and across educational levels, from early childhood through to tertiary level. These explorations include considerations of assessment as, of, and for learning in relation to science pedagogy, science curriculum, and science education policy. The explorations also examine the relationships between assessment and learning, particularly in terms of the goals, values, beliefs, attitudes, and purposes of science education and how these are interpreted by different audiences, such as policymakers, teachers, and researchers.

Assessment philosophies, practices, and possibilities clearly do not exist in isolation from the broader educational context. In Fig. 1.1, we highlight some of the key elements of this broader context that interact with, influence, and are influenced by assessment. On the left-hand side, policy includes both general educational policies and policies related directly to the science curriculum and its assessment. These system-level assessment and curriculum policies have a reciprocal impact on each other. For example, the national or state curriculum sends signals about what is considered to be important to teach and learn in science—what is the ‘knowledge of worth’. However, it is often the external assessment that sends even stronger signals about which knowledge is most greatly valued, driving what is actually taught.

Both curriculum and assessment frameworks are of course interpreted and enacted at the micro-level by schools and individual teachers. In the actual delivery of the curriculum, therefore, it is the school’s and teacher’s interpretations of ‘knowledge of worth’ that influences the implementation of system-level policies. These interpretations are played out in terms of the content that is prioritised, the pedagogical approaches adopted by the teacher, and the teacher’s associated pedagogical content knowledge (PCK). For example, while there may be a considerable emphasis on the nature of science in national curriculum documents, unless there are requirements for this to be assessed, particularly at high-stakes levels, teachers may choose to pay only scant attention to teaching about how science as an area of knowledge differs from other fields of knowledge. Where assessments exist that call for formulaic-type responses, this may result in pedagogy focused on ‘teaching the formula', with the concomitant development of PCK in which nature of science is considered to be a set of pre-specified, decontextualised notions.

As reflected in Fig. 1.1, in an ideal situation, assessment frameworks represent a valid reflection of the requirements set out by the intended curriculum frameworks written at the national/state and school level (the dotted upward arrow). Where this is not the case, we know that assessment frameworks—particularly high-stakes assessments—are likely to significantly influence the curriculum implemented in the classroom and experienced by students (the downward arrow). In addition, how a teacher interprets and responds to curriculum and assessment policy substantially depends on that teacher's values and beliefs about the purposes and nature of science, science education, and assessment (the grey box). The teacher's interpretation and enactment of curriculum and assessment policy, then, influences the pedagogical decisions that are made and how these are enacted in the classroom. Teacher pedagogy, in turn, influences the teacher's PCK and vice versa. Both pedagogy and PCK work together to influence student learning. Student learning, in its turn, has a reciprocal influence on teacher pedagogy and PCK.

In this book on assessment, examples of each of these complex interactions are explored: how policy might influence assessment practices in science; how assessment policy and frameworks are likely to determine the knowledge of worth presented in the intended curriculum and the importance of aligning the development of assessment policy with curriculum development; how teachers' values and beliefs can influence the ways in which they interpret assessment and curriculum frameworks; and how teachers' interpretations of these frameworks can impact on classroom practice and student learning.

Within any educational jurisdiction, assessment policy addresses formative and, particularly, summative agendas determined by that jurisdiction. In practical terms, these assessment agendas are often primarily focused on the single agenda of measuring student achievement for accountability purposes. This includes schools reporting to parents and the community on student achievement relative to predefined parameters, interschool comparisons using these measures, and system-wide and national comparisons. Any of these forms of accountability concerns can, and very frequently do, drive education intervention, monitoring, and reform (depicted somewhat simplistically as the influence of policy on assessment and curriculum in Fig. 1.1). Assessment policy also interacts with wider education policy. For instance, the current focus in many countries on 'key competencies' as well as numeracy and literacy outcomes signals a more marginalised position for science as a contributor to general education, rather than as a subject having equal status with language and mathematics. These key competency and literacy-numeracy policy directions might be played out in the classroom, even at secondary level, through teacher-student interactions that focus on literacy and numeracy or behavioural outcomes rather than science learning outcomes.

Policy motivations relating to intercountry comparisons and identifying best practice have led to two large-scale international measures of student achievement in science—the Trends in Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA). Each of these has considerable political influence in many of the participating countries. Aspects of the significance of this political influence are explored in Chap. 2 of this volume by Peter Fensham. He draws on his experiences as a member of both the TIMSS Science Advisory Group and PISA Science Expert Group to offer a comparative critique of the studies, including their content, modes of administration, and reporting of findings. He goes on to consider the potential impacts of the intercountry comparisons on the science education of participating countries and economies. Building on Fensham's critique, Alister Jones and Cathy Buntting demonstrate in Chap. 3 how PISA and TIMSS have influenced the science education landscape in New Zealand. They go on to consider the influences of national education and assessment policies on the operationalisation of school science at both primary and secondary levels, providing examples of policies that at once broaden and constrain science pedagogy. In this sense, they recognise the potential for a policy disconnect between what governments might say they want in terms of science education and what they signal through policies, including science-specific education/assessment policies and general education and assessment policies.

Science education policy, as articulated within curriculum documents, often seeks to send clear messages about what is to be regarded as knowledge of worth (shown as a dotted line in Fig. 1.1). However, in real terms, it is the ways in which these curriculum intentions are assessed—the assessment philosophies and practices that are used—that show what counts as the 'knowledge of worth' (the solid downward arrow in Fig. 1.1). For example, while curriculum documents may assert the intention that students develop an appreciation of science and its usefulness in helping to explain their ever-changing world, assessment of this intent is extremely difficult when policy dictates external summative assessments such as examinations, and such assessment almost never occurs. As a consequence the intent is not realised in either classroom pedagogy or student learning. The assessment programme therefore impacts profoundly on what science is assessed, and how, and therefore on what science is learned, and how.

Robin Millar, in Chap. 4, considers an ideal where assessments are developed alongside and at the same time as curriculum statements, arguing that this is the best way to clarify what is meant by the intended learning outcomes. He illustrates this by presenting an insider perspective from his involvement with the *Twenty First Century Science* project, a major curriculum development in the UK. Although this did not adopt a 'backward design' approach, he writes that he 'came later to wish that it had'. He argues, for example, that the development of assessment instruments that could provide evidence of the intended learning is the best way to clarify what is meant by the learning outcomes. The challenge of achieving alignment between curriculum and assessment frameworks is explored further by Peter Fensham and Léonie Rennie in Chap. 5. To demonstrate their position that assessment of learning needs to be authentic, they offer several examples of how curriculum innovations such as context-based science education, socio-scientific issues, and integrated science curricula might be 'authentically assessed'. In doing so, they highlight the trust that needs to be placed in the professional knowledge of teachers in terms of interpreting the curriculum and assessing student learning in ways that are valid and reliable. This is also picked up later by Paul Black in Chap. 11.

Two country-based case studies, one Dutch and one North American, continue the discussion about the interplay between curriculum and assessment. Wilmad Kuiper, Elvira Folmer, and Wout Ottevanger in Chap. 6 consider three senior secondary science curriculum pilot projects in the Netherlands. As with the earlier chapters, they highlight the importance of aligning the intentions of the new curricula with how these intentions are assessed. They point out, however, the difficulty in achieving appropriate clarity and agreement between stakeholder groups: the curriculum writing groups, the Examinations Board, and the pilot teachers. There was also, understandably, considerable teacher uncertainty about the intents of the new curriculum and how these might be assessed. In addition, the evidence suggests that ownership of change is not unproblematic and takes time. Chapter 7 by Audrey Champagne presents another perspective on curriculum-assessment interactions, providing a unique case study of the National Assessment of Educational Progress (NAEP) programme in the USA. From her position as an insider involved in the generation of each of the policy statements referred to in the chapter, Audrey analyses the science knowledge and abilities to be assessed by the programme over the approximately 40 years of its existence. Both her chapter and the previous one therefore focus specifically on the particularly powerful influence of high-stakes assessment on the science curriculum as it is enacted in the classroom. In spite of the hundreds of thousands of hours spent defining the content that needs to be taught in science at the formal education level, there is still the need to either reduce content to 'must know' principles that can be assessed within limited examination timeframes or change the ways in which high-stakes assessments are conducted (as advocated by Fensham and Rennie in Chap. 5).

Obviously, the intentions of any curriculum, signalled through curriculum and assessment documents, are interpreted by teachers, who translate the intended curriculum and enact it in their classrooms. How this interpretation occurs, and the consequent impacts on classroom practice, is highly dependent on the teacher's beliefs and value systems. These include the beliefs and values teachers hold about science, science education, and its place in their particular sociocultural setting, their role as a teacher and science teacher, and the role and place of assessment. In many ways, the beliefs and value systems teachers hold can therefore be seen as a filter teachers use to either accept or reject ideas presented in the intended curriculum (see the grey box in Fig. 1.1). This 'filtering' is explored in Chap. 8 by Deborah Corrigan and Rebecca Cooper, who compare the values of secondary school science teachers in two states in Australia—Victoria, where examinations are externally administered in the final year of schooling, and Queensland, where all work is set and assessed by the classroom teacher. Using a case study approach, they demonstrate that teachers' values and beliefs about science as well as their pedagogical practices are often challenged, or indeed sidelined, by summative assessment policies and frameworks. By way of example, Mike Askew in Chap. 9 explores the sidelining of creativity as an important way of thinking and as a value to be promoted and developed in science education. He argues, for example, that while many curricula and teachers value the incorporation of creativity in science education, a significant shift is needed in both classroom culture and assessment practices in order for creativity to become embedded in science pedagogy and PCK. Providing another example of how summative assessment can sideline the learning of science skills and the development of scientific attitudes, Hongming Ma in Chap. 10 reports on the educational outcomes promoted by China's College Entrance Examination and how these outcomes directly influence teachers' pedagogical practice by strongly emphasising content. This occurs in spite of educational reforms designed to foster process skills and attitudes in addition to traditional knowledge outcomes.

The type of assessment that has the most positive influence on learning is widely recognised to be that where the information gained from the assessment is used formatively to shape future learning, that is, assessment of, as, and for learning. Paul Black, one of the world leaders in this area, points out in Chap. 11 that the last decade has seen a growth in teachers' interest in formative assessment. In this chapter, he considers ways in which summative assessments can be used for formative purposes to benefit learning. In a sense, his chapter acts as a bridge in this volume between the earlier chapters, which focus on the powerful influences of summative assessment frameworks at the system level, and the later chapters focusing on the classroom, including teacher pedagogy, PCK, and more formative approaches to assessment. Black also argues that high-stakes assessment can only be valid, and support learning, if it gives weight to teacher assessment. This builds on arguments introduced earlier in the book by Fensham and Rennie (Chap. 5), Corrigan and Cooper (Chap. 8), and Askew (Chap. 9), maintaining that classroom teachers are the best placed to provide overall judgements of student capabilities in science.

Within the field of early childhood education, Marilyn Fleer and Gloria Quiñones offer in Chap. 12 an account of assessment *perezhivanie* or the holistic relationship between assessment and the lived experiences of the child. This includes how the assessment environment affects the course of the child's development and how the physical and social environment is understood by the child, matters very rarely considered in later years of education. Within their framework, Fleer and Quiñones highlight the need for early childhood assessments that are embedded in pedagogy and which take into account the connections between a child's cognition and emotion, the need for dynamic 'in-the-moment' assessment, and 'potentive' aspects of that assessment. They argue that a pedagogy that involves assessment for, with, and of early childhood science content knowledge and capability provides a way forward for early childhood teachers who wish to engage more intentionally with enhancing children's learning of science.

At the primary level of schooling, Bronwen Cowie in Chap. 13 builds on Black's focus on classroom formative assessment. Teachers' formative assessment practices shape what students come to understand science is about; what it means to learn, know, and do science; and how students come to conceptualise their own relationship with science as a knower, learner, and user. The increasing diversity of students in science classrooms offers a richness that can be harnessed for the benefit of all students, as demonstrated in this chapter through several vignettes in which cultural and everyday knowledge are deliberately incorporated into the rich tapestry of classroom interactions. If teachers are to make space for diversity, they need to know their students and their students' communities well and have both breadth and depth of science content and pedagogical content knowledge.

In Chap. 14, Desmond Lee Hang and Beverley Bell also consider notions of cultural diversity by exploring ways in which formative assessment practices might be put into practice in secondary science classrooms in Samoa. In this context silence (or *le-tautala*) is a cultural practice, making question-answer interactions between teacher and students problematic. The chapter explores the possibilities of using written communication as an opportunity for providing formative assessment as part of a culturally responsive pedagogy. Hang and Bell suggest that written formative assessment sheets may also be useful to 'bridge the gap' in classrooms beyond Samoa where students who practice *le-tautala* may be learning alongside those who do not.

In another exploration of the influence of culture on student engagement and achievement in school science, Jinwoong Song in Chap. 15 considers the apparent paradox between the high science achievement and low engagement identified by TIMSS and PISA among students in East Asian countries. While this disparity is not unique to East Asia, Song explores possible cultural influences on science learning by providing a brief overview of societal and classroom cultures in East Asia with a particular focus on his home country, South Korea. For example, he points out the Confucian influence on teacher-student and student-student interactions, a tendency for students to express ideas indirectly, and a group-oriented approach to learning. The culture of East Asian classrooms, like in Samoa and other Pacific Islands, therefore does not offer opportunities for orally mediated formative assessment practices in the way that Western education systems do. In concluding, Song posits that cultural influences on classroom interactions likely impact on student engagement in science, even where science achievement—according to TIMSS and PISA—is ostensibly in good health.

Moving back to a Western education context, Angela Fitzgerald and Richard Gunstone provide in Chap. 16 an in-depth case study of one teacher's extensive use of formative assessment interactions in a primary classroom. The range of examples of formative assessment throughout this chapter demonstrate the possibilities that can be created to enhance science learning when the teacher pays attention to accessing and evaluating student learning and then uses this information to make decisions about the next learning steps to be presented to students. The teacher's ability to do this appeared in this case to be enhanced through the use of a comprehensive teacher guide for the science unit being taught. The guide, widely used in primary schools in Australia, includes science background content as well as information about likely progression of learning and suggestions for further pedagogical steps in response to particular ideas offered by students.

A further extension of considering assessment in science learning is related to the learning science teachers themselves undertake, in both their initial professional training and their ongoing professional development. In Chap. 17, Pernilla Nilsson and John Loughran offer another in-depth analysis of the potential for formative feedback to enhance student learning, this time of pre-service teachers participating in an undergraduate science education course in Sweden. Throughout the detailed transcripts provided in the chapter, persuasive evidence is presented about the value of formative assessments for enhancing pre-service science teachers' pedagogical knowledge and PCK.

Christine Harrison in Chap. 18 also explores formative assessment in the context of teacher development. She provides details of the King's Researching Expertise in Science Teaching (KREST) project, focusing on evidence-based professional development of science teachers in England and Israel. The project design was premised on teacher professional development from a personal growth perspective, with a concomitant change in agency. In other words, the drive for change originated with the teacher. Of relevance to this volume is the fact that many of the self-selected issues explored by the teachers were related to formative assessment practices. As a result of the professional development, the teachers were able to strengthen their formative assessment practices at the same time as they carried out periodic summative assessments. In order to do this, they needed to recognise good pedagogical practice within the relevant science domain, make sense of its complexities, and understand the effects and synergies of various aspects of their own practice.

Assessment is therefore not uncontested, as these chapters indicate. It comes in a range of guises—for example, as, of, and for learning—and exists at multiple levels—from policy frameworks to classroom practice. In Fig. 1.1, we attempted to illustrate the multilevel, multidirectional interactions between assessment and policy, curriculum, and pedagogy. These interactions include the ways in which particular aspects of scientific understanding, or characteristics of science, may be more valued than others depending on the emphasis in both formative and summative assessment. In Fig. 1.1, we also highlighted the interplay between pedagogical knowledge, PCK, and student learning, with several of the chapter descriptions above hinting at how such interplay may occur in relation to assessment.

The concluding messages expressed in each of the chapters in this volume also provide a basis for consideration of where the gaps might be in thinking about assessment in science education practice and research. In the final chapter of this volume, we therefore offer our analysis of what these gaps are and suggest possible fruitful areas for further investigation in order to enhance assessment's role in relation to science education policy, curriculum, and pedagogy.

Reference

Loughran, J. (2006). *Developing a pedagogy of teacher education: Understanding teaching and learning about teaching*. Milton Park: Routledge.

Chapter 2
International Assessments of Science Learning: Their Positive and Negative Contributions to Science Education

Peter J. Fensham

Introduction

Currently, two large-scale international assessments of science learning are in wide use. The IEA's Trends in Mathematics and Science Study (TIMSS) and the OECD's Programme for International Student Assessment (PISA) provide very different, comparative assessment information to a large number of countries, many of which participate in both projects. The pervasive influence of these projects on the practices and policies of science education means that they must be considered in any discussion of assessment in science education.

Until the 1990s, interest in international comparisons in science education was not great, being largely driven by the IEA and its friends, affiliated psychometric bodies for education measurement in a number of the more developed countries. The IEA's First International Science Study was conducted in 19 countries in 1970–1971, using a written test of 10- and 14-year-olds and of those in the final year of schooling. The Second International Science Study in 1983–1984 assessed the same populations of students in 24 countries.

By the 1990s, the ideology of the market and of global competition was taking hold in education, and national interest in a Third International Mathematics and Science Study (TIMSS) quickened. Forty-five countries signed up for this study which included written tests of learning for 9-year-olds and 13-year-olds and one based on physics for students in the final year of secondary schooling. The testing in 1994–1995 did include a performance test for a smaller sample of the 9-year-old population, but its findings were not included in the main report, ostensibly because of their lower reliability. Since then performance assessment has been quietly dropped.

P.J. Fensham (✉)
Monash University & Queensland, University of Technology,
Building 6, Wellington Road, Clayton, Victoria 3800, Australia
e-mail: p.fensham@qut.edu.au; peter.fensham@monash.edu

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_2,

At the end of 1995 when the results of the study of mathematics and science achievement were published, accountability was becoming a pressing issue for national expenditure on education, and, internationally, notions of what global competitiveness meant for education and for science and mathematics education in particular were being discussed. In the absence of any other reliable performance findings, the TIMSS findings assumed more prominence and attention than the two earlier science studies had enjoyed. This heightened interest supported a TIMSS Repeat study in 1998 when the original 9-year-old cohort would be another 13-year-old cohort. This repeat study was followed by the decision, supported by even more countries, to continue this TIMSS-type study of these two student populations every 3 years, now known as the Trends in Mathematics and Science Study with the framework described by Mullis et al. (2003).

The OECD in Paris had for a number of years gathered structural and economic data about education in its member countries and in the early 1990s launched a project that gathered reports of successful innovations in the fields of mathematics, science and technology education (Black et al. 1995). A natural next step, perhaps encouraged by the response to TIMSS, was for the OECD to embark on its own comparative studies of educational performance but with a different intention from TIMSS. The result was the ambitious PISA study which would run over a 3-year cycle enabling each of reading, mathematics and science to be the major domains for testing—2000, reading; 2003, mathematics; and 2006, science—with the other two as minor testings. The detailed Science Framework for the 2006 testing is set out in OECD (2006). Before the end of the first PISA cycle, the interest in it among OECD members and among other non-member countries—who were allowed and encouraged to participate—ensured that PISA, too, would become an ongoing juggernaut with a second cycle for 2009–2012–2015.

Despite the fact that there is no international agreement about the purposes, curriculum and learning standards for school science education, TIMSS and PISA have separately made identifications of key aspects of the purposes of science education. With so many financial resources being spent each year, internationally and in the many participating countries, on these international assessments, this chapter sets out to provide an appraisal of how these projects are contributing, positively and negatively, to the assessment and good health of science education (see Jones and Buntting, this volume, who do just this for New Zealand).

Insider and Outsider Perspectives

In 1993 I was invited (as a known critic of the IEA's Second Science Study) to join the small Science Advisory Group for TIMSS and served in that role through the initial development of the test instruments, the first data collection and analysis, and to the publication of its findings in 1996–1997. In 1998, I was invited to join the small Science Expert Group which was charged with designing, analysing and reporting the testings for science as a minor domain of PISA in 2000 and 2003. I remained on that Expert Group when it was enlarged to 14 members for the 2006 testing of science as the major domain and through to its publication of findings in 2007–2008. Accordingly, I write this chapter from the privileged position of an insider, with experience of both projects' decision making, but not of their day-to-day management. Furthermore, as the only Australian in the PISA Expert Group, I have been consulted for advice (rarely directly implemented) since 2000 on a number of occasions when developments in science education were happening in Australia at the state or national levels.

My outsider perspective is like that of other interested science educators. We have access to the projects' reports of their findings and to their aftermath influence (in so far as this is published) on the policy and practice of science education, nationally and internationally.

Overview

In this chapter I want to make clear that each project has its own distinctive intentions about assessing science learning, and these intentions highlight and limit their usefulness to science education, both internationally and nationally. Furthermore, both projects have had the opportunity for data collection that exceeds by an order of magnitude what is possible for an individual or even a cross-national group. Accordingly, each of the projects had considerable potential to be an innovative influence on school science about three key aspects: *the science learning to assess, the approach to assessment* and *the presentation and discussion of the comparative findings*. The choices each project made about these three aspects then determined whether this potential has been positive or negative in practice. After these key aspects are discussed, the projects' wider influence on the policy and practice of science education is considered, and the chapter concludes with some comments on the projects as research studies and as stimuli for research by others.

The Science Learning to Assess

Roberts' (2007) distinction between Vision I and Vision II scientific literacy is a convenient way of identifying the science knowledge being assessed in these two projects. A curriculum aiming for Vision I emphasises science knowledge that is an introduction to the canon of the orthodox natural sciences. The science knowledge for Vision II is drawn from situations involving science and technology (S&T) that students are likely to encounter in everyday life.

IEA/Science Assessment Intentions

The assessment studies of the IEA are unavowedly concerned with providing information about the current curriculum for science that forms a common core of science knowledge across the participating school systems. These curricula in 1993, and still in most cases, define their content for learning in terms of Vision I. Accordingly, the TIMSS project with its tests seeks to retrospectively measure, as an attained curriculum, the science learning among the chosen student populations using the common core of science knowledge as the intended curriculum. The intended curriculum for science in each of the participating countries is identified through an analysis of textbooks and curriculum statements. Data about the implemented curriculum is gathered by separate questionnaires.

This approach means that the science knowledge relating to other emphases about science learning that are not reflected in most countries are excluded. The TIMSS test can, thus, have only a marginal role in suggesting innovations or new dimensions for science teaching and learning. For example, in the early 1990s, there was considerable interest in some countries (i) in students' alternative conceptions about scientific phenomena and concepts and (ii) in using Science/Technology/Society (STS) as an approach to teaching science. Early in the history of TIMSS, an international team reviewing its progress pointed out that it would be more valid and helpful to science education overall if there was a core test of the common curricular content, plus optional tests of innovative dimensions like the above that were emerging. This was quickly ruled out by the controlling body, largely on the grounds of reliability—the dominant statistic in these international projects. As a result, the status of TIMSS and the power associated with its findings do have an effect of reinforcing the status quo of science education.

OECD/PISA Science Assessment Intentions

Rather than replicate TIMSS as a measure of retrospective curriculum learning, the OECD's commission for the PISA project was to provide information to participating countries about how well prepared their 15-year-old students were for twenty-first-century life in reading, mathematics and science—an unusually prospective brief for the assessment of learning. This population of students was chosen because in a number of countries, it is the age when compulsory study of science and mathematics can cease. Future preparedness, as an assessment intention was quite unknown in 1998 among the OECD countries. Hence, there were no existing models for their testing, and one had to be developed that would lead to measures of the students' capability to apply their science knowledge to novel situations. While the innovative intention to measure preparedness could not but be applauded and endorsed by the member countries of the OECD, there was widespread scepticism about what would be found by such a study.

Since science was a minor domain in PISA 2000 and 2003, the Expert Group had the opportunity to explore several approaches to its task before settling for the testing as a major domain in 2006 on a definition of scientific literacy as an individual's:

- Scientific knowledge and use of that knowledge to identify questions, to acquire new knowledge, to explain scientific phenomena and to draw evidence-based conclusions about science-related issues
- Understanding of the characteristic features of science as a form of human knowledge and inquiry
- Awareness of how technology shapes our material, intellectual and cultural environments
- Willingness to engage in science-related issues and with science as a reflective citizen (OECD 2006)

With this definition, PISA science was firmly committed to a Vision II approach to science knowledge. The scientific literacy definition was operationalised as three cognitive and three affective scientific competences—*identifying scientific issues, explaining phenomena scientifically and using scientific evidence* and *interest in science, support for science and responsibility towards resources and environments*.

Levels of Science Learning

This comparison of the science learning emphasised in these two international projects leads to a clear distinction about the goals for science teaching and learning that is often blurred in a national curriculum. It is quite common to find the science content knowledge for teaching and learning to be listed in a curriculum's statement under a dual heading of Knowledge and Understanding. It is as if these two English words are synonymous, since there is usually no explanation that they may be intended to refer to different learning of the same content, such as different levels of learning. I am not suggesting that such a national curriculum intends only one level of learning. Indeed, most would endorse an intention for different levels of learning from shallow recall to deeper application, but this difference is often lost under the dual heading.

The science content knowledge that is tested in TIMSS is overwhelmingly at the shallower end of this spectrum, namely, recall of science knowledge that has been taught or absorbed from a textbook. This is in keeping with the basic interest the participating countries have in TIMSS, namely, to know how well their students can reproduce content knowledge that is common to them and the other countries' intended curricula at the chosen levels of schooling.

The countries' interest in PISA, on the other hand, is its claim to primarily measure how well their students can apply the science knowledge they have learned to science and technology (S&T) contexts that are novel to them and hence go beyond how it was taught or presented in textbooks. Thus, it can be said that, between them, the two projects provide powerful models for curriculum authorities to define different types and/or different levels of learning for science education (see Millar, this volume, on the importance of exemplary items). They also offer assessment tools that provide authentic measures of the recall of common science knowledge and of the application of science knowledge in new situations.

It is of interest to note that it was non-English members of the PISA Science Expert Group who pointed out that the distinction between recall of knowledge and its application in novel situations can be served well by the two English words, *Knowledge* and *Understanding*.

Assessment of Affect About Science

In the years since the first TIMSS testing, there has been an accelerating stream of reports from international and national studies that indicate a decline in student interest in science and in science careers. Schreiner and Sjøberg (2007) from the data in the ROSE project showed that this decline was particularly prevalent across the more developed countries but was generally absent in more developing countries. There were also marked differences in the science topics of interest, especially in terms of gender.

Both projects have set out to measure student attitude to science but again with distinctive difference in how this attitude is conceived. TIMSS took the view in 1994–1995 (subsequently maintained) that interest in science is a personal attribute and hence focused its measurement of this construct in the student questionnaire that also provides measures of a number of other personal and family-related variables. The students' responses to several items lead to scale measures of several aspects of the construct, *attitude to science*. The positive outcome of these attitudinal measures is that they provide participating countries with representative and highly reliable indications of their students' attitudes to science. Furthermore, when these measures are compared across the successive testings of TIMSS, they provide countries with important indications of trends in affect about science that would be expensive and difficult to achieve in locally based studies.

On the negative side the TIMSS approach to affect assumes that the words "science" and "technology" are well-understood terms for their young student respondents. When the items to measure affect are placed in the student questionnaire, these words are used in a decontextualised way. It is well known in the research literature on affect and science that students of the age of the two TIMSS populations hold a large range of meanings for these two words and that their affect can vary considerably when these meanings are clearly differentiated.

With the advantage that science was not to be the major domain in PISA until 2006, the Science Expert Group was able to recast its preliminary definition of scientific literacy to include a strong statement about affect and science (see above). This led to PISA Science in 2006 assessing two affective constructs, *interest in science* and *support for science*, by embedding items for them in the contextual units of the overall test alongside the items for the cognitive competences. A third affective construct, *responsibility towards resources and environments*, was placed, for political reasons within the project, in the student questionnaire, but its items were still related to environmental issues. The student questionnaire also included sets of items to capture these somewhat older students' more generic attitudes towards science.

The embedding of affective and cognitive items in the main assessment test was a major innovation and contribution to science education in two ways. Firstly, it signalled very clearly that both types of learning were natural expectations from compulsory school science. This practice threw out a challenge to the designers of science curricula to think hard about the role of interest in choosing content for teaching and learning. For example, a project to establish a new national science curriculum was launched in Australia not long after the PISA report in late 2007. It was relatively easy to ensure that this new curriculum begins with the words: "Science is a way of answering interesting questions about the natural world" (ACARA 2011, p. 1). Secondly, the embedding meant that students could respond positively to the specific science in one contextual unit and negatively to what underlay another contextual unit. A much richer portrayal of the meaning of the students' *interest in science* and *support for science* thus emerged, paralleling the different responses to science topics that the ROSE project has reported.

The Approach to Assessment

The Mode of Assessment

The use by both TIMSS and PISA of a paper-and-pencil mode of assessment has brought both positive and negative outcomes for science education. This mode made the testings generally a familiar activity to many (but not all) of the countries' students for whom responding to paper-and-pencil testing is the dominant mode of assessment to which they are accustomed, both within school and externally. For the reason that a wide coverage of the science knowledge in the intended curriculum be achieved, the commonest form of item in TIMSS is a time-efficient, simple multiple choice one, and this, too, is commonly used among most (but not all) of the participating countries. TIMSS does include some complex multiple choice and free-response items that can provide a greater sense of validity. PISA uses fewer multiple choice items and hence more of the other two more valid types of items. Although both projects occasionally go further by having items where a second level of response is sought (e.g. give your reasons to support this answer or choose your reason from a set of possible reasons), neither has managed to solve how to score the first answer/second answer combination adequately.

The inclusion of the range of item types in the projects should encourage countries and their schools to also use a wider range of assessment items since the more precise and open ones can then offer diagnostic as well as formative indications of student learning.

The development of the achievement tests for both TIMSS and PISA has involved procedures to ensure validity and reliability that go beyond those used in most countries. They include extensive face validity of the items among panels of experts, linguistic and cultural analyses for bias and statistical analysis of extensive trials with student samples in several countries to establish each item's discriminating power (for TIMSS, see Garden and Orpwood 1996; and for PISA, see OECD 2009; McCrae 2009).

These thorough approaches to test development now stand as exemplary models for the development of similarly intended assessment instruments at a national, regional or local level of education. At these levels, good models of assessment procedures are very much needed, since only some of those responsible for extra-school tests, and even fewer of the science teachers setting intra-school tests, have expertise in designing valid and reliable tests. Given the increasing influence in a number of countries of testing at all these levels, it is alarming that much paper-and-pencil testing of science does not have the safeguards of good item design that have pertained in TIMSS and PISA. The recent Essential Secondary Science Assessment (ESSA) initiated for the full population of Grade 8 students in the state of New South Wales is, however, an example that has largely followed these processes (Department of Education NSW 2012).

Unfortunately, the practice of both these large-scale projects is to release only a small fraction of the items from any one testing so that their elegance as scales is never publicly evident. Nevertheless, enough items have been released over the years for them to be seen as reliable "item banks" for the types of science learning each project intends. Unfortunately, these "banks" are seriously underused at the intra-national levels.

The national findings on individually released items are drawn on by curriculum authorities to gain a more reliable sense of what their system of schools is achieving compared with the reports of the schools' own assessments. In addition some school and teacher use of these items has occurred, but their potential as well as designed indicators of science learning in relation to specific topics has not been encouraged as much as it might have been, given the familiarity that TIMSS and PISA now have. Furthermore, because the associated difficulty level of each released item is usually given, they have diagnostic usefulness for teachers when teaching the associated topic. Such use seems to be rarely promoted by the responsible national authorities.

The negative aspect of the projects' use of the paper-and-pencil mode lies in what it cannot test. There are now a number of commonly agreed curriculum goals and intentions for school science education that are not amenable to this mode of testing. The classic and abiding example of these is the assessment of practical performance in science, but now decision making about socio-scientific issues, context-based science and science project work in and outside school can be added as not amenable to this mode of testing (see Fensham and Rennie, this volume). Neither TIMSS nor PISA acknowledges the absence of any testing of the science learnings associated with these newer goals. Such high-status silence can easily be interpreted as suggesting they are not of worth.

For PISA Science 2000 an attempt was made to include items testing the scientific competence of communicating within and about science, another of these newer goals, but it proved too difficult to design authentic short items for it. Then in association with the 2003 testing, a successful albeit expensive set of computer-based items was developed that presented the dynamic character of scientific phenomena as the stimulus, a feature that cannot be well represented in paper-and-pencil testing. Alas, only four countries took up the challenge of this exploratory test and it was not developed further for 2006.

It is intended, I believe, that the second testing of Science as the major domain of PISA in 2015 will be delivered entirely by computer, but this will simply be a computerised version of the paper-and-pencil mode. It will be a pity if the opportunity of such a computer setting is not also used to demonstrate and explore the role that multivariate modelling is playing in science. Many excellent learning objects of modelling software are now available for use in school science, and the inclusion in PISA of even an optional test for this learning would assist this type of science to become more mainstreamed in the curriculum.

Unexpected Findings

Despite the psychometric rigour of the test development process, some items reach the tests that lead to quite intriguing and unexpected results. As an insider I know that the archives of both projects contain a number of items that throw light on widespread assumptions that underpin school science education. One classic example was included in the TIMSS physics test for final year students in 1994–1995. It involved a diagram (see Fig. 2.1) of a 2-l plastic bottle containing water in which there were three holes down the side from near the top to near the bottom. This, at the time, was very like a familiar diagram in older physics textbooks under the topic *Pressure*. Incorrect relative trajectories of three jets of water were displayed and the item required the students to respond to the question, *What is wrong with the diagram?*

Some European students scored well on this item, but most from other countries were incorrect. What was expected to be a relatively easy item proved to be very difficult. The difficulty arose in two ways. Firstly, the bottle in the test was sitting on a shelf, not free floating in space as in the textbooks. The shelf added a further dimension to the problem that could be resolved by careful thought, or by the use of calculus, to show that the middle hole (rather than the bottom hole in the text book case) should have the furthest trajectory. Secondly, the invitation in the question to comment on the diagram led to a number of very reasonable comments, such as *If the heavy line on top of the bottle is indeed a lid then no water will come out*, or *2l plastic bottles don't have flat bottoms*. A small percentage of these "incorrect" answers stated: *The three trajectories will hit the shelf at same point.* Careful rereading of these gave the clue that these students had seen the three holes as being vertically directly above each other, leading to this interaction between the trajectories.

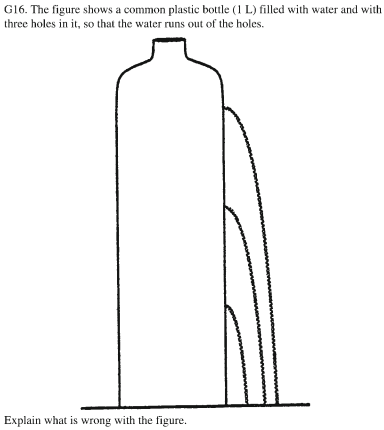


Fig. 2.1 A TIMSS item for Physics, final year population, 1994

Contextualised Assessment

Apart from the important statistical quality of its items and its limited use of free-response and complex multiple choice items, TIMSS offers little that is innovative to the assessment of science learning. PISA Science, on the other hand, uses a style of item that is not yet familiar in most countries, namely, contextualised items, which could well be promoted in the numerous systems around the world that are encouraging their science teachers to add some everyday and real-world relevance to their teaching. Their novelty warrants a fuller account of their use in PISA, and this leads on to a critique of their usefulness in contemporary school science.

Given its charter of providing a measure of student preparedness in science for the twenty-first-century world, it was not surprising that PISA would use contemporary S&T situations as presenting contexts for its test. Measuring its defined scientific competences in such contexts is also consistent with Vision II science knowledge. This use of real-world S&T contexts was a novel feature in a science assessment instrument, and it came 20 years after it might have strengthened the case for the STS movement's flirtation in a number of countries with real-world S&T situations in the later 1980s and early 1990s.

The use of real-world contexts in PISA enabled the project's brief to report on how well students could apply their science knowledge, rather than simply recall it. Almost all the S&T contexts used in the PISA testings would be novel to the great majority of the students, albeit having intrinsic interest for large numbers. In this way PISA has modelled that a competence is only evident in relation to a context—a truism that is not always recognised in all context-based, science teaching and learning.

PISA's priority was to measure several well-defined scientific competences, not to explore a student's mastery of an S&T context. Several items measuring two or more of these were asked for each presenting context. Each student, in this way, had three cognitive scores, which could be summed to give a scientific literacy score. Once again, PISA offers a model for assessing students in relation to different ways of learning science, a lesson few countries have yet adopted in their obsession with a single score for each subject area.

The association of several items measuring different scientific competences with each presenting context idea was also innovative—a radical departure from the usual paper-and-pencil test of science where the items appear singly, often measuring the same type of learning about the different content topics in the curriculum, albeit some being presented in terms of an idealised or contrived context. This PISA approach to context-based assessment of science learning has now been adopted in some national science tests in several countries and in some state jurisdictions in Australia. This innovation in science assessment has had an immediate effect of encouraging science teachers to make more use of real-world contexts, such as media reports, to teach the curriculum's science topics.

On a negative note, PISA's use of context is, nevertheless, disappointing as an assessment model for the strong interest in the last decade in what is called context-based science teaching. Context-based science teaching has been interpreted in a variety of ways, as occurred earlier with STS science (Solomon and Aikenhead 1994), but in its richest form, it is based on the assumption that the context provides exemplary meaning for the science concepts, principles and scientific ways of thinking it invokes, and in learning this science knowledge, a fuller appreciation of the context is achieved. There is thus learning of science and learning about the context. Both these learnings need to be assessed. In a more minimalist way, context-based science teaching can treat the presenting contexts, as PISA does, as mere vehicles to test the presence and application of science knowledge. A more complete approach to assessing context-based science education is discussed in Fensham and Rennie, this volume.

There are a number of reasons why the PISA Science tests, despite their pioneering use of real-world contexts, fall so short of what might have been an assessment stimulus for the richer end of the context-based science spectrum.

Nentwig et al. (2009) analysed the 111 items used in PISA Science 2006 in terms of their degree of contextualisation, i.e. how dependent the item task was on the information in the presenting context. They found that only about half the item tasks were dependent on the presenting information. Traditional conceptual science knowledge was more common in the low contextualised items, whereas knowledge about the nature of science was more common in the high contextualised ones. Such an uneven distribution of the knowledge types limited the analysis that could be done, but in Germany and ten other OECD countries, students perform equally in high contextualised and low contextualised items. In the Netherlands, however, students (especially boys) do better on high contextualised items, a finding that may reflect the reported emphasis in the Dutch curriculum on knowledge about scientific processes and real-life issues, alongside canonical knowledge (Roth et al. 2006).

Fensham (2009) attempted to reanalyse the PISA data in terms of context mastery but found the test design did not ensure that equal numbers of items for each competence were asked for each S&T context and that each context contained competence items with a similar range of comparable difficulty.

More damningly, Sadler and Zeidler (2009) acknowledged several features of PISA Science that mark it out as an innovative international test—*variety of items, contextualised prompts, emphasis on scientific processes and use of scientific evidence*. They go on, however, to point out how little attention has been given to the now common references in curriculum statements to equipping students in *decision making in real-life S&T situations*. They point out that scant attention is given in the tests to the third and fourth aspects of its own definition of scientific literacy (see above). They conclude that, at a broad level, the PISA Science Framework has much in common with their and others' versions of science education that take socio-scientific issues (SSI) seriously. Only the items that involve the competence of using scientific evidence come close to the seriously discursive ways these authors wish students to engage with SSIs.

The Presentation and Discussion of Comparative Findings

Both projects convert the test responses of their total student sample to a mean value of 500 (with a standard deviation of 100). The national scores are calculated in relation to this arbitrary mean of 500. They also acknowledge the error possibilities in their measures. It is politically understandable that these projects reported the mean score of each country's sample of students in their tables of comparative ranking, but this is statistically and educationally untenable, given the multiple sources of error that such tests and testing inevitably have. To be fair, the tables do indicate blocks of countries between whose students there is no statistically significant difference in the mean scores. Thus, the mean score of a country's students may be ranked at say 12, but in fact they lie in the second highest block of countries. There are numerous reports of journalists and educational authorities seizing on the crude ordinal ranking, and in some cases, the latter have acted precipitously in response.

The reporting practice is further compounded by the projects' use of the standard error of the mean in terms of its confidence interval as the measure of statistical difference. This measure which indicates differences when mean scores differ by as little as 5–10 is somewhat of an exaggeration, compared with, say, the use of the standard deviation, which would require a difference in mean score of between 20 and 49 to be rated as a modest.

Differences Between Groups

There are a number of positive outcomes of the reports from the projects. Clear indications of gender differences in many countries emerge that may not otherwise be known or so well documented. Another is the reporting of the spread of performance within the large samples from each country. Thus, the difficulty of the items is used to indicate the size of this spread at a grade level or at an age level. It is not uncommon for differences of more than three standard deviations to be found, and this information needs to be taken very seriously by educational authorities. There is also a spread of performance when the socioeconomic status (SES) of students is used as a covariate. The higher performing countries in general have a low SES slope, but some like Australia with quite high mean scores have steeper slopes indicating serious inequities in their school systems. Similar analyses of differences for other demographic indices, like rurality, indigeneity and immigrant status, are reported from the projects.

Assessment Profiles

A different positive contribution of the projects' findings is that both of them, although providing a composite Science score as just discussed, present science learning in terms of a rich profile of scores. In TIMSS there are scores for the different science content areas and for environmental issues and the nature of science. In addition the national performance on particular items in each of these areas is used to illustrate even finer details in the science that has been taught and learnt. In PISA there are scores for each of three cognitive scientific competences.

This exemplification of science learning as a multifaceted process in which different foci in science and different cognitive demands are recognised is much needed in how school science is assessed, formatively in the classroom and summatively at the end of a period of teaching and learning. Fensham and Rennie discuss assessment as a profile at some length and how it can be exemplified in practice.

Influence on National Science Education

The many countries participating in TIMSS and PISA have invested considerable numbers of personnel and sums of money in these projects. Accordingly, it is almost to be expected that these countries will be influenced by the findings of the projects with respect to their own science education. The extent and manner of this influence is, in fact, very variable and it is probable that TIMSS has had more influence overall than PISA. This is due to the closer relation that the science content of TIMSS and its approach to assessing learning has to a country's existing practices. Furthermore, being scores for populations in the mid primary years and the early secondary years, these TIMSS scores can be interpreted as ones that can be attended to with short-term responses such as professional development workshops. The PISA population, for good reason, was chosen to be at or near the end of the compulsory study of science, and, hence, its terminal character means that the findings about it do not have the same immediate relevance for educational authorities. If the findings of PISA are to have influence, much more radical changes in the science curriculum need to be contemplated.

I have commented above and written elsewhere about some negative influences of the TIMSS test (Fensham 1995). One is its restraining influence on curriculum reform in science education. Because the IEA itself and its TIMSS committees have to test only the essentially common content across many countries' school science curricula, their tests act to reinforce the maintenance of the existing view of the science curriculum. Likewise, the project's use of assessment modes that limit the learning intentions that are measured does nothing to suggest other modes that may lead to more valid assessment of the possibilities for science education.

Despite these limitations of the TIMSS findings, preventing them from validly and authentically reflecting a given country's intentions for science education, a reported performance can lead to national and local decisions that can have significant impact on present practices. For example, Australian students' performance in TIMSS (albeit located in the second highest group of countries) was used to justify the introduction of national paper-and-pencil tests at Grades 3, 5, 7 and 9, which are much more like TIMSS than they are like PISA. One state authority went further, commissioning practice test items and making a content knowledge test, beyond their university degree, the ultimate basis for primary teacher registration. These additional tests of students and future teachers will now act in a quite contrary direction to the aims for science teaching and learning that much Australian effort has been supporting.

The PISA Science test is less prone to have these sorts of influence because (a) it is not assessing a whole curriculum for school science, and (b) it is so different in intent from those in traditional or even recently revised national curricula. Nevertheless, the Framework of PISA Science and the more varied forms of items used in its testing have been used to promote innovations in science education in a number of countries. Examples in Australia, including the appearance of the new content strand, *Science as Human Endeavour,* in the new national curriculum, and the use of context-presenting assessment items in the ESSA test in NSW at Grade 8, can be fairly attributed to the influence of PISA Science (Department of Education NSW 2012).

Some European authors have expressed considerable concern that these projects, coming as they do out of one tradition of educational research and assessment, may have deleterious effects on the natural evolution of science education in countries with other traditions.

Sjøberg (2004) questions the limited aspects of scientific literacy that TIMSS and PISA Science have chosen to assess and views with suspicion the link the latter has with the OECD's economic agenda. He is also critical of PISA's assumption that a paper-and-pencil test can serve to assess "the important knowledge and skills needed in adult life" (OECD 2006, p. 8). Fischler (2011) goes further, seeing both TIMSS and PISA as emphasising an approach to assessment of learning and its conditions that are restricted to and can be expressed, positivistically, in quantitative terms. He is also philosophically critical of PISA's choice of scientific competences as the only units for assessing student learning, because they are underpinned by a utilitarian view of education. PISA Science does define three cognitive scientific competences—*using evidence, explaining phenomena and investing phenomena*—in a utilitarian way, even though the last two can be found among the traditional aims of science education almost universally. Fischler argues that the focus on these competences detracts from other purposes science education ought to contribute like Mündigkeit (learners who aspire towards autonomy and self-activity) and Selbsttätigkeit (learners as free and sovereign beings)—goals that are strongly present in the central and northern European holistic tradition of Didaktik and Bildung (Shirley 2008, p. 38). Dolin (2007) and Hopman (2008), from the same tradition, see the restrictions that each project imposes with a singular view of science learning, an approach to test construction and testing procedures, and by the way teachers and students will perceive these tests as defining their role. All of these they see as antithetical to the intentions for learning that Bildung encourages with its more open-ended approach to the teaching of science.

An indication of the reality of these concerns can be drawn from two studies of the impact of TIMSS and PISA, which have been reported by Waddington et al. (2007) and by Achieve Inc. (2010). The first of these was based on experts from 11 European countries and from the USA, Israel, Australia and Taiwan, discussing the possible development and introduction of science content standards. The second was a commissioned study for US policymakers of the differences and similarities between science standards in ten high-achieving countries on the tests, as a possible basis for setting the next generation of science standards in the USA.

De Boer (2011) draws on these two reports to explore contemporary interests in the globalisation of science education. The pressure to improve science education undoubtedly has an economic competitive edge, and international league tables of achievement feed this aspect. The globalisation of science itself, particularly in the fields of health and environment, also provides pressure, and PISA's style of contextualised assessment encourages this (Fensham 2011). Setting content standards for science learning can be tools for both encouraging these pressures and assisting their implementation in the science curriculum, but a more critical question is to which sort of science education are these standards to be related?

De Boer (2011) found two trends. In countries where standard setting has been a state matter (e.g. Switzerland, Germany, Australia and the USA), there is now active consideration of more precisely defined outcome standards at the national level that would be supported by common assessment. Other countries are considering more holistic and integrated interpretations of their science learning outcomes, which would require broader competency models to describe the student outcomes, instead of discrete knowledge and skill statements. Denmark, Germany, Portugal and Taiwan each reported interest in competency models but these were markedly different. The Taiwan model refers to "cultivating students with competences that they can carry with them for a life time" (Chiu 2007, p. 304). This conjures up an interesting balance between utilitarianism and Didaktik. De Boer concludes from these most recent reports that it would be hard to claim that any general response is emerging in science education, either in specifying content, in extending national testing or in relating to the autonomy of science teachers. In some countries these features are increasing, but in others they are decreasing.

Contribution as Research

Any one of the TIMSS reports and the PISA 2006 report can be rated as a successful very large-scale comparative study of science learning internationally. The extent to which the overall focus for science learning in each project is nationally significant depends on how close these foci are to the national interests.

There are a number of countries for which the projects have provided hitherto unavailable national data. Some of these are countries that have federal systems of education like Germany, Canada, the USA and Australia. Another group is the European countries, like Norway and Denmark, where national testing is seen as quite contrary to the democratic nature of the education system, but since only light sampling was required, these nations' participation was possible. Then there are a number of countries that do have a national curriculum but, until they joined the projects, had been unable to carry out such thorough national monitoring.

The PISA project has made some other, more specific contributions as research. The decision about the embedding of affective items enabled the PISA project team to explore, in their large trials, the use of both *Likert scale statements* and *Identify your opinion* to probe affect about science. While the former style of item was chosen for the 2006 test, some positive and negative features of each approach were clarified thanks to these trial studies.

As mentioned above, items for six environmental issues to explore the affective scientific competence, *responsibility towards resources and environments*, were included in the student questionnaire in a manner that led to measures of *awareness, importance, optimism* and *responsibility*. Their analysis led to what are probably the first cross-national findings of the significance on gender differences in environmental issues, namely, that, in the majority of a wide spread of countries, boys express more *awareness* and *optimism* about these issues whereas girls see them as more *important* and with *less optimism.*

As pieces of research the projects do provide countries with baseline data on the strengths and weaknesses of their students in different aspects of science learning. TIMSS, and PISA in 2015, will also report on trends in this baseline. They also identify on their key science performance measures the achievement of subgroups of students—girls/boys, indigenous/nonindigenous, migrants/nonmigrants and family socioeconomic backgrounds. PISA, by giving a priority to the important issue of family socioeconomic background and science learning, has contributed substantially to the wider literature on how this critical background element should be conceived and measured.

Stimuli for Further Research

At the national level of both projects, it has been possible for researchers to gain permission to gather additional data that addresses particular local issues. This opportunity has been taken up in a number of cases (e.g. Hong Kong, home influences; and

Australia, migration), and it has enabled quality representative data to be collected which would have been very difficult and expensive otherwise. National bodies have also been encouraged to over sample subpopulations of students that are of special local interest, such as indigenous students in Australia.

After each testing has been officially reported, both projects have encouraged researchers to make use of their large data bank, and an increasing number of scholars have taken up this offer. With the limitations indicated in this chapter, the quality of these assessment data is very high, and it exceeds much of the corresponding data being reported publicly and in the research literature. The instruments are developed rigorously, and the strict sampling of the target student populations leads to national and more local representativeness that is not usually possible for individual researchers. Despite TIMSS' much longer life, the PISA project seems to have stimulated further research on the science assessment findings. Among the more interesting studies of the TIMSS data have been the attempts to understand the comparative results in terms of the country's tradition of science education. For example, Valverde and Schmidt (1998) compared the levels of expected science learning across countries, and Ramseier (2001) found that a number of central European countries that have traditionally valued declarative science knowledge did perform well on those items but did less well on scientific procedures, compared with other group of countries in which there has been a long emphasis on practical science.

Even before science was the major domain in PISA, the novelty in its variety of item formats attracted the attention of Yip et al. (2004), who reanalysed the Hong Kong data to explore the gender differences in relation to these formats.

In 2009, a well-attended conference on PISA-stimulated research was held in Kiel (Leibnitz Institute for Science Education, IPN 2009), and soon afterwards a special issue of the Journal of Research on Science Teaching (2009, 46[8]) had the theme of the role of S&T contexts in the cognitive assessment; three of these studies were referred to above in the section on *Contextualised items*.

The International Journal of Science Education (2011, 33[1]) has recently devoted a special issue to PISA's assessment of the affective competences, with studies that reanalyse and extend the project's original data and analysis. The bold and pioneering decision referred to earlier to embed affective items within the S&T contextual stimuli of its units not only threw down the gauntlet that science learning was intended to be affective as well as cognitive but also provided new theoretical and methodological directions for research on interest in science (Krapp and Prenzyl 2011).

The contextual- and domain-related information of the test units act as concrete and illustrative stimuli so that students have a clear idea of the science about which their interest is being questioned, in stark contrast with so many attitudinal studies that use generic words like "science" and "technology". Methodologically, Drechsel et al. (2011) showed how multidimensional item response models can be applied to the embedded responses to disentangle students' interest in science. They illustrate their approach in terms of the students' differences in interest in living systems compared with physical and technological systems, which revealed interesting gender differences, and in relation to the measures of science achievement itself. Kjærnsli and Lie (2011) made use of the same division in interest and employed regression and correlation analysis to link this division with the students' interest in future science careers, across groups of culturally similar countries.

The cross-national character of the PISA data also enabled the recognised role that culture has in a student's interest in science to be much more thoroughly analysed and discussed (Ainley and Ainley 2011; Basl 2011; Olsen and Lie 2011). Finally, gender differences in the students' responses to the embedded items were found by Buccheri et al. (2011) to relate to their preferences for future careers.

Conclusion

The presence of TIMSS and PISA, as external assessments of what is happening nationally in science education, has challenged countries to face a number of questions about the nature and quality of their current policy and practices and what can be done to improve these. It needs to be recognised, as Baker (1997) cogently pointed out after the first TIMSS report, that the answers to these questions do not lie in these projects. Rather, the data the projects produce should be seen as opportunities for country authorities to ask and reflect on interesting questions about their own familiar practices. Each country needs to respond to these questions in their own context and in their own way.

The projects' challenges about the nature and quality of science teaching and learning are posed by the findings about science learning that stem from their testings. The fact that the common standards that are the basis of each project's tests are very different—TIMSS, a Vision I view of scientific literacy, and PISA, a Vision II view—means that countries are challenged about these alternatives. Some seem to see it as a choice of one or other set of science standards and hence one or other view of scientific literacy. Rather it should be clear that since each project only explores a single perspective of what it means to learn science, neither perception can possibly be the model for school science education for the digital years of the twenty-first century ahead. Science itself is changing and this means its needs for future professionals differ from those required in the single disciplinary pathways that were so powerful in the last century. The ways in which science and technology are influencing modern society are also changing, not least because so many critical issues are demanding political responses before the science is understood in the experimental manner of the physical sciences that provided so much confidence and authority in the past. The twenty-first-century society requires more, not fewer, scientifically informed citizens who are able to make personal and social decisions about these issues, and these as discussed earlier go well beyond the science that underpins either TIMSS or PISA. Their science bases retain importance, but science education will need to encourage some quite different science learning intentions as well and adjust its practice to give them adequate attention.

One danger that must be avoided is the allure that these projects offer to curriculum authorities with their high quality items and one test measures. Hopman (2008) has drawn attention to the particular attraction that such measures have when education across the world is increasingly seen in terms of markets, competition and accountability. The science education research community needs to accelerate its efforts to demonstrate that some of the these different science learnings can also be assessed authentically and responsibly (see Fensham and Rennie, this volume)

A second set of challenges come from the model both projects have used to measure and discuss the societal conditions that may relate to science performance and hence that may be manipulated to bring about improvement. The model is a positivistic one that defines these conditions or social context in which science education occurs in terms of a set of social constructs that are assumed to have common meaning across the participating countries. Sets of items are then devised that will provide measures for these constructs that can then be compared cross nationally. The findings are discussed in the projects' reports as having policy implications. Thus, it is suggested that a country's relative science performance could be changed by altering one or more of these social constructs.

In the case of 20 years of TIMSS and a decade of PISA, there is scant evidence that such simplistic social manipulation is possible or effective. These social constructs are not simply there to frame school science education. Together they make up a much more complex web of society that has been conditioned by each country's unique history and accordingly is not simple to alter. An example of a social construct that cannot be simply altered is the student's SES background. It has been shown in many countries to correlate significantly with science achievement. But the nature of this relationship is complex, not simple. Finland and South Korea have low SES slopes but for quite different reasons. Australia and Canada, on the face very similar countries, have significantly different SES slopes. Since this slope relationship was first reported, no country has yet altered it substantially.

There is now much anecdotal and reported evidence that participating countries interpret the projects' findings in their own idiosyncratic way and respond to them, usually in minor ways within education itself, rather than by the wider social revolution that would be required to change the context of education or of society. These responses are consistent with what I have argued in detail elsewhere that the questions about both the quality of science education and its improvement are, in reality, related to a holistic sense of national and educational culture than is implied in the contextual model the OECD or the IEA have been using (Fensham 2007).

References

ACARA. (2011). *Science: Rationale* (p. 1). http://www.australiancurriculum.edu.au/Science/rationale. Accessed 1 Sept 2011.

Achieve Inc. (2010). *Taking the lead in science education: Forging next-generation science standards.* Washington, DC: Author. www.achieve.org/international-science-benchmarking-report. Accessed 1 Sept 2011.

Ainley, M., & Ainley, J. (2011). A cultural perspective on the structure of student interest in science. *International Journal of Science Education, 33*(1), 51–71.

Baker, D. P. (1997). Surviving TIMSS, or everything you have forgotten about international comparisons. *Phi Delta Kappa, 79*(4), 295–300.

Basl, J. (2011). Effect of school on interest in natural sciences: A comparison of the Czech Republic, Germany, Finland and Norway based on PISA 2006. *International Journal of Science Education, 33*(1), 145–157.

Black, P., Aitkin, M., & Pevsner, D. (1995). *Changing the subject: Innovation and change in science, mathematics and technology education*. New York: Routledge.

Buccheri, G., Gürber, N. A., & Brühwiler, C. (2011). The impact of gender on interest in science topics and the choice of scientific and technical vocations. *International Journal of Science Education, 33*(1), 159–178.

Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practice*. Portsmouth: Heinemann.

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. 303–346). Munster: Waxmann.

De Boer, G. (2011). The globalisation of science education. *Journal of Research in Science Teaching, 48*(6), 567–591.

Department of Education NSW. (2012). *Essential Secondary Science Assessment*. http://www.schools.nsw.edu.au/learning/7-12assessments/essa/consultation.php, http://www.schools.nsw.edu.au/learning/7-12assessments/essa/index.php. Accessed 11 Oct 2012.

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). Munster: Waxmann.

Drechsel, B., Carstensen, C., & Prenzyl, M. (2011). The role of content and context in PISA interest scales: A study of embedded interest items in the PISA 2006 science assessment. *International Journal of Science Education, 33*(1), 73–95.

Fensham, P. J. (1995). One step forward…. *Australian Science Teachers Journal, 41*(4), 24–29.

Fensham, P. J. (2007). Context or culture: Can TIMSS and PISA teach us about what determines educational achievement in science? In B. Atweh, A. C. Barton, M. C. Borba, N. Gough, C. Keitel, C. Vistro-Yu, & R. Vithal (Eds.), *Internationalisation and globalization in mathematics and science education* (pp. 151–172). Dordrecht: Springer.

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.

Fensham, P. J. (2011). Globalisation of science education: Comment and commentary. *Journal of Research in Science Teaching, 48*(6), 698–707.

Fischler, H. (2011). Didaktik—An appropriate framework for the professional work of science teachers. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching* (pp. 31–50). Dordrecht: Springer.

Garden, R. A., & Orpwood, G. (1996). TIMSS test development. In M. O. Martin & D. L. Kelly (Eds.), *Third international mathematics and science study technical report* (Vol. 1). Chestnut Hill: Boston College.

Hopman, S. T. (2008). No child, no school, no state left behind: Schooling in the age of accountability. *Journal of Curriculum Studies, 40*(4), 417–456.

Kjærnsli, M., & Lie, S. (2011). Students' preferences for science careers: International comparisons based on PISA. *International Journal of Science Education, 33*(1), 121–144.

Krapp, A., & Prenzyl, M. (2011). Research on interest in science: Theories, methods and findings. *International Journal of Science Education, 33*(1), 27–50.

Leibnitz Institute for Science education. (2009). *PISA Research Conference.* Kiel, September 2009. www.pisaresconf09.org. Accessed 1 May 2012.

McCrae, B. J. (2009). PISA 2006 test development and design. In R. W. Bybee & B. J. McCrae (Eds.), *PISA Science 2006: Implications for science teachers and teaching* (pp. 27–38). Arlington: NSTA.

Mullis, I. V. S., Martin, M. D., Smith, T. A., Garden, R. A., Gregory, K. D., Gonzalez, E. J., Chrostowski, S. J., & O'Connor, K. M. (2003). *TIMSS assessment frameworks and specifications*. Chestnut Hill: I.E.A.

Nentwig, P., Roennebeck, S., Schoeps, K., Rumann, S., & Carstensen, C. (2009). Performance and levels of contextualisation in a selection of OECD countries in PISA 2006. *Journal of Research in Science Teaching, 46*(8), 897–908.

OECD. (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: OECD.
OECD. (2009). *PISA 2006 technical report*. Paris: OECD.
Olsen, R. V., & Lie, S. (2011). Profiles of students' interests in science issues around the world: Analysis of data from PISA 2006. *International Journal of Science Education, 33*(1), 97–120.
Ramseier, S. J. (2001). Scientific literacy of upper secondary students: A Swiss perspective. *Studies in Educational Evaluation, 27*(1), 47–64.
Roberts, D. (2007). Scientific literacy/scientific literacy. In S. K. Abel & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 125–171). Mahwah: Lawrence Erlbaum Associates.
Roth, K. J., Druker, S. L., Garnier, H. E., Lemmens, M., Chen, C., Kawanka, T., et al. (2006). *Teaching science in seven countries. Results from the TIMSS video study*. Washington, DC: National Center for Educational Statistics.
Sadler, T. D., & Zeidler, D. L. (2009). Scientific literacy, PISA, and socioscientific discourse: Assessment for progressive aims of science education. *Journal of Research in Science Teaching, 46*(8), 909–921.
Schreiner, C., & Sjøberg, S. (2007). Science education and youth's identity construction – two incompatible projects? In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *Re-emergence of values in science education* (pp. 231–247). Rotterdam: Sense.
Shirley, D. (2008). The coming of post-standardisation in science education: What role for the German Didaktik tradition? *Zeitschrift für Erziehungswissenschaft, 10*(9), 35–45.
Sjøberg, S. (2004). Science and technology in the new millennium—Friend or foe? In *Proceedings of the 11th IOSTE Symposium*, Lublin, July 25–30.
Solomon, J., & Aikenhead, G. (Eds.). (1994). *STS education: International perspectives on reform*. New York: Teachers College.
Valverde, G. A., & Schmidt, W. H. (1998). 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. (Eds.). (2007). *Making it comparable: Standards in science education*. Munster: Waxmann.
Yip, D. Y., Chiu, M. M., & Ho, E. S. (2004). Hong Kong student achievement in OECD-PISA study: Gender differences in science content, literacy skills, and test item formats. *International Journal of Science Education, 26*(1), 91–106.

Chapter 3
International, National and Classroom Assessment: Potent Factors in Shaping What Counts in School Science

Alister Jones and Cathy Buntting

Introduction

In this chapter, we examine the potential influences of international, national and classroom assessment programmes on what counts in school science. To do this, we propose a multilayered framework that considers the interactions between levels of assessment (international, national and classroom) and how science is articulated at the national, school and classroom levels. International and national science assessments as well as broader assessment policies will therefore be considered in relation to their potential influences on school science. Our aim is to illustrate, drawing on the New Zealand context, how assessment policies and procedures might act to broaden—but more often, constrain—what counts in school science.

For example, two large international assessments—Trends in Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA)—may impact on school science in different ways. The close adherence of TIMSS to curriculum content common across participating countries, and government desires for high achievement in TIMSS, means that a strong content focus needs to be retained in primary classroom teaching. On the other hand, PISA is designed to assess a broader range of students' understandings of and about science and has the potential to broaden what gets emphasised within science curricula and therefore what counts as secondary school science.

Nationally, high-stakes assessments and exit qualifications at the secondary level indicate very directly what is valued and therefore what is emphasised in the classroom as science. While this has the potential to create opportunities for innovation,

A. Jones (✉) • C. Buntting
Faculty of Education, University of Waikato, Gate 1 Knighton Road,
Hamilton 3240, New Zealand
e-mail: a.jones@waikato.ac.nz; BUNTTING@waikato.ac.nz

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_3,

it more often constrains what and how science is taught. The use of exit assessments in determining university entrance presents an additional constraint. At the primary level, national initiatives, such as a focus on 'key competencies', and numeracy and literacy strategies have de-emphasised science in favour of more process-orientated outcomes. In addition, the last decade has seen an increased emphasis on assessment for learning practices (Black and Wiliam 1998) at both the primary and secondary level. In this process, increased attention has been paid to effective pedagogies such as feedback and negotiated and agreed learning intentions. However, particularly in the primary school level where teacher knowledge in science is often somewhat fragile, there can be emphasis on the pedagogy rather than the science ideas (Cowie et al. 2008).

Here, we consider how and why international and national assessment initiatives such as those outlined above have broadened or constrained what counts in New Zealand school science, both at the macro (policy-curriculum) level and at the level of the classroom. Both primary and secondary schooling will be considered.

International Assessments and What Counts in School Science

New Zealand, like many countries, currently invests in two international studies in science: TIMSS administered under the auspices of the International Association for the Evaluation of Educational Achievement (IEA) and PISA set up by the Organisation for Economic Co-operation and Development (OECD). These studies provide researchers, policymakers and practitioners with rich comparative data and are a potentially valuable resource for educators interested in seeking ways to understand student learning in science with the aim of enhancing both what is learnt and how it is learnt (see Fensham, this volume).

In spite of the rich potential made available through these large studies, in New Zealand and elsewhere, there appear to be few attempts at secondary analysis to provide wider insights into how and what science learning is occurring and why. Instead, the data are too often captured for political rather than educational purposes, politicians and the media using the findings to either celebrate or castigate curriculum initiatives and teachers. On a more positive note, country rankings may also direct resourcing towards strategies aimed at enhancing science education. For example, New Zealand's disappointing TIMSS rankings in the 1990s resulted in the production of the 'Building science concepts' series, a set of 64 books published by the Ministry of Education to support primary teachers in their teaching of science.

Trends in Mathematics and Science Study (TIMSS)

TIMSS measures trends in mathematics and science achievement at the fourth and eighth grades (10- and 14-year-olds) and also claims to be 'monitoring curricular implementation and identifying the most promising instructional practices from around the world' (Martin et al. 2008, p. 14). In addition to assessing students' conceptual knowledge in the areas of mathematics and science on a four-yearly cycle, a rich array of background information is collected related to the intended, implemented and attained curriculum (Robitaille and Garden 1996). This includes an extensive curriculum mapping exercise used to align the degree of 'fit' between the TIMSS test items and each country's intended curriculum.

Items in the Science Achievement Test are constrained to content largely common across the participating countries' curricula. Unfortunately, this deliberate decision removes possibilities for identifying how more isolated curriculum innovations have fared (Fensham, this volume). Furthermore, evidence available in the reports and published items indicates a narrowing in the content of the assessed items—in contrast to intentions worldwide to broaden the intended curriculum. For example, 1998 subscales included earth science, life science, physics, chemistry, environment and resources, and scientific inquiry and the nature of science. In 2003 these were reduced to earth science, life science and physical science in Grade 4 and earth science, life science, chemistry, physics and environmental science in Grade 8. In addition, hands-on tasks administered to a subsample of students in the 1994/1995 cycle were subsequently dropped.

TIMSS, through its choice of assessment items both in terms of content and delivery, therefore places a strong emphasis on 'what science is' and defining scientific knowledge to be tested rather than investigating more closely what many countries are trying to achieve through their science curricula, that is, a broader notion of scientific literacy. According to TIMSS, therefore, what counts in science is the conceptual frame; attributes such as engagement, relevance and an inquiry stance remain largely unarticulated.

Grade 4 TIMSS Achievement in New Zealand

In New Zealand, 2007 achievement at Grade 4 level was on a par with the TIMSS scale average, with NZ ranking 22nd out of 36 countries (compared with 12th out of 25 for Grade 4 in 2003). Systems-level information collected by TIMSS is telling with regard to both the time spent teaching science and teacher education and ongoing professional learning. For example, at Grade 4 level there was a significant decrease in the amount of time spent teaching science each week between 2003 and 2007, with an average of 45 h over the year (down from 66 and compared with the 2007 international average of 67 h; see Martin et al. 2008, p. 206). Although teachers were generally well qualified compared to many of the other participating countries, only 10 % of the 2007 TIMSS students were taught by a teacher with primary education qualifications *and* a major in science (compared with the international average of 24 %). In addition, students' teachers were more likely to have participated in professional development targeting improving students' critical thinking or inquiry skills (47 vs. 33 % internationally) than in science content (14 vs. 34 %), science pedagogy (12 vs. 35 %), science curriculum (17 vs. 31 %), integrating information technology into science (19 vs. 24 %) and science assessment (11 vs. 28 %)

(Martin et al. 2008, p. 270). This appears to echo New Zealand's interpretation of science as inquiry, as evident in curriculum documentation, although it is notable that despite the greater focus on critical thinking and inquiry in teacher development, students did not achieve better than the international averages in either the 'applying' or 'reasoning' cognitive domains.

Taken together, the disappointing performance of New Zealand students in the 2007 Grade 4 TIMSS assessment, the findings associated with teacher education and professional development and the paltry amount of time actually spent teaching science suggest that what counts in New Zealand primary science classrooms is actually very little. Put another way, *science* counts very little despite its place as one of eight key learning areas in the national school curriculum and attempts by successive governments to address the resourcing needs to change this. As already indicated, for example, one of the outcomes of New Zealand's lower performance in the 1990s was the commissioning of the 'Building science concepts' series. While valued and used by teachers (Cowie et al. 2004), the series nevertheless emphasises concept development rather than wider aspects of scientific literacy. This suggests that the focus in primary science—as signalled by nationally developed resources—is aligned with the content outcomes measured by TIMSS. An alternative approach, proposed by Bull et al. (2010), is to place emphasis on stimulating students' interest and curiosity and on developing literacy skills.

Grade 8 TIMSS Achievement in New Zealand

In New Zealand, the Grade 8 TIMSS test is administered in the first year of secondary school. Unfortunately no data are available for 2007 as the government chose to administer only PISA in that year rather than run two international tests at the secondary level in a small country. However, 2003 TIMSS results place New Zealand significantly higher than the international average, with a ranking of 13 out of 45 countries. Chemistry was the weakest content area, followed by physics. This is in contrast to the percentage of students taught the content covered in TIMSS, with 59 % reportedly having been taught the chemistry topics, 48 % physics, 46 % life science and 36 % earth science (Martin et al. 2004, p. 210). In terms of attitudes to science, both students' self-confidence (41 % had high self-confidence) and valuing of science (40 %) were lower than the international averages (48 and 57 %, respectively), although their actual achievement was higher (520 vs. 474) (Martin et al. 2004, pp. 36, 160, & 166). Students' low valuing of science but high achievement was also observed in countries like Chinese Taipei, Republic of Korea, the United States and Australia. The proportion of students indicating high levels of enjoyment in learning science was also lower than the international average. Although it is pleasing that this had increased significantly from 1999—as had the international average—the overall low level of enjoyment is concerning because of the longer-term impacts on lifelong learning, scientific literacy and informed citizenship (Gluckman 2011).

Unlike at the primary level, secondary teachers had generally undertaken teacher education specialising in a science content area, and 90 % of students had been taught by such a teacher. In addition, teachers tended to participate in science-related professional development. For example, 84 % of students' teachers had undertaken professional development in science assessment (compared with 47 % internationally), 79 % in the science curriculum (cf. 52 %), 72 % in science content (cf. 58 %), and 52 % in integrating ICT into science (cf. 45 %).

Thus, the 2003 TIMSS data at the Grade 8 level are overall more positive than at the Grade 4 level in both 2003 and 2007, most likely reflecting the greater emphasis on science as a specialist subject in New Zealand secondary schools in terms of both time allocated to its teaching and teachers' more extensive science pedagogical content knowledge (Shulman 1987). The positive overall placing of New Zealand compared to other countries suggests that, on curriculum-derived items, Grade 8 students have a robust conceptual knowledge. What counts in science—at least at the level of the attained curriculum—therefore appears to be an emphasis on the conceptual domain. Monitoring the impact of curriculum shifts towards greater relevance and an emphasis on the nature of science (Ministry of Education 2007b) is of considerable interest but is unlikely to be uncovered by TIMSS in its current format. We are also concerned that a continued desire by the government to continue to do well in TIMSS will potentially limit the interpretation of the curriculum to a focus on content. In addition, students' low positive attitudes to and enjoyment of science are alarming because of the potential negative impact on long-term engagement beyond the compulsory school years. This is unlikely to be addressed by an atomistic approach to content presentation.

Programme for International Student Assessment (PISA)

In contrast to TIMSS, which is designed to link with school science curricula, PISA seeks to investigate scientific literacy in a much more general sense, assessing how well education systems are preparing 15-year-olds for life in twenty-first century society. This position is much more in line with widespread international interest about the public understanding of science (Fensham 2007). It also offers, as Bybee and McCrae (2011) argue, 'a unique perspective for the science education community. Most assessments look back at what students were expected to learn and whether they attained the knowledge and skills … PISA's perspective looks ahead…' (p. 8).

In 2006, PISA expanded its 2000 definition of scientific literacy—'the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity' (OECD 1999)—to include four interrelated features that include an individual's:

– Scientific knowledge and ability to use that knowledge to identify questions, acquire new knowledge, explain scientific phenomena, and draw evidence-based conclusions about science-related issues

- Understanding of the characteristic features of science as a form of human knowledge and inquiry
- Awareness of how science and technology shape our material, intellectual, and cultural environments
- Willingness to engage in science-related issues, and with the ideas of science, as a constructive, concerned, and reflective citizen (OECD 2006)

In line with this, PISA's assessment framework has four dimensions. Firstly, each assessment task is situated within a real-world application of science associated with five broad contexts: health, frontiers of science, natural resources, environment and hazards. Secondly, scientific knowledge of and about science is assessed. Thirdly, scientific competencies are considered, including identifying scientific issues, explaining phenomena scientifically, and using scientific evidence. Finally, some of the tasks include questions related to students' attitudes to the issues being tested.

Fensham (2007) points out that it is the valuing of a more coherent set of understandings applied to scientific situations of relevance in the students' lifeworlds that is a key difference between PISA and TIMSS, which tends to reward broad but superficial scientific knowledge. However, as Harlen (2001) observes:

> [It] is not that the concepts involved [in PISA] would not have been taught, since these do not go beyond the widely accepted core of the curriculum, but that [students] are required to combine this knowledge with thinking processes which reflect an approach to education going beyond the knowledge of concepts and principles. (p. 86)

New Zealand Achievement in PISA

In comparison with the relatively low TIMSS ranking of New Zealand compared with top-performing countries, our 15-year-olds achieved remarkably well in PISA science in both 2006, when science was the major domain, and again in 2009 when it was a minor domain. (The focus of PISA, which is administered in a three-yearly cycle, alternates between reading, mathematics and science as the major domains.) For example, in 2006 New Zealand was ranked 7th out of 57 participating countries and economies and only Finland and Hong Kong-China achieved mean scores that were statistically higher. We also had the second highest number of students in the top-performing categories, and it is possible that this positive outcome may in some part be due to the professional development undertaken by teachers (see the TIMMS data above). However, the large proportion (14 %) of students with a low level of scientific proficiency must not be ignored.

As with TIMSS, 2006 attitudinal data indicate that students were less interested in learning science topics than the OECD average (461 points vs. 500 points), although this is in keeping with the international trend for students in low-performing countries to show relatively high levels of interest in science and students in high-achieving countries showing relatively lower levels of interest (Bybee and McCrae 2011). However, within New Zealand—as within nearly all other countries—students with higher engagement in science generally had higher achievement in science than those with lower engagement. Bybee and McCrae draw on Buckley (2009) in pointing out that this apparent contradiction—that attitude influences achievement within countries but not between them—can be attributed at least in part to cross-national differences in response style.

In 2009, New Zealand again achieved very highly in scientific literacy (a minor domain in PISA), with Finland still the only OECD country with a significantly higher performance. However, there was again a relatively large proportion of students with very low levels of scientific literacy when compared with other top-performing countries, demonstrating the very broad abilities among New Zealand students. It remains unclear whether this reflects a continued classroom emphasis on knowing knowledge rather than knowing how to find the knowledge and then how to use it or whether it is linked to other factors such as students' low engagement in science and/or the PISA assessment tasks. For example, perhaps it is the assessors' choice of contexts in PISA that impact on Māori and Pasifika students' levels of engagement and achievement.

That Māori and Pasifika students are over-represented in the lower tail and under-represented among higher performers when compared with their European/Pākehā and Asian counterparts is concerning. Raising the achievement of these ethnic groups is a specific goal of the education system (Ministry of Education 2007a), and PISA provides a measure of progress towards this. PISA results on within-school variability of student achievement also reinforce the need for initiatives that support teaching and learning of diverse students who are learning within the same school environment rather than assuming that diverse students will be in different schools or classes (Satherley 2006).

In addition, the contextualising of assessment items within PISA means that a high level of reading literacy is required in addition to understandings of and about science. It therefore seems likely that New Zealand's high reading literacy is a contributing factor to our relative success in scientific literacy. For example, in 2006 only three of the 57 participating economies achieved significantly better than New Zealand in reading. Balanced against this is—again— the relatively large proportion of students with very low reading literacy, with 20 % at or below Level 1. (The three top-performing countries in reading literacy—Finland, Korea and Hong Kong-China—had 6–8 % at this level.)

The expanded definition of scientific literacy and the use of contextualising information at the start of each PISA task send a clear signal to policymakers and practitioners about the importance of context and broader notions of the nature of science and the relevance of evidence when considering scientific literacy. This is in line with international rhetoric supporting the use of real-life contexts designed to engage and motivate students in science education by emphasising the relevance of science concepts to everyday living (e.g. Jones, 2012). The role of contexts within assessment items is, however, not as clear, and it seems to us that in some cases the additional information required to establish the context in PISA items may even act as a distractor. For example, in 'Mary Montagu, the history of vaccination' (see Fig. 3.1), the released questions require an understanding of vaccines, and the information in the newspaper article seems superfluous. This item would likely have been classified by Nentwig et al. (2009) as having a low level of contextualisation

Read the following newspaper article and answer the questions that follow.

The history of vaccination

Mary Montagu was a beautiful woman. She survived an attack of smallpox in 1715 but she was left covered with scars. While living in Turkey in 1717, she observed a method called inoculation that was commonly used there. This treatment involved scratching a weak type of smallpox virus into the skin of healthy young people who then became sick, but in most cases only with a mild form of the disease.

Mary Montagu was so convinced of the safety of these inoculations that she allowed her son and daughter to be inoculated.

In 1796, Edward Jenner used inoculations of a related disease, cowpox, to produce antibodies against smallpox. Compared with the inoculation of smallpox, this treatment had less side effects and the treated person could not infect others. The treatment became known as vaccination.

QUESTION 2

What kinds of diseases can people be vaccinated against?

A. Inherited diseases like haemophilia.

B. Diseases that are caused by viruses, like polio.

C. Diseases from the malfunctioning of the body, like diabetes.

D. Any sort of disease that has no cure.

QUESTION 3

If animals or humans become sick with an infectious bacterial disease and then recover, the type of bacteria that caused the disease does not usually make them sick again.

What is the reason for this?

A. The body has killed all bacteria that may cause the same kind of disease.

B. The body has made antibodies that kill this type of bacteria before they multiply.

C. The red blood cells kill all bacteria that may cause the same kind of disease.

D. The red blood cells capture and get rid of this type of bacteria from the body.

QUESTION 4

Give one reason why it is recommended that young children and old people, in particular, should be vaccinated against influenza (flu).

Fig. 3.1 Mary Montague. A 2009 PISA assessment item (OECD 2010, pp. 143–144)

in that the additional text is irrelevant to answering the questions. However, this does not weaken the case for using narrative and real-life examples within a teaching and learning programme.

We propose that contexts may be loosely or tightly coupled to the science content and that tight coupling is the aim if the contexts are to effectively support students in making meaning of concepts of and about science. Thus, the story of Mary Montagu is only of educational value if the teaching and learning of immunity and vaccinations is tightly linked to this example and if the narrative is used as the basis for exploring relevant scientific concepts, principles and ways of thinking. How to authentically assess understandings of both the science *and* the context needs to be given careful consideration (Rennie and Fensham, this volume), but it seems to us that additional contextual information is only worthwhile in assessment if it is closely coupled to the content, that is, it provides information that must be extracted in order to answer the questions or to stimulate further thinking and analysis. Contexts that are loosely coupled to content, for example, to demonstrate a simple application of a concept or to engage students in the content, also have educational value when used in teaching and learning programmes, but their place in assessment items is far more contentious. Nevertheless, the use of even loose coupling in PISA does convey a strong message about the usefulness of contextual approaches in teaching for—but not necessarily assessing—broader notions of scientific literacy.

Despite the potential for PISA to influence classroom practice through its signalling of what science learning is important, much is lost by maintaining the secrecy of a large number of items so they can be used in the next round of testing. In adopting this approach, Fensham (2007) argues that PISA—and TIMSS—denies to curriculum authorities and teachers the 'most immediate feedback the projects could make, namely, the release in detail of the items, that would indicate better than framework statements, what each means by "science learning"' (p. 217).

National Assessment Programmes and What Counts in School Science

Two New Zealand assessment projects—the National Education Monitoring Project (NEMP) and the Assessment Resource Banks (ARBs)—provide a seemingly more immediate influence on classroom practice than TIMSS and PISA are likely to achieve. In addition, broader assessment policy related to numeracy and literacy initiatives as well as an increased emphasis on 'key competencies' impacts significantly on school timetabling and emphases, including what counts in science, particularly at the primary level. At the senior secondary school level, the National Certificate in Educational Attainment (NCEA) at Years 11, 12 and 13—and its links with university entrance criteria—dominates the way science programmes are designed and delivered. The influence of each of these policies on what counts in school science is considered in turn.

National Education Monitoring Project (NEMP)

NEMP commenced in 1993 at the directive of the New Zealand Ministry of Education and is tasked with assessing and reporting on the achievement of Year 4 and Year 8 students in all areas of the school curriculum on a four-yearly cycle. The aim is to provide information that 'allows successes to be celebrated and priorities for curriculum change and teacher development to be debated more effectively' (Crooks and Flockton 2004, p. 5). Assessment is undertaken with a 2.5 % random national sample, with students assessed in their schools by teachers specifically seconded and trained. This creates a significant opportunity for professional development of primary teachers, with many having reported major changes in their teaching and assessment practices (Crooks and Flockton 2004).

Because NEMP is not constrained by the need for cross-country comparisons, a relatively sophisticated administration has been developed with task instructions being given orally by teacher facilitators, through video presentations, on laptop computers, or in writing, reducing the reading and writing required by wholly paper-and-pencil assessments. Many of the assessment tasks also involve the use of equipment and supplies. Such a varied approach allows for the inclusion of tasks that interest students so that results are more likely to represent their capabilities rather than their motivation.

Like PISA, NEMP uses everyday contexts as the central organising theme for tasks, although these are illustrative rather than rich and there is generally a straightforward relationship between the context and the concept or variable. However, the contexts can be displayed in a range of different forms and appear to be consistently more tightly coupled to the questions, as defined above. For example, in an item investigating students' understandings of motion and resistance to motion, a video clip is displayed showing a person riding a scooter on different surfaces. Questions asked by the facilitator include the following:

1. What surface was the most difficult for this person to ride on?
2. Why do you think that was the most difficult surface?
3. How could you tell the sand was the most difficult to ride on?
4. What was the easiest surface for the person to ride on?
5. How could you tell that surface was the easiest to ride on? (Crooks and Flockton 2004, p. 33)

Because NEMP is a national monitoring project focusing on identifying shifts in achievement from Year 4 to Year 8, and within year levels over time, student responses are reported for each question without providing averages across all items such as is reported by TIMSS and PISA. Our analysis suggests that Year 8 students are more likely to identify a greater range of variables in each context, as well as provide more detailed explanations of those variables. However, Year 4 students were generally more positive than Year 8 s about doing science at school. While this pattern is common internationally, with older students generally more discerning and critical, and also more realistic about their own abilities, it is concerning that the percentage of Year 8 students enjoying science at school dropped from 37 % to 24 % from 1999 to 2007.

A further difference between NEMP and TIMSS or PISA is that the level of analysis is more discriminatory, with greater reporting of the complexities of student thinking. There is also a strong emphasis on subgroup analysis for each task, and tasks can be identified where Pākehā/European and Asian students tend to perform significantly better than Māori and Pasifika students. This level of analysis therefore becomes a useful guide to teachers. In addition, the tasks demonstrate not only what counts in science in terms of key ideas but also how contexts might be tightly coupled to the teaching.

Many of the released NEMP tasks are accessed by teachers for use in informing and modifying classroom teaching (Cowie et al. 2004), including using the tasks as models of good practice (Crooks and Flockton 2004). This is most likely because the tasks closely reflect curriculum aims and are designed to replicate much more closely what might happen in a normal classroom environment rather than a testing situation, making them suitable for more general use. However, Gilmore (2001) noted the challenge of helping teachers to 'build bridges' between existing NEMP tasks and their actual classroom pedagogy. One effort to address this was a project by Hipkins and Kenneally (2003) in which they analysed 200 taped episodes of student groups carrying out NEMP investigation tasks. This, coupled with a review of the literature, was used to develop and trial strategies for the active teaching of investigative skills. Hipkins and Kenneally subsequently recommended to the NEMP Board: 'It would help teachers to think more widely about types of investigations if NEMP investigation tasks modelled these' (p. 86), clearly indicating the possible links between NEMP and what counts in the science classroom. This is demonstrated even more powerfully in the ARBs, where extensive effort is directed towards developing informative teacher guides as outlined below.

Assessment Resource Banks (ARBs)

The ARBs are large collections of science, mathematics and English assessment resources made available since 1997 by the New Zealand Council for Educational Research under contract to the Ministry of Education. Catering for Years 3–10 (7–14-year-olds) they offer a substantial resource for New Zealand teachers. Although initially designed to assist in the development of classroom and school-wide assessment, the emphasis of the ARBs since 2003 has been on supporting formative assessment, which is viewed as any task or interaction where 'the information gained is used to inform what happens next in the classroom' (Joyce and Darr 2008, pp. 3–4). Many of the resources therefore include support for teaching and learning discussions and in some cases examples of students' work. For example, the focus of the assessment is clearly described, and relevant background knowledge, diagnostic information (including misconceptions, gathered from research literature and student trials), and next learning steps tied to particular responses are provided. Each item is designed to be carried out as part of normal classroom activities to help teachers make sense of what students are saying, doing and thinking; to make decisions about what to do next; and to help students to reflect on their learning as they are learning.

Many of the ARBs items are drawn from PISA, TIMSS and NEMP—illustrating one route by which international and national assessment projects can make their way into classroom practice. There is also an extensive selection of resources associated with the 'Building science concepts' series developed in response to the early TIMSS rankings (see above). As with NEMP, the contexts are tightly coupled to the questions. For example, in 'Ideas about rolling' Year 4 (8-year-olds) students are asked to choose which toy vehicle will roll further in different conditions (see Fig. 3.2). The teacher notes recommend looking for explanations that include reference to the slope of the ramps, the floor surfaces, and/or the type of vehicle (including the number of wheels, condition of tyres, aerodynamics of shape, weight and size). Links are provided to other ARB tasks focusing on fair testing and investigating in

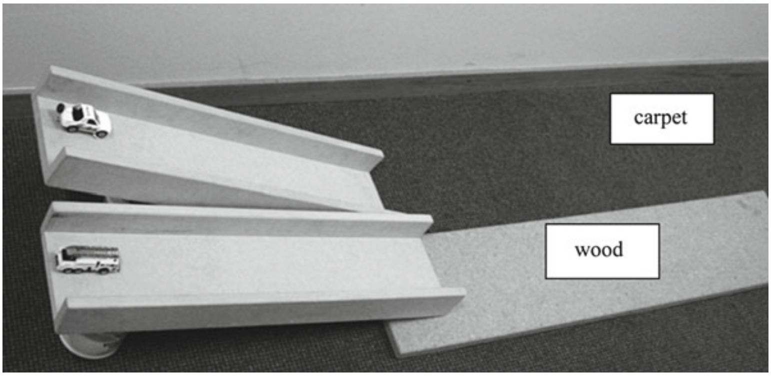


Fig. 3.2 Assessment Resource Bank (ARB) task: ideas about rolling. © Crown, 2010. Used with permission'

science, and appropriate pedagogical strategies are suggested. It is pointed out that a key feature of science explanations is the focus on improving current explanations as we gather more evidence. Examples of everyday versus scientific vocabulary used by students who trialled the resource are provided, and teachers are encouraged to focus on helping students to develop simple explanations using *their* terminology rather than worrying too much about abstract ideas about forces. It is also recommended that teachers ask questions that encourage students to think further about what they notice and what they might want to try next.

The teacher notes therefore go a considerable way towards demonstrating what counts in science when using particular contexts with particular year groups. Although it is not known how many New Zealand teachers actually access the ARBs, it does not seem surprising that a recent survey of users (Dingle and Joyce 2011) found that over 80 % were positive about the tasks and considered the teacher pages to be useful.

Broader Assessment Policies

New Zealand government initiatives such as the formal identification of five key competencies—'capabilities for living and lifelong learning' (Ministry of Education 2007b, p. 12)—in the national curriculum and the emphasis on numeracy and literacy strategies have tended to reduce the time spent on science, particularly in primary schools, and de-emphasised science content in favour of more process-orientated outcomes.

For example, both primary and secondary schools are now specifically required to consider ways to enhance the development of students' key competencies—thinking; using language, symbols and texts; managing self; relating to others; and participating and contributing. One of the significant issues raised by this is how to assess these competencies and indeed whether this needs to be done (Hipkins 2007).

Closely associated with this is teacher concern about 'how they might rationalise 'content', given the need to make space for new types of curriculum goals such as the focus on learning to learn, or on development of the key competencies' (Cowie et al. 2009, p. 13). There is concern, therefore, that an emphasis on key competencies may result in reduced time teaching traditional content. When considering this, we see some irony in the observation that the basis for the key competencies is the OECD's Definitions and Selection of Competencies (DeSeCo) Project—the same project that was used as a framework in developing PISA to assess competencies in reading, mathematics and science (OECD 2005). Despite the potential for synergy between emphasising the key competencies and meeting science education aims as articulated by PISA (and aligned with the New Zealand science curriculum), in New Zealand and elsewhere, the two objectives may be viewed at the school level as being competitive in terms of priorities rather than synergistic.

At the same time that primary school teachers are having to contend with the emphasis on key competencies in the 2007 *New Zealand Curriculum*, national standards for literacy and numeracy have been mandated. The controversial legislation requires that every primary-aged student is assessed against these standards several times per year, with regular reporting to parents about how their child is doing compared to the national standards and other children their age. While ostensibly introduced to help improve learning outcomes for all students, but particularly those who are currently leaving school without qualifications, the policy 'constitutes a major break from current practice in New Zealand, and needs to be implemented with care and consideration to both intended and possible unintended consequences' (Crooks et al. 2009, p. 1). Of particular relevance to this chapter is the fact that the requirements of the standards send a strong message to schools and the public about what is valued in education, making it important that the standards do not undermine the provision of a balanced curriculum. This might be difficult to achieve, given the requirement for class and school reports by school Boards of Trustees and the Ministry of Education and the potential for the standards to become the most significant criterion by which teaching quality and school reputation is judged. In addition, we have already seen shifts in funding away from professional development in learning areas such as science in order to focus on literacy and numeracy. In fact, science gets very little mention in current professional development strategies in spite of a call by the Prime Minister's Science Advisor for increased engagement in science education (Gluckman 2011). Herein lies a significant policy—and funding—disconnect.

In addition, the introduction of key competencies and the literacy and numeracy standards occurred when primary teachers were already feeling the pressures of what they consider to be an overcrowded curriculum, with science often viewed as an 'add-on' subject competing for attention. As one teacher reported in an initial interview:

> When you have an overcrowded curriculum, as you do, people will water things down to fit everything in and that's just how it goes. The numeracy project, it's in your face all the time, so it's a 'have-to'. Science is in a green book on the shelf. (Cowie et al. 2008, p. 17)

Primary school teachers are therefore under considerable pressure as generalists to meet curriculum demands, fulfil national reporting requirements and provide a range of valid learning opportunities for students. Science is but one learning area in the midst of many other competing priorities, including those determined by general assessment policies—which also impact on what counts in science. At secondary school, and particularly the senior secondary level, school exit examinations tend to dominate the way programmes are designed and delivered.

National Certificate in Educational Achievement (NCEA)

NCEA is the main secondary school qualification in New Zealand. Introduced in 2002, it is standards based and schools are able to choose from a range of internal and external assessments to measure how well students meet these standards. When a student achieves a standard, they gain a number of credits. A certain number of credits are needed to gain an NCEA certificate, which is awarded at Levels 1, 2 and 3 and which students generally work through in Years 11–13. Work for each standard is judged as being achieved, achieved with merit, or achieved with excellence, and high achievement is recognised at each level by awarding NCEA with Merit or NCEA with Excellence.

Science has an extensive range of standards that can be used to design and assess senior courses, especially when compared with other subjects. However, the fragmentation of the subject into discrete internal and/or external standards can be both a strength and a weakness. As a strength, it allows flexibility in course design. This was 'sold' as providing schools with opportunities to deliver programmes that not only engage students but also reflect the multi- and interdisciplinary nature of modern science, such as biotechnology or forensic science. In addition, the variety of standards allows for greater customisation of individual learning programmes and qualifications that can be tailored to meet the diverse needs of students.

Balanced against the potential for this flexibility to allow expanded notions of what it means to do science in school is the possibility that a particular department—or individual teacher—can design a legitimate science programme where key themes may not be taught because they are not being assessed. Some, of course, would argue that perhaps this does not matter—particularly if the courses foster student engagement. Students, too, may 'work' the system, and a project investigating the impact of NCEA on student motivation (Meyer et al. 2006) found that students were able to deliberately miss parts of a course that they did not like, thought they would do poorly at or thought were too challenging. Students also reported not wanting to participate in external examinations once they had achieved the required number of credits to be awarded the certificate. That far fewer merits and excellences are awarded in physics than in other science subjects may also be driving student subject choices within science, distorting what it means to do science at school (Benson, S., personal communication, June 3, 2011).

Another aspect—which of course plagues all assessment regimes—is the strong influence of high-stakes testing on classroom practice and the curriculum experienced by students. Hume and Coll (2010) found, for example, that extensive use of planning templates and exemplar assessment schedules to evaluate the Level 1 standard ‘Carry out a practical investigation with direction’ meant that ‘what students came to perceive and experience as scientific investigation was the single, linear and unproblematic methodology of fair testing’ (p. 56), with little opportunity to come up with original questions and solutions to experimental design—or even to integrate this with other parts of their science course(s). Moeed (2010) similarly found that the NCEA templates were used to present students repeatedly with one type of investigation, thus promoting a very narrow view of science investigation. However, she points out the tension teachers experience between balancing their reservations about the assessment and their responsibility to help students achieve academic success. In many schools, she suggests teaching, learning and motivation to learn ‘science investigation’ are being overwhelmed by the assessment regime.

University entrance requirements also impinge on the types of science courses schools choose to offer. Herein lies a dichotomy, with university science increasingly reflecting the multi- and interdisciplinary nature of modern science endeavours but university administration continuing to require a large number of credits in discrete, traditional bundles like chemistry and biology in a conservative effort to ‘maintain standards’. The value of this is further questioned by the finding that higher performance at university is more closely related to how well students performed at school rather than to the particular subjects they studied (Engler 2010).

In theory, therefore, the opportunities for flexibility provided by the large number of science assessment standards broaden the scope for developing engaging, context-based science courses. However, insufficient professional development and the ongoing pressures from university entrance requirements have in practice constrained what many schools offer to their students in terms of science.

Assessment for Learning and What Counts in School Science

Assessment for learning practices has gained prominence around the world because of the positive influence on student achievement (Black and Wiliam 1998). This is also true in New Zealand, where two decades ago the primary purpose of assessment in New Zealand schools was defined as being to ‘suggest actions which may be taken to improve the educational development of students and the quality of education programmes’ (Ministerial Working Party on Assessment for Better Learning 1990). It was mentioned explicitly in the first *New Zealand Curriculum Framework* (Ministry of Education 1993) and is emphasised in the revised *New Zealand Curriculum* (Ministry of Education 2007b) and other more recent documents (e.g. Absolum et al. 2009). Even the introduction of national standards in literacy and numeracy described earlier has maintained a focus on the use of professional teacher judgement to draw on a range of appropriate data and is underpinned by principles of assessment for learning (Ministry of Education 2010).

Significant professional development has been provided to primary and secondary teachers since 2002 through the Ministry of Education’s Assess to Learn (AtoL) project, which has offered in-depth learning for teachers in the use of assessment for learning principles. This includes a facilitator meeting with staff in a school to negotiate an aspect of practice to trial in the classroom, joint planning, and facilitator observations and feedback. Teachers and schools report positive and sustainable changes in teaching, learning and assessment processes, practices and systems, with an increased emphasis on giving feedback focused on learning and next steps rather than managerial aspects (Poskitt and Taylor 2008). However, evidence of shifts in student achievement is provided only for reading and writing, literacy being the chosen focus area in the majority of primary schools. At the secondary level, too, reported findings are generic rather than subject specific, and although professional support tended to be provided at the departmental rather than school level, the focus appears to be on general strategies rather than approaches embedded within a particular subject domain facilitated by someone with appropriate subject expertise. This assumption—that assessment for learning practices can be 'learned' as general strategies—is, in our view, problematic. For example, in some of our classroom research, we have noted that some teachers have identified intended learning outcomes with students (as an assessment for learning strategy), but the intended learning outcomes were not science outcomes but rather language or self-management learning outcomes.

This is reinforced by Jones and Moreland (2005) who highlighted the influence of teachers' pedagogical content knowledge (PCK) on their ability to interact formatively with students, emphasising how teachers cannot provide experiences that guide student progress towards the understanding of science ideas if they themselves do not know what the ideas are. A pertinent example is provided by the Interactions in Science and Technology Education (InSiTE) project (Cowie et al. 2008), where a Year 7/8 teacher—who was also the AtoL coordinator in her school—found the dynamic, unpredictable nature of a class discussion on physical and chemical change problematic. For example, while she could confidently explain water freezing as a physical change, using the criterion that the process is reversible, she struggled to respond to students' suggestions, such as a seed growing, a glass breaking, an egg boiling and chocolate melting in your mouth. Cowie et al. reflected:

> This is an example of a primary school teacher who was self-reflective enough to actively improve her own content knowledge prior to teaching, yet her reliance on definition and a restricted set of examples along with her limited experience in applying the concepts restricted her ability to respond to student ideas. Tayla's acknowledged lack of PCK may have been what led her to focus more on activities than ideas. (p. 19)

It needs to be recognised, however, that any science exploration has potential to initiate unexpected questions from students and that responding to these is not unproblematic.

An understanding about the curriculum and its goals, as well as how students might make progress, is important (e.g. Black and Wiliam 1998). Messages about what counts in students' science classroom experiences are therefore dependent on what the teacher focuses on. This, in turn, is influenced by the teacher's views of science and the goals of the school curriculum, as well as his or her understanding of science concepts, science-specific teaching and assessment practices, and how students might progress in their understandings of these concepts. Jones and Moreland (2005) and Moreland and Cowie (2008) demonstrate the use of a subject-specific planner to help teachers articulate the links between intended learning outcomes, learning activities and likely interactions with and between students. They suggest that such an approach helps teachers shift the focus from facilitating classroom activities to providing opportunities for developing particular scientific understandings.

If teachers' understandings of the nature and purpose of the discipline strongly influence their PCK, which in turn influences their formative (and summative) assessment strategies, student learning is likely to be significantly impacted by teachers' views of what counts in science. It is for this reason, we suggest, that signals about the importance of context-based approaches to science education, for example, in PISA, are so critical.

So What *Does* Count in School Science?

The rhetoric internationally highlights science and science education as a key element for building national capability and enhancing knowledge-based economies. However, science education curriculum and policy does not sit in isolation from other educational policies and political decisions. It is therefore important that we do not ignore the interrelatedness of wider educational policy, including assessment policy, with what it means for the operationalisation of school science. This appears to be true both at the macro (curriculum) level and also at the micro level, that is, in the classroom and in individual teacher-student interactions.

In our consideration of New Zealand's investment in international and national science assessments and in generic assessment policies, such as those for literacy and numeracy, we have attempted to illustrate how these might act to broaden—but often constrain—what counts in school science from the perspectives of politicians, curriculum developers, school leaders, teachers and students. To do this we used a multilayered framework for examining the impact of assessments on the construction of science in schools, that is, interactions between levels of assessment (international, national and classroom) with the articulation of science at the curriculum, school and classroom level.

Although international assessments such as TIMSS and PISA provide only snapshots of student performance in science, they have the potential to impact significantly on what counts in school science. Directly, New Zealand teachers can access the reports and/or adapted tasks such as the ARBs. More indirectly, TIMSS and PISA, as international performance indicators, get the attention of politicians seeking to raise their country's ranking. With the strong content focus of TIMSS, however, there is a conflict between wanting to be seen to do well internationally and the national desire to broaden the notion of learning science. Herein lies a policy disconnect: wanting creativity in course design while also wanting to do well in content.

In contrast, PISA's contextual approach is more closely aligned with international and national rhetoric around developing a wider notion of scientific literacy through school science. In an effort to help extend thinking about how this might be achieved, we have referred to contexts as being either tightly or loosely coupled to the content. For example, it did not seem to us that the context provided at the start of some of the released PISA items was essential for students to be able to respond accurately to the questions, and we suggest that in these cases the context is only loosely linked to the content. We do not deny, however, that such a loose coupling may still be useful within a teaching and learning situation, and as such PISA continues to issue a significant signal about the potential of context in terms of both assessment and teaching and learning.

Closer coupling between context and content is evident in two of New Zealand's national assessment programmes, NEMP and the ARBs. NEMP uses a rich approach to assessment, with items subsequently accessed by teachers for both formative and summative purposes. NEMP also offers teachers professional development opportunities through being seconded to act as assessors. The ARBs are essentially summative tasks that are designed to have formative impacts on teaching programmes, and extensive teacher notes provide valuable guidance to teachers about possible 'next steps'. Like NEMP, the ARBs link content tightly with contexts, and both NEMP and the ARBs provide exemplars of how science education might be expanded at the primary and junior secondary levels.

This optimistic outlook needs to be considered, however, in the light of other assessment policy initiatives such as the emphasis on numeracy and literacy, including the introduction of national standards at the primary level. The focus on key competencies in the latest curriculum revision may also have shifted classroom emphasis from conceptual development and cognitive outcomes towards more process-orientated outcomes, such as co-operation and working together for its own sake rather than in order to learn science more effectively. At the senior secondary level, NCEA offers opportunities for designing and assessing innovative, engaging science programmes. Unfortunately, university entrance requirements and lack of teacher professional development and planning time mean that NCEA has in reality more often constrained what schools offer in terms of science programmes.

At the classroom level, formative assessment is heavily influenced by whether the teacher is focused on conceptual or procedural outcomes. New Zealand's Assess to Learn project offers a potent example of generic teacher development that has changed the nature of classroom interactions in the education landscape. However, findings from the InSiTE project (Cowie et al. 2008) highlight the need for professional development that considers formative assessment practices *within science* if teachers are to be supported in enhancing the *science* learning of their students. One way to do this is by promoting programme planning in which teachers are encouraged to articulate their own and students' knowledge of and about science. In addition, a consideration of how real-life contexts, tightly coupled to content, can be integrated in science classroom programmes seems to us to continue to offer a way forward in engaging students and enhancing their science learning.

In conclusion, we have attempted to illustrate the multiple interactions between international, national and classroom assessment and the influences these might have on what counts in school science. The picture is complex, but emerging evidence suggests that these multilayered assessment initiatives have significant impact on the way in which science education is operationalised at the national, school and classroom level, including the nature of the curriculum, classroom approaches and the time spent on science.

References

Absolum, M., Flockton, L., Hattie, J., Hipkins, R., & Reid, I. (2009). *Directions for assessment in New Zealand: Developing students' assessment capabilities*. Wellington: Ministry of Education. http://www.tki.org.nz/r/assessment/research/mainpage/directions/. Accessed 3 June 2011.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Buckley, J. (2009). *Cross-national response styles in international educational assessments: Evidence from PISA 2006*. https://edsurveys.rti.org/PISA/documents/Buckley_PISAresponsestyle.pdf. Accessed 3 June 2011.

Bull, A., Gilbert, J., Barwick, H., Hipkins, R., & Baker, R. (2010). *Inspired by science. A paper commissioned by the Royal Society of New Zealand and the Prime Minister's Chief Science Advisor.* www.nzcer.org.nz/pdfs/inspired-by-science.pdf. Accessed 3 June 2011.

Bybee, R., & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education, 33*(1), 7–26.

Cowie, B., Jones, A., & Harlow, A. (2004). Primary and secondary teachers experiences of science in the New Zealand Curriculum. *NZ Science Teacher, 104*, 29–34.

Cowie, B., Moreland, J., Jones, A., & Otrel-Cass, K. (2008). *The classroom InSiTE project: Understanding classroom interaction to enhance teaching and learning in science and technology in Years 1–8*. Wellington: Teaching and Learning Research Initiative. http://www.tlri.org.nz/assets/A_Project-PDFs/9215-Cowie/9215_finalreport.pdf. Accessed 3 June 2011.

Cowie, B., Hipkins, R., Boyd, S., Bull, A., Keown, P., McGee, C., et al. (2009). *Curriculum implementation exploratory studies. Report to the Ministry of Education*. Wellington: New Zealand Council for Educational Research and Wilf Malcolm Institute of Educational Research, University of Waikato.

Crooks, T., & Flockton, L. (2004). *National Education Monitoring Project. Science assessment results 2003*. Wellington: Ministry of Education.

Crooks, T., Darr, C., Gilmore, A., Hall, C., Hattie, J., Smith, J., & Smith, L. (2009). *Towards defining, assessing and reporting against national standards for literacy and numeracy in New Zealand*. http://www.eshd.canterbury.ac.nz/Documents/Towards%20defining,%20assessing%20and%20reporting%20against%20national%20standards%20for%20literacy%20and%20numeracy%20in%20nz.pdf. Accessed 3 June 2011.

Dingle, R., & Joyce, C. (2011). *ARB survey*. http://arb.nzcer.org.nz/pdf/ARBsurvey.pdf. Accessed 3 June 2011.

Engler, R. (2010). *Are particular school subjects associated with better performance at university?* Wellington: Ministry of Education. http://www.educationcounts.govt.nz/__data/assets/pdf_file/0007/79342/School-subjects-tertiary-performance-28072010.pdf. Accessed 3 June 2011.

Fensham, P. (2007). Values in the measurement of students' science achievement in TIMSS and PISA. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The re-emergence of values in science education* (pp. 215–229). Rotterdam: Sense.

Gilmore, A. (2001). *National Education Monitoring Project: Assessment resources for teachers. Report to the Ministry of Education*. Christchurch: University of Canterbury.
Gluckman, P. (2011). *Looking ahead: Science education for the twenty-first century. A report from the Prime Minister's Chief Science Advisor*. Auckland: Office of the Prime Minister's Science Advisory Committee. http://www.pmcsa.org.nz/wp-content/uploads/2011/03/Looking-ahead-Science-education-for-the-twenty-first-century.pdf. Accessed 3 June 2011.
Harlen, W. (2001). Scientific literacy in the OECD/PISA project. *Studies in Science Education, 36*, 79–103.
Hipkins, R. (2007). *Assessing key competencies: Why would we? How could we?* Wellington: Ministry of Education. http://www.nzcer.org.nz/research/publications/assessing-key-competencies-why-would-we-how-could-we. Accessed 3 June 2011.
Hipkins, R., & Kenneally, N. (2003). *Using NEMP to inform the teaching of science skills. A report to the NEMP board*. Wellington: NZCER/NEMP/Ministry of Education. http://www.nzcer.org.nz/pdfs/12736.pdf. Accessed 3 June 2011.
Hume, A., & Coll, R. (2010). Authentic student inquiry: The mismatch between the intended curriculum and the student-experienced curriculum. *Research in Science & Technological Education, 28*(1), 43–62.
Jones, A. (2012). Technology in science education: Context, contestation, and connection. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 811–821). Dordrecht: Springer.
Jones, A., & Moreland, J. (2005). The importance of pedagogical content knowledge in assessment for learning practices: A case study of a whole school approach. *The Curriculum Journal, 16*(2), 193–206.
Joyce, C., & Darr, C. (2008). *The role of an online repository of assessment tasks and resources: The place of the Assessment Resource Banks*. Paper presented at the 6th conference of the International test Commission: The impact of testing on people and society: Enhancing the value of test use, Liverpool. http://www.nzcer.org.nz/pdfs/16705.pdf. Accessed 1 June 2011.
Martin, M. O., Mullis, I. V. S., Gonazalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international science report. Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill: TIMSS and PIRLS International Study Centre, Boston College.
Martin, M. O., Mullis, I. V. S., & Foy, P. (2008). *TIMSS 2007 international science report. Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill: TIMSS and PIRLS International Study Centre, Boston College.
Meyer, L., McLure, J., Walkey, F., McKenzie, L., & Weir, K. (2006). *The impact of NCEA on student motivation. Final report to the Ministry of Education*. Wellington: Victoria University of Wellington. http://www.educationcounts.govt.nz/__data/assets/pdf_file/0007/29356/student-motivation-report-june-2006.pdf. Accessed 1 June 2011.
Ministerial Working Party on Assessment for Better Learning. (1990). *Tomorrow's standards*. Wellington: Ministry of Education.
Ministry of Education. (1993). *The New Zealand curriculum framework*. Wellington: Learning Media.
Ministry of Education. (2007a). *Statement of intent 2007–2012*. Wellington: Author.
Ministry of Education. (2007b). *The New Zealand curriculum*. Wellington: Learning Media.
Ministry of Education. (2010). *Ministry of Education position paper: Assessment*. Wellington: Author.
Moeed, A. (2010). *Science investigation in New Zealand secondary schools: Exploring the links between learning, motivation and internal assessment in Year 11*. Unpublished doctoral dissertation, Victoria University of Wellington.
Moreland, J., & Cowie, J. (2008). *Supporting teachers to develop their pedagogical content knowledge in service of assessment for learning*. Paper presented at the British Education Research Association Conference, Edinburgh.
Nentwig, P., Roennebeck, S., Schoeps, K., Rumann, S., & Carstensen, C. (2009). Performance and levels of contextualisation in a selection of OECD countries in PISA 2006. *Journal of Research in Science Teaching, 46*(8), 897–908.

OECD. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris: Author.

OECD. (2005). *The definition and selection of key competencies. Executive summary*. http://www.oecd.org/dataoecd/47/61/35070367.pdf. Accessed 3 June 2011.

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

OECD. (2010). *PISA 2009 results: What students know and can do. Student performance in reading, mathematics and science* (Vol. I). Paris: Author.

Poskitt, J., & Taylor, K. (2008). *National Education Findings of Assess to Learn (AtoL) Report.* http://www.educationcounts.govt.nz/publications/schooling/27968/5. Accessed 1 June 2011.

Robitaille, D. F., & Garden, R. A. (Eds.). (1996). *Research questions and study design* (TIMSS monograph, Vol. 2). Vancouver: Pacific Educational Press.

Satherley, P. (2006). *Student outcome overview 2001–2005*. Wellington: Ministry of Education.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1–22.

Chapter 4
Improving Science Education: Why Assessment Matters

Robin Millar

The Purposes of Assessment in Education

The purposes of assessment in education divide into three broad categories: summative, formative and quality assurance. Black and Wiliam (2007), for example, begin a recent review of large-scale assessment systems by saying that 'The main purposes of assessment are concerned respectively with the *support of learning*, with *certification*, i.e. with reporting the achievements of individuals, and with satisfying demands for public *accountability*' (p. 4, emphases in original).

Assessment designed to support learning is often termed *formative*. Assessment plays a formative role in almost every educational transaction, in providing evidence that informs and influences the subsequent actions of the learner and/or the teacher. This influence may be conscious and explicit or largely tacit. The outcomes of assessment of this sort are typically not reported beyond the immediate teaching and learning context. There is evidence from research that formative assessment is more effective when it avoids comparing one learner with another, or with a group norm, but instead focuses on providing task-related feedback that helps the learner to better understand the relationship between an aspect of their current performance and the desired performance (Black and Wiliam 1998a, b).

Assessment is used summatively when it is concerned with certifying learning. Summative assessment leads to statements that are communicated to a student and to others about the nature and extent of that student's learning. Contextual factors will determine who these 'others' are in any given instance. The outcome of summative assessment is often a mark or a grade or a set of these for distinct strands of learning within a topic or course. This outcome may have relatively minor consequences for the student and/or their teacher and/or school—or more substantial ones—in which case the assessment may be termed 'high stakes'.

R. Millar (✉)
Department of Education, University of York, York, YO10 5DD
email: robin.millar@york.ac.uk

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_4,

A third purpose of assessment in education, often linked to its summative purpose, is to generate data that can be used to make judgements about education systems or their components (e.g. schools, departments, teachers) and hence provide a measure of accountability for the use of the substantial (and often mainly public) funds that are spent on education in most countries.

I want, in this chapter, to argue that there is a purpose of assessment in education that is more fundamental than any of these three, because it precedes and underpins all of them—yet is curiously neglected in most discussions of assessment. It is the role of assessment in clarifying the learning that is intended in any given situation. Any statement of intended learning outcomes of a programme, or course, or module is inevitably ambiguous and open to (often quite wide) variation in interpretation. Assessment instruments and practices are the tools by which this ambiguity is reduced, perhaps even removed. Assessment operationalises outcomes and hence defines them. Unless we know what we will accept as evidence of the achievement or non-achievement of any given learning objective, we do not really know what that objective is or means. The job of an assessment instrument (a question or task or a set of these) is to generate this evidence. In doing so, it is more than merely a tool for carrying out a task that is already fully and clearly defined. Rather the assessment instrument *becomes* an operational definition of the objective.

Reading lists of educational objectives, as they are normally set out in curricula and in specifications (or syllabi) for courses and examinations, often feels like walking in the fog. You have a rough idea where you are and with care can avoid serious collisions with external reality but are not really sure where you are heading. In areas of school science where there are well-established assessment practices, we have become largely unaware of the fog. But whenever we try to extend the reach of school science and incorporate less familiar objectives, for instance, those associated with practical capability in science or with understanding of the nature of science, the poor visibility becomes harder to ignore. The sense that instruments and practices are unsatisfactory becomes more apparent to everyone involved, though the response is more often to criticise the methods and instruments that are being used than to ask more fundamental questions about learning objectives and how (perhaps indeed, if) their attainment could be evidenced.

This chapter is a plea to science educators to take assessment tools and instruments more seriously, to recognise them as operational definitions of the outcomes we value and are striving to achieve, to value them more highly as critically important 'tools of the trade' and to put considerably more effort than hitherto into developing, discussing, debating, critiquing, adapting, validating and testing them.

What Do We Want, Do We Really Really Want?

Education is about changing people, in particular changing what is inside their heads. After most teaching and learning episodes, we want learners to know and understand some things that they previously did not. So how do we know if a given activity, or lesson, or sequence of lessons, or course is achieving this? The central problem is that the outcomes we seek are not directly observable. It is almost a commonplace to say we want to teach science 'for understanding' rather than mere recall—but how do we know if a student 'understands' a specific idea or does not? Mulhall et al. (2001) raise this same question when they ask, 'What, in detail, do we expect students to learn when we talk of 'conceptual understanding' in electricity?' (p. 583). Their view is that 'we [the science education community] do not have even the beginnings of systemic answers' (p. 583). They go on to say—and I would agree—that 'some justified response to [this question] is a necessary, if not sufficient, condition for any helpful advances in the thinking about and practice of teaching electricity' (p. 583). The problem does not apply only to electricity.

As we cannot observe understanding directly, we must infer it from things we *can* observe: what the student says, or writes, or does usually in response to a given stimulus, such as a question or task. In other words, we must operationalise the objective: 'understands X' becomes 'is able to make the desired response to a given question or task involving X'. For example, the English National Curriculum for Science (DfEE/QCA 1999) says that pupils between the ages of 11 and 14 'should be taught that unbalanced forces change the speed or direction of motion of objects and that balanced forces produce no change in the movement of an object' (p. 34). The outcome of successful teaching and learning of this point might be termed 'an understanding of Newton's first law of motion'. But what would this mean in practice?

The clearest answer we can give to this would be to identify a question or task, or a small set of these, that someone who had 'an understanding of Newton's first law of motion' would be able to do—perhaps also making explicit some of the features we would expect to see in their responses. It is not surprising that many science teachers use past questions from summative tests, especially where these are high stakes, as their best guide to what to cover when teaching a topic and the depth of treatment to aim for. This is a tacit recognition that examples speak louder and more clearly than abstract general statements of outcome—that the most useful definitions are operational ones.

I am not arguing that there is one operationalisation of a construct such as 'understands Newton's first law of motion' that is correct while others are not—or even that there is one which is unequivocally better than others. It is clear that judgements are inevitably involved and that any choice made is not value-free. So, for example, when Mazur (1997) discusses the poorer performance of his Harvard physics undergraduates on qualitative conceptual questions on electric circuits than on apparently more advanced conventional questions requiring the use of Kirchhoff's laws (pp. 5–7), he is making a judgement that the former are a better indicator of 'understanding' than the latter. This is a judgement with which many would agree but it is, nonetheless, a judgement, not a conclusion that follows inexorably.

The central issue here is not correctness but clarity. Greater clarity will not of itself remedy deficiencies or resolve disputes, but it might help to make the issues involved more visible and more precisely defined. To that extent, it might facilitate progress towards improvement and consensus or the acceptance of different viewpoints as legitimate. Assessment is contentious, as Wiliam (2010) notes, precisely because it reduces ambiguity. It pins things down in a way that looser statements of objectives avoid. It forces the issue. It makes you ask yourself what you really want from an episode of teaching and how you will know if you have got it or not. This is why it is important.

Assessment and Curriculum Specification

The argument presented in this chapter has grown out of personal involvement over the past three decades in several large-scale curriculum development projects (*Salters Science*, *Science for Public Understanding*, *Twenty First Century Science*), the work of the PISA 2006 Science Expert Group and research carried out within the *Evidence-based Practice in Science Education (EPSE)* Research Network (Millar et al. 2006). These have, in different ways, convinced me that it is not good enough to specify intended learning outcomes in the usual kind of curriculum or programme specification language—statements of what students should know, or understand, or be able to do. These, even the latter, are almost always too vague and unspecific—and so people interpret them differently. The problem with statements of objectives or learning outcomes is like the general problem of rules: they always require further and more elaborated rules to explain them, which in turn require … and so on ad infinitum. We learn more, and more quickly, from examples.

To make clear what we expect students to learn from following a teaching programme or curriculum, we need to generate, probably alongside (rather than instead of) a statement of objectives in conventional form, a large collection of questions and tasks that operationalise the intended learning outcomes. These would provide clear indications of the sorts of things we would like students to be able to do at the end of a teaching intervention or at different stages of their trajectory through school. And this needs to be done for *all* of the content that we deem essential, as these operationalisations are content specific. Relating this to the example used above, I am arguing that better decisions can be taken about what students should learn about Newton's first law of motion, and when they should learn it, if the discussion takes place around a set of questions designed to provide evidence of 'understanding' rather than in isolation from the tools and instruments that might provide evidence of learning. The effectiveness of current science education reform initiatives around the world (such as the revision of the US National Science Education Standards and the review of the English National Curriculum) could be significantly enhanced by going beyond statements of learning objectives in the usual language of such documents and also providing examples of questions and tasks that could elicit good evidence of the intended learning at each stage of a student's school career.

The process of operationalising objectives by developing instruments that could be used to assess them is important for several reasons. First, it helps anyone involved in planning a teaching intervention, of whatever size and scale, to clarify the learning objectives *for themselves*. It enables ('forces' might be a better word) them to put their ideas and intentions to the test—and hence become clearer about what their objectives mean—or discover they don't really know what they mean and need to be revised, or reviewed, or dropped. Second, it makes it much easier to communicate the objectives of a teaching intervention clearly to others. This is what the practice of many teachers in using past examination paper questions to guide their teaching is telling us. Third, it brings the discussion of learning objectives out into the open and enables the intended learning outcomes to be discussed, critiqued and improved in a much more focused and transparent way. The discussion is focused on the definition of the construct being assessed rather than on the technical adequacy of the assessment methods used.

All of this becomes even more acutely important if our learning objectives are broader and more diverse than understanding canonical science content. If, say, we want to develop students' ability to construct a sound argument based on evidence, or to deconstruct an argument put forward by someone else, how do we identify those students who can do this in the way that we wish and those who cannot? Or if we say that we want to improve students' understanding of some aspect of scientific enquiry, or of scientific reasoning, or of the nature of science, how do we know if a student has learned what we wanted them to learn or has not? Only by operationalising these learning outcomes—by developing tasks and instruments that might provide evidence of students' understanding or capability—do we ourselves come to an understanding of what we had in mind, of what it really is that we want students to learn. It is, of course, possible that we will discover that what we said we wanted them to learn really doesn't make much sense, and we should drop the idea. But, more optimistically, we may on at least some occasions be able to arrive at a clearer understanding of the learning we are aiming for and of how we might recognise students who had achieved this learning.

For example, one of the things we have learned, or at least should have learned, from the past 20 years of the English National Curriculum for science, is that it is very difficult to specify intended learning outcomes in the area of scientific enquiry. Attainment target 1 (scientific enquiry) has not been a success. It did not promote the kind of practical enquiry that those who first proposed it intended. Rather it led to a routinisation of practical activity and a reduction in the kind of illustrative practical work that can help students gain knowledge of natural phenomena and develop understanding of concepts and principles. The story is well documented by Donnelly et al. (1996). The same pattern has been played out again more recently in the introduction of a curriculum strand called 'How science works'. Again this was rather incompletely articulated in the National Curriculum, in words that were open to wide variation in interpretation (DfES/QCA 2004, p. 37). A satisfactory way of assessing it has yet to emerge. Yet many would argue that an understanding of the central features of scientific enquiry and an ability to analyse and present arguments based on empirical evidence are critically important outcomes of a science education—for all students. The solution is not to omit these learning outcomes from the science curriculum but rather to work towards greater clarity about what exactly we wish students to be able to do and in what situations—and how they can provide evidence of this.

This example should make clear that I am not arguing for shorter, sharp inventories of multiple-choice questions to assess learning in specific science domains, though these do have a role to play. Our learning objectives may be much broader and include things that multiple-choice questions, or indeed written questions of any kind, could not assess. We may, for example, want to improve students' ability to take part in well-informed discussion about a science topic—and see this as a better measure of their understanding than the ability to answer some written questions. Fine—but then we need to specify in some detail how we will recognise those students who have achieved the intended learning—what the observable characteristics of their contributions to discussion will be. The same would apply if we said our objectives were largely, or partly, in the affective domain—to improve students' interest in a science topic or encourage a more positive attitude towards it. In all of these cases, we need to ask: how will we recognise the outcome(s) we desire? What will count as evidence of successful learning? Developing the tools to measure a learning outcome we desire clarifies the intended outcome and facilitates its communication to key others.

The Role of Assessment in Science Curriculum Development

The development of new teaching approaches and interventions is usually seen by those who do it as a means of improving the learning experience of students and the learning outcomes they achieve. While many agree that improving practice depends on the interrelated triad of curriculum, pedagogy and assessment, the curriculum development process typically begins with the design of lesson activities and instructional sequences, followed later by consideration of the training that might be needed to help teachers implement these as intended and finally by the instruments to assess outcomes. The idea of 'backward design' (Wiggins and McTighe 2006) turns this sequence on its head. Essentially it says, when planning a teaching intervention on any given topic, begin by developing the tasks (written, oral, practical) that you would use to find out at the end if students have learned what you want them to learn. Only then ask yourself how you would plan a teaching sequence to help as many of them as possible to get to the point where they can do such tasks. Bybee (2006) indicates that this approach has been used in BSCS curriculum projects since the late 1990s, but this is the exception in international perspective rather than the norm. The advocates of 'backward design' argue that the significantly greater clarity about intended outcomes that follows from designing the assessment tasks enables teaching to be better focused and more purposeful.

I want to illustrate the issues involved here through an example from a major curriculum development project in which I am currently involved, which did not adopt the 'backward design' approach—but came later to wish that it had. The *Twenty First Century Science* project (Millar 2006, 2009) is an attempt to implement the key recommendations of the *Beyond 2000* report (Millar and Osborne 1998). It set out in 2002 to develop and test a curriculum structure that recognised, and tried to reduce, the tension in the secondary school curriculum between 'science for all' and 'science for future scientists'. When the project began, the standard time allocation for science in secondary years 4–5 of the English school system (students aged 14–16) was 20 % of the school week, making science a 'double subject' in the General Certificate of Secondary Education (GCSE) examination. *Twenty First Century Science* proposed splitting this into two equal halves (each the size of a 'normal' GCSE subject)—and developing a core GCSE Science course with a scientific literacy emphasis—plus two optional GCSE courses in Additional Science and Additional Applied Science. This offered students (and schools) three science course options: GCSE Science alone or in combination with either of the two Additional Science courses. The intention was to enable better matching of the science programme to individual students' interests, aspirations and abilities.

The design of the core GCSE Science course was based on a view of 'scientific literacy' similar to that of the *National Science Education Standards* (NRC 1996). It emphasised the ability to read with understanding textual material relating to science of the sort that people encounter outside school and aimed to develop students' ability to evaluate scientific knowledge claims and participate in discussions with their peers and others about scientific explanations and applications of science. This emphasis led the developers to the view that the learning objectives of the course should include some important 'ideas about science', as well as some of the 'big ideas' of science (science content knowledge). Drawing on previous experience in developing a course for senior high school students, *Science for Public Understanding* (Millar 2000), the *Twenty First Century Science* project team recognised that it was essential to list the 'ideas about science' that the course would try to teach in as full and clear a manner as possible. One criterion for selecting these ideas was that they should be useful for accessing and engaging with science-related issues in the public domain. An assessment of a sample of typical newspaper articles on science highlighted topics such as evaluating evidence of correlational and causal claims, awareness of uncertainty in data, risk and risk assessment and the issues that arise around the adoption of new technologies and artefacts based on scientific understanding. It also seemed important that students should begin to think about and to understand the relationship between empirical data (or evidence) and proposed explanations.

An epistemological framework proposed by Giere (1991) for helping people to analyse everyday accounts of scientific work offered a useful model. The version of this framework which we used is shown in Fig. 4.1. The numbers indicate the order in which to read the diagram:

1. The starting point of any scientific enquiry is a question about some aspects of the real world.
2. By observation and perhaps experimentation, we can collect data (descriptive or numerical).
3. At some point, someone proposes an explanation to account for the available data. This cannot be deduced from the data. Proposing an explanation always involves imagination and is essentially a creative step.

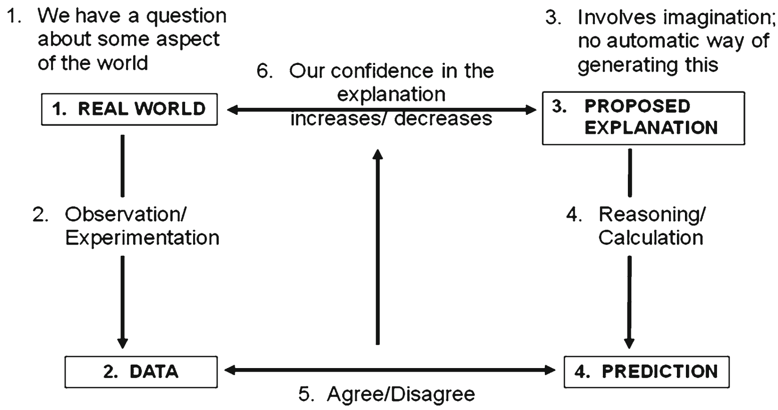


Fig. 4.1 A framework for evaluating a proposed scientific explanation (Based on Giere 1991)

4. A proposed explanation allows us to make some predictions about what will happen in new situations that we have not previously looked at.
5. To test the proposed explanation, we collect data (or look again at existing data) to see if it is in agreement with these predictions.
6. Agreement or disagreement increases or decreases our confidence that the proposed explanation might be right.

Giere's framework emphasises the distinction between the 'real world' which we want to explain (on the left-hand side) and the world of ideas, out of which explanations are made (on the right-hand side). It highlights two important points:

– That an explanation cannot be deduced from the available data, by logical reasoning, but has to be imagined' or conjectured (there is no arrow from Box 2 to Box 3).
– That a proposed explanation cannot be tested directly but only indirectly by comparing data from observation or experiment with predictions from the explanation. So agreement or disagreement may increase or decrease our confidence in an explanation but do not unequivocally 'prove' or 'refute' it. Disagreement might equally well be due to faulty data or incorrect reasoning or calculation in making predictions.

In Giere's original diagram, Box 3 (of Fig. 4.1) was labelled 'theoretical model'. While scientific explanations are often based on theoretical models, many of the science issues that enter the public domain concern proposed causal explanations, linking a factor to an outcome or (more often) to the probability of an outcome. In a scientific literacy course, we wanted a framework that could apply to explanations of all kinds—both causal and those based on an underlying mechanism or theoretical model. The version of Giere's framework in Fig. 4.1 acted as a 'checklist' for identifying specific points about data collection and interpretation that we thought were important and for organising them. These points dealt with:

Table 4.1 Some of the 'ideas about science' taught in the *Twenty First Century Science* GCSE courses (OCR 2005)

Students should learn that …
A scientific explanation is a conjecture (a hypothesis) about how data might be accounted for. It is not simply a summary of the data but is distinct from it
An explanation cannot simply be deduced from data but has to be thought up imaginatively to account for the data
Scientific explanations are tested by comparing predictions made from them with data from observations or measurements

– Empirical data (e.g. the inevitability of measurement uncertainty, repeatability of data and issues of sampling and sample size)
– Investigation design (e.g. control of variables and the distinction between correlation and cause)
– The generation and nature of explanations (in particular that explanations cannot be deduced from data but are always conjectures which are to some extent tentative)
– The logical implications of a match or mismatch between prediction and data (in particular, that a match does not 'prove' that an explanation is correct or a mismatch that it is wrong)
– The role of the scientific community in scrutinising and checking claims (in particular, the importance of peer review)

In addition to the epistemological 'ideas about science' that are implicit in Fig. 4.1, analysis of media reports about scientific matters suggested that two other groups of 'ideas about science' should be included, relating more specifically to the technological applications of science. These were risk and risk assessment and individual and social decision-making about scientific and technological developments.

This resulted in a list of 'ideas about science' to be taught and developed through the GCSE Science course, grouped under six headings:

– Data and their limitations
– Correlation and cause
– Developing explanations
– The scientific community
– Risk
– Making decisions about science and technology

To illustrate the kinds of ideas in this list, three of those in the 'developing explanations' group are shown in Table 4.1.

We were, however, aware that teachers and those who write examination questions might welcome further guidance as to what we had in mind in stating such objectives, so the 'ideas about science' were presented in the course specification as a two-column list (Table 4.2). The right-hand column was not intended to add more content but simply to help teachers and examiners to see what might provide evidence that the objective had been achieved.

Table 4.2 Three representative 'ideas about science', as they were presented in the *Twenty First Century Science* GCSE Science specification (OCR 2005)

Students should learn that …	A student who understands this …
A scientific explanation is a conjecture (a hypothesis) about how data might be accounted for. It is not simply a summary of the data but is distinct from it	Can identify statements which are data and statements which are (all or part of) explanations Can recognise data or observations that are accounted for by, or conflict with, an explanation
An explanation cannot simply be deduced from data but has to be thought up imaginatively to account for the data	Can identify imagination and creativity in the development of an explanation
Scientific explanations are tested by comparing predictions made from them with data from observations or measurements	Can draw valid conclusions about the implications of given data for a given explanation, in particular: • Recognises that an observation that agrees with a prediction (derived from an explanation) increases confidence in the explanation but does not prove it is correct • Recognises that an observation that disagrees with a prediction (derived from an explanation) indicates that either the observation or the prediction is wrong and that this may decrease our confidence in the explanation

The *Twenty First Century Science* GCSE courses have now been in use for eight school years, three as a pilot programme available only to 80 schools which opted to take part in the trial and five as one of the four suites of approved GCSE Science specifications from which schools in England can choose. Throughout this time, the 'ideas about science' which students should be taught have been specified in the form shown in Table 4.2. Revised specifications came into effect from September 2011 and have the same format. Interactions with teachers using the courses, and more acutely with the examiners (mostly practising or former school science teachers) who write questions for the summative examination, have, however, convinced us that, while the right-hand column in Table 4.2 is helpful, it is not clear enough. As a result, we have embarked on a programme to develop additional resources for teachers and examiners to clarify this further. The goal is to indicate more clearly the sorts of tasks that might be used to generate evidence that a student understands the things listed in the right-hand column of the 'ideas about science' tables. The first step is to develop question descriptors for every 'idea about science'. Table 4.3 shows how this might look for the first 'idea about science' in Table 4.2.

The second step is then to write a sample question corresponding to each description. This work is in progress. We see the process as essentially iterative; we anticipate having to make changes to the question descriptions as questions are drafted, critiqued, modified and accepted as 'fit for purpose'. New ideas for additional, or better, question descriptions may emerge from the question development process. It might become apparent that Table 4.3 is incomplete and that other types of question are needed to provide evidence of understanding of this 'idea about science'. It is

Table 4.3 Possible question descriptions for one 'idea about science'

	Description of possible questions to provide evidence of understanding
(a)	Provide a list of statements. Ask student to identify those which are data and those which are explanations or parts of explanations
(b)	Provide a short textual account of a scientific episode. Ask student to identify a sentence that reports data and a sentence that states a possible explanation
(c)	Provide an explanation. Present a list of pieces of data. Ask student to say whether each is consistent with the explanation, conflicts with it, or is neutral
(d)	Present information about a situation and several statements about it. Ask student which can be asserted from the data and which are inferences that go beyond it

possible that this work might lead us to review the wording of some of the 'ideas about science' themselves, as we become more aware of what an operationalisation of them might look like. This is precisely the process I have described in general terms in the earlier sections of the chapter.

We can now see that we should have done this work at a much earlier stage—or at least that having done it earlier would have helped us to sort out our own thinking and to communicate our ideas and intentions more effectively to others. Unfortunately the time scale for curriculum development at national level in England rarely allows things to proceed at an ideal pace or in an ideal order. And it is possible that only by getting involved have we come to see how we should have gone about it. But that learning is surely transferable. Would it not be sensible for anyone involved in major curriculum development work to give priority to the development of instruments that would provide evidence of successful learning of the intended outcomes of the programme or course—and to develop these before, or at least alongside, materials to support the teaching itself? In future curriculum development projects, we plan to adopt this 'backward design' approach.

The Role of Assessment in Science Education Research

The discussion so far in this chapter has focused on curriculum issues. A similar blindness to the critical importance of assessment instruments and tools has also limited the impact of research on learning in science. A large number of research studies have been carried out over the past half century on students' ideas and thinking in specific science domains, before, during and after instruction. Learning is a central educational concern, so the extent of the research effort on learning in science is no surprise. What *is* surprising is that it has not yet led to the development of tools and instruments for assessing specific aspects of science learning, at specific ages or stages, that are broadly accepted by the research community and widely used in specific studies on more local or particular issues. Instead, new studies typically develop their own instruments for assessing learning outcomes. These are rarely tested for reliability, or validated by peer review, to the extent that they should be. Often the outcome measure is the weakest part of a reported study.

One of the characteristics of a mature research community is a measure of agreement about good questions to ask and about appropriate tools and methods to use to try to answer them. Science education research does not yet meet this criterion. In Kuhnian terms it is 'pre-paradigmatic' (e.g. Kuhn 1996). Changing this state of affairs would require the community of science educators to put more effort into developing better and more publicly agreed measures of the learning outcomes we value. The need for such an effort is perhaps most clearly indicated by the response of science educators to the small number of instruments of this sort that do exist. One of the best known is the Force Concept Inventory (Hesteneset et al. 1992). This is a set of 30 multiple-choice questions on Newtonian mechanics, each offering five answer options of which one is correct, designed for use with students taking physics courses at senior high school or first-year university level. It was quickly taken up and used by several physics educators in the USA to collect data across a range of institutions (see, e.g. Hake 1998) and subsequently used in studies at both school and university level in other countries (e.g. Savinainen and Viiri 2008). Its citation score on Google Scholar (1,453 on 16 November 2011) is another indication of its impact. It has also led to the development of similar 'inventories' for other topics in physics and in other sciences. And, in line with general points made earlier about the effects of developing assessment instruments, it has also generated discussion and debate about exactly what it measures (Huffman and Heller 1995) and about what the learning objectives of elementary mechanics courses should be (Mazur 1992).

As a research community, we need to give higher priority to the development of assessment instruments of a wide range of types—with the ultimate aim of having a set of carefully tested and validated instruments that can be used in specific studies, enabling comparisons to be made more confidently and reliably and raising the quality and credibility of conclusions and claims. This need is perhaps most acutely apparent in studies that involve an experimental comparison of two teaching methods or approaches or interventions to see if one is decisively better than another. Such studies are often requested by policymakers and research funders and are often seen as the 'gold standard' as regards the quality of evidence produced. The applicability of randomised experiments to education is strongly contested, with strong views held on both sides (for contrasting views, see Hammersley 2002; Slavin 2002). The quality of the evidence that a randomised experiment produces does not, however, depend only on its research design; it also depends on the quality of the instruments used to measure the learning outcome. The advocates of experimental designs, however, frequently appear blind to the critical importance of the outcome measure, instead focusing almost exclusively on the study design. Yet the outcome measure is often the weakest element of such studies—and its weakness undermines the entire study. If we are seriously interested in comparing different teaching approaches and methods (and it is hard to see how this could not be of interest to science educators), then it is essential to develop widely accepted outcome measures before putting effort into the design of studies to compare interventions.

Not Perfect, Just Good Enough

This chapter is a plea for more attention to be paid—by those who write curriculum statements, develop teaching interventions for their own classroom use or for wider publication and plan and carry out research on the learning and teaching of science—to the design and development of assessment instruments. It argues that development of the actual instruments that could be used to provide evidence of the intended learning is the best way to clarify what we mean by those outcomes and that concrete exemplars are the most effective way to communicate these intentions to others.

One further word of caution is perhaps necessary, in conclusion. Criticism of tests and examinations in science—as in many other subjects—is endemic. Perhaps, however, it is too often a lazy response. It is often easy to find fault with assessment instruments and procedures; it is rather harder to design alternatives that are clearly better. Criticism is easy because assessment instruments are never perfect. No operationalisation captures all of the construct it purports to assess or captures it exactly. Nothing I have said in this chapter should be taken to imply that I think perfect assessment instruments can be developed. What we need are instruments that are 'good enough', in the sense in which this phrase was used by the eminent child psychologist Bruno Bettelheim (1987). They should point in the right direction, get enough things right, and above all avoid causing damage. Developing science assessment instruments that are 'good enough' would be a major step forward, both for research and for practice and policy.

References

Bettelheim, B. (1987). *A good enough parent: A book on child-rearing*. New York: Knopf.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7–74.

Black, P., & Wiliam, D. (1998b). *Inside the black box. Raising standards through classroom assessment*. London: School of Education, King's College London.

Black, P., & Wiliam, D. (2007). Large-scale assessment systems: Design principles drawn from international comparisons. *Measurement: Interdisciplinary Research & Perspective, 5*(1), 1–53.

Bybee, R. (2006). Boosting science learning through the design of curriculum materials. Paper presented at the ACER Conference Boosting Science Learning—What will it take? http://research.acer.edu.au/research_conference_2006/7. Accessed 7 July 2011.

DfEE/QCA. (Department for Education and Employment/Qualifications and Curriculum Authority). (1999). *The National Curriculum for England. Science*. London: QCA.

DfES/QCA. (Department for Education and Skills/Qualifications and Curriculum Authority). (2004). *The National Curriculum for England. Science*. London: QCA.

Donnelly, J., Buchan, A., Jenkins, E., Laws, P., & Welford, G. (1996). *Investigations by order. Policy, curriculum and science teachers' work under the Education Reform Act*. Nafferton: Studies in Education Ltd.

Giere, R. (1991). *Understanding scientific reasoning* (3rd ed.). Fort Worth: Holt, Rinehart and Winston.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74.

Hammersley, M. (2002). *Educational research, policymaking and practice*. London: Paul Chapman.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher, 30*, 141–157.

Huffman, D., & Heller, P. (1995). What does the Force Concept Inventory actually measure? *The Physics Teacher, 33*, 138–143.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago.

Mazur, E. (1992). Qualitative versus quantitative thinking: are we teaching the right thing? *Optics and Photonics News*, *3*, 38. Republished in *The Pantaneto Forum*, 11 (July 2003). www.pantaneto.co.uk/issue11/mazur.htm. Accessed 7 July 2011.

Mazur, E. (1997). *Peer instruction*. Upper Saddle River: Prentice Hall.

Millar, R. (2000). Science for public understanding: Developing a new course for 16–18 year old students. *Melbourne Studies in Education, 41*(2), 201–214.

Millar, R. (2006). Twenty First Century Science: Insights from the design and implementation of a scientific literacy approach in school science. *International Journal of Science Education, 28*(13), 1499–1521.

Millar, R. (2009). Designing a science curriculum to enhance students' scientific literacy. In R. W. Bybee & B. J. McCrae (Eds.), *PISA Science 2006: Implications for science teachers and teaching* (pp. 203–214). Arlington: NSTA.

Millar, R., & Osborne, J. (Eds.) (1998). *Beyond 2000. Science education for the future.* London: Nuffield Foundation. www.york.ac.uk/media/educationalstudies/documents/staff-docs/Beyond%20 2000.pdf. Accessed 7 July 2011.

Millar, R., Leach, J., Osborne, J., & Ratcliffe, M. (Eds.). (2006). *Improving subject teaching. Lessons from research in science education*. London: Routledge.

Mulhall, P., McKittrick, B., & Gunstone, R. (2001). A perspective on the resolution of confusions in the teaching of electricity. *Research in Science Education, 31*(4), 575–587.

NRC. (National Research Council). (1996). *National Science Education Standards*. Washington, DC: National Academy.

OCR. (Oxford Cambridge and RSA Examinations). (2005). *GCSE specification: Science A. Twenty First Century Science suite. J630*. Cambridge: OCR.

Savinainen, A., & Viiri, J. (2008). The Force Concept Inventory as a measure of students' conceptual coherence. *International Journal of Science and Mathematics Education, 6*(4), 719–740.

Slavin, R. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*(7), 15–21.

Wiggins, G., & McTighe, J. (2006). *Understanding by design* (2nd ed.). Upper Saddle River: Pearson Education.

Wiliam, D. (2010). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. *Review of Research in Education, 34*, 254–284.

Chapter 5
Towards an Authentically Assessed Science Curriculum

Peter J. Fensham and Léonie J. Rennie

Introduction

Some years ago, one of us wrote that two things were clear about formal educational systems:

- What is assessed in these systems determines what teachers and students recognise as knowledge of worth.
- Teachers in general are conscientious in doing their best to ensure that their students will learn this knowledge of worth (Fensham 2006, p. i).

Very recently we were both involved as advisors to the development of a national science curriculum for use in the Australian states and territories, across the compulsory school grades of 1–10. This curriculum followed an agreement, in 2007, by the Ministerial Council for Employment, Education, Training and Youth Affairs (MCEETYA) to pursue greater national consistency in school curriculum, with common standards to be developed initially in English, mathematics and science. A national body, the Australian Curriculum, Assessment and Reporting Authority (ACARA), was established to oversee the development of the curriculum, testing and reporting. MCEETYA had previously endorsed Statements of Learning in science structured around three aspects: Science as a Human Endeavour, Science as a Way to Know and Science as a Body of Knowledge. Building on these three aspects, the national curriculum in science has three strands, Science Understanding, Science as a Human Endeavour and Science Inquiry Skills (ACARA 2010).

P.J. Fensham (✉)
Monash University & Queensland, University of Technology,
Building 6, Wellington Road, Clayton, Victoria 3800, Australia
e-mail: p.fensham@qut.edu.au; peter.fensham@monash.edu

L.J. Rennie
Curtin University, Kent Street, Bentley, Perth 6102, Australia
e-mail: L.Rennie@curtin.edu.au

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_5,

There was no shortage of ideas among the advisors as to ways in which this new curriculum could be an improvement on what existed in the independent jurisdictions. Among the suggestions were the following:

- Choose three or four broad topics or themes for each year within which the content for learning would be listed.
- Focus on developing big ideas in science.
- Encourage context-based teaching of science.
- Develop students' interest in science, foreshadowed in the statement in the curriculum rationale that science is a way of answering interesting questions.
- Engage students in an individual science inquiry in each year and share with others in larger extended investigations either in-school or out-of-school.
- Extend Science Inquiry Skills to include unfamiliar aspects of the nature of science, actively initiating investigative questions, selecting data collection methods, analysing data to provide evidence, making decisions and communicating findings.
- Give special attention to the new strand, Science as a Human Endeavour, so that it is not diminished by the attention given to the other more familiar strands.

For a variety of reasons, many of these positive suggestions did not bear fruit in the final draft (ACARA 2010). Not least among these reasons was the issue that the suggestions would lead to aspects of science learning that are not amenable to assessment within the narrow confines of the term that underpinned the political decision to move to a national curriculum. It is also the case that some of what was proposed, while familiar as innovative science teaching at the research level, has not yet been so developed in terms of assessment. This mismatch between improved quality learning in science education and the established modes for assessing learning, either in school or externally, is not a uniquely Australian problem. In this chapter, we consider how some current ideas for improved science education might be assessed and so become more likely to be incorporated into the mainstream of science teaching and learning. We illustrate our discussion with examples of exemplary practice and conclude with some notes on what these broader kinds of assessment approaches mean for teachers.

Principles of Assessment

Throughout this chapter, we have adopted six principles about assessment in science education:

- Students are individually entitled to assessment of their learning, both formative assessment that contributes to the learning process and summative assessment for a significant period of teaching.
- Assessment in science education should involve knowledge of science and knowledge about science.
- Each key aim in the curriculum should be assessed, and be assessed as authentically as possible.

- A variety of modes of assessment will be necessary to achieve this authentic assessment.
- A profile of achievement of these aims will result for each student. Such a profile will provide much more information than a single score, enabling appropriate advice for further education or for employment.
- The classroom teacher will be the person most likely to be able to make some of these assessments in an authentic way, but many teachers will need support to do so.

Before continuing, we note that it is not within the brief of this chapter to discuss the interactions between pedagogical practices and assessment for the science curriculum intentions we discuss. Black (see this volume) provides a more complete model of instruction.

In suggesting possibilities for assessing some new science learning outcomes, we are assuming that in the course of a significant period of science education, the teachers will have provided their students with what Roberts (1982) described as a curriculum emphasis—a succession of learning experiences that are designed to develop knowledge and processes involved in the intended outcome. There are appropriate pedagogies for each of these curriculum emphases. The research literature has described these sufficiently, and enough supportive curriculum materials are now available to make teaching for these learning outcomes possible.

The Nature of Curriculum

We begin with a historical perspective on the structure of curriculum and the nature of assessment approaches traditionally found. Early analyses of curriculum structure recognised four essential components. In his classic text, *Basic Principles of Curriculum and Instruction,* Tyler (1949) posed his analysis around four questions that related to the educational purposes or aims of the school curriculum, the learning experiences to be selected to achieve these aims, how those experiences can be organised for effective instruction and how their effectiveness in achieving the educational aims can be evaluated. Tyler emphasised that these four components needed to be internally consistent.

From at least the 1960s, it has been common for the curriculum for science in schooling to be set out in two parts—a rationale and/or set of aims and a detailed spelling out of the content that is intended for teaching and learning. The latter may be in strands that are distinguished by science discipline or by different aspects of science, such as science content knowledge and science process knowledge (or, more recently, nature of science). These two parts answer Tyler's first question quite thoroughly. The second and third questions, relating to the nature and organisation of the learning experiences that provide students with the kinds of educational experiences that enable the aims to be achieved, are usually left to teachers to choose and arrange. There are many textbooks and teachers' guides to assist them to perform this role.

The fourth question, relating to evaluating the extent to which the educational aims have been achieved, is rarely addressed. Tyler's focus was on evaluating curriculum, but his source of data was the student: What changes in the students occurred as a result of the learning experiences? Our focus is on the assessment of these changes, and it is clear, as it was for Tyler, that effective assessment must match the nature of the learning experiences designed to achieve the curriculum aims. As a number of chapters in this volume demonstrate, teachers are often not prepared well in assessment, and so it is often poorly done. Further, more emphasis is placed on summative assessment, rather than formative assessment that can assist students' learning on the way. In addition, curricula rarely provide enough detail about the form that summative assessment of students' learning will take. Not surprisingly, there is considerable scope for slippage between the aims, the knowledge, understanding and skills students are expected to achieve and the assessment of their learning. It is essential that it is clearly specified how each of a curriculum's aims and strands of learning is to be assessed, if they are to be attended to by teachers during their teaching and recognised by students as important for learning. Two examples, relating to senior chemistry and science as inquiry, will demonstrate the slippage between intended and summatively assessed outcomes.

An Example of Senior Chemistry

The Victorian Curriculum and Assessment Authority has, for many years, listed eleven aims for the grades 11 and 12 chemistry curriculum (VCAA 2010). It states that the curriculum has been designed to enable students to:

1. Develop an understanding of major ideas of chemistry and develop an ability to apply these ideas in both everyday and hypothetical situations.
2. Use theoretical models as aids to explaining chemical phenomena and appreciate that such models are subject to constant scrutiny and necessary modification.
3. Develop a knowledge of the language and methods of chemistry.
4. Reflect on their developing understanding of chemistry and its role in their lives.
5. Develop the practical skills necessary to undertake experimental work.
6. Explore the wider social, economic, technological and environmental aspects.
7. Develop perspectives on ethical questions.
8. Consider the role of chemistry in other areas of science.
9. Develop skills in cooperative work and communication of ideas.
10. Develop an interest in and enjoyment of the study of chemistry.
11. Develop an understanding of the procedures required for the safe use of chemical equipment and safe handling of chemicals, both in the laboratory and in everyday situations. (VCAB, 1984, p. 3)

Students' learning in the face of this set of aims was assessed using only an externally set, single paper and pencil examination comprising items that were almost entirely associated with Aims 1 and 3, with some that might be classified as related to the first part of Aim 2. This led to a quantitative score for each student that contributed to tertiary entry. Teachers were expected to certify that their students had undertaken appropriate practical chemistry to fulfil the requirements of Aims 5 and 11, but the criteria for doing this were vague. There was no assessment, either internally or externally, of the second part of Aim 2 or of Aims 4, 6, 7, 8, 9 and 10.

This senior chemistry example highlights three common and continuing features of current summative assessment practice: the high value given to external, paper and pencil tests as legitimate and seemingly objective and reliable measures of what students know and can do; the lower value given to teacher assessment; and the failure to assess all the intended aims.

The Case of Science as Experimental Inquiry

'The process of building this knowledge is as important as the knowledge itself because the journey of discovery reveals much along the way' (National Curriculum Board n.d., p. 5).

For much of the last century, the Achilles heel for a science curriculum was its strong emphasis on the role of 'the practical' in science and its neglect in most cases of any adequate mode of assessment. Even when its primary purpose was the development of a set of scientific skills, the complications of establishing arrangements to assess students individually in these skills were too much for most authorities. As demonstrated in the senior chemistry example, the examiners contented themselves with simply an assurance from the teachers and concentrated their attention on paper and pencil testing of the student's recall and application of the intended science content knowledge.

As recently as 1998, 16 authors could write chapters for a book, *Practical Work in School Science*, without one chapter addressing the critical point of its assessment. To be fair, the editor does refer in a list of further readings to eight papers on assessment (Wellington 1998). The dilemma about assessing practical work has been summarised by Black (2004) in his forceful statement, 'If you wish to know how students are doing in practical science investigations, there are no surrogates. The only valid source of this assessment is the students' teachers whether they like or not' (p. 173).

In the normal passage of schooling, the classroom teacher regularly plays the roles of both teacher and assessor. It is only when the assessment has a high-stakes quality that the assessment role becomes more contentious. Teachers, in general, are less comfortable when their assessment of the students has a judgemental consequence of high significance, like access or not to a level of further education. In these instances the role places a higher level of demand on the teacher. One of these is the issue of the fairness or evenness of the teachers' assessments across teachers and schools. The subtitle of Yung's (2006) book, *Assessment Reform in Science,* is *Fairness and Fear*, and he provides evidence for how his set of biology teachers faced and handled this issue.

We shall return to Black's truism throughout this chapter. It is easy to state that this or that assessment responsibility lies with the teacher. Inferring this responsibility, however, implies that the curriculum has made quite clear why its different aspects of science are important in school science and that there is common understanding among teachers of these aspects. Furthermore, if teachers are to undertake this assessing role, they must be prepared for it and also be supported in appropriate ways by the education system of which their school is a part. We return to this critical point of teacher support later.

Assessment in Contemporary Science Curricula

Since the 1990s the purposes for science education in schooling have expanded considerably beyond acquiring basic conceptual knowledge in one or more science disciplines and some facility with the skills of using scientific equipment to carry out, usually prescribed, laboratory investigations. Such expansions include the following:

- Using interesting contexts to give relevance to science concepts and principles which, in turn, enable these contexts be more fully comprehended.
- Experiencing scientific investigation—the engagement of students in extended open-ended scientific investigations, both individually and with others, either in-school or out-of-school.
- Extending the focus of science as an inquiry process to include hitherto unfamiliar aspects of nature of science, such as initiating investigative questions, selecting data collection methods, analysing data to provide evidence, and communicating findings.
- Paying attention to the skills of argumentation and decision-making that socio-scientific issues involve.
- Introducing the human dimensions of science—humanistic science education, science as a human and cultural endeavour.

Variety of Modes of Assessment and a Profile of Achievement

It is clear that the expansion of purposes of science curriculum will require a diversity of assessment techniques. Clearly, a single test of science knowledge and understanding will not assess adequately the kinds of outcomes embodied in the expansions listed above. Many of the possible outcomes of science education are not amenable to a one-session paper and pencil test, and an authentic assessment will only be achieved when several modes of assessment are used. Rather than a single score, a profile of each student's achievement will be necessary to capture the broad range of outcomes that are now expected. It is important that the students develop and keep portfolios of their work over time, not only to allow teachers to assess their progress but to enable them to learn to reflect upon their own performance and in doing so develop skills of self-assessment.

Authentic assessment tasks can include oral presentations; written, drawn and otherwise created artefacts; group activities, including problem solving; evaluating evidence and documentation of processes culminating in decisions; and well-crafted written assignments. All of these activities provide valid ways of assessing students' performance beyond just the recall of knowledge. It is up to the teacher, with help and guidance, to devise assessment tasks that elicit and encourage the appropriate behaviour to demonstrate learning of the intended outcomes. The use of a variety of formative assessment tasks will serve the student with differentiated information about the progress of their learning, compared with what they can know of this when there is only a summative assessment at the end of learning. Interestingly, authentically assessing a range of outcomes expected in the curriculum eliminates the temptation for teachers to 'teach for the test', because students now have multiple ways to be helped by their teachers and to demonstrate what they know and can do.

The Effect of High-Stakes Assessment

Although it is clear that a single examination cannot capture the broad outcomes of students' achievement, it is not easy to move away from a well-established regime of high-stakes, external testing. External assessment has high status and is seen to be a reliable and objective summative measure of student performance. It also has a strong influence on parents, students, and what is taught (see Black, this volume for a recent example of this). However, as we saw in the earlier case of senior chemistry, external testing usually assesses only a narrow part of a curriculum's aims.

Over the years there have been efforts to broaden the scope of testing with an additional school-based component, but these have had to struggle with the issue of the validity of this assessment. For example, it is now quite common (see Black, this volume), even at the high-stakes levels, such as the end of secondary schooling, for there to be an external assessed component via a paper and pencil test and an internally assessed component, based on portfolios or other evidence. The latter has been used to enable assessment tasks such as an extended investigation or large assignment (see below in the section titled Context-Based Science) or a task that involves accessing scientific reports and literature beyond the textbook level. Although these external and internal components should be treated as providing a richer and more informative assessment of a student's learning, the high-stakes purpose of producing a single score for each student too often undermines the representation of the internal assessment, especially if, as is not uncommon, internal assessment is statistically moderated by the student's relative performance on the external test. Such procedures are usually justified on the grounds that these different performances are positively correlated. However, this correlation is by no means sufficient to justify such a loss of assessment information. Teachers will be very aware that it is common to find students who flower in an investigation, in orally communicating their research or in arguing the worth of scientific data, compared with their written performance. Achievement represented as a single score cannot fully represent what students know, understand and can do.

In the futurist assessment situation to which this chapter is directed, we explore a variety of assessment modes that match some new aspects of science learning that would provide students with both formative information and a more comprehensive and authentic summative assessment expressed as a profile of performances.

Towards Authentically Assessed Achievement in Science Education

In the previous sections of this chapter, we have made the case that the purposes of science education foreshadow much broader learning outcomes for students than are captured by common assessment practice. In the next sections, we discuss approaches to assessment practice in science education that are more congruent with three of these intended outcomes of contemporary science curricula: context-based science education, decision-making processes and socioscientific issues and integrated science education. We chose these because they are increasingly referred to in curriculum statements as human dimensions of science and because they are well supported in the research literature and in reports of innovative practice.

We assume in each case that the specific science knowledge involved will have already been taught earlier or in situ. Furthermore, we assume that key discourse words for the intellectual processes that are involved in these approaches to teaching science will have been regularly articulated by the teacher and required of students when engaging in discussion. The power of such modelling of key discourse words in the classroom discourse cannot be overemphasised (see Fitzgerald and Gunstone, this volume). With these assumptions a key concern of formative assessment will be how well this social discourse in the classroom has been 'accommodated' by individual students.

Incidentally, one of us (Peter) when visiting a Canadian school observed a lesson introducing stoichiometric calculations to 16-year-olds. To his surprise (until he learnt the class had just visited a local factory making superphosphate fertiliser), three quite complicated, balanced equations were set out as possibilities for this production reaction. The teacher pointed out that each equation provided different quantitative evidence that could be used to decide which equation represented the actual reaction. The words 'decide' and 'gather evidence' were used at least a dozen times during that lesson. The culmination was the next day in a laboratory replication of the reaction that led to the weight of calcium sulphate as a solid product being the evidence for deciding which equation was appropriate. When Peter asked one student who was carrying out the filtration process, 'What are you doing?' the reply was 'Gathering evidence!!'

Context-Based Science Education

Historical Background

By the 1990s, there was general realisation that the school science curriculum in its traditional academically oriented form was not meeting the needs of the growing numbers of students now continuing longer in schooling. In a number of countries, there were calls for a curriculum that would provide a Science for All (Fensham 1985). It was argued that school science should offer more for all students than was evident in its previous focus on preparing the next generation of science-based professionals.

The first attempts to achieve such a new and more relevant approach have been described as the Science/Technology/Society (STS) movement (Solomon and Aikenhead 1994). In its most developed form, STS science education envisaged the idea of *Concepts in Contexts*, that is, there were two interdependent learnings. First, science concepts would be taught in relation to a number of real-world contexts in which they had application. In this way, the significance and power of the concepts and principles of science would become more apparent. Second, this learning would, in turn, enhance the students' awareness and understanding of these real-world contexts. Numerous modules for teaching STS science were developed with those for the PLON physics project in the Netherlands being a sophisticated example (Eijkelhof and Koortland 1988).

For a complex web of reasons, when major reforms of the science curriculum occurred in the 1990s, they were not STS based and still a science discipline-oriented approach was adopted across all the years of schooling. The STS movement died before it was sufficiently established to require formal assessment procedures (Fensham 2009; Layton 1994).

Since 2000, widespread evidence has emerged in the more industrialised countries of a serious decline of interest among students in science and science careers (Relevance of Science Education [ROSE], Trends in International Mathematics and Science Study [TIMSS], European Commission 2007), and this has resurrected interest among science educators and curriculum authorities in the many social contexts in which science and technology are involved. A relevant initiative has been the OECD's PISA Science project. Since 2000, as discussed by Fensham (this volume), PISA has drawn attention to science learning in relation to real-world contexts. Its ways of assessing a student's learning ability involve going beyond the recall of science knowledge to its application in relation to novel, real-world science and technology contexts.

This PISA approach to the assessment of science learning has certainly been regarded as a pioneering effort for most of the participating countries. It has not, however, as a number of critics have pointed out (see Fensham, this volume), taken its presenting contexts as intended targets of learning. These contexts were simply vehicles for achieving a profile of measures of scientific competences that were then added to give a score for the project's definition of scientific literacy. In this sense PISA Science turns out to be a weak form of assessing context-based science teaching and learning.

Research, Practice and Assessment

Under the title of Context-Based Science, the idea of using real-world contexts that involve science and technology (S&T) has revived as a now popular movement for research and classroom innovative practice (see, e.g. Gilbert 2006; Gilbert et al. 2011; Stocklmayer et al. 2010). The outcomes of these research studies suggest that goals of such teaching are likely to:

– Increase students' interest and engagement
– Provide a deeper sense of conceptual understanding through relevant applications
– Increase appreciation of the use of science in society
– Encourage transfer of science learning to novel contexts
– Expand awareness of S&T contexts

To be authentic the assessment of context-based science should include tasks that relate to these outcomes.

The school science curriculum for the various levels of schooling has recently been defined in terms of a context-based approach, for example, the Senior Years Science Courses in Queensland (http://www.qsa.qld.edu.au/1941.html) and *Twenty-First Century Science* in England and Wales (www.21stcenturyscience.org/). In each case, where context-based science is the mainstream curriculum, the issue of its authentic assessment has been raised. One way in which it has been dealt with is through the requirement that in each semester in the grades 11 and 12 science subjects, one extended response task (ERT) or extended response investigation (ERI) must be included. An example of strong context-based assessment, an ERT in grade 11 *physics*, is discussed. We have reproduced this extended response task in full in Appendix 1, for three reasons. Firstly, it is no longer in the state of Queensland an unusual example for assessing science. Indeed in Queensland, where intra-school assessment has been the primary assessment for these high-stakes years for three decades and context-based science has been practised for a decade, there is a large and growing bank of such tasks for each of the four disciplinary senior sciences and for an integrated science course, *Science 21* (http://www.qsa.qld.edu.au/11362.html). When these curriculum emphases become mainstream, good teachers begin to devise and share such tasks in order to assess the learning criteria that these emphases intend. Secondly, for persons who have hitherto been concerned only with more traditional mainstream curricula and its summative assessment, the format of the task may be new and instructive. Thirdly, it embodies several features that relate to an intra-school assessment of students' learning in a system committed to a context-based approach. Among these are:

1. The overall topic of the task is likely to be of intrinsic interest to large numbers of students at this senior level of schooling because within a year some will be preparing and applying for a driver's licence.
2. Each of the three parts of the task emphasises the broad criteria or aims for this subject—Knowledge and Conceptual Understanding, Investigative Processes,

and Evaluating and Concluding. It should be noted that these criteria apply both to the science and technological content in the task and to the contextual appreciation to which it leads.

3. The intended concepts and processes are listed and they include both scientific and technological ones. There is a clear expectation of conceptual transfer.
4. The intellectual processes for the task are differentiated, with bold type, reinforcing the ways students are being taught to think and act in this science subject.
5. Short answers, longer pieces of purposeful writing, and a variety of levels of calculations are required.
6. The nature of science data as primary or secondary is designated.
7. Tasks 3 and 5b (see Appendix 1) extend the students' awareness of the accident context but very much because of the knowledge of its science and technology.

These seven features give a detailed structure to the task, and they can be used to shape assessments that could be used to promote and gauge the science learning expected in the earlier stages of schooling. The S&T contexts need to be chosen to have a real degree of interest for at least a majority of the students. The intended science concepts and their technological counterparts at each level of schooling can be set in several scenarios and can be associated with keywords that help the students identify how and what they are to intended to learn. The range of subtasks and their different levels of answer enable the student to become aware of the different expectations about what it means to communicate in science education. The seventh feature above is important as it illustrates how learning about a context and applying science ideas to it expands one's knowledge of it.

Incidentally, the Accident Task (see Task 5 in Appendix 1) also illustrates how easily context-based teaching integrates humanistic elements into the teaching of a traditional physics topic. The traffic engineer and the forensic police make direct use of the conceptual science relationships to gather the data they need to refer their consequences to other professionals in the story. The student (as public) is informed about avoiding accidents and their legal possibilities. In other such extended research tasks, the integration could as easily be between the public and the different disciplinary scientists whose knowledge is inevitably involved in any real-world S&T context.

Decision-Making Processes and Socioscientific Issues

As long ago as 1984, *Science for All Canadians* listed participating in socioscientific decisions as one of four purposes for school science education (Science Council of Canada, 1984). Since that time there have been regular references to it in many curriculum statements for science education. For example, the *National Science Education Standards* in the USA expected that 'A scientifically literate person can use appropriate scientific processes and principles in making personal decisions … and engage intelligently in public discourse and debate about matters of scientific and technological concern' (NRC 1996, p. 13). More recently, the *Framework for K-12 Science Education* has as one of its overarching goals that '*all* students … possess sufficient knowledge of science and engineering to engage in public discussion on related issues' (NRC 2011, p. ES-1, original emphasis). The recent *Australian Curriculum: Science* states that the strand, Science as a Human Endeavour:

> highlights the development of science as a unique way of knowing and doing, and the role of science in contemporary decision making and problem solving. It acknowledges that in making decisions about science practices and applications, ethical and social implications must be taken into account. (ACARA 2010, p. 3)

Despite the consistent inclusion of this intended outcome among the mainstream aims for school science, little support for it is to be found in the lists of detailed science content and scientific processes of the many curricula that have been endorsed since 1984 as the agenda for teaching and learning. Furthermore, as a mainstream intention of the curriculum, 'decision-making' has been dogged by a lack of means of assessing it. Decision-making and some of its component abilities, 'recognising alternatives', 'scientific argumentation', 'identifying values', 'assessing risk' and 'choosing between the alternatives', are not terms that are regularly used in the language of science classrooms. Some other components, like 'gathering and applying evidence', are, as instanced above, more familiar, even in the primary years where discussion of a 'fair test' is commonplace.

Making decisions about S&T issues and the sometimes contentious science involved implies that students should acquire the abilities that underpin recognising alternatives and making choices between them. This choice may be about the science itself, about alternate technologies that derive from that science or about the likely consequences of these applications. They may be made in terms of self-interest or the interest of society more largely.

The PLON physics project in the Netherlands, *STSE Chemistry* and *Science Plus* in Canada, and *Science and Society* in England and Wales all appeared in the 1980s as examples of innovative science teaching at the secondary level that involved elements of decision-making. Aikenhead (1994) developed *Logical Reasoning in Science and Technology (LoRST)*, resource materials in a text for a year of study that included a specific unit on *Decision Making*.

Aikenhead (1989) had earlier pioneered the research interest using decision-making theory to explore the ways students approach scientific inquiry. Ratcliffe (1997), Driver, Newton and Osborne (2000), Kolstø (2001), Ryder (2001) and Sadler and Zeidler (2005) are others who have worked with students in classrooms on S&T issues requiring argumentation and decision-making. The case for argumentation in science education has now been very strongly supported in the book by Erduran and Jiménez-Aleixandre (2007), but explicit attention to its assessment in individual students was not included.

These authors all acknowledge the importance of the science content knowledge of the specific issue but, even more strongly, point to the importance of a broad knowledge about nature of science (NOS). Allchin (2011) has recently mounted a strong criticism of presenting nature of science as a static set of beliefs as has been the case in some well-known NOS tests. He proposes a more functional view, Knowledge of the Nature of Whole Science (KNOWS), that emphasises the reliability of scientific claims and goes on to suggest ways of assessing it (see below).

With respect to the decision-making process itself, the research studies of it in science classrooms have led to a number of abilities as components:

i. Distinguishing between scientific questions and those that relate to other types of knowledge
ii. Gathering relevant data
iii. Assessing the quality of data
iv. Interpreting the data (distinguishing between correlation and causation), considering alternative explanations and integrating empirical data and nonempirical ideas
v. Using scientific models
vi. Distinguishing between observation and theory, recognising the conjectural nature of theory and distinguishing between evidence and theory
vii. Appreciating uncertainty in science (the consequential risk probability)
viii. Recognising and balancing personal and social values that impact on decision-making in science

This would be a formidable list of skills to learn and assess if each is seen separately. Fortunately, studies in secondary school classrooms, where more complex socioscientific issues (SSIs) have been the context for science teaching, have found that these components arise fairly naturally as a holistic set of perspectives about any particular SSI. This means their assessment can itself be combined as outlined below.

In a school science curriculum that takes decision-making seriously, a number of the above components, individually and together, will have been experienced in simpler and appropriately relevant contexts during the earlier years of schooling. In each of the following reports of earlier studies, the examples of the discourse used in the classrooms could just as easily have been couched in the components of decision-making. For example, Biddulph et al. (1986) showed that young primary students readily learn to ask questions about natural phenomena and to differentiate them into investigable (scientific) ones and non-investigable (non-scientific) ones. Erickson et al. (1992) found that 4th grade students could gather data and apply it as evidence to solve problems like ‘Which of these three magnets is the strongest?’ and ‘Which of these three brands of tissue holds the most water?’ In a quite different type of study, Ritchie et al. (2011) found that identifying and using science data as evidence could so successfully engage 4th and 6th grade students that they co-authored science mystery stories good enough to be published.

The first two of these studies suggest how simple but direct tests of the decision-making component can be constructed using appropriate phenomena and fair-test tasks that will enable the teacher to easily provide formative feedback. The observation of the student’s engagement and discourse acquisition in the repeated experiences of one of these components would provide the primary evidence for his/her summative assessment.

For teachers in the middle secondary years, a large number of quality items testing students' use of evidence in S&T contextualised tasks are now freely available from the PISA Science project (OECD 2006, 2007). These can be used formatively to focus teaching and learning on the role of evidence or to provide achievement information on applications of science knowledge that contributes to a summative assessment.

In the later secondary years when the teaching of complex SSIs provide regular practice at decision-making, the assessment of the students' learning could be conceived of as follows:

1. *Establish a baseline position of each student's views about science/technology/society* Aikenhead and Ryan (1992) developed a multidimensional instrument with a large number of items relating to Views on Science Technology and Society (VOSTS). Selection of a dozen or so of its items, suitably validated for a particular national context, would provide teachers with a useful profile of each student's thinking about STS issues prior to teaching a course that included decision-making. Such a baseline would provide helpful data for the teachers' responsibility to assess the decision-making development of each student during and after the course of study.
2. *Participation and development during the course of study* The class teacher will have regular opportunities to observe and hear how the students contribute and engage with each other in working on several SSIs and, in so doing, hear them practise using the component discourses of decision-making. Students should be required to keep an account under key discourse word headings of their experiences with the SSIs during the course as part of their personal portfolios.

Another method of assessing students' development of the skills s used in the discussion of socioscientific issues has been suggested by Simonneaux and colleagues (e.g. Morin and Simonneaux 2010; Simonneaux 2010). Based on a synthesis of various perspectives described by other authors, Morin and Simonneaux (2010) developed a matrix of six dimensions characterising reasoning about socioscientific issues in science education. These dimensions are points of view (*provides perspectives from different points of view in terms of the actors and stakes involved*), scale (*envisions spatial and temporal interactions and feedback*), types of knowledge (*refers to the kinds of knowledge relating to the controversial issue*), uncertainty *(identifies the validity and uncertainty of knowledge*), values (*discusses values and principles underpinning the issue*) and regulation (*understands the regulatory procedures relating to the issue*). The matrix is shown in Table 5.1.

As shown in Table 5.1, Morin and Simonneaux (2010) devised four levels at which students might show awareness and understanding of each of these dimensions, using sustainable development as an example. A student's score on each of these factors can be mapped onto a six-point star, such as shown in Fig. 5.1 in Appendix 4, thus providing a profile of performance and an indication of strengths and weaknesses in reasoning skills. The descriptors for each level could be reworded according to a specific SSI that is to be taught and then used to assess the depth to which students have come to grips with the issue. For example, after the SSI has

Table 5.1 Characterisation of socioscientific reasoning in an educational perspective for sustainable development (From Morin and Simonneaux 2010)

Dimension	Level 1	Level 2	Level 3	Level 4
Points of view				
Are situation and/or actors' points of view approached according to various angles?	He/she approaches the issue and its context under one angle only OR perceives no differences in the various actors' points of view	He/she considers the issue under several angles, envisages diverse aspects of the context OR realises that the various actors have not the same points of view without considering environmental, economic, social aspects	He/she considers the issue, the context or the various points of view of the actors according environmental, economic, social aspects	He/she perceives a controversy about the stakes and puts in perspective the various assertions in the interests of the actors
Scale				
Are changes of spatial or temporal scales made?	Envisages only the local, or only the global, only a distant future or only the short term	Considers different spatial scales OR the effects within short or long term, questions the sustainability of choices	Considers different spatial scales AND the effects within short or long term, questions the sustainability of choices	Conceives dynamic systems (spatial interactions with diverse scales, temporal feedback)
Knowledge What is the reference knowledge?	Considers only one type of knowledge, academic or not (vernacular, media)	Juxtaposes different knowledge, academic or not	Combines different socioscientific knowledge	Perceives a controversy about socioscientific knowledge

(continued)

Table 5.1 (continued)

Dimension	Level 1	Level 2	Level 3	Level 4
Uncertainty				
Are the conditions of validity of knowledge and uncertainties identified?	Does not perceive lack of information, knowledge is considered as truth	Expresses the necessity for information (supposed existing) OR finds differences in the assertions of actors OR questions risks (conditions of acceptability of an option)	Expresses the necessity for production of new knowledge OR questions certain assertions	Discusses the conditions of validity of the reference knowledge (epistemological doubt) OR discusses the impacts of technoscientific knowledge
Values				
Is there an awareness of the underlying values?	Is not aware of the weight of the values or of the beliefs in the advanced arguments	Expresses values or principles which underlie the personal commitment OR identifies the actors' values	Discusses the values or the principles engaged	Perceives possible conflicts within the values
Regulation				
The relations between the particular and collective interests – are they thought?	Considers that there is already a solution (law, deontological, technoscientific) OR does not consider the interactions between the various actors	Envisages a new regulation, only inside a category of actor	Envisages a new regulation, between various categories of actors	Discusses regulation procedures between the categories of actors or the governance (the modalities of decision-making)

been developed and discussed in class, either the teacher or perhaps groups of students (under the teacher's guidance) could score their performance on each of the six dimensions and then map the scores on to the six-armed star, providing a visual image of overall performance. In Appendix 4, Fig. 5.2 shows the shape of a particular set of scores that indicates more work is required in understanding that *scientific knowledge is uncertain*. Students would quickly realise that the shape to aim for would be a hexagon, with each score at level 4.

Either the folios of keywords or the maps of reasoning dimensions become ready means for the class teacher to make formative suggestions about the development of each student in relation to decision-making overall and also its component abilities.

An interesting point for observation in the move from one SSI to another would be the transferability of the component abilities from one SSI to another. Sadler and Zeidler (2009) have suggested that there may be a small set of personal orientations that students need for SSIs in general. These are:

(a) Appreciating the inherent complexity and multifaceted nature of SSIs
(b) Analysing issues from multiple perspectives
(c) Recognising the need for information relating to the tentative nature of science in SSIs
(d) Employing scepticism in the review of information presented by parties with vested interests

The emergence of these orientations in their students should also inform teachers when they are assessing their engagement in, and learning of, the decision-making process.

3. *Interviews with individual students* Ratcliffe (1997) and Zeidler et al. (2009) have each reported student assessment procedures for decision-making. Both procedures involve the classroom teacher (or another teacher familiar with decision-making) interviewing students individually. As a summative procedure, such oral testing is not commonly used in education systems of the Anglo-American tradition, but it is a quite familiar procedure in systems of the European tradition.

The interview schedule used by Sadler and Zeidler (2009) has been adapted from King and Kitchener's (1994) model of reflective judgement. A short paragraph reporting a pair of conflicting scientific reports about an S&T issue is given to the student to read and think about. The interview then proceeds, using a set of questions that should already be familiar to the student from the classroom experiences. A copy of the schedule is included in Appendix 2.

4. *Written test* Allchin (2011) argues convincingly that any test of the decision-making process and its component abilities must be set in the context of a real S&T case. He lists some exemplary cases, together with a scoring rubric for the students' written responses. This rubric elaborates in a more detailed point form the component abilities for decision-making listed earlier. Suitable and similar

topical and historical cases can be drawn from newspaper and other media reports and from popular science journals. The particular abilities to be tested can be selected from the rubric as appropriate for the students' stage level of schooling. One of Allchin's cases is presented in Appendix 3.

Integrated Science Education

Curriculum integration of science has had a long and chequered history. In the 1970s, when integrated science education was gaining popularity, Haggis and Adey (1979) were able to collect data on 130 integrated science curricula worldwide. Most of these curricula included two or more science subjects, such as physics and chemistry, but an increasing number also included a nonscience subject, or 'social component'. Integrated curricula connect a range of subjects in a diversity of ways, varying from parallel teaching of concepts, such as proportional reasoning in mathematics and science, but the subjects remain separate, to fully integrated curricula that retain little, if any, distinction between the subjects we usually see listed in school timetables. Often courses described as integrated are very similar to those we have referred to as 'science in context'. The underlying rationale for integrating the curriculum is that the real world is not divided into separate subjects; therefore, students are better served by a curriculum that is more like the real world because it would seem to be more relevant to their needs.

Curriculum integration has met strong opposition from those adhering to the primacy of the discipline, arguing that students need to learn the established, canonical disciplinary knowledge that is at the core of each subject curriculum. Proponents argued that the content of a curriculum should start with the needs of the learner, not the discipline, and hence the curriculum is directed towards the issues, problems and concerns students will need to deal with in real life. However, this does not obviate the need for some disciplinary knowledge. Beane (1995), a notable proponent of curriculum integration, explained it this way:

> Curriculum integration…calls forth those ideas that are most important and powerful in the disciplines of knowledge—the ones that are most significant because they emerge in life itself. And because they are placed in the context of personally and socially significant concerns, they are more likely to have real meaning in the lives of young people, the kind of meaning they do not now have. (p. 620)

Assessment in an integrated curriculum poses problems for those committed to testing only for discipline-related knowledge. Rennie et al. (2011) described how researchers studying integrated curriculum are apt to reduce their measurement of learning outcomes to traditional content-based measures, ignoring the proposed affective and other noncognitive outcomes that underpinned the original purpose of integration. Rennie et al. (2011) demonstrated that a much broader conception of learning outcomes, and therefore a comprehensive approach to measuring them, is required to assess what is learned from an integrated approach. Assessment procedures in integrated curriculum are easily embedded in real-world contexts because the real world itself is integrated with respect to the disciplines. Assessment in integrated contexts allows simultaneous assessment of several disciplines in the one task, and it also provides opportunities for authentic assessment. In the following discussion, we draw from two accounts of authentic assessment requiring students to demonstrate a range of knowledge and skills, all underpinned by their immersion in real-world issues.

Venville et al. (2008) described how Kentish Middle School used an integrated program for its learning community of 120 6th and 7th grade students (aged 11–12 years) during which the science, mathematics, English and social studies teachers used the local lake as a focus to achieve the curriculum objectives over a term. Aligned with the content outcomes described in the local jurisdiction's curriculum framework were five clusters of 'core shared values': a pursuit of knowledge and a commitment to achievement of potential, self-acceptance and respect of self, respect and concern for others and their rights, social and civic responsibility, and environmental responsibility. (Rennie 2007 described these values in more detail.) The curriculum framework statements referring to these values were displayed in a notice on the classroom wall, and these values infused the term's program relating to the preservation of the lake and its ecology. Mr Keane, the social studies teacher, explained in an interview how students would work together to produce a model city and the kinds of foci related to how urbanisation affected the lake:

> They've got to work together in groups of four or five. And for instance with the city they've got to produce a model of the city, so there's probably a bit of technology, there's a lot of maths in that. Actually a model of the city. But then to actually illustrate five ways that city will actually save the lake. That's based on the outcomes that they're going to study in science, to apply them to a city of their own creation. So, yeah they can talk about food webs, they can talk about habitats, and then how a city can actually retain habitats or destroy habitats. … Salinity they can talk about. How a city can actually contribute or reduce salinity. Extinction or feral—the whole life of a living [thing], so what does a cat in a city do? What does it mean for a lake? So on the life and living side [the science topic for the term] there's a real focus in that. Pollution, recycling's probably another. So there's six or eight things there they can apply.

The curriculum activities over a term in one class of the learning community were described by Venville et al. (2008, p. 865) and included several excursions to the lake and to a water treatment plant, guest speakers, doing fieldwork in the school grounds, students collecting their own data about water quality, pollution and recycling, effects of land clearing and so on. The activities were built around the lake, the home, the city, the water and the garden, and all activities culminated in the two major summative assessment items, an 'expo' of students' model houses (based on the English subject requirement for an expository description of work done) and a role play about further development around the lake. Mr Keane explained:

> And in an expo, our expo, each of the learning areas from the lake, the home, the city, the water, and the garden, they come together as a sharing thing. So this is how we save the lake from the city and they [students] actually give a short speech with a poster or a model. This

> is our model of a city. This aids the lake by doing this. So, okay, that's for five minutes. Then here's our model of a home, or a poster of a home, or, and this is how this saves the lake. Here's our diagrammatic—and these are the maths guys—here's our map of a garden that saves the lake and here's where we've planted natives [plants] and here's where the reticulation goes in and there's seventeen metres of it. So they can provide their area of brilliance for a real problem. … We're going to actually assess speaking and listening as an English assessment. … In the afternoon we've actually got a formal assessment on that, so again, because of the expo, we'll actually derive an expo, they'll be given a short question, a short statement, and they can either take the positive or the negative [view] on that and write an expo of that.

The role play's explicitly stated outcomes were: 'Students will experience and understand the competing interests that surround every development decision. Students will review the science and social issues relevant to the development of a wetland area'. Again working in groups, students developed arguments for either the developers, the existing townspeople, the Friends of the Lake, the Lake Sailing Club or the Kentish City Council. Students were provided with background notes for each group; for example, the Friends were concerned about pressure placed on the colony of blue-eared turtles due to tree clearing. Students were able to build on their term's work to develop their argument for or against development and a strategy forward and then, in a town meeting, present it to the Minister for Planning—a role played by an 8th or 9th grade student (aged 14–15 years and in a leadership program)—who listened to the arguments, gave a considered decision and was questioned by the students about it.

Clearly, students' presentations in both the expo and the town meeting gave scope for teachers to assess progress in aspects of the core shared values as well as some of the other content outcomes achieved. But these were not the only assessment tasks; assessment was a continuous process at Kentish Middle School. Mr Keane explained how teachers developed a profile of work for each child:

> We've never expected Kentish to have an assessment that actually gives everything in one time. We try and devise a lot of options for assessment and that, so the kids having a bad day, they don't get a bad mark on that [aspect] and we can pick up that they leave with the dominant skills, that they weren't real strong on speaking and listening last week but with what they've done there we can actually step [up] their [mark]. … Don't expect that Thursday, week nine everyone's going to do a test and they're gonna walk away; that's your assessment piece for the year. We … try and build in these holistic, these small assessment pieces but collect some data, collect some data, collect some data, collect some data. At the end of it, of this, we can actually say something about the kid.

Analysis of a local problem involving the whole class in science-based activities, but integrating other subject areas, has also been used for the straightforward purpose of motivating students. Yoon (2005) used a town hall meeting in association with staff from an outdoor education centre to help demonstrate to students that science is, 'among other things, a socially constructed enterprise, where decisions are based on critical evaluation of multiple points of view and influences on society' (p. 55). In this Canadian location, a family of beavers had moved into the area and as their presence was creating some environmental changes, a decision was to be made about whether or not to relocate them. Yoon's students were 9th graders (15 years) and although this was their first role play, she gave only minimal instructions to help students focus on the beaver issue. Pairs of students were placed into six special interest groups, such as farm owners, residents, and news reporters. Three other students were assigned to be the Town Hall Council, and they were to have no prior opinion but to hear arguments about the relocation of the beavers and make a reasoned decision. During a class visit to the outdoor education centre, students were treated by staff as members of their special interest groups and given a 90-min presentation of historical background, statistical information and unique characteristics of the forest ecosystem. After a period of collaboration to gel the points relevant to their special interest, students were guided through the area, noted flora and fauna and examined changes induced by the beavers. Back at school, students further researched the issue and at the town hall meeting, held one afternoon, positions were put, questions asked and answered, and the Town Hall Council gave their verdict.

Yoon does not record how assessment was made, but she described 'the atmosphere in class was one of sheer jubilance … students laughing and talking about the meeting on their way out' (p. 62). Her aim to motivate the unmotivated was clearly achieved. In addition, she pointed out the salience of this learning approach in enabling usually low achievers to demonstrate their learning in ways other than through the traditional testing regime.

The kinds of tasks described above fulfil all of the criteria earlier enumerated for context-based tasks and also subsume the components underpinning decision-making. Specifically, the above tasks demonstrate the following features:

1. The task is set in a context familiar to the students and has real, rather than contrived, significance to the students' community.
2. Completing the task draws upon knowledge and skills from more than one subject area, hence demonstrating cross-subject application of concepts and also providing opportunities for assessment across areas on the same tasks.
3. Collecting data first-hand provides opportunities to practise investigative skills and gives immediacy to the findings and the context provides meaning to the outcomes.
4. Teachers have opportunities to observe students' progress on a range of skills, attitudes and values not amenable to pencil and paper testing.
5. The decision-making culmination of the activity provides closure to a sequence of tasks, demonstrates the interaction between concepts from more than one learning area and gives students a sense of achievement.

The results seem to be a much greater understanding of the human and contextual issues surrounding decision-making. As Mr Keane pointed out when interviewed about the Kentish students after their assessments:

> They've just changed their attitudes. They're not reactive anymore. They're actually active. … They are actively making decisions and changing their world … they need to have that sort of personal power, and in a way, on a global scale, they're empowered.

Implications for Teachers of an Authentically Assessed Curriculum

The sets of assessment examples described above place responsibility very clearly on teachers. For competent teachers like Mr Keane, this was no problem. He firmly believed in authentic assessment (although he did not call it that) and the development of a profile of achievement for each student. However, Mr Keane taught in a middle school, and it is at the senior school level where the outcomes of assessment have the most significant consequences for students' future. Such high-stakes demand more than just a teacher-produced assessment score; they demand that such assessment be not only fair but comparable across all schools where the same course is offered and comparable across courses (see also Black, this volume).

There are numerous mechanisms to ensure evenness of judgement across teachers and schools. The most successful of these has long been in operation in some other subject areas like Art and Technology. In these subjects students are required to produce a series of artefacts from which the quality of their effort is to be inferred. Small teams of peer teachers visit several schools to provide a commentary on the assessments that individual teachers have made. This type of peer moderation has not been commonly used in science, but, as we are arguing, more and more desirable aspects of science teaching and learning require the classroom teacher to be the prime assessor. Even though in science artefacts are not produced, as in Art, Design, and Technology, it is possible for teachers to present and explain to their peer teachers the bases for their students' assessments.

The process of moderation is highly developed in the Australian state of Queensland (see http://www.qsa.qld.edu.au/586.html), where there has been no external assessment associated with entry to higher education since 1972. Wilson and Sloane (2000) define moderation as 'the process in which teachers discuss student work and the scores they have given that work, making sure that the scores are being interpreted in the same way by all teachers in the moderation group' (p. 201). Clearly, this is an active process for teachers who must discuss samples of student performance, analyse the basis for scoring and come to agreement on the level of that performance. This kind of consensus, or social, moderation is a means of achieving consistency in qualitative judgements of student performance and hence contributes to quality assurance, that is, 'a process for establishing confidence in the quality of procedures and judgements' (Maxwell 2010, p. 457). When these assessments are demonstrated to be consistent before they are used or reported, then the moderation serves the purpose of quality control.

Moderation also serves an additional role, and that is the professional development of teachers taking part in the moderation process. In his discussion of summative assessment, Daugherty (2010) emphasised the importance of teacher training and support in a systematic approach to summative assessment where teachers are to have an active role. An acknowledged outcome of peer moderation has been that, besides some minor adjustments to an individual student's performance grade in the process of developing consistency across performances, there is a great deal of learning by the teachers involved. Not only does their appreciation of the meaning of the explicit criteria become sharpened but the seeds of new possibilities and criteria emerge when exceptional student performance is encountered. One might say that the process of summative assessment for the students becomes a process of formative assessment for the teachers. Furthermore, teachers report great satisfaction when they experience this type of professional peer-peer interaction. In a case study involving teachers in several schools, Black (this volume) describes the enthusiastic response of the teachers to their first experience of moderation despite it being an additional demand. In contrast, Kuiper, Folmer and Ottevanger (this volume) provide examples of how a lack of clarity in assessment guidelines resulted in confusion for pilot teachers involved in efforts of curriculum renewal, who were then not sure just what they were expected to teach. Professional interaction, or moderation of the kind described above, then promotes both confidence and commonality in teachers' interpretation of the assessment process.

Peer moderation by teachers is an expensive form of assessment when applied across a large school system, but this cost and its proven effectiveness needs to be compared with the millions that are regularly spent on ineffectual and less effective activities for professionally developing teachers. Teachers cannot become skilled in formative and summative assessment unless they understand very clearly the aims of the science curriculum they are teaching. Millar (this volume) points out the valuable role that the set of assessment tasks plays in clarifying for teachers the meaning and purposes embedded in curriculum statements.

Wiggins and McTighe (2005) discuss authentic assessment as an outcome of using the 'backward design' process, wherein the desired learning outcomes (or results) are carefully documented first. The appropriate performance tasks and other evidence required to assess whether those outcomes have been achieved are identified, and only then, by working backwards, are the various teaching strategies and learning opportunities chosen that would allow students to succeed on those tasks. By designing the performance tasks before the learning program, assessment will become more congruent with the breadth of desired learning. Further, because assessment is such a driving force for what is taught, the implemented curriculum will more closely reflect course aims, because, as we said at the beginning of this chapter, teachers generally do their best to ensure their students learn what has been declared to be learning of value.

Developing the skills of science teachers for their roles as formative and summative assessors of learning should be a high priority for all education systems that wish to promote a rich set of learning outcomes for their science curriculum. Only then can they hope to see the principles of assessment set out earlier in this chapter being enacted and supported by the day-to-day practices that occur in their science classrooms.

Appendix 1: Extended Response Task for Context-Based Assessment

(Source: Fensham, 2011, personal communication, adapted from an Extended Response Task used in Queensland schools)

Vehicular Motion

Introduction

Fundamental to the performance of any motor vehicle is how well it grips the road. Tyre and brake performance, friction and road surface conditions all have a part to play in determining how well a vehicle performs. In this extended response task, you are required to investigate these and other aspects relating to the performance and safety of modern motor vehicles. By the end of this task, you should have a good understanding of how tyre skid marks left at a motor vehicle accident scene can help determine who is at fault in an accident.

Part A: Knowledge and Conceptual Understanding

Task 1a. Explanation of Concepts Associated with Vehicular Motion

Investigate *and provide written explanation of the concepts and processes listed below. Your explanation should include a scientific description of the concepts and an example of how they relate to vehicular motion.*

Acceleration	Equations of motion	Reaction time
Coefficient of sliding friction	Newton's First Law	Reaction distance
Friction force	Newton's Second Law	Braking distance
Tyre tread	Newton's Third Law	Vehicle stopping distance

Task 1b. Scenario

A driver undergoing safety training brings a vehicle initially travelling at a constant velocity to rest over 55.2 m. The driver's reaction time is 1.2 s and the coefficient of sliding friction is 0.7. Tyre marks are left on the road. ***Calculate*** *(a) the vehicle's initial speed, (b) the vehicle's reaction time, (c) the vehicle's braking distance and (d) the length of tyre skid marks. (Show all working.)*

Task 2. Explanation of Concepts Associated with Vehicular Motion

Reflecting *upon your understanding of the concepts associated with vehicular motion,* ***formulate*** *a list of factors that would affect the braking performance of a modern motor vehicle. For each factor in your list, justify how it relates to the braking performance of the vehicle (250 words maximum).*

Task 3. Scenario

As a leading traffic engineer, you are commissioned to write an article titled ***Factors Affecting Motor Vehicle Braking Performance in Wet Weather on Concrete Pavements*** *for a leading scientific publication. Use the information you have gathered during the research task above to write the article. The article must be between 600 and 700 words in length, contain tables and/or graphs and be referenced as for your school's standard reference policy.*

Part B: Investigative Processes

Task 4a. Analysis of Secondary Data

Refer to the experimental data in the table below for ***skid mark length*** *and* ***initial speed*** *at various coefficients of friction. Use Microsoft Excel or your Casio Class pad calculator to graph the* ***initial speed*** *versus* ***skid mark length*** *at each of the coefficients of friction.*

Qualitatively describe *the relationship between initial speed and skid mark length. Is a linear relationship a good approximation? What* ***effect*** *does the coefficient of friction have? Justify all conclusions.*

Length of skid mark and initial speed data for various coefficients of friction

$\mu = 0.4$		$\mu = 0.6$		$\mu = 0.8$	
Length of skid mark (s) in m	Initial speed (u) in m/s	Length of skid mark (s) in m	Initial speed (u) in m/s	Length of skid mark (s) in m	Initial speed (u) in m/s
3.0	4.8	3.0	5.9	3.0	6.9
7.0	7.4	7.0	9.1	7.0	10.5
10.0	8.9	10.0	10.8	10.0	12.5
20.0	12.5	20.0	15.2	20.0	17.7
30.0	15.0	30.0	18.9	30.0	21.6
40.0	17.9	40.0	21.6	40.0	25.1
50.0	19.8	50.0	24.2	50.0	28.1

Task 4b. Analysis of Secondary Data

Investigate the theoretical relationship between the initial speed of a vehicle which slides to a halt and the length of its skid marks. **Show** how this relationship is **derived.**

Reflect upon the answer that you provide in Task 4a. Does the theoretical relationship agree with your conclusion?

Task 4c. Formulation of a Justifiable Question

Using your research, **formulate a hypothesis** that could be used as a basis for an investigation into the relationship between initial speed of a vehicle and length of skid marks.

Part C: Evaluating and Concluding

Task 5. Exploration of a Scenario

Introduction

An important aspect of automobile accident reconstruction is the **analysis** of tyre track skid marks.

The length of skid marks, along with data on a car's tyres and the road surface, allows an accident reconstruction engineer to make a good estimate of a car's speed just before the driver hits the brakes.

This information can assist in determining who is at fault in an accident.

The Scenario

You are a senior traffic incident investigation officer with the Queensland Police Force and have been called to the scene of a serious accident.

The incident involves a motorcycle and a car at the intersection of Charles and Camilla Streets.

Although the two vehicles narrowly avoided colliding, the sole passenger riding in the car was unrestrained and was flung from the front passenger seat into the dashboard, sustaining injuries and was rushed to hospital. The motorcycle rider and the driver of the car were uninjured.

Witness statements were recorded by uniformed police immediately after the incident.

From these statements the following is clear:

1. The motorcycle was heading south along Charles St and went through the set of traffic lights just as they were turning amber.
2. The car was heading east along Camilla St.
3. Both vehicles screeched to a halt and narrowly avoided colliding.

You have also received a call from a senior emergency doctor at the hospital. She is concerned as the female passenger has sustained internal injuries.

In order to have a better understanding of the stresses imposed on her internal organs during the collision with the dashboard, the doctor needs to know the relative speed with which the female passenger hit the dashboard.

Task 5a. Exploration of a Scenario

Calculate the relative speed with which the female passenger hit the dashboard (assume that the distance from the front seat to the dashboard is 1,500 mm). The diagram provided below may assist you in answering this question (show all your working). Assume that the coefficient of sliding friction for the car is 0.92 and for the motorcycle is 0.85.

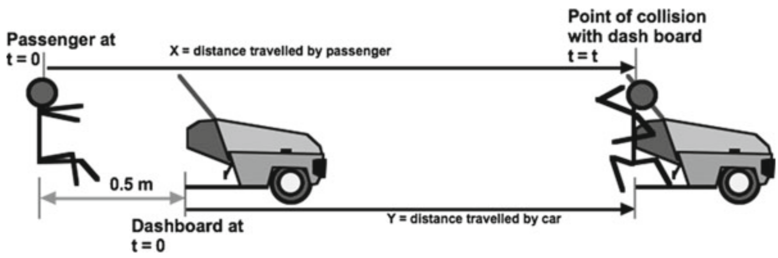
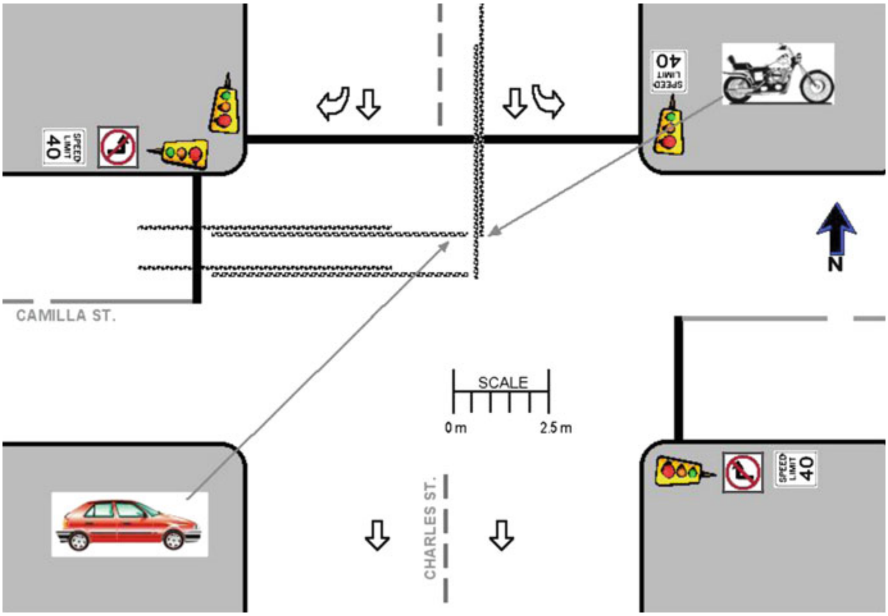


Diagram of collision as witnessed by a ground-based observer

Task5b. Exploration of a Scenario

In a report of no more than 300 words, state whether you believe charges need to be laid on either driver. **Justify** your conclusion with scientific reasoning (you will need to refer to the diagram of the incident scene shown below).



Appendix 2: Reflective Judgement Interview Standard Probe Questions

(Source: Zeidler et al. 2009, p. 97)

Probe question	Purpose
1. What do you think about these statements? Note: If no particular point of view is endorsed, ask 1a) Could you ever say which was the better position? How? Why not? How would you go about making a decision about this issue? Will you ever know for sure which is the better position? How? Why not?	To allow participants to share an initial reaction to the problem presented. Most state which point of view is closer to their own
2. How did you come to hold this point of view?	To find out how the respondent arrived at the point of view and whether and how it has evolved from other positions on the issue
3. On what do you base that point of view?	To find out about the basis of the respondent's point of view, such as personal evaluation of the data, consistency with the expert's point of view or a specific experience. This provides information about the respondent's concept of justification

(continued)

Probe question	Purpose
4. Can you ever know for sure that your position on this issue is correct? How or why not?	To find out about assumptions concerning the certainty of the knowledge (e.g. whether issues like this can be known absolutely and what the respondent would do in order to increase the certainty, or why that would not be possible)
5. When two people differ about matters such as this, is it the case that one opinion is right and one is wrong? If yes, what do you mean by 'right'? If no, can you say that one opinion is in some way better than the other? What do you mean by 'better'?	Assesses the adequacy of alternative interpretations: to see if dichotomous either/or view of the issue (characteristic of the early stages) is held; to allow the participant to give criteria by which she/he evaluates the adequacy of arguments (information that helps differentiate high- from middle-level stage responses)
6. How is it possible that people have such different points of view about this subject?	To elicit comments about the respondent's understanding of differences in perspectives and opinions (what they are based on and why there is such diversity of opinion about the issue)
7. How is it possible that experts in the field disagree about this subject?	To elicit the respondent's understanding of how she/he uses the point of view of an expert or authority in making decisions about controversial issues (such as whether experts' views are weighted more heavily than others' views, and why or why not)

Appendix 3: Sample NOS Assessment Question

(Source: Allchin 2011, p. 520)
Revised Mammogram Recommendations, November, 2009

A female acquaintance of yours is just turning 40. Concerned about the possibility of breast cancer, she had planned to get a mammogram in the next few months, despite her fears about excessive radiation. She has heard that a major national task force now advises waiting until 50 yet finds reassurance in *Women's Health* Magazine about still following the old guidelines.

You both knew another woman who was diagnosed unexpectedly with breast cancer at age 43 and died last year. Your acquaintance is unsure how to interpret the apparently conflicting information and asks your help. What analysis of this reported change in scientific consensus would you provide to inform her decision?

Resource documents: (web addresses provided for each document)

- Women's Health magazine article (Feb. 6, 2010)
- New York Times article (Nov. 17, 2009)
- US Preventive Services Task Force report, recommendation & supporting statement (Nov. 2009).
- Editorial published in Annals of International Medicine (Feb.15, 2009)
- Editorial published in Annals of International Medicine (Feb.15, 2010)

This question links to nine of the ten Dimensions of Reliability in Science in Allchin's (2011, p. 525) scoring rubric.

Appendix 4: Star Diagrams of Dimensions of Assessment

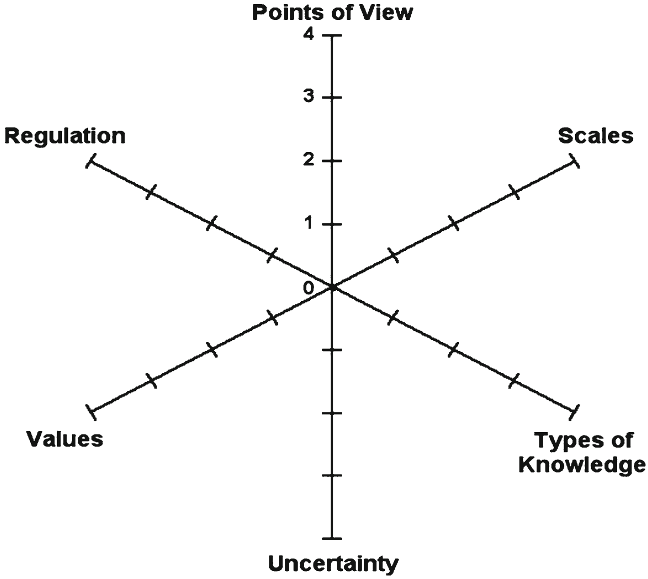


Fig. 5.1 Six dimensions of assessment for reasoning about socioscientific issues (After Simonneaux 2010)

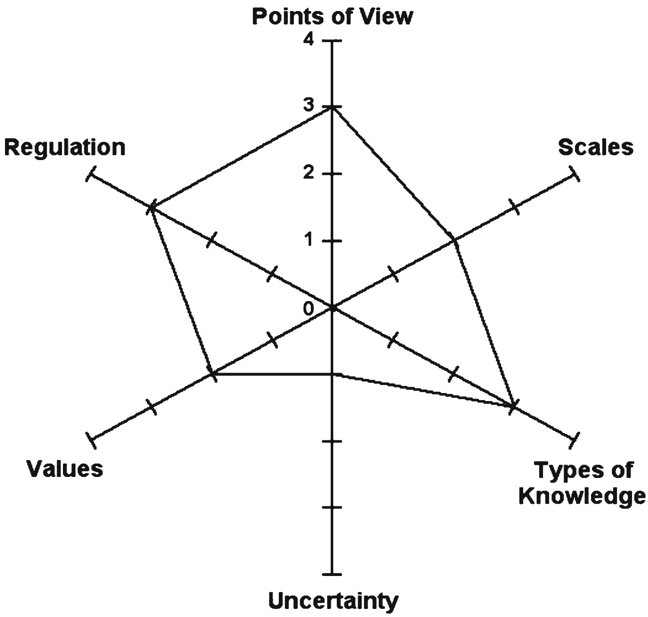


Fig. 5.2 Star diagram assessing performance on six dimensions of assessment for reasoning about socioscientific issues

References

ACARA. (2010). *The Australian curriculum: Science*. http://www.australiancurriculum.edu.au/Science. Accessed 17 Dec 2010.

Aikenhead, G. S. (1989). Decision-making theories as tools for interpreting student behavior during a scientific inquiry simulation. *Journal of Research in Science Teaching, 26*(2), 189–203.

Aikenhead, G. S. (1994). Collaborative research and development to produce an STS course for school science. In J. Solomon & G. Aikenhead (Eds.), *STS education: International perspectives on reform* (pp. 216–227). New York: Teachers College.

Aikenhead, G. S., & Ryan, A. G. (1992). The development of a new instrument: "Views on Science-Technology-Society" (VOSTS). *Science Education, 76*(5), 477–491.

Allchin, D. (2011). Evaluating knowledge of the nature of (whole) science. *Science Education, 95*, 518–542.

Beane, J. A. (1995). Curriculum integration and the disciplines of knowledge. *Phi Delta Kappan, 76*(8), 616–622.

Biddulph, F., Symington, D., & Osborne, R. (1986). The place of children's questions in primary science education. *Research in Science & Technological Education, 4*(1), 77–88.

Black, P. (2004). Personal communication with Peter Fensham. Cited in P. J. Fensham (2004). *Defining an identity: The evolution of science education as a research field* (p. 173). Dordrecht: Kluwer.

European Commission. (2007). *Science education now: A renewed pedagogy for the future of Europe*. Brussels: European Commission, Directorate-General for Research.

Daugherty, R. (2010). Summative assessment by teachers. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopaedia of education* (Vol. 4, pp. 384–391). Oxford: Elsevier.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*(3), 287–312.

Eijkelhof, H. M. C., & Koortland, K. (1988). Broadening the aims of physics education. In P. J. Fensham (Ed.), *Development and dilemmas in science education* (pp. 182–305). London: Falmer.

Erduran, S., & Jiménez-Aleixandre, M.-P. (Eds.). (2007). *Argumentation in science education: Recent developments and future directions*. New York: Springer.

Erickson, G., Bartley, T., Carlisle, R., Meyer, K., & Stavy, R. (1992). *The 1991 British Columbia assessment of science: Part II: Student performance component*. Victoria: Queen's Printer.

Fensham, P. J. (1985). Science for all: A reflective essay. *Journal of Curriculum Studies, 17*, 415–435.

Fensham, P. J. (2006). Foreword. In B. H. W. Yung (Ed.), *Assessment reform in science: Fairness and fear* (p. i). Dordrecht: Springer.

Fensham, P. J. (2009). Teaching science to achieve scientific literacy. In R. W. Bybee & B. J. McCrae (Eds.), *PISA SCIENCE 2006: Implications for science teachers and teaching* (pp. 187–202). Arlington: NSTA.

Gilbert, J. K. (2006). On the nature of "context" in chemistry education. *International Journal of Science Education, 28*(9), 957–976.

Gilbert, J. K., Bulte, A. M. W., & Pilot, A. (2011). Concept development and transfer in context-based science education. *International Journal of Science Education, 33*(6), 817–837.

Haggis, S., & Adey, P. (1979). A review of integrated science education worldwide. *Studies in Science Education, 6*, 69–89.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey-Bass.

Kolstø, S. D. (2001). Scientific literacy for citizenship: Tools for dealing with the science dimension of controversial socioscientific issues. *Science Education, 85*(3), 291–310.

Layton, D. (1994). STS in the school curriculum: A movement overtaken by history. In J. Solomon & G. Aikenhead (Eds.), *STS education: International perspectives on reform* (pp. 32–44). New York: Teachers College.

Maxwell, G. (2010). Moderation of student work by teachers. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopaedia of education* (Vol. 3, pp. 457–463). Oxford: Elsevier.

Morin, O., & Simonneaux, L. (2010). *Caractérisation de les raisonnements socioscientifiques dans une perspective d'éducation au développement durable: Apports et limites du traitement d'une* *controverse environnementale par le débat en formation professionnelle d'enseignants*. Geneva: Actes du congrès international de l'Actualité de la Researche en Education et en Formation.

National Curriculum Board. (n.d.). *National science curriculum: Initial advice*. http://www.pdfdocspace.com/docs/37294/science-initial-advice-paper.html. Accessed 17 Dec 2010.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academies.

National Research Council. (2011). *A framework for K-12 science education; Practices, cross-cutting concepts, and key ideas*. Washington, DC: National Academy.

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

OECD. (2007). *PISA 2006: Science competencies for tomorrow's world*. Paris: Author.

Ratcliffe, M. (1997). Pupils' decision-making about socio-scientific issues within the science curriculum. *International Journal of Science Education, 19*(2), 167–182.

Rennie, L. J. (2007). Values in science portrayed in out-of-school contexts. In D. Corrigan, D. Gunstone, & J. Dillon (Eds.), *The re-emergence of values in science education* (pp. 197–212). Rotterdam: Sense.

Rennie, L. J., Venville, G., & Wallace, J. (2011). Learning science in an integrated classroom: Finding balance through theoretical triangulation. *Journal of Curriculum Studies, 43*(2), 139–162.

Ritchie, S. M., Tomas, L., & Tones, M. J. (2011). Writing stories to enhance scientific literacy. *International Journal of Science Education, 33*(5), 685–707.

Roberts, D. (1982). Developing the concept of 'curriculum emphases' in science education. *Science Education, 66*(2), 243–260.

Ryder, J. (2001). Identifying science understanding for functional scientific literacy. *Studies in Science Education, 36*, 1–44.

Sadler, T. D., & Zeidler, D. L. (2005). The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engineering issues. *Science Education, 89*, 71–93.

Sadler, T. D., & Zeidler, D. L. (2009). Scientific discourse: Assessment for progressive aims of science education. *Journal of Research in Science Teaching, 46*(8), 909–921.

SCC. (1984). *Science for every student: Educating Canadians for tomorrow's world (Report No. 36)*. Ottawa: Science Council of Canada.

Simonneaux, L. (2010). Students' reasoning on socioscientific issues and socially acute questions. In D. J. Boerwinkel & A. J. Waarlo (Eds.), *Genomics education for decision-making: Proceedings of the second invitational workshop on genomics education* (pp. 73–81). Utrecht: CD-β.

Solomon, J., & Aikenhead, G. (Eds.). (1994). *STS education: International perspective on reform*. New York: Teachers College.

Stocklmayer, S. M., Rennie, L. J., & Gilbert, J. K. (2010). The roles of the formal and informal sectors in the provision of effective science education. *Studies in Science Education, 46*, 1–44.

Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. Chicago: The University of Chicago Press.

VCAA. (2010). *Chemistry study design*. Melbourne: Victorian Curriculum and Assessment Authority.

VCAB. (1984). *Chemistry*. Melbourne: Victorian Government Printer.

Venville, G., Sheffield, R., Rennie, L. J., & Wallace, J. (2008). The writing on the wall: Classroom context, curriculum implementation, and student learning in integrated, community-based science projects. *Journal of Research in Science Teaching, 45*, 857–880.

Wellington, J. (Ed.). (1998). *Practical work in school science: Which way now?* London: Routledge.

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). New York: Merrill Education/ASCD College Textbook.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education, 13*(2), 181–208.

Yoon, S. A. (2005). Motivating the unmotivated: Relevance and empowerment through a town hall debate. In S. Alsop, L. Bencze, & E. Pedretti (Eds.), *Analysing exemplary science teaching* (pp. 53–62). Berkshire: Open University.

Yung, B. H. W. (Ed.). (2006). *Assessment reform in science: Fairness and fear*. Dordrecht: Springer.

Zeidler, D. L., Sadler, T. D., Applebaum, S., & Callahan, B. (2009). Advancing reflective judgment through socioscientific issues. *Journal of Research in Science Teaching, 46*(1), 74–101.

Chapter 6
Aligning Science Curriculum Renewal Efforts and Assessment Practices

Wilmad Kuiper, Elvira Folmer, and Wout Ottevanger

Curriculum, Assessment and Change

Change in education is easy to propose, hard to implement and extraordinarily difficult to sustain. Pilot projects show promise but are rarely converted into successful system-wide change (Hargreaves and Fink 2006). Unfortunately, these general observations are equally valid for renewal initiatives in the domain of science education (Atkin and Black 2003; Kuiper 2009). The fate of improvement initiatives in science education should prevent us from being overly optimistic about the chances for success, especially preventing us from expecting quick results (van den Akker 1998). Most changes hardly ever pass the stage of intentions, as the implemented curriculum tends to be very durable. There is far more continuity in what teachers actually do in their classroom than is suggested by changes in the intended curriculum (Cuban 1992).

The curriculum refers to the content and purpose of an educational programme together with the organisation of these (Walker 1990). The core of a curriculum generally concerns the aims and content of learning. Changes to this core usually presuppose changes to many other aspects of [the plan for] learning. A clarifying way to visualise the relationship between the various aspects of a curriculum is the so-called curricular spiderweb (van den Akker 2003; Thijs and van den Akker 2009; see Fig. 6.1). At the core of this conceptual model is the rationale (the overall

W. Kuiper (✉)
Netherlands Institute for Curriculum Development (SLO)
and Utrecht University, 3508 TC, Utrecht
email: w.kuiper@slo.nl

E. Folmer
Netherlands Institute for Curriculum Development (SLO),
Monash University, Clayton, VIC, Australia

W. Ottevanger
Netherlands Institute for Curriculum Development (SLO) and Vrije University of Amsterdam,
Monash University, Clayton, VIC, Australia

Vrije University of Amsterdam

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_6,

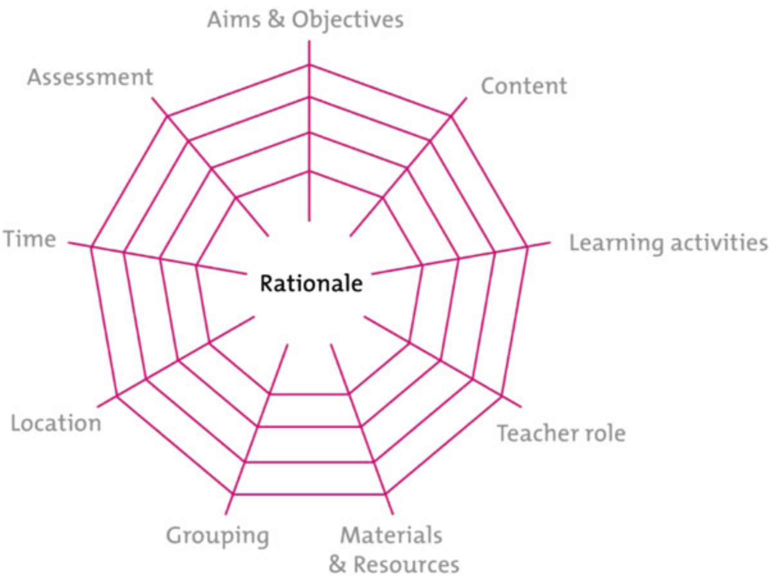


Fig. 6.1 Curricular spiderweb (van den Akker 2003)

principles or the central mission of the curriculum), which connects all the other components: aims and objectives, content, learning activities, teacher role, materials and resources, grouping, location, time and—last but not least—assessment. The spiderweb metaphor emphasises how, within any one curriculum, component aspects of that curriculum may change over time but that any dramatic shift in balance will pull the entirety out of alignment. Though it may stretch for a while, prolonged imbalance will cause the system to break. Efforts to reform curricula must therefore devote attention to balance and linkages between all ten components (McKenney et al. 2006). The relevance of the various components varies, among others things, across the levels of the curriculum (macro, meso, micro). Curriculum documents at the macro level (e.g. the examination programmes mentioned later in this chapter) usually focus on aims/objectives, content and also sometimes a rationale (often in rather broad terms), on occasion accompanied by an allocation of time. When one takes textbooks, generally all ten components are addressed and have to be coherently addressed in order to expect successful implementation. The component of assessment deserves separate attention at all levels, since careful alignment between assessment and the other curriculum components is critical for successful curriculum change (van den Akker 2003).

As regards science education, assessment is perceived as essentially an afterthought of curriculum change and as an undervalued and underinvested aspect of curriculum development (see Harrison in this volume) or as the most neglected dimension (Millar 2010). There is ample evidence that failure lies ahead when curriculum renewal aims and objectives are only inadequately reflected (or not reflected at all) in assessment approaches, procedures and instruments to be linked with the renewal. Changes in curriculum goals and content require concurrent changes in assessment and examination policies and practices. Or, to put it differently, effective assessment must match the nature of the learning experiences designed to achieve the curriculum aims (see Fensham and Rennie in this volume). Because of its 'backwash' effect (cf. Watkins et al. 2005), assessment can be either an obstacle preventing or a lever assisting curriculum change efforts. Assessment may be an obstacle when assessment and examination remain unaffected by changes in the intended curriculum and/or the implemented curriculum. An example of this is when teachers teach inquiry and problem solving while their students suffer on mandated tests that focus mainly on terminology and computation (Kulm and Stuessy 1991, p. 74). Assessment as a lever, described by Kulm and Malcolm (1991) in the title of their well-known book 'Science assessment in the service of reform', may occur when intended changes do concur with changes in assessment and examination. An example of this is when an intended and implemented focus on reasoning skills is adequately reflected in a test or in an examination. So, assessment is a change agent (Strobart 2005) and for that reason, restructuring assessment is one of the most significant challenges facing science education curriculum reform (Kulm and Stuessy 1991).

Assessment may be viewed from various perspectives:

– Assessment *of* learning (Black 1998; Black et al. 2003), emphasising assessment designed to demonstrate whether or not students have met curriculum goals, with the purpose either to help make schools accountable or to enable schools to provide certificates. The latter purpose 'calls for assessment methods which can be reliable, in that they are comparable across different schools, and indeed across the country as a whole, and also valid in that they give the users what they really need to know about each student' (Black et al. 2003, p. 1).
– Assessment *for* learning (Black 1998; Black et al. 2003; Cowie and Bell 1999; Wiliam 2011; see also in this volume: Black; Nilsson and Loughran; Lee Hang and Bell), with the purpose of promoting students' learning. 'Assessment for learning is usually informal, embedded in all aspects of teaching and learning, and conducted by different teachers as part of their own diverse and individual teaching styles' (Black et al. 2003, p. 2).
– Assessment *as* learning (Earl 2003), emphasising assessment as a process of metacognition for students.

Whatever perspective is taken, assessment may pertain to various levels of curriculum and curriculum development: the level of the individual student, the classroom level (written and oral tests, combined with alternative modes of teacher-initiated assessment), the school level (school-wide assessment and examination policy, including, for instance, the way internal, school-based examinations are organised and administered) and the system/national level (external, high-stakes examinations).

This chapter focuses on the importance of the alignment between science curriculum renewal efforts and assessment *of* learning by means of external end-of-school (exit) examinations, and the traps involved in seeking this alignment. Three recently completed science curriculum renewal pilots in senior secondary education in the Netherlands (2007–2010) are used as exemplifying cases.

Biology, Chemistry and Physics Curriculum Renewal Pilots

In an attempt to cope with, among other things, poor coherence within and across subjects, lack of content relevance for large groups of students (especially girls in the case of physics), and fragmentation and content overload, in 2005, the Dutch Ministry of Education mandated subject-specific curriculum renewal committees to develop and pilot new examination programmes for senior secondary biology, physics and chemistry education. Each committee was chaired by a university professor in the respective area of expertise and consisted of a number of science education experts, science educators, and teachers. Students did not participate in the committees.

Senior secondary education in the Netherlands encompasses grades Secondary 4 and 5 of the senior general education track (SGE, a 5-year programme preparing students for higher vocational/professional education) and grades Secondary 4, 5 and 6 of the preuniversity education track (PUE, a 6-year programme preparing students for university and higher vocational/professional education). On the initiative of the renewal committees, the intended renewals embodied a context-based approach. In this approach, students learn chemical, physical, biological and mathematical concepts 'in context' (CVS 2003; Goedhart 2004) and are taught to recontextualise concepts in contexts other than the context in which concepts were originally taught and learned. Transfer of a concept from one context to another occurs via abstraction (cf. van Oers 2001, 2004). Originally, the approach was inspired by reform initiatives like *Salters' Chemistry* in the UK and *Chemie im Kontekst* in Germany, which in turn were based on an analysis carried out in the late 1990s by Millar and Osborne (1998). The key premise, rooted in empirical evidence gathered in the UK about the appealing and motivating effect of context-based science education on students, is that such an approach is beneficial in addressing at least some of the issues mentioned at the beginning of this paragraph (Kuiper 2009). The context-based approach was defined as a curricular model for selecting, arranging and rearranging objectives and contents (i.e. the 'what', based on contexts and concepts rather than only on concepts) as well as a pedagogical approach (the 'how').

Starting from a context-based approach, the renewal committees developed a subject-specific *vision document* outlining their views on science education and how the programmes would help to support these views (comparable with a 'rationale' depicted in Fig. 6.1). The renewal committees also developed draft *examination programmes* for each of 'new biology', 'new chemistry' and 'new physics', outlining the goals to be attained and tested in high-stakes external and internal exit examinations. The external examinations pertain to about 60 % and the internal examinations to about 40 % of the goals. Contrary to the external examinations, the internal examinations are school based. Based on the pilot examination programmes, enacted as 'working versions' by the Ministry of Education, the Examinations Board (a government agency) elaborated pilot *syllabi* (including some sample examination items), further specifying the goals to be tested in the external exit examination.

The committees also developed pilot *teaching and learning materials* that were intended to exemplify the intended curriculum reform at the classroom level (that is, materials that outlined the ‘how’ of the new curriculum). Based on the pilot syllabi (also enacted as ‘working versions’ by the Ministry), external written examinations were developed by the National Institute for Educational Measurement (Cito). The starting point for the development of these pilot examinations was the regular written examinations developed concurrently (see below). The external pilot examinations were administered to pilot school students in April 2009 (Secondary 5 SGE) and April 2010 (Secondary 5 SGE and Secondary 6 PUE). As for the regular examinations, these took 3 h. The procedure applied to the development of external pilot examinations was in many ways identical to the regular development cycle used for the regular examinations (examination programme development, syllabus development, external examination construction and administration). However, there were two exceptions. Firstly, in this particular case the Ministry mandated subject-specific curriculum renewal committees—and not the Netherlands Institute for Curriculum Development (SLO)—to develop and pilot new examination programmes. Secondly, whereas a regular development cycle takes two years, only 1 year was available for the pilot cycles.

The intended reforms were organised around subject-specific pilots. The pilots were conducted from 2005 through 2010, with the involvement of 7–14 secondary schools per subject from August 2007 through June 2010 (three consecutive school years). The renewal committees advised the Minister positively about the feasibility of the nationwide implementation of their new examination programmes after completion of the pilots at the end of 2010.

Longitudinal Curriculum Evaluation Study

The three pilots were supported by an independent, longitudinal evaluation study conducted by SLO from fall 2006 through 2010 (Bruning et al. 2011; Kuiper et al. 2011a, b; Ottevanger et al. 2011). As a stepping-stone in the evaluation, the conceptual typology of curriculum representations shown in Table 6.1 was used (Goodlad 1979; Kuiper 1993; van den Akker 2003). For a (renewal-inspired) curriculum to be effective, it is a prerequisite that the various curriculum representations closely cohere and are aligned. Of course, there is never a linear, top-down transformation from curriculum intentions via implementation in teaching and learning settings to students’ outcomes. It is a complicated process in which much elaboration and adaptation may be needed and may occur. Also a lot of ‘noise’ may arise. Indeed, there is ‘considerable scope of slippage’ (see Fensham and Rennie in this volume) as a result of what original intentions can be blurred, distorted or even devastated (Kuiper et al. 2005).

All curriculum representations were addressed in the evaluation, including the tested curriculum and the learned curriculum at the level of external pilot examinations for new physics, new chemistry and new biology for both SGE and PUE.

Table 6.1 Typology of curriculum representations

Intended curriculum	*Ideal*	Tenets and guiding principles underlying the curriculum renewal
	Formal	Intention specified in examination programme, syllabus, guidelines and teaching and learning materials
Implemented curriculum	*Perceived*	Interpretations and perceptions of pilot teacher as regards the curriculum renewal
	Operational	Curriculum-in-action
Attained curriculum	*Experiential*	Learning experiences and perceptions of pilot teachers
	Tested	Curriculum as tested and reflected in external exit examination
	Learned	Performances of pilot students on external examination

The evaluation aimed at finding out to what extent the intended curriculum renewals resulted in programmes that are *viable and feasible* (for pilot teachers and their students) and *assessable* (by means of external pilot examinations). It was meant (i) to generate suggestions for improvements of the proposals during the pilots and (ii) to inform a policy decision after completion of the pilots, to be taken by the Minister of Education, about a countrywide upscaling of the reforms. This decision was to be based on research findings from the pilots about positive and negative factors and conditions seen as likely to impact on any whole-country implementation. Data about viability and feasibility were collected by means of teacher questionnaires, student questionnaires and interviews with teachers and students during school visits in grades Secondary 4–6, during three consecutive school years (from school year 2007–2008 through 2009–2010).

In this chapter we confine ourselves to the topic of *assessability*. Partly in cooperation with the Examinations Board and Cito, data have been collected about (i) the *content* and *format* of the external pilot examinations, focusing particularly on whether or not the content and format of the external pilot examinations adequately reflected the renewal specifications and requirements (the intended curriculum), and (ii) the *process* of constructing the external pilot examinations administered in spring 2009 and 2010. This second part focused on the role and the perceptions of the various actors in this process. The research group for this part of the evaluation consisted of, for each subject, the curriculum renewal committee (the 'booster' and 'owner' of the curriculum renewal process), the Examinations Board's subject department (responsible for syllabus development as well as formulation of the confidential 'examination construction order' for Cito), examination construction teams from Cito (for SGE and PUE) and pilot teachers. Data were collected by means of interviews with each group along with a whole-group interview, for SGE in September 2009 and for SGE and PUE in September 2010.

Furthermore, the focus was on the *performances* of the pilot students on pilot examinations in comparison with those of students who took the regular examinations (based on data gathered by Cito). In order to make meaningful comparisons between performances of non-pilot and pilot students, the pilot examinations consisted of three roughly equal parts: regular items ('overlapping items'), slightly adapted regular items (adapted in order to make them better fit the pilot context), and pilot-specific items. The latter set of items was especially intended to be context oriented. Questions were in either the free-response format—requiring students to generate and write answers in terms of calculations, interpretations, explanations and derivations—or in the multiple-choice format.

The focus on performances in pilot examinations was also prompted by empirical evidence gathered in the UK. Based on an in-depth review of five studies dealing with controlled evaluation of context-based courses, Bennett et al. (2003) concluded that there is (some) evidence that context-based science education has an appealing and motivating effect on students but that there is also 'good evidence to support the claim that context-based approaches do not adversely affect pupils' understanding of scientific ideas' (p. 3). Empirical evidence of the possible gains of context-based approaches in conceptual understanding was therefore, at best, inconclusive. The question was what the evaluation of student performances on the pilot examinations could contribute to this knowledge.

The Assessability of the Renewal-Inspired Examination Programmes

The main findings from the evaluation of the external pilot examinations for new physics, new chemistry and new biology as regards content and format, process and performances of pilot students are now outlined.

Content and Format

An overall finding is that it has indeed been possible to develop external examinations based on the draft examination programmes, although as yet not all the renewal features have been reflected in the pilot examinations. When this is considered in terms of Table 6.1 (above), then a range of discrepancies is suggested. These involve discrepancies within the formal curriculum and between the formal curriculum and each of the perceived, the operational and the tested curriculums. A summary of these discrepancies for each of the subjects is now given.

New Biology

For biology, the pilot examinations for both GSE and PUE were considered to be fairly well aligned with the examination programmes. However, a general observation was that there was not a large difference between the pilot examinations and the regular examinations, partly due to the gradual trend over about 10 years towards including contexts in external biology examinations. However, in the opinion of the curriculum renewal committee, the examinations could and should have been more innovative, more in line with the original intentions of the renewal. This could have been realised by not using regular examinations as the starting point for the development of pilot examinations (see also the ‘Process’ section below). According to most stakeholders interviewed, the following renewal features were visible in the pilot examinations:

- The relevance and appeal of the topics for students.
- The increased focus on reasoning skills. Nevertheless, the curriculum renewal committee indicated that a stronger focus would not have been out of place. Cito, however, emphasised that it was quite troublesome to formulate univocal scoring rubrics for open-ended questions requiring students to demonstrate reasoning skills.
- The consistency between questions pertaining to one topic/context.

The following aspects were assessed as not (sufficiently) reflecting the intended renewal:

- The focus on reproduction (factual knowledge) rather than on production.
- The redundancy in the information provided as part of the contexts used (in the opinion of the renewal committee and pilot teachers). However, both the Examinations Board biology section and Cito saw that some information had been appropriately added in order to guarantee content coverage with the syllabi.

Also to be improved were the following:

- The alignment between the examination programmes, syllabi, teaching and learning materials and the content tested.
- The consistency between, on the one hand, the pilot curriculum and learning materials for GSE focusing on the use of life contexts and professional contexts, and, on the other hand, the pilot examinations for GSE, mainly focusing on scientific contexts. The renewal committee strongly and persistently adhered to the use of professional contexts rather than scientific contexts in GSE examinations, as the 5-year GSE programme is intended to prepare students for higher professional (and not scientific) studies. However, the Examinations Board and Cito argued just as persistently that it was quite tricky to operationalise this requirement into examination questions and, in addition, that there was nothing wrong with using ‘scientific’ contexts in GSE (examinations).
- The degree of specification of the intended curriculum renewal. Cito in particular expressed a clear need for more detailed specification.

Another striking observation was that pilot teachers experienced a clear tension between the broadly formulated draft examination programmes and the level of detail required to answer pilot examination questions. This inconsistency made pilot teachers insecure about how to prepare their students for the pilot examination. Further, the degree of freedom offered prompted some teachers to try to teach ‘anything’ (which, in turn, contributed to a feeling of curriculum overload).

New Chemistry

According to all actors interviewed, the pilot examinations for both GSE and PUE seemed to be fairly aligned with the examination programmes. According to most stakeholders, the following renewal features were visible in the pilot examinations:

- The use of contexts as well as the kind of contexts used
- The notice taken—especially in the 2010 examinations—of the micro-macro concept that was an important and substantial aspect of the renewal
- The increased focus on reasoning and giving arguments rather than on reproduction
- The chain of reasoning within the clusters of questions pertaining to specific topics/contexts
- The slightly more open scoring rubrics for the 2010 examinations

An aspect that was assessed as not (sufficiently) reflecting the intended renewal was the large amount of text pilot students were confronted with, especially with regard to the contextually embedded questions. In the opinion of pilot teachers, this made questions much more complex for their students, particularly in comparison with the kinds of questions regular students had to answer.

An aspect that was also criticised was the lack of alignment between the draft examination programmes and the syllabi. The curriculum renewal committee designated the syllabi for GSE and PUE 'as obstacles preventing change'; on the other hand, the Examinations Board chemistry section and the chemistry educators from Cito saw inadequate specification of the intended curriculum renewal as an impediment to the construction of examination questions.

Other striking findings were as follows:

- That pilot students from some schools had not had the opportunity to learn some of the concepts tested because their pilot teacher (for whatever reason) had not been able to teach the relevant module. This points to a lack of alignment between the operational curriculum (teaching practice, including use of teaching and learning materials) and the formal curriculum (especially syllabi).
- That, while pilot teachers said they very much appreciated the more open scoring rubrics, they at the same time asked for scoring guidelines to be clearer and to be able to have more practice in dealing with these.

New Physics

The intended renewal for physics pertains to both a content-related reform (i.e. more attention to technological developments/innovations and topics like quantum mechanics and relativity theory) and a pedagogical one (more 'learning by doing' instead of whole-class teaching). The renewal committee deliberately aimed for 'evolution rather than revolution', among other things because of concerns expressed by physics teachers and physics educators about a gradually increasing focus over the last 15 or so years on contexts at the expense of the assessment of conceptual understanding in the external physics examinations (Kuiper 2009).

A perceived strength was that from the interviews it appeared that cautious experimenting had occurred with making some of the renewal features visible in the external pilot examinations for GSE and PUE, namely, less reproduction and more production and less working with formulae and more emphasis on a qualitative approach. In general, most actors interviewed were more or less pleased with the quality of the examinations, in particular because of the proportion as well as the quality of context-oriented questions in the exams. However, most actors agreed that the insertion of such questions followed an already existing trend. Because of that, the difference between the pilot examinations and the existing examinations was considered to be rather small.

Aspects that were criticised were the following:

– The lack of clarity, especially in the opinion of the Examinations Board's physics department and Cito, about what was new about new physics. This was particularly so at syllabi level. The curriculum renewal committee, conversely, styled parts of the syllabi as 'obstacles preventing change'.
– The too little attention in the external pilot examinations paid to other important renewal features, like the flexible application of concepts to different contexts as well as investigating, designing and modelling.
– The predominant use of rather technical contexts in the 2009 GSE pilot examination; those contexts were considered as less appealing for girls.

Process

An overall finding is that there was a lack of ownership among the various actors towards the intended curriculum renewal. In addition, the communication among the various stakeholders left very much to be desired.

New Biology

Although the curriculum renewal committee did not bear formal responsibility for the construction of pilot examinations, measures were taken by the committee to secure the involvement of pilot teachers in the development process (as constructors/advisors in the Cito examination construction team and as members of the Examinations Board's biology group). In addition, one renewal committee member appraised the appropriateness of regular examination questions for insertion into pilot examinations. The curriculum renewal committee strongly preferred starting the development of the pilot examinations from scratch rather than from regular examinations. Such a procedure was expected to result in more renewal-inspired pilot examinations.

A controversy gradually emerged between, on the one hand, the curriculum renewal committee and pilot teachers, and, on the other hand, the Examinations Board biology team and the Cito biology group. The point at issue was the committee's strong plea for the explicit use of professional contexts in the GSE pilot examinations. The Examinations Board and Cito took the position that a too strong emphasis on those contexts was unrealistic and undesirable (see above). Unease with the situation prompted the latter two actors to ‘some civil disobedience’.

A bottleneck was the initial lack of clarity about the role of the various actors in the pilot. This caused a delay in the development of both sample examination questions and the 2009 GSE pilot examination. Another problem was that syllabus development and lesson materials development were undertaken quite separately. Guiding the development of the curriculum-in-action at the pilot schools were the teaching and learning materials developed by the pilot teachers (under the auspices of the committee), while guiding the development of the pilot examinations were the syllabi developed by the Examinations Board. Interaction between these quite separate groups was almost out of the question. In addition, the phasing in of examinations development did not match well with what happened at the pilot schools. The renewal committee expressed a strong preference for syllabus development as an inextricable part of the examination pilot. However, the Examinations Board (which was, as already noted, responsible for syllabus development) was strong on ‘cooperation based on independency’ and acted accordingly by involving the renewal committee in syllabus development, examination construction and examination evaluation. Another finding is that the renewal committee would have appreciated more ‘renewal creativity’ from Cito. Cito, on the other hand, complained about a lack of specification of the renewal provided by the committee. The inevitable overall conclusion is that, among the various actors, there was little shared ownership of the curriculum renewal as propagated and specified by the curriculum renewal committee.

New Chemistry

Also in the case of the new chemistry pilot, a major bottleneck was caused by the separate routes of syllabus development and lesson materials development. Guiding the curriculum-in-action in the pilot schools were the teaching and learning materials developed with the involvement of pilot teachers. Materials development in regional networks of two or three schools supported by a coach—as well as teachers’ professional learning along with school development—received much attention in the bottom-up renewal approach initiated by the renewal committee for new chemistry. The syllabi were considered to be guiding the pilot examinations (developed by Cito), but because of the predominant role teaching and learning materials (development) played in this particular pilot, the Cito construction team gradually became increasingly frustrated about which of the two possibilities should give the stronger guidance for examination construction—syllabus (in Cito’s opinion ‘yes’, as a matter of principle) or teaching and learning materials (‘no’). The consequence of this was (i) a lack of interaction between syllabus development and the directions of development in the pilot schools, (ii) that syllabus and pilot examination development were formally not part of the pilot, and (iii) that, because of time constraints, there were hardly any opportunities for reflection on and adaptation of pilot examinations. The commonalities with what occurred in the new biology pilot are striking.

Generally speaking, the communication between, on the one hand, the curriculum renewal committee and the pilot teachers, and, on the other hand, the Examinations Board chemistry team and Cito left much to be desired. Even when they were in touch (e.g. curriculum renewal committee and Cito about how to formulate contexts, the need-to-know principle advocated by the committee and the micro-macro concept at the level of pilot examinations), they barely seemed to speak the same language.

New Physics

The renewal committee for new physics, just like the committee for new biology and contrary to how the committee for new chemistry proceeded, applied a top-down curriculum renewal approach: first, the development of a vision document and draft examination programmes for GSE and PUE; next, within the confines of these frameworks came the programme-wide development of teaching and learning modules.

The findings about the process for new physics can be summarised as follows:

- It was regretted by the renewal committee that the development of the pilot examinations did not come within the compass of the committee's responsibility.
- Due to a perceived lack of item construction expertise and difficulties in finding available people, it was not possible to involve a pilot teacher or a committee member in the development of all pilot examinations. In the initial pilot phase, there was poor communication and a lack of mutual appreciation between the curriculum renewal committee and Cito. Cooperation was initially quite forced but improved gradually.
- The intended tripartite structure of the pilot examinations (pilot-specific, adapted regular, and regular items) was realised, although it was not feasible to stick to three equal parts.
- Initially, it was not clear which of the curriculum renewal committee and Cito was considered to be responsible for operationalising the renewal in sample pilot examination questions. Sample questions brought up by the committee were considered not appropriate by Cito.
- Also in the new physics pilot, materials development occurred independent from syllabi development.
- As in the biology pilot, the phasing of examinations development did not match well with what happened in the pilot schools.

Student Performance

As noted earlier, the pilot examinations for all three subjects consisted of three roughly equal parts: regular items ('overlapping items'—ones common to the previous and pilot examinations), slightly adapted regular items (adapted in order to

Table 6.2 Student performances on external exit examinations (pilot vs. regular) for biology for GSE 2009 and 2010 and PUE 2010

	Secondary 5 GSE 2009		Secondary 5 GSE 2010		Secondary 6 PUE 2010	
	Pilot	*Regular*	*Pilot*	*Regular*	*Pilot*	*Regular*
n students	307	20,081	372	13,735	165	12,798
Average p-value	56.4	58.3	50.9	55.3	52.2	55.1
Average p-value overlap	60.8	59.3	45.7	47.4	55.0	61.2

Table 6.3 Student performances (average p-values pilot vs. regular students) on overlap items for biology, chemistry and physics GSE 2009, GSE 2010 and PUE 2010

	Secondary 5 GSE 2009		Secondary 5 GSE 2010		Secondary 6 PUE 2010	
	Pilot	*Regular*	*Pilot*	*Regular*	*Pilot*	*Regular*
Biology	56.4	58.3	50.9	55.3	52.2	55.1
Chemistry	53.0	55.0	56.0	60.0	51.0	57.4
Physics	54.2	60.0	54.3	57.7	60.0	59.0

Note: n pilot students varies from 166 (physics GSE 2009) to 465 (physics PUE 2010); n regular students varies from 10,279 (physics GSE 2010) to 20,081 (biology GSE 2009)

make them better fit the pilot context) and pilot-specific items. This format, more in particular the overlapping items, offered the possibility of making comparisons between performances of non-pilot and pilot students. Table 6.2 shows the results for biology (based on Bruêns 2010; Bruggeman 2009, 2010; Kuiper et al. 2011a), exemplifying more or less comparable findings for chemistry (Ottevanger et al. 2011) and physics (Bruning et al. 2011). Student overall scores were corrected for examination difficulty. From the test and item analysis conducted by Cito, it appeared that, on a 1–100 scale, for Secondary 5 GSE 2009, the pilot examination as a whole was 1.9 points more difficult than the regular examination as a whole (see row 'Average p-value': 56.4 vs. 58.3). Pilot students slightly outperformed regular students (by 1.5 points) on the overlapping items (60.8 vs. 59.3). For Secondary 5 GSE 2010, the pilot examination as a whole was 4.4 points more difficult than the regular examination (50.9 vs. 55.3). This time, regular students outperformed pilot students by 1.7 points on the overlapping items (45.7 vs. 47.4). For Secondary 6 PUE 2010, the pilot examination as a whole was 2.9 points more difficult than the regular examination as a whole (52.2 vs. 55.1). The regular students outperformed the pilot students by 6.2 points on the overlapping items (55.0 vs. 61.2).

When we focus more specifically on the average p-values for the set of overlapping items for biology, chemistry and physics (see Table 6.3), it appears that in seven out of nine cases, the average p-value for the pilot students is lower than the average p-value for the regular students. The only exceptions are biology GSE 2009 (60.8 for pilot students vs. 59.3 for regular students) and physics GSE 2010 (60.0 for pilot students vs. 59.0 for regular students). Based on these findings, it might be tempting to conclude that, by and large, regular students have developed a better understanding of scientific ideas (as measured by the overlapping items) than pilot students. However, one should be very cautious in drawing such a conclusion, for two reasons. First, the number of pilot students that took the examinations is rather small, varying from 166 for physics GSE 2009 to 465 for physics PUE 2010. For this reason, we have refrained from conducting additional statistical analyses. Second, interview and observation data show that performances of pilot students have been strongly coloured by the fact that not all pilot students had the opportunity to learn all content tested. This is due to the pilot context in which pilot teachers ran out of time in using (for the first time) part of the teaching and learning materials that were meant to prepare students for the examination, did not have all materials to available to them in time, made their own choices in teaching mandatory topics in more or less depth and/or were insecure about what content would be tested in the external examination. This 'pilot effect' also appears from an analysis at an item level. For biology, chemistry and physics, substantial differences appeared in p-values for overlapping items between pilot students and regular students. For biology, for example, variations in p-values were noted from +.4 (a higher score for pilot students on a specific item) to −.23 (a lower score for pilot students on a specific item). With reference to the typology of curriculum representations in Table 6.1, these substantial variations point to a poor overlap between the tested curriculum (i.e. the external examination) and the curriculum as perceived and put into practice by pilot teachers (i.e. how pilot teachers prepared their students for the external examination).

Concluding Remarks

In the introductory section, it was noted that changes in curriculum goals and content require concurrent changes in assessment and examination policies and practices. Using three science curriculum renewal pilots in senior secondary education in the Netherlands as exemplary cases, we focused on the importance of the alignment between science curriculum renewal efforts and assessment of learning by means of external exit examinations for new biology, new chemistry and new physics. The main findings presented above demonstrate the many traps involved in seeking this alignment as well as the lessons that can be learned from these.

The good news from the evaluation of the three examination pilots is that it has indeed been possible to develop external examinations based on the draft examination programmes, and reflecting at least part of the renewal intentions. Also based on other evaluation findings about the viability and feasibility of the renewal-inspired programmes not presented here, this somewhat promising finding provided enough reason to positively inform, albeit with some reserve, a policy decision about countrywide upscaling. The reserve was mainly dictated by the as yet poor alignment between curriculum documents at the level of the formal curriculum (examination programmes, syllabi, teaching and learning materials) and between those formal curricular documents and assessment instruments and practices at the level of external pilot examinations (see Table 6.1). This poor alignment appears to be rooted in a combination of five problems.

First, none of the three renewal committees were able to create sufficient clarity about the essential features of the intended curriculum renewal, and/or they underestimated the need to specify these essential features. Clarity about goals and means is a perennial problem in the change process and is—along with the perceived need for change, the difficulty and extent of change required and the quality/practicality of the programme—a critical factor affecting adoption and implementation (Fullan 2007). The lack of clarity also diminished the capacity to generate sample examinations questions that could exemplify the essential features. Against the background of the past decade or more of gradually increasing focus on contexts in the external examinations, it is understandable that for some stakeholders it was not clear how or to what extent the intended renewal was different from what had already been happening.

Second, attempts to achieve the intended changes reflected in the syllabi went far from smoothly. As syllabi further specify the goals to be tested in external exit examinations, these are very influential steering documents for both examination construction teams from Cito and teachers as they prepare their students for the external examinations. So, a syllabus is a pivotal component of the set of documents that comprise the formal curriculum (see Table 6.1). The difficulties observed were due to differences among stakeholders in beliefs about science education and in ownership of the intended renewal (see the third problem below). There were also difficulties in the necessarily limited possibilities of making a curriculum renewal visible in a syllabus. It would have been very helpful if syllabus development and examination construction had explicitly been part of the pilots and had been more closely linked with the development and use of materials. It is likely that such a format—with the Examinations Board and Cito in the (enforced) role of co-owners of the renewal—could have prevented some of the problems.

Third, there was a lack of shared need for and ownership of the intended curriculum renewal among the various actors who played a role in 'translating' the renewal features into assessment/examination. Partly as a consequence, the communication among the various stakeholders left very much to be desired. A mandate from the Ministry of Education to jointly develop and pilot renewal-inspired examination programmes (pilot examinations included) could have provided an effective basis for deliberation (Walker 1990, 2003), interaction and negotiation (Pinar et al. 1995) among stakeholders.

The fourth problem was the great difficulty of operationalising renewal features in sample external examination questions and, in doing so, of finding a proper balance between the curriculum renewal perspective and psychometric concerns about reliability and comparability. The evaluation findings about the assessability of the pilot programmes demonstrate the delicacy of finding a proper balance between the science education reform perspective and psychometric concerns. The experiences with constructing sample examination questions in the three pilots show that too strong a focus (in this instance by Cito) on the psychometric issues of reliability and comparability, something clearly relevant in cases of high-stakes examinations, tended to be a cause of some renewal ambitions and ideas being in danger of being overlooked. Overlooking ideas important in the renewed curriculum is a significant issue of validity, which is defined by Strobart (2006) as a judgement about how effectively an assessment samples a construct or domain and about the appropriateness of the inferences generated by the assessment data. From a curriculum renewal perspective, validity is at the centre of a test or an examination. This problem was clearly reinforced by the decision to develop the pilot examinations from the regular examinations. From a quality assurance perspective, this decision makes sense, but in practice it restrained the renewal process. It resulted in both rapid acceleration and slamming on the brakes simultaneously.

Fifth, the pilots focused solely on the external examinations rather than also on the internal ones. An external examination is a (high-stakes) paper and pencil test, and the weakness of such a test lies in what it cannot test. In other words, by its very nature and due to the fact that in these specific cases it pertains to only about 60 % of the goals outlined in an examination programme, the external examination assesses only some of the curriculum aims. Indeed, in general such external examinations can only ever assess a part of curriculum aims. Hence, it needs 'two to tango' in such exit examinations: the external, centralised component *and* the internal, school-based component. In the specific context that is the focus of this chapter, it is a lost opportunity that no conditions were created to align the development of both parts of the exit examination with one another. Given this lost opportunity and also the fact that the development of the internal examinations was utilised as a kind of an 'experimental garden' for the development of the external examinations as well as a means for operationalising and hence defining outcomes that cannot be assessed in a paper and pencil test (see Millar in this volume), it would have been better to have brought the breadth and comprehensiveness of the renewal intentions and efforts into the limelight at the level of the exit examination. This could have afforded major impetus to the science curriculum renewal process.

References

Atkin, J. M., & Black, P. (2003). *Inside science education reform. A history of curricular and policy change*. New York: Teachers College.

Bennett, J., Hogarth, S., & Lubben, F. (2003). *A systematic review of the effects of context-based and Science-Technology-Society (STS) approaches in the teaching of secondary science*. London: EPPI-Centre, Institute of Education.

Black, P. (1998). *Testing: friend or foe? Theory and practice of assessment and testing*. London: Routledge.

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

Bruêns, J. (2010). *Verslag pilotexamen biologie eerste tijdvak* (intern werkdocument). Arnhem: Cito.

Bruggeman, A. (2009). *Evaluatie pilotexamen biologie eerste tijdvak* (intern werkdocument). Arnhem: Cito.

Bruggeman, A. (2010). *Evaluatie pilotexamen biologie eerste tijdvak* (intern werkdocument). Arnhem: Cito.

Bruning, L., Folmer, E., Ottevanger, W., & Kuiper, W. (2011). *Curriculumevaluatie bètaonderwijs tweede fase. Examenpilot nieuwe natuurkunde havo/vwo 2007-2010*. Enschede: Netherlands Institute for Curriculum Development (SLO).

Commissie Vernieuwing Scheikunde Havo en Vwo. (2003). *Chemie tussen context en concept. Ontwerpen voor vernieuwing*. SLO: Enschede.

Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Principles, Policy and Practice, 6*(1), 32–42.

Cuban, L. (1992). Curriculum stability and change. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 216–247). New York: Macmillan.

Earl, L. M. (2003). *Assessment as learning. Using classroom assessment to maximize student learning*. Thousand Oaks: Corwin.

Fullan, M. (2007). *The new meaning of educational change* (4th ed.). New York: Teachers College.

Goedhart, M. (2004). *Contexten en concepten: Een nadere analyse*. Amsterdam: Universiteit van Amsterdam/Amstel Instituut.

Goodlad, J. I. (1979). *Curriculum inquiry. The study of curriculum practice*. New York: McGraw-Hill.

Hargreaves, A., & Fink, D. (2006). *Sustainable leadership*. San Francisco: Jossey-Bass.

Kuiper, W. (2009). *Curriculumevaluatie en verantwoorde vernieuwing van bètaonderwijs*. Orational address. Enschede/Utrecht: Netherlands Institute for Curriculum Development/ Freudenthal Institute for Mathematics and Science Education.

Kuiper, W. A. J. M. (1993). *Curriculumvernieuwing en lespraktijk* [Curriculum reform and teaching practice]. Unpublished doctoral dissertation, University of Twente, The Netherlands.

Kuiper, W., Boersma, K., & van den Akker, J. (2005). Towards a more curricular focus in international comparative studies on mathematics and science education. In K. Boersma, M. Goedhart, O. de Jong, & H. Eijkelhof (Eds.), *Research and the quality of science education* (pp. 41–54). Dordrecht: Springer.

Kuiper, W., Folmer, E., Ottevanger, W., & Bruning, L. (2011a). *Curriculumevaluatie bètaonderwijs tweede fase. Examenpilot experimentele biologieprogramma havo/vwo 2007–2010*. Enschede: Netherlands Institute for Curriculum Development (SLO).

Kuiper, W., Folmer, E., Ottevanger, W., & Bruning, L. (2011b). *Curriculumevaluatie bètaonderwijs tweede fase. Samenvattend eindrapport*. Enschede: Netherlands Institute for Curriculum Development (SLO).

Kulm, G., & Malcolm, S. M. (Eds.). (1991). *Science assessment in the service of reform*. Washington, DC: American Association for the Advancement of Science.

Kulm, G., & Stuessy, C. (1991). Assessment in science and mathematics education reform. In G. Kulm & S. M. Malcolm (Eds.), *Science assessment in the service of reform* (pp. 71–88). Washington, DC: American Association for the Advancement of Science.

McKenney, S., Nieveen, N., & van den Akker, J. (2006). Design research from a curriculum perspective. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 67–90). London: Routledge.

Millar, R. (2010). Using research to improve practice in science education: Where should we begin and what should we aim to produce. In K. Kortland & K. Klaassen (Eds.), *Designing theory-based teaching-learning sequences for science education. Proceedings of the symposium in honour of Piet Lijnse at the time of his retirement as professor of physics didactics at the University of Utrecht.* Utrecht: Freudenthal Institute for Science and Mathematics Education.

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

Ottevanger, W., Folmer, E., Bruning, L., & Kuiper, W. (2011). *Curriculumevaluatie bètaonderwijs tweede fase. Examenpilot nieuwe scheikunde havo/vwo 2007-2010*. Enschede: Netherlands Institute for Curriculum Development (SLO).

Pinar, W. F., Reynolds, W. M., Slattery, P., & Taubman, P. M. (1995). *Understanding curriculum. An introduction to the study of historical and contemporary curriculum discourses*. New York: Peter Lang.

Strobart, G. (2005). Assessment and change. *Assessment in Education, 12*(3), 215–216.

Strobart, G. (2006). Valid assessment. *Assessment in Education, 13*(1), 1–3.

Thijs, A., & van den Akker, J. (2009). *Curriculum in development*. Enschede: Netherlands Institute for Curriculum Development (SLO).

van den Akker, J. (1998). The science curriculum: Between ideals and outcomes. In B. J. Fraser & K. G. Tobin (Eds.), *International handbook of science education* (pp. 421–447). Dordrecht: Kluwer.

van den Akker, J. (2003). Curriculum perspectives: An introduction. In J. van den Akker, W. Kuiper, & U. Hameyer (Eds.), *Curriculum landscapes and trends* (pp. 1–10). Dordrecht: Kluwer.

van Oers, B. (2001). Contextualisation for abstraction. *Cognitive Science Quarterly, 1*(3–4), 279–306.

van Oers, B. (2004). *The recontextualization of inscriptions: An activity-theoretical approach to the transferability of abstractions.* Paper for the invited symposium 'Rethinking abstraction and decontextualization I relationship to transfer dilemmas', AERA, San Diego.

Walker, D. F. (1990). *Fundamentals of curriculum*. San Diego: Harcourt Brace Jovanovich.

Walker, D. F. (2003). *Fundamentals of curriculum. Passion and professionalism* (2nd ed.). Mahwah: Lawrence Erlbaum Associates.

Watkins, D., Dahlin, B., & Ekholm, M. (2005). Awareness of the backwash effect of assessment: A phenomenographic study of the views of Hong Kong and Swedish lecturers. *Instructional Science, 33*(4), 283–309.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*, 3–14.

Chapter 7
Content to be Assessed Across the History of the National Assessment of Educational Progress

Audrey B. Champagne

Introduction

Performance of US students on large-scale assessments influences the nation's educational policy and curriculum. Both performance changes over time and comparisons of US students with students from other nations influence federal and state policies and ultimately the curriculum enacted in the classroom. Policy makers are aware of performance differences and explanations for differences posited by policy analysts but are generally ignorant of the science knowledge and skills measured by the assessment instruments that produce the data on which performance comparisons are based. For the most part, teachers are familiar with the science knowledge and skills measured by their state's assessment but not with the science knowledge and skills measured by the US National Assessment of Educational Progress (NAEP) or international assessments such as the Programme for International Student Assessment (PISA) and the Trends in Mathematics and Science Study (TIMSS) (e.g., Mullis et al. 2005).

This chapter is based on the premise that data on which progress is measured and international comparisons are made should be evaluated using the knowledge and abilities measured by the items as well as statistical criteria that define measurement quality. It describes and compares the science knowledge and abilities to be assessed by Science NAEP from its first administration in 1969 to the present time. This chapter begins with the science knowledge and abilities described in the document *Science Objectives* (Committee on Assessing the Progress of Education 1969) which provided the specifications for the first Science NAEP assessment and ends with the *Science Framework for the 2009 National Assessment of Educational Progress* (National Assessment Governing Board 2008). Changes in the knowledge

A.B. Champagne (✉)
State University of New York at Albany
e-mail: audreychampagne@cs.com

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_7,

and abilities and the factors such as the US economy and the status of US science that influenced the changes as well as federal/state policy debates surrounding where responsibility for the school curriculum resides are included in this chapter. (See Appendix 1, NAEP Objective and Framework Documents for complete references to the seven documents and how they will be referenced in this chapter.)

The development of a large-scale assessment is a complex process beginning with description of the knowledge and abilities (science content) to be assessed, development and field test of items/tasks to measure the science content, administration of the assessment, and finally reporting of student performance. This chapter focuses on the first step, the description of the science content to be assessed. The arguments put forth in this chapter are based only on the science content described in the objective and framework documents, not on the degree to which the science content to be assessed is the science content actually assessed.

The System: Political and Educational Interactions

Unlike most nations (Germany and Switzerland are exceptions), the primary responsibility for education in the USA resides with the states. State governors and chief education officers have legal authority. Local control is maintained by school boards, which oversee finances, monitor compliance with federal and state regulations, and define local educational practices consistent with state regulations. Funding for K-12 education comes from taxes, the revenue from which is administered primarily the states and to a limited extent by local school boards. Federal funds are administered by the states. Because federal funds carry with them strict requirements for their use, acceptance by the states of these funds provides the federal government with considerable influence on the states' education policies, which in turn limits local control.

Of all the education services managed by local districts, the most passionately guarded is curriculum. Inroads on local control of the curriculum increased dramatically with the federally mandated No Child Left Behind (NCLB 2001) which required state standards and assessments. Science followed reading and mathematics in the NCLB implementation schedule. Despite the existence of science standards developed under the aegis of highly respected scientific organizations, the American Association for the Advancement of Science and the National Research Council, each state developed its own science standards. Furthermore, while NCLB proclaimed NAEP as the gold standard for assessment, each state designed its own science assessments. Development of state science assessments followed the approval of science standards by state legislatures.

External reviews have been highly critical of the quality of states' standards and assessments for reading, mathematics, and science, which has led to efforts toward establishing national standards, assessments, and curricula. This Common Core State Standards Initiative is headed by the National Governors Association and Council of Chief State School Officers. While most states have adopted common core standards in English language arts and mathematics, translation of the standards into common curricula and assessments is already under intense criticism (Gewertz 2011).

The states' fierce control of the K-12 curriculum and testing of it are reflected in the early history of the NAEP, which begins in 1963 with Francis Keppel, then US Commissioner of Education. As Commissioner of Education, Keppel was required to report to the US Congress on the progress of education in the states (Fitzharris 1993). Keppel noted that information required for such a report was not available and asked Ralph Tyler to consider the possibility of a measure of student learning. Tyler was instrumental in forming the Exploratory Committee on Assessing the Progress of Education (ECAPE) and directed its work. The charge to the committee was to "confer with teachers, administrators, school board members, and others concerned with education to get advice on the way in which a national assessment of educational progress can be designed to be constructively helpful to the schools and to avoid possible injuries" (Merwin 1966, p. 5). Active participation by educators and education societies characterized the early work of the committee. Initial discussions concluded that "the assessment should be under the direction of a private commission and not be a project of Federal or State Governments … it was recommended that financial support of the Committee's initial work come from private sources" (Merwin 1966, p. 5). Consistent with the recommendation, the committee approached the Carnegie Corporation of New York for funding, which the corporation granted.

Tyler's presentation of the objectives and plans for a national assessment of educational progress contains lengthy discussions of the purposes of evaluation at the individual student and school levels and describes assessment of education's effectiveness as a general public concern (Tyler 1966) addressing fears regarding a nationwide testing program. These are that such an assessment:

- "would influence the direction and amount of effort of pupils and teachers,"
- "enables the Federal government to control the curriculum," and
- "would stultify the curriculum by not allowing changes over the years." (p. 2)

Tyler addresses these fears explaining how the plan for the assessment will avoid them. In 1969, the Exploratory Committee on Assessing the Progress of Education changed to the Committee on Assessing the Progress of Education (CAPE) and the administration of the assessments was given to the Education Commission of the States. The National Assessment Policy Committee was organized to take responsibility for the governance of NAEP.

Currently, the management of NAEP is the responsibility of the National Center for Education Statistics, a federal government entity, and governance for NAEP is the responsibility of the National Assessment Governing Board whose members are appointed by the US Secretary of Education.

The influence of Science NAEP on the school science curricula is an empirical question that has not been investigated. However, the NAEP Science Objective and Framework documents provide a systematic look at the science content deemed worthy of assessment over the years and its relationship to US science curricula. The relationship envisioned by the members of the ECAPE was that objectives

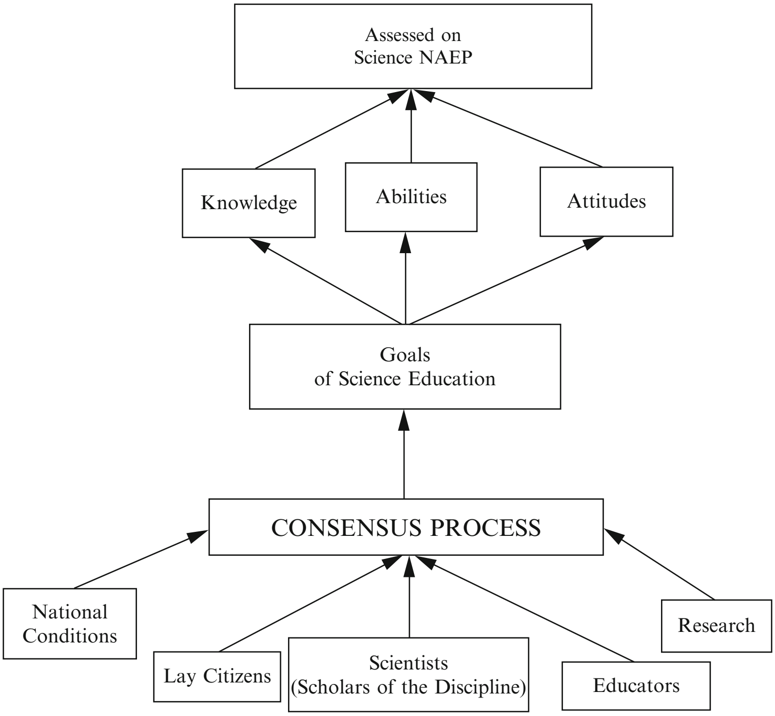


Fig. 7.1 NAEP consensus process

specifying what would be measured by the assessment "reflect what schools currently consider important goals of science education" (Lombard and Owen 1965 Introduction). This relationship was further elaborated in the policies of the National Assessment Policy Committee, one of which specified that objectives for the assessment be acceptable to three groups of people.

> First, they [objectives] must be considered important by scholars in the discipline of a given subject area, Scientists, for example, should generally agree that the Science objectives are worthwhile, Second objectives should be acceptable to most educators and be considered desirable teaching goals in most schools. Finally, and perhaps most uniquely, National Assessment objectives must be considered desirable by thoughtful lay citizens. Parents and others interested in education should agree that an objective is important for youth of the country to know and that it is of value in modern life. (Mastie et al. 1972, p. 2)

A consensus process was set in place by the National Assessment Policy Committee to identify objectives for the assessment. Continued by National Assessment Governing Board, this process has ensured that the assessments reflect perspectives of science educators, including teachers, administrators, curriculum developers, and higher-education-based science educators. Figure 7.1 depicts the consensus process. The consensus process results in the specification of the knowledge, abilities/skills, and attitudes to be assessed.

Over the years the consensus process has been influenced by other factors including the economy, perceived threats to the preeminence of US science, commissioned reports, national standards and benchmarks, US students' performance on international assessments, and social science research. Even so, a consequence of the consistency of NAEP policy is that the science NAEP objectives and framework documents provide a consensus perspective on the science content that educators, scientists, the lay public, and policy makers believe is worthwhile and appropriate. Consequently, the analysis that follows provides one perspective, that of Science NAEP, on the content of school science in the USA, its consistency and changes over the time period from 1963 to the present.

Science Literacy

Science literacy is the pervasive construct in the NAEP objectives and framework documents. The rationale for science literacy, the attributes of science literate individuals, distinctions between science literacy expectations for ordinary people and for scientists, the challenges posed by the measurement of the attributes, and the constructs and verbs used to describe them are all topics addressed in the NAEP objectives and framework documents.

The NAEP Rationale for Science Literacy

The statement containing the rationale for science literacy that has influenced Science NAEP throughout its history was drafted by one of the contractors engaged by ECAPE to provide guidance to the committee on its work.

> A comprehensive program in science education must consider two numerically unequal groups of students: those who may eventually pursue scientific or technical careers, and the great majority: those who will not. As our population increases and our economy expands, the national need for scientific and technical personnel will continue to increase. Thus, science education must give a realistic introduction to scientific and science-related work for those students capable of it, while at the same time, encouraging, developing, and testing their interest and enthusiasm in careers in science and technology.
>
> Even more attention, however, must be given to the great majority of students who will not eventually pursue scientific or technical careers. In a free society it is the citizens from all walks of life who make the public decisions, and in an increasingly technological age more of these decisions can be intelligently made only with an understanding of the scientific considerations that bear upon them. Also, the success of the entire scientific enterprise depends to a great extent on the atmosphere of its surrounding culture. If the general attitude is to be favorable to the scientific enterprise, the public must have some degree of awareness of the place of the scientific endeavor in man's heritage, the importance of basic research in science, and the potential contributions of the applications of science to a better, or worse, way of life. Thus the schools attempt to produce a scientifically literate citizenry generally favorably disposed toward supporting the work of scientists. (Lombard and Owen 1965, pp. 5–6)

This statement appears without reference to Lombard and Owen in the *Science Objectives* (Committee on Assessing the Progress of Education 1969) and is mentioned either with or without reference to the 1969 Science Objectives document in the 1972–1973, 1976–1977, 1985–1986, and 1990 Objectives documents. The *Science Framework for the 1996 National Assessment of Educational Progress* (National Assessment Governing Board n.d.) mentions science literacy only once in passing, and *Science Framework for the 2009 National Assessment of Educational Progress* (National Assessment Governing Board 2008) contains new language regarding science literacy and its importance:

> In the rapidly changing world of the 21st century, science literacy is an essential goal for all of our nation's youth. Through science education, children come to understand the world in which they live and learn to apply scientific principles in many facets of their lives. In addition, our country has an obligation to provide young people who choose to pursue careers in science and technology with a strong foundation for their postsecondary study and work experience. The nation's future depends on scientifically literate citizens who can participate as informed members of society and as a highly skilled scientific workforce, well prepared to address challenging issues at the local, national, and global levels. Recent studies, including national and international assessments, indicate that our schools still do not adequately educate all students in science. (National Assessment Governing Board 2008, p. v)

There are significant differences in the rationales for science literacy in the two statements. Careers and the importance of a science-literate citizenry are acknowledged in both. However, careers in science only are mentioned in the earlier document, while the importance of a highly skilled, scientifically literate workforce is mentioned in the later document. Differences include a shift from a national perspective to a global one. Another is the way in which "two groups of students" are referenced in the earlier statement. One group is of students capable of scientific and science-related work, and the other group is the great majority. The text leaves the reader to infer that students in the second group are either not capable or chose not to go into science-related work. A third difference relates to attitude. The earlier text can be interpreted to be putting forth the following argument: the goal of science education is a science-literate citizenry, science-literate persons have favorable attitudes toward science and scientists, and thus a goal of school science is to instill favorable attitudes toward science and scientists.

These changes are significant. The idea that science literacy is important to the nation's place in the world's economy has gained acceptance as we have become more aware that performance in the workplace requires some knowledge of science and the ability to reason scientifically. This realization has been the rallying cry for strengthening STEM education in the USA. However, acceptance of the idea that all young people are capable of learning science at some level is taking longer. The doubt lingers that some segments of the population, including women and members of certain ethnic groups, have the capacity to become scientists or to engage successfully in work requiring scientific knowledge and abilities. Because the doubt is contrary to the nation's social and philosophical perspectives, the education community is challenged to identify reasons other than innate capacity to explain reasons for the underrepresentation of certain groups in science and to address these so as to increase participation in science.

The influence of attitudes toward science and scientists on students' decisions to study science and consider careers in science and attitude as a goal of science education are matters currently under discussion. At issue is whether attitudes should be instilled or should develop though experience. Central to consideration of this issue is whether students' attitudes toward science should be a part of science assessment (Loveless 2009).

Characteristics of Science-Literate Individuals

From rationales arguing the importance of science literacy, the NAEP objectives and framework documents typically turn to descriptions of the characteristics of science-literate individuals. For instance, in their report to ESCAP, Lombard and Owen follow their rationale for science literacy with the following text:

> To achieve these goals (giving students a realistic introduction to science…a public favorable to the scientific enterprise) science education programs attempt to bring about measurable improvements in the following characteristics of students:
>
> 1. Their ability to apply, in appropriate situations, the methods, techniques, and rational processes associated with scientific work.
> 2. Their understanding of the major conceptual schemes that currently interrelate, and form the core of, the various scientific disciplines.
> 3. Their understanding of the position, limitations, and potential of science and its applications in today's society, and their attitudes toward scientists and their work.
> 4. Their interests in science, which for some students are manifested by movement toward scientific or technical careers, and for all students are displayed outside formal schooling by their continued learning in, and attention to, scientific subjects.
> 5. Their realization of science as a human intellectual activity. (Lombard and Owen 1965, p. 6)

Science Objectives for the 1969 Science NAEP, developed by Educational Testing Service for ECAPE and subsequently vetted by the consensus process are:

1. Know fundamental facts and principles of science.
2. Possess the abilities and the skills needed to engage in the processes of science.
3. Understand the investigative nature of science.
4. Have attitudes about and appreciations of scientists, science, and the consequences of science that stem from adequate understandings (Committee on Assessing the Progress of Education 1969).

The illustrations above are consistent with the approach in most of the NAEP objective/framework documents. Science literacy is defined in terms of the characteristics of science-literate individuals most often their knowledge, understanding, and abilities. While science literacy is defined in terms of the characteristics of adults/high school graduates, the statements are often introduced in terms of student characteristics. Only later in the documents are these broadly stated characteristics redefined explicitly in terms of characteristics of students at Grades 4, 8, and 12.

Science Literacy for Life and Profession

The characteristics of science-literate individuals described in NAEP are those that enable meeting personal, social, and civic responsibilities. These characteristics are different from the characteristics that enable scientists to meet their professional responsibilities. While science literacy for life is different from science literacy for professional practice, each derives from the characteristics of the enterprise of science and its practitioners, scientists. Often the characteristics of professional literacy are conflated with those of science literacy for life. How does science literacy for life as defined by NAEP compare with science literacy for the professional scientist?

Venezky (1990) defines three levels of literacy: learned, competent, and capable of minimal function. These levels are not distinct but identify points along a continuum. From a societal perspective, literacy levels range from that necessary for functioning as a contributing member of society—earning a living, voting regularly and intelligently, and attending to health matters—to that level necessary for functioning as a learned participant, for example, a Supreme Court Justice or a Nobel Prize winner. Ordinary literacy (somewhere between competent and capable of minimal function) allows us to "get by" in the daily activities of life; other types of literacy are characteristic of the professions and the academic disciplines. The ordinary literacy goals of US K-12 education are high. As well as being science literate, high school graduates are expected to be language literate, literature literate, mathematically literate, historically literate…. How does the ordinary literacy expected of the high school graduate compare with the ordinary literacy of academics and professionals?

There is no easy answer to this question. Some would argue that individuals in the academic disciplines and the professions are broadly literate in the ordinary sense. Others would argue that not all professionals or academicians are necessarily literate in the ordinary sense, if being literate in the ordinary sense includes being science literate. Furthermore, each discipline or profession has its own knowledge and practices that define the literacy of the profession. Even within the professions, there are different levels of literacy. For instance, physicians' levels of medical literacy range from "capable of sufficient function in the profession," including a history of practice with no malpractice suits, to "learned," that is, holding a distinguished chair of medicine.

An analysis of the characteristics of science-literate high school graduates in NAEP documents motivates some questions about the relationship between the professional and ordinary literacy of scientists and the expectations for the characteristics of high school graduates who are science literate in the ordinary sense. Figure 7.2 provides a structure for the analysis of that relationship. The left column of Fig. 7.2 lists components of the science enterprise, the right column, literacy components of students, the knowledge, abilities/skills, and attitudes of students who have reached the goal of science education.

All of the knowledge, abilities/skills, and attitudes of students in Fig. 7.2 are present in one or more of the NAEP science objectives/framework documents. The

Science Enterprise Components	Components of Student Science Literacy
Goals	**Goals** Can distinguish the goals of science from the goals of other human endeavors, technology, for instance
Disciplines	
Physics	
Chemistry	
Earth	
Life	
…	
Interdisciplinary - Physical Chemistry Biochemistry	
Knowledge Products	**Knowledge Products**
Facts	Can demonstrate knowledge of the knowledge products of science
Concepts	Can demonstrate understanding of the knowledge products of science
Principles	Can apply the knowledge products of science to scientific problems, practical problems, and personal problems
Conceptual Schemes	Can apply the knowledge products of science to personal, social, and civic decision making.
Practices Inquiry Reasoning patterns	**Practices** Can demonstrate knowledge of the practices of science Can demonstrate knowledge of science inquiry Can demonstrate abilities that are components of science inquiry Can apply scientific reasoning to scientific, practical, personal, social and civic problems Can apply scientific reasoning to practices, personal, social, and civic decision making. Can communicate scientific information and ideas
Attributes of Practitioners (Scientists) Ethics Habits of Mind Values	**Attributes of Practitioners (Teachers)** Can demonstrate knowledge of the professional attributes of scientists Values Science Has positive (informed) attitudes of science and scientists
Philosophical Perspectives	**Philosophical Perspectives** Can demonstrate knowledge of the philosophy of science
History	**History** Can demonstrate knowledge of the history of science
Relationships with other Disciplines and Professions Engineering Mathematics	**Relationships with other Disciplines and Professions** Can demonstrate knowledge of the interactions of science with other dis ciplines and professions
Interactions with Society	**Interactions of Science and Society** Can demonstrate knowledge of the interactions of science and society

Fig. 7.2 Components of the science enterprise and expectations for students

components of science literacy in the NAEP objectives and framework documents are numerous and may well be more numerous than the characteristics that defined practicing scientists' level of science literacy. This is a conjecture that can be tested empirically. However, though empirical evidence is absent, personal observation and knowledge of most doctoral programs provide weak evidence that the conjecture has some validity. Practicing scientists' depth of knowledge in their specialty is considerable; however, their knowledge of science outside their disciplines may have less breadth than that expected of high school graduates or science teachers. My personal experience derives from the opportunity over the past 3 years of working with a group of expert research scientists to help K-12 science teachers integrate certain "big ideas" in their curricula. As we worked with teachers, our disciplinary knowledge was often challenged when the topic under discussion was outside our discipline of expertise.

Our discussions over wine revealed gaps in our knowledge of the nature of science, its values, ethics, and philosophical perspectives as well as its historical development. Because most doctoral programs do not require course work in the history and philosophy of science, practicing scientists typically have not had the opportunity to become acquainted with these components of science. This matter of the breadth of the US school science curriculum has been recognized (Schmidt et al. 1977) and used to explain the science performance of US students on TIMSS.

Measurement of Science Literacy

The challenge of measuring science literacy is acknowledged in the NAEP objectives and framework documents. For instance, the 1976–1977 Objectives document supports the goal of science literacy while acknowledging the challenge posed by measuring it:

> The changes in science education over the last 20 years have been dramatic, yet the general goals of science education have not changed radically. Science educators still support the statement during the development of the 1965 National Assessment of Educational Progress (NAEP) science objectives: "The major purpose of science education is to develop scientifically literate individuals". (Merwin and Tyler 1965, p. 30)
>
> While few argue that assertion, translating it into a structure that satisfactorily measures the attainments in science of American youth has been a difficult process. Over the years, National Assessment has responded to the need for increasingly precise definitions and measurements of scientific literacy. The 1972–73 science objectives …were a step in this clarification process, and this, the most recent compendium of objectives is yet a further refinement of the basic task. (Education Commission of the States 1979, p. 1)

The 1990 Objectives document provides some interesting insights into the process of defining characteristics of the science-literate person as measureable constructs:

> Science educators generally agree that the primary purpose of school science is to cultivate scientific literacy: however, there is far less agreement as to what constitutes scientific literacy or how such a definition might be used to guide the development of meaningful

Learner's Internal State and Experience		Outcomes
Science Knowledge	Scientific Habits of Mind	Solving Problems
Thinking Skills		Conducting Inquiries
Laboratory Skills		Personal and Civic Decision Making
Attitudes		Life-long Learning and Sense of Control Over One's Life

Fig. 7.3 Elements of scientific literacy and their interactions—1990 (National Assessment of Educational Progress 1989, p. 9)

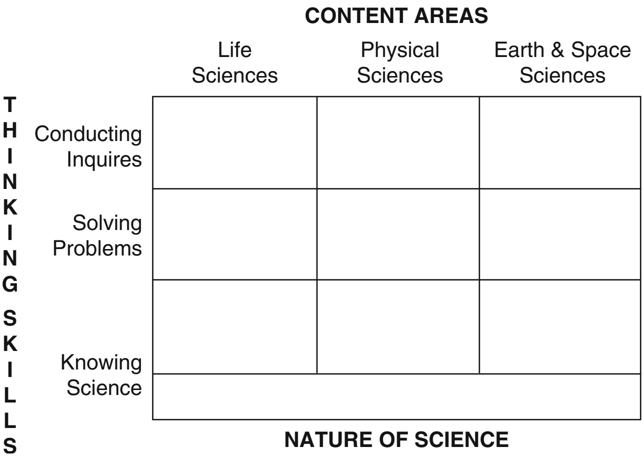


Fig. 7.4 Framework for the 1990 science assessment (National Assessment of Educational Progress 1989, p. 12)

> curriculum and instructions. At the very least, there appears to be consensus among educators that school science should help students acquire the knowledge, skills and understandings necessary to fulfill their human, social, and economic responsibilities. (National Assessment of Educational Progress 1989, p. 8)

The following "interactive model" of science literacy and discussion of it appear in the document (Fig. 7.3):

> [A]ccording to the interactive model …, scientific habits of mind serve as a filter between features of the learner--including his or her cognitive abilities, attributes, and home and school experiences--and the outcomes of science learning. This analysis is followed by a disclaimer: "Given limited resources, the 1990 science assessment can cover only some of the many elements of scientific literacy. Those elements of the model to be included in the assessment [are] science knowledge, scientific habits of mind, and the ability to solve problems and conduct inquiries." (National Assessment of Educational Progress 1989, p. 8)

That said, the Framework for the Assessment is presented as shown in Fig. 7.4. Notice that certain elements contained in the interactive model are not included in the framework and the way in which elements in the interactive model are restructured in the framework. Attitudes, for instance, are not included in the framework but are sampled as background variables. While the text says that scientific habits of mind will be included, they do not appear in the framework. Solving problems, conducting inquiries, and knowing science are classified as thinking skills and included in the framework as such. Personal and civic decision making, essential elements of science literacy, are not included in the framework.

Following the framework is a discussion of the Nature of Science, Thinking Skills, and the Science Content Areas. It is informative to compare the detail in which each of these is discussed in the document with the percent of assessment questions to be devoted to each in the assessment. The discussion of the Nature of Science occupies approximately 10 pages; Thinking Skills, approximately 4 pages; and Science Content Areas 7 pages. However, just 10 % of the items at Grade 4, 10 % of the items at Grade 8, and 12 % of the items at Grade 12 address the Nature of Science. Combined, Thinking Skills and Science Content Area are allotted 90 % of the assessment questions at Grades 4 and 8 and 88 % of the assessment questions at Grade 12 (National Assessment of Educational Progress 1989 p. 16). Just one more page is devoted to Thinking Skills and Science Content Areas combined than to the Nature of Science. The 1990 Objectives document illustrates the tension between the value placed on certain components of science literacy such as the nature of science and restrictions placed on their assessment by psychometric and time resource considerations.

Components of Science Literacy

Framework Summary Structures describing the components of science literacy contained in each of the seven NAEP objectives and framework documents are contained in Appendix 2. In these summary structures, the content to be assessed is described in terms of science disciplines, constructs (nouns or noun phrases such as principle, nature of science, or inquiry), and verbs (such as know, understand, or explain). Figure 7.5 summarizes the disciplines and constructs (nouns) contained in each Science NAEP Framework Summary Structures. Figure 7.6 summarizes the verbs contained in each year's Framework Summary Structure.

A few of the documents provide the reader with definitions of terms used in the document. In others the reader is left to assume the intended meaning or to infer it from the objectives or exemplary items. The 1972–1973 Objectives document contains definitions of both constructs and verbs:

> *Know.* To "know" is to recognize or recall phraseology, textual material, or other content. "Knowing" does not include or imply "understanding," "applying," or "appreciating."
>
> *Understand and Apply*. To understand is to be able to explain in one's own words, to recognize when stated in phraseology different from that in the textbook, to interpret, such as interpreting or drawing inferences from tables or graphs of data. To "apply" is to make actual use of previously acquired knowledge. One must have knowledge and understand that knowledge before it can be applied. For example, one must understand the information on a weather map before he can use it to describe weather conditions in various parts of the country or predict the movement of high and low pressure areas for several successive days.

1969	1972-73	1976-77	1985-86	1990	1996	2009
Fundamental Facts and Principles	**Facts and Simple Concepts** **Laws(Principles)**	**Content** Biology Physical Science Earth Science	**Life Science** **Physics** **Chemistry** **Earth and Space Sciences**	**Life Sciences** **Physical Sciences** **Earth and Space Sciences**	**Earth** **Physical** **Life**	**Science Principles** Physical Science Life Science Earth and Space Sciences **Scientific Inquiry** **Technological Design**
Abilities and Skills Needed to Engage in the Processes of Science		**Processes** Process/methods Decisions Making				
	Conceptual Schemes	**Integrated Topics (Multidisciplinary)**			Themes	
	Inquiry Skills					**Scientific Inquiry**
The Investigative Nature of Science	**Scientific Enterprise**		**Nature of Science**	**Nature of Science**	**Nature of Science**	
Scientists, Science, and the Consequences of Science that stem from Adequate Understanding		**Science and Society** Societal problems Science and self Applied science				
			History of Science			

Fig. 7.5 Disciplines and constructs

1969-70	1972-73	1976-77	1985-86	1990	1996	2009
Know	Know	Knowledge	Knows	**Thinking Skills** Knowing Science	Conceptual Understanding	
Possess Abilities and Skills		Comprehension	Uses	Solving Problems Conducting Inquiries	Scientific Investigation	Identifying Using **Scientific Inquiry**
Understand	Understand and Apply	Application			Practical Reasoning	**Technological Design**
		Synthesis	Integrates			
Have Attitudes and Appreciations	Appreciate					

Fig. 7.6 Verbs contained in each year's Framework Summary Structure

Appreciate. To "appreciate," is to value objects, ideas, and processes. Appreciation is expressed behaviorally in willingness to attend, to give up resources for, to take a stand in favor of, etc. One must "know about" in order to appreciate, although he need not necessarily "understand."

Fact. A simple description, definition, or observation that many people can make (for example, "the grass is green"; the dog has four legs.")

Simple Concept. An internalized representation that is, an individual's generalized idea or mental picture of a property or set of objects, events or natural phenomena (for example, the idea of "dogginess" as a generalized category that includes specific dogs such as poodles, beagles, and collies). A concept includes at least two facts with an internalized relationship and a least one which is not in agreement.

Law (principle). A law states a relationship between two or more concepts, (For example, the law "force equals mass times acceleration," tells someone with adequate understanding of "mass" and "acceleration" and knowledge of mathematical symbols how the two terms are related.) "Laws" are authoritative, can be agreed upon by others, and can be tested. "Law" is presently the preferred term: "principle" is included because it appears in textbooks and reference sources.

Conceptual Scheme. A broad theme of science, a synthesis of a group of interrelated laws (principles). (For example, the theme of complementarity of organism and environment includes relationships between laws of energy resources in the environment, energy pathways, food webs, populations and communities and ecosystems.)
Inquiry Skills. Those skills needed to engage in the process of science (for example, observations, hypothesis testing, data collection, and ordering of knowledge).

The Scientific Enterprise. The contextual situation in which all science exists, it encompasses everything involved in science: the nature of the scientific method, relationships between science, technology, and society, limitations of science, roles of scientists, and the like. (Mastie and Johnson 1972, pp. 10–11)

The 1985–1986 Objectives document provides definitions for the verbs: knows, uses, and integrates. These verbs are labeled cognitive processes and described in terms of the cognitive processes required to respond to certain types of assessment exercises:

Knows: These exercises test primarily factual knowledge, Successful performance depends on the ability to recall specific facts, concepts, principles, and method of science: to show familiarity with scientific terminology; to recognize these basic ideas in different contexts; and to translate information into other words or another format. This category generally involves a one-step cognitive process.

Uses: These exercises test the ability to combine factual knowledge with rules, formulas, and algorithms for a specified purpose, Successful performance depends on the ability to apply basic scientific facts and principles to concrete and/or unfamiliar situations; to interpret information or data using the basic ideas of the natural sciences; and to recognize relationships of concepts, facts, and principles to phenomena observed and data collected. This category generally involves a two-step cognitive process.

Integrates: These exercises test the ability to organize the component processes of problem solving and learning for the attainment of more complex goals. Successful performance depends on the ability to analyze a problem in a manner consistent with the body of scientific concepts and principles, to organize a series of logical steps, to draw conclusions on the basis of available data, to evaluate the best procedure under specified conditions, and to employ other higher-order skills, needed for reaching the solution to a problem.

This category generally involves multi-step cognitive processes, In particular, it requires such mental processes as generalizing: hypothesizing: interpolating and extrapolating: reasoning by analogy, induction and deduction; and synthesizing and modeling. (National Assessment of Educational Progress 1986, p. 9)

Analysis of the constructs and verbs contained in the NAEP objectives and framework documents is a challenge because each document either defines or uses the constructs and verbs in a unique way. Even relatively simple constructs such as concept and principle are used differently in the documents. In some documents, concept is used synonymously with principle. In others the relationship between concept and principle is explicitly defined. More complex constructs, comprised of multiple elements such as the nature of science and inquiry, have different meanings attached to them resulting in different definitions of the construct. The bottom line is that many of the same constructs appear in several of the documents with different meanings associated with them. Inconsistencies of the sort seen in the NAEP documents are rampant in the science education literature and in science education textbooks. These differences make it difficult to have meaningful conversations about these important constructs much less to design assessment tasks to measure them. This is especially true regarding inquiry, the nature of science, and their components.

Verbs

These are as difficult to analyze as are the constructs. Know and understand in their various grammatical forms (knowing, understanding, knowledge, understanding) are especially challenging. Knowing and understanding are inferred from behaviors. One infers that a person knows a fact such as the boiling point of water when in response to the question, “What is the boiling point of water?” the person states that the boiling point is 100 °C or selects 100 °C from a list of temperatures. Understanding of boiling point might be inferred if a person responds that the boiling point of water is variable, depending on atmospheric pressure, or selects a statement to that effect from a list of statements. In the NAEP objectives and framework documents, these verbs are further elaborated by statements that describe observable behaviors or by items designed to elicit a behavior.

The most evident change in verbs relates to the relative strengths of behavioral and cognitive perspectives on learning and performance. The influence of the Bloom’s Taxonomy (Bloom 1956) on verbs is evident in the 1972–1973 and 1976–1977 Objectives documents, which contain a one-page condensed and simplified version of the Bloom’s Taxonomy. The transition to a cognitive perspective is evident in the 1985–1986 Objectives document. The influence of Bloom’s Taxonomy is evident in the titles and emphasis on objectives in the early documents and the shift to a more cognitive perspective in the use of the terms “cognitive” and “cognition” in the later documents. The 1985 Objectives document has features of both perspectives. It is titled *Science Objectives*, but a major heading on the framework contained in the document is “Cognition.” The major differences across the documents are the descriptions of the characteristics of science-literate individuals at ages 9, 13, and 17 or at Grades 4, 8, and 12. Broadly speaking the characteristics of science-literate individuals at different ages are described in terms of what they know about science, their abilities or skills, and their attitudes toward science.

Verb categories change from thinking skills in the 1990 Objectives document to knowing and doing in the 1996 Framework document and to identifying and using in the 2009 Objectives document. However, no matter the terminology used, the verbs contained in the more detailed statements of the behaviors to be assessed are remarkably similar. Recognize, explain, predict, identify patterns, design, hypothesize, interpret, differentiate, and conduct are examples of verbs used in many of the documents. Despite policy documents, standards, and prevailing research perspectives, the constructs and verbs have remained remarkably consistent from 1965 to the present.

Constructs

Constructs come in different forms—single words, phrases, and sentences. Some have been recognized as components of science literacy for many years; others have been nascent and now are emerging as important components. Nature of science, scientific inquiry, and science conceptual schemes are examples of familiar constructs related to science literacy. Despite their familiarity, there is little consensus on their meaning or how they might be measured. Learning progressions and cross-disciplinary constructs have gained recent attention in science education and present similar challenges to definition and measurement.

Learning Progressions

Learning progressions are a verb-related construct that appears in the 2009 Framework document as a construct with potential transformative power when applied to the curriculum and to assessment. Is this construct as revolutionary and potentially transformative as the research and policy literature suggest? Review of the NAEP objectives and framework documents suggests that the ideas are not as revolutionary as their proponents suggest and that effecting transformation of assessment takes more than a new construct.

Definitions of learning progressions in the current literature differ primarily on the issue of empirical evidence. The examples below reflect this difference:

> Learning progressions are descriptions of the successively more sophisticated ways of thinking about a topic that can follow one another as children learn about and investigate a topic over a broad span of time (e.g., 6 to 8 years). (NRC 2005, p. 219)

> Learning progressions in science are empirically-grounded and testable hypotheses about how students' understanding of, and ability to use, core scientific concepts and explanations and related scientific practices grow and become more sophisticated over time, with appropriate instruction (NRC 2007). These hypotheses describe the pathways students are likely to follow to the mastery of core concepts. They are based on research about how students' learning actually progresses—as opposed to selecting sequences of topics and learning experiences based only on logical analysis of current disciplinary knowledge and on personal experiences in teaching. (Corcoran et al. 2009, p. 15)

These definitions suggest that at any given point in time, a child's knowledge (facts, concepts, and principles) about a topic, the organization of that knowledge, and the child's reasoning with that knowledge might be assessed and that over time the knowledge, its structure, and reasoning change—the knowledge increases, and the structure and thinking become more sophisticated. Assessment strategies capable of measuring these changes would provide empirical evidence supporting a learning progression.

The 2009 Framework document applies research (Smith et al. 2004) to design learning expectations at Grades 4, 8, and 12, which, if empirically verified by student performance, would be a learning progression for some portion of the topic and states of matter (Fig. 7.7).

Thus, the current view is that learning progressions are a new idea. However, the 1969–1970 and 1972–1973 Objectives documents contain descriptions of science objectives that have many attributes in common with the performance expectations for States of Matter. For instance, see Fig. 7.8 taken from the 1972–1973 Objectives document.

The performance expectations and objectives are similar in the structure of the science contained in them even though the educational descriptors (performance expectations and objectives) and level (grade or age) designations are different. The performance expectations are empirically based; however, no basis is given for the objectives. Thus, the objectives fail the empirical-base criterion contained in the Corcoran definition above.

However, just as with the performance expectations, data from assessments of these objectives would provide data that would either verify or refute the posited progression.

Cross-Disciplinary Constructs

Conceptual schemes, integrated topics, and themes are constructs mentioned in three NAEP objectives and framework documents. These constructs appear, disappear, and reappear in science education. Stated as principles or concepts and labeled in a variety of ways, these constructs apply across disciplines and serve two distinct functions: one as principles or concepts characteristic of the knowledge of science-literate individuals, the other as structures that can provide curricular coherence. These cross-disciplinary constructs appear in the Framework Summary Structures in the 1972–1973, the 1976–1977 Objectives documents, and the 1996 Framework document which references themes contained in *Science for All Americans* (AAAS 1990). In one form or another, cross-disciplinary constructs are mentioned in publications related to the 1969 assessment and in the Benchmarks for Science Literacy (AAAS 1993) and the National Science Education Standards (National Research Council 1996). The 2009 Framework document mentions themes in reference to the Steering Committee Guidelines; however, themes do not play a significant role in the framework or the Framework Summary Structure.

Grade 4	Grade 8	Grade 12
P4.3: Matter exists in several different states; the most common states are solid, liquid, and gas. Each state of matter has unique properties. For instance, gases are easily compressed while solids and liquids are not. The shape of a solid is independent of its container; liquids and gases take the shape of their containers.	**P8.1:** Properties of solids, liquids, and gases are explained by a model of matter that is composed of tiny particles in motion.	**P12.1:** Differences in the physical properties of solids, liquids, and gases are explained by the ways in which the atoms, ions, or molecules of the substances are arranged and the strength of the forces of attraction between the atoms, ions, or molecules
Identifying Science Principles Classify samples of material as solid, liquid, or gas.	**Identifying Science Principles** Given an animation of molecules in motion, identify the substance that is being illustrated as a solid, liquid, or gas.	**Identifying Science Principles** Explain why ice is harder than liquid water in terms of the strength of the force between the molecules.
Using Science Principles Infer that a change of state (e.g., freezing or melting) affects the identity of an object but not the identity of the material of which it is made.	**Using Science Principles** Predict how the mass of a sample of iodine will change after sublimation. Justify the prediction based on what occurs during sublimation at a molecular level.	**Using Science Principles** Use the concept of molecular arrangements and bonds to explain why graphite is very soft and diamond is very hard, even though they are both made of pure carbon.
Using Scientific Inquiry Collect, display, and interpret data showing how the temperature of a substance changes over time as it cools and becomes a solid.	**Using Scientific Inquiry** Plan and conduct an investigation to determine the melting point and boiling point of an unknown substance.	**Using Scientific Inquiry** Using molecular theory, explain the results of experiments showing how the volume of three different liquids changes when they are heated.
Using Technological Design Propose a method for determining for certain whether holiday chocolates that have been shaped by different processes (melting, freezing, reshaping, or breaking into pieces) have the same amount of chocolate in them.	**Using Technological Design** Choose the best solution for increasing the altitude of a hot air balloon based on an understanding of the macroscopic and microscopic changes that occur when the gas inside the balloon is heated.	**Using Technological Design** Design an instrument to measure temperature as accurately as possible, taking into account both the thermal properties of liquids and solids to be used in the device and the structural shape and dimensions of the device.

Fig. 7.7 Performance expectations for states of matter (National Assessment Governing Board 2009, p. 91)

II. Understand and Apply the Fundamental Aspects of science in a Wide Range of Problem Situations

B. Understand and apply laws (principles).

4. Apply some principles of classical mechanics.

Age 9
Understands that strong pushes or pulls produce greater changes in motion than weak ones: understands the operation of the lever as it applies to teeterboards and scissors; demonstrates that heavier objects require stronger supporting structures than lighter objects; applies ideas of floating, for example, to boats.

Age 13
Determines average speeds of objects when given distances and times; selects and/or obtains the data needed to determine speeds. Applies ideas such as gravity, friction, and kinetic energy, although understanding the ideas is probably qualitative rather than quantitative; solves mathematical problems involving these quantities

Age 17
Distinguishes between pairs of selected idea such as mass and weight, speed and acceleration, and kinetic energy and potential energy; works simple mathematical problems involving force, velocity, acceleration, and energy,

Adult
Demonstrates a functional understanding of ideas such as gravity, force, velocity, acceleration, weight, equilibrium, inertia, and friction; may apply enough information about gravitation, celestial mechanics, and laws of motion to understand, for example, the fundamental mechanical problems of space flight. (p.32)

Fig. 7.8 Objectives relating to the fundamental aspects of science (Mastie and Johnson 1972, pp. 31–32)

Two projects currently underway include cross-disciplinary constructs, one the Michigan State University PROM/SE project (n.d.) and the other the NRC Committee on Conceptual Framework for the New K-12 Science Education Standards. PROM/SE with support from the National Science Foundation developed a set of Eight Fundamental Concepts (Schmidt et al. 2011). The NRC *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas* (National Research Council 2011) identifies:

> … two kinds of crosscutting elements of science and engineering: disciplinary core ideas and crosscutting concepts **…** designed so that students continually build on and revise their knowledge and abilities over multiple years, and support the integration of such knowledge and abilities with the practices needed to engage in scientific inquiry and engineering design. (NRC 2011, p. ES1)

Crosscutting concepts are "scientific concepts that bridge across disciplinary boundaries and have explanatory value throughout much of science and engineering" (National Research Council 2011, p. 41). These cross-disciplinary elements are similar to the cross-disciplinary constructs in the NAEP objectives and framework documents.

DOCUMENT	CONSTRUCT
National Assessment of Educational Progress, Science Objectives for the 1972-73 Assessment	Conceptual Schemes
Science Objectives for the Third Assessment 1976-77	Integrated Topics
Science for All Americans (1990)	Common Themes Systems Models Constancy and Change Scale
Benchmarks for Science Literacy (1993)	CommonThemes from SFAA mentioned but not included as benchmarks
NSES (1996)	Unifying Concepts Systems, order, and organization Evidence, models, and explanation Constancy, change, and measurement Evolution and equilibrium Form and function
NAEP Framework (1996)	Themes Models Systems Patterns of Change
NAEP Framework (2009)	Cross cutting content is not represented by abstractions such as “models, “ constancy and change, “or “form and function, “ but is anchored in the content statements themselves
A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas (2011)	Cross-cutting Scientific Concepts: Patterns, similarity, and diversity Cause and effect: mechanism and prediction Scale, proportion, and quantity Systems and system models Energy and matter: flows, cycles and conservation Form and function Stability and change

Fig. 7.9 Cross-disciplinary constructs in national documents

Figure 7.9 contains the cross-disciplinary constructs contained in documents developed in the USA. These are either single concepts or a group of related concepts. Many of the US states’ standards contain concepts drawn from *Science for All Americans* or documents that came after it.

Not all cross-disciplinary constructs are concepts. Some are stated as relationships between two or more concepts—as principles. Examples of this type are the conceptual schemes identified by Lombard and Owen as the content emphasis of science courses in 1965 and the eight PROM/SE Fundamental Concepts identified by expert scientists and educators (Figs. 7.10 and 7.11).

Despite the considerable time difference between these statements, they have notable similarities. Four of the statements match and the majority of statements in

Conceptual Schemes

1. The basic scheme of human physiology and its relation to health, nutrition, and reproduction (9 years)

2. The present model of the universe and celestial mechanics (9 years)

3. The particle nature of matter and the interaction of units (13 years)

4. The kinetic-molecular theory (13 years)

5. The conservation laws (9 years)

6. Electromagnetic radiation and wave phenomena (17 years)

7. The evolutionary character of geographical changes and biological development (13 years)

8. Equilibrium systems and the concepts of minimal energy content and random energy distribution (13 years)

9. The relationship between and organism and its environment (9 years)

The number after each minor objective represents the lowest age level at which the majority of reviewers felt the objective should be assessed. All the objectives were believed to be continuous and cumulative: that is, any objective appropriate for nine-year-olds is also appropriate for the other age levels.

Fig. 7.10 Conceptual schemes from Lombard and Owen (1965, p. 12)

Prom/se Fundamental Concepts

Everything is made of atoms and atoms are composed of subatomic particles.

Cells are the basic units of organisms.

Electromagnetic radiation pervades our world.

Evolution: Systems evolve and change with time according to simple underlying rules or laws.

Parts of a system move and interact with each other though forces.

Parts of a system can exchange energy and matter when they interact.

Physical concepts like energy and mass can be stored and transformed, but are never created or destroyed.

Life Systems evolve through variation.

Fig. 7.11 PROM/SE fundamental concepts (Schmidt et al. 2011, p. 7)

each list are about physical science. Differences are that the PROM/SE Fundamental Concepts are stated more abstractly than the Conceptual Schemes and there is an Earth and space statement in the Conceptual Schemes but not the PROM/SE. Granted, not all the statements are cross-disciplinary; however, most of the physical science statements are cross-disciplinary. What is remarkable about this similarity is consistency over time. The PROM/SE Fundamental Concepts were developed without knowledge of earlier work on cross-disciplinary constructs.

With few exceptions, constructs remain consistent over time:

- Disciplinary structure is explicit in all but the 1969–1970 Objectives document. Biology is referred to as life sciences in later documents, and chemistry and physics are combined into physical sciences in all but the 1985–1986 Objectives document where chemistry is mentioned specifically. (The presence of two chemists on the 1969–1970 Science Learning area Committee might account for this, the only specific mention of chemistry.).
- Facts and concepts are mentioned only in the earlier document summaries; principles are mentioned in the two earliest documents and appear again in the most recent document.
- Cross-disciplinary constructs, conceptual schemes, integrated topics, and themes appear in three documents.
- The nature of science and science and society appear in all but the latest document.

Consistency and Change Over Time

Over the years the rationale for science literacy has remained remarkably consistent. However, changes have occurred in descriptions of the characteristics of science-literate individuals. Changes are often be explained by the state of the nation's economy, the global standing of US scientific endeavors, US student's performance on international assessments, prevailing socio-philosophical perspectives, and learning research. Certain changes were the result of particularly influential individuals on NAEP committees. Examples of these are the appearance of chemistry in the 1969–1970 Objectives document, of practical reasoning in the 1996 Objectives document and technological design in the 2009 Framework document.

Knowledge Product Changes

With respect to knowledge products, the topics, principles, and evidence supporting the validity of the principles have remained the most consistent across the history of NAEP. The knowledge products are structured by discipline, life, physical, Earth, and space science, and stated as principles. Over the years, with few exceptions the principles have remained the same. Conservation of mass and energy is the most consistent. Relatively few new principles have appeared; those related to plate tectonics and DNA are the most evident newcomers.

While the presence of the knowledge and skills related to the nature of science has waxed and waned, the presence of the knowledge products has been almost without change.

Interdisciplinary constructs have a history in NAEP science and may gain more importance with the development of core standards for science based on the new NRC framework (NRC 2011). However, designing strategies for assessing these constructs has proven to be a challenge. It is relatively easy to assess students' understanding of a particular system, such as the digestive system or the solar system. It is more difficult to design assessments that measure students' understanding of systems in the abstract. Without tasks that measure the construct, what it means to understand highly abstract concepts such as system, radiation, or inquiry is left to individual interpretation.

Learning progressions and science assessment introduced in the 2009 Framework document seem to be new ideas in science education. However, the earlier NAEP objectives and framework documents that trace students' understanding of a topic, such as classical mechanics at different ages was clearly an assessment strategy to gain data about the progression of learning the topic in a population of students.

Factors Influencing Change

An important, if weak, influence is that of NSF-sponsored science curricula, which is mentioned in the 1972–1973 Objectives document but not by name—BSCS biology or PSSC physics, for instance—but as the "newer" science curricula. However, their influence is not evident in this document. The 1976–1977 Objectives document notes "[T]he science curricula during the last two decades have placed increased emphasis on methods of thinking and acting that are applicable beyond the narrow limits of the science classroom" (Education Commission of the States 1979, p. 8). This perspective is reflected in the "new emphasis in the 1976–1977 assessment" on decision-making components of the objectives.

The influence of post-Sputnik documents on science literacy is evident in the 1985–1986 Objectives document:

> The objectives for 1985–86 build on what has gone before, taking into account recent developments and new emphases.
>
> A number of prestigious reports critical of the schools sparked unprecedented public debate and calls for improvement across the country. *A Nation at Risk*, the report of the President's Commission on Excellence, called for increased attention to science and for emphasis on the development of higher-order thinking and problem-solving skills, *Educating Americans for the 21st Century*, the report of the National Science Board Commission on Precollege Education in Mathematics, Science and Technology, stated, "We must return to the basics, but the 'basics' of the 21st century are not only reading, writing, and arithmetic. They include communication and higher problem solving skills, and scientific and technological literacy—the *thinking* tools that allow us to understand the technological world around us." (National Assessment of Educational Progress 1986, p. 6)

The response of the Assessment Policy Committee was to call for increased emphasis on problem-solving and higher-order skills in the 1985–1986 assessment. The emphasis on higher-order thinking and higher problem-solving skills was not only acknowledged but also addressed in a NAEP special project, A Pilot Study of Higher-Order thinking skills assessment techniques in science and mathematics (Blumberg et al. 1986).

The 1996 Framework document mentions science literacy only in passing but responds to criticisms in the Alexander-James study, *The Nations' Report Card: Improving the Assessment of Student Achievement* (Alexander and James 1987), that the national assessments of science were "almost exclusively devoted to assessing factual information," and "did not attempt to assess abilities to organize and transform a body of information into a coherent scientific account" (p. 2).

> To address this criticism the study group recommended that NAEP broaden its scope to include collection of information on whether students are able to design, perform, and analyze experiments; on whether they have acquired complex thinking abilities essential to various fields of science; and on whether they can perceive fundamental relationships. To measure these kinds of knowledge and abilities, NAEP was urged to include open-ended items and performance tasks in its assessment techniques. (Alexander and James 1987, p. 2)

The influence of publications of the American Association for the Advancement of Science, the National Center for Science Education, and the National Research Council is evident in the 1996 and 2009 Framework documents both in terms of policy set by the Steering Committees and in the content of the frameworks. For instance, Steering Committee Guidelines in the 2009 Framework required that the framework be informed by the *National Standards and Benchmarks*.

NAEP Document Changes

The seven NAEP objectives and framework documents have similarities and differences. The similarities include some reference to the history of NAEP and the details of the consensus process as it was applied in the development of the particular document. The documents differ considerably in length and in reference to curriculum and the psychological perspective that influenced the approach taken to the description of the characteristics of science-literate individuals. From the first to the most recent document, the trend is increasing length—the 1969 document being a concise 33 pages (6.5 × 22 cm) and the 2009 Framework document being 155 pages (22 × 28 cm). Beginning with the 1985–1986 assessment, two documents were developed, one the objectives or framework document and the other a specifications document containing greater details regarding the assessments instruments including topics to be assessed, the time allotted to the topics, and the kinds of items that would be used.

The increased lengths of the documents as well as the new specifications documents reflect the developing importance of NAEP to US educational policy. As the policy use of NAEP data increased, so did the need for greater specificity for the design of the assessments and psychometric quality.

Assessing valued science knowledge and abilities while maintaining psychometric quality is challenging. The development of an assessment system is a complex process involving the translation of a complex idea, science literacy, into valid and reliable instruments and the presentation and analysis of the data from their administration into reports that are meaningful to policy makers, the general public, and most importantly to educators. My venture into the history of NAEP science convinced me that information about psychometrics and trends across NAEP's history is much more plentiful than information about the science that was tested. Neither the National Center for Education Statistics nor the National Assessment Governing Board had copies of two of the earliest objectives documents. I obtained copies of these only with the help of an outstanding education librarian, Linda Fitch of the Northwest Regional Laboratory.

The psychometric quality of NAEP was the object of a study by the National Research Council, reported in *Grading the Nation's Report Card* (NRC 1999). The executive summary notes that "the frameworks and assessment materials do not capitalize on contemporary research, theory, and practice in ways that would support in-depth interpretations of student knowledge and understanding" and recommends that

> [i]n this assessment development process, multiple conditions need to be met: (a) NAEP frameworks and assessments should reflect subject-matter knowledge; research, theory, and practice regarding what students should understand and how they learn; (b) assessment instruments and scoring criteria should be designed to capture important differences in the levels and types of students' knowledge and understanding both through large-scale surveys and multiple alternative assessment methods; and (c) NAEP reports should provide descriptions of student performance that enhance the interpretation and usefulness of summary scores.

Overall the report is long on psychometrics and learning theory but short on principles for selection of content or ways to assess valued goals of education with psychometric vigor. This emphasis on psychometrics is well intentioned and points to needed improvements in NAEP psychometrics, but content selection is important also. The NAEP objectives and framework documents claim that current best thinking and research were taken into consideration in the development of objectives and frameworks. However, the documents do not contain research supporting the content selected. Consensus, not research, determines the content to be assessed. The research basis for science content selection is as important as research-based strategies for the measurement of that content.

What Have We Learned?

The conclusions drawn here must be regarded with caution. Broadly speaking science literacy can be considered as being comprised of knowledge and abilities related to the knowledge products of science and knowledge and abilities that produce those knowledge products. Features of the scientific enterprise include science practices, attributes of scientists, the philosophy of science, the history of science, and the interactions of science and society. Each feature of the scientific enterprise has many component parts. Inquiry, the complex practice that generates the knowledge products of science, is a case in point. Inquiry has a philosophical foundation, takes many different forms, and has many component parts. Descriptions of the knowledge and abilities that characterize science-literate students contain abstract constructs, such as inquiry, the nature of science, and understanding. Differences in the meaning of these constructs have implications for the consensus process and translating them into assessment tasks. Ultimately it is the tasks that comprise assessment instruments that provide definitive definition. However, items are not available so we must rely on individual interpretation from the information contained in the NAEP objectives and framework documents.

Over the 40 odd years of NAEP Science, the importance of science literacy and the knowledge and abilities that characterize science-literate persons have remained remarkably constant despite the many factors that influence the consensus process. However, issues persist, including how that content is communicated, the breadth of knowledge and abilities expected of science-literate high school graduates, and challenges to the measurement of the most valued knowledge and abilities.

The science knowledge and skills to be assessed have remained remarkably constant over the history of Science NAEP and across contemporary documents describing the science all US students should know: *Benchmarks for Science Literacy* (Benchmarks)(AAAS 1993), The *National Science Education Standards*(NSES) (NRC 1996), and the most recent document to appear on the science education scene, *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (K-12 Framework) (NRC 2011). The consistency is not accidental. The charges to the committees responsible for the *Science framework for the 2009 national assessment of educational progress* (National Assessment Governing Board 2008) and the K-12 Framework explicitly direct the committees to attend to the NSES and the benchmark documents.

What is notable about these documents is that each uses different terminology to describe science constructs and structures these in different ways. For instance, statements of science principles are called content statements and principles and presented as hierarchical lists in the 2009 NAEP Framework. Principles are called core ideas in the K-12 Framework and are organized in paragraphs. In the 2009 NAEP Framework and the K-12 Framework, familiar verbs such as know and understand are called practices, extensive descriptions of behaviors that are indicative of knowing and understanding.

Whatever the constructs are called, the sheer quantity is so large that expecting students to learn it all in K-12 science is unreasonable and quite contrary to the expectations in other countries (Schmidt et al. 1977). Even so the experts writing the K-12 Framework added engineering practices to the already lengthy requirements for K-12 science.

Large-scale assessments provide important data to policy makers and send strong messages to curriculum developers and teachers about what science is important. However, designing and implementing such assessments to provide valid and reliable data are resource-intensive endeavors, and assessing the most valued knowledge and abilities is frequently the most resource intensive. For instance, a 50 min/student limit on assessment time, as exists in some parts of the globe, means that important performance objectives such as sustained research work cannot be adequately measured. In addition, strategies are yet to be developed for the measurement of student understanding of certain constructs such as big ideas. It is possible to assess the idea that energy is conserved in specific instances, but we have yet to figure out how to assess students' understanding of the power of energy conservation to explain phenomena across many contexts.

What Do We Make of It?

Ultimately, the greatest challenge facing science education is limiting the science content all students are expected to learn. Once identified, these most powerful ideas and abilities must be communicated to policy makers and educational practitioners so that they can be translated into classroom and assessment practices.

Appendix 1: NAEP Objectives and Framework Documents

Document title	Assessment year(s)	Referenced in text as
Science Objectives	1969–1970	1969–1970 Objectives document
National Assessment of Educational Progress, Science Objectives for 1972–73 Assessment	1972–1973	1972–1973 Objectives document
Science Objectives for the Third Assessment	1976–1977	1976–1977 Objectives document
Science Objectives, 1985–86 Assessment	1985–1986	1985–1986 Objectives document
Science Objectives. 1990 Assessment	1990	1990 Objectives document
Science Framework for the 1996 National Assessment of Educational Progress	1996	1996 Framework document
Science Framework for the 2009 National Assessment of Educational Progress	2009	2009 Framework document

Appendix 2: Framework Summary Structures

1969 Assessment

Four primary objectives

Sub-objectives

Age 9, 13, 17, Adult

Primary objectives:

I. Know fundamental facts and principles of science
II. Possess abilities and skills needed to engage in the processes of science

III. Understand the investigative nature of science
IV. Have attitudes about and appreciation of scientists, science, and the consequences of science that stem from adequate understandings

(Committee on Assessing the Progress of Education 1969, pp. 9–23)

1972–73 Assessment

Elements of scientific literacy a two-dimensional grid used in the organization of the science objectives

	Fundamental aspects of science				
Objectives	Facts and simple concepts	Laws (principles)	Conceptual schemes	Inquiry skills	The scientific enterprise
Know	(I A)	(I B)	(I C)	(I D)	(I E)
Understand and apply	(II A)	(II B)	(II C)	(II D)	(II E)
Appreciate	(III A)			(III B)	(III C)

Mastie and Johnson (1972)

Third Assessment: 1976–1977

		Knowledge	Comprehension	Application	Synthesis	Percent weighting
Content	Biology					15
	Physical science					15
	Earth science					10
	Integrated topics					10
Processes	Process/methods					18
	Decision making					12
Science and Society	Societal problems					7
	Science and self					7
	Applied science					6
	Percent weighting	20 %	20 %	40 %	20 %	100 %

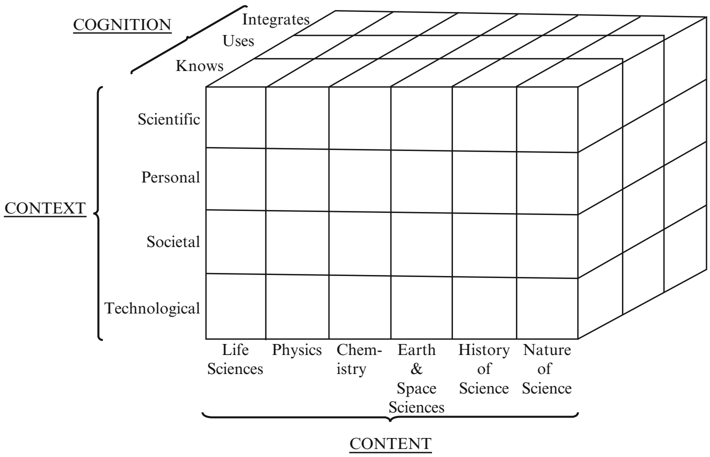
Cognitive Objectives Matrix

(Education Commission of the States 1979, p. 3)

1985–86 Assessment

Exhibit 1

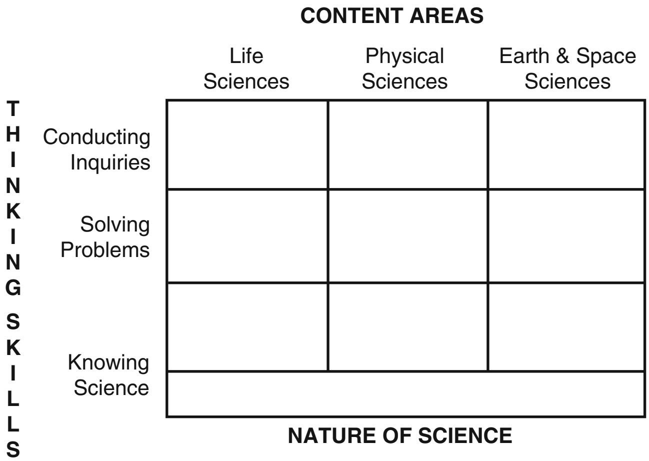
Framework for Science Assessment Exercises



(National Assessment of Educational Progress 1986, p. 8)

1990 Assessment

Framework for the 1990 Assessment



(National Assessment of Educational Progress 1989, p. 12)

1996 Assessment

1996 Science Framework Matrix

	Fields of Science		
Knowing and Doing	Earth	Physical	Life
Conceptual Understanding			
Scientific Investigation			
Practical Reasoning			
	Nature of Science		
	Themes Models, Systems, Patterns of Change		

(National Assessment Governing Board n.d., p. 13)

2009 Assessment

Performance Expectations Matrix

		Science content		
		Physical science principles	Life science principles	Earth and space principles
Science practices	Identifying science principles Using science principles Using scientific inquiry Using technological design			

(National Assessment Governing Board 2008, p. 82)

Item Distribution

	Grade 4	Grade 8	Grade 12
Distribution of items by content area and grade as percentage of student response time			
Physical science	33.3	30.0	37.5
Life science	33.3	30.0	37.5
Earth and space science	33.3	40.0	25.0
Distribution of items by science practice and grade as percentage of student response time			
Identifying science principles	30	25	20
Using science principles	30	35	40
Using scientific inquiry	30	30	30
Using technological design	10	10	10

(National Assessment Governing Board 2008, p. 113)

References

AAAS. (1990). *Science for all Americans*. New York: Oxford University.

AAAS. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Alexander, L., & James, H.T. (1987). *The nation's report card: Improving the assessment of student achievement.* (With a review of the report by a Committee of the National Academy of Education, by Glaser, R.). Washington, DC: National Academy of Education.

Bloom, B. B. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1. Cognitive domain*. New York: McKay.

Blumberg, F., Epstein, M., MacDonald, W., & Mullis, I. (1986). *A pilot study of higher-order thinking skills assessment techniques in science and mathematics*. Princeton: National Assessment of Educational Progress.

Committee on Assessing the Progress of Education. (1969). *National assessment of educational progress: Science objectives*. Ann Arbor: Committee on Assessing the Progress of Education.

Corcoran, T., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform*. New York: Consortium for Policy Research in Education. Center on Continuous Instructional Improvement. Teachers College-Columbia University.

Education Commission of the States. (1979). *Science objectives for the third Assessment*. Denver: Education Commission of the States.

Fitzharris, L. H. (1993). *An historical review of the National Assessment of Educational Progress from 1963–1991*. Unpublished doctoral dissertation, University of South Carolina.

Gewertz, C. (2011). 'Manifesto' proposing shared curriculum draws counterattack. *Education Week, 30*(31), 9.

Lombard, J. W., & Owen, W. B. (1965). *Objectives of science education: Compiled as a guide to the development of an assessment program in science education*. Ann Arbor: Committee on Assessing the Progress of Education.

Loveless, T. (2009). *The 2008 brown center report on education: How well are American students learning?* (Vol. II, No. 3). Washington, DC: Brookings Institution. http://www.brookings.edu/reports/2009/0225_education_loveless.aspx. Accessed 1 Sept 2011.

Mastie, M. M., & Johnson, G. H. (Eds.). (1972). *National Assessment of Educational Progress, science objectives for 1972–1973 assessment*. Denver: National Assessment Office.
Mastie, M., Johnson, M., & George, H. (Eds.). (1972). *National Assessment of Educational Progress, science objectives for 1972–73 assessment*. Denver: Education Commission of the States.
Merwin, J. C. (1966). The progress of exploration toward a national assessment of educational progress. *Journal of Educational Measurement, 3*(1), 5–10.
Merwin, J., & Tyler, R. (1965). *Working paper on objectives. Education Commission of the States*. Ann Arbor: Exploratory Committee on Assessing the Progress of Education.
Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A., & Eberber, E. (2005). *TIMSS 2007 assessment frameworks*. Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
National Assessment Governing Board. (2008). *Science framework for the 2009 national assessment of educational progress*. Washington, DC: National Assessment Governing Board.
National Assessment Governing Board. (2009). *Science framework for the 2009 National Assessment of Educational Progress*. Washington, DC: National Assessment Governing Board.
National Assessment Governing Board. (n.d.). *Science framework for the 1996 National Assessment of Educational Progress*. NAEP Science Consensus Committee. Washington, DC: National Assessment Governing Board.
National Assessment of Educational Progress. (1986). *Science objectives, 1985–86 assessment*. Princeton: National Assessment of Educational Progress, Educational Testing Service.
National Assessment of Educational Progress. (1989). *Science objectives, 1990 assessment*. Princeton: National Assessment of Educational Progress, Educational Testing Service.
National Research Council. (1996). *National science education standards*. Washington, DC: The National Academies Press.
National Research Council. (1999). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress*. Washington, DC: The National Academies Press.
National Research Council. (2005). *How students learn in the classroom*. Washington, DC: The National Academies Press.
National Research Council. (2007). *Taking science to school*. Washington, DC: The National Academies Press.
National Research Council. (2011). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Committee on a Conceptual Framework for New K-12 Science education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
No Child Left Behind Act of 2001, Pub. L. No. 107–110, Stat. 1425 (2002). American Recovery and Reinvestment Act of 2009 (ARRA), Section 14005–6, Title XIV, (Public Law 111–5).
PROM/SE. (n.d.). *Promoting Rigorous Outcomes in Mathematics and Science Education (PROM/SE)*. Michigan State University Mathematics and Science Partnership project supported by the National Science Foundation cooperative agreement EHR-0314866 (2003–12).
Schmidt, W. H., Raizen, S. A., Britton, E. D., Bianchi, L. J., & Wolfe, R. G. (1977). *Many visions, many aims, Volume 2: A cross-national investigation of curricular intentions in school science*. Dordrecht: Kluwer.
Schmidt, W. H., Leroi, G., Billinge, S., Lederman, L. Champagne, A., Hake, R., Heron, P., McDermott, L., Myers, F., Otto, R., Pasachoff, J., Pennypacker, C., & Williams, P. (2011). Towards coherence in science instruction: A framework for science literacy. *PROM/SE Research Report*, *Vol. 8*. Michigan State University.

Smith, C., Wiser, C. W., Anderson, A., Krajcik, J., & Coppola, B. (2004). *Implications of research on children's learning for assessment: Matter and atomic molecular theory.* Paper prepared for the National Research Council, Center for Education, Board on Testing and Assessment, Committee on Test Design for K-12 Science Achievement, Washington, DC.

Tyler, R. W. (1966). The objectives and plans for a national assessment of educational progress. *Journal of Educational Measurement, 3*(1), 1–4.

Venezky, R. L. (1990). Definitions of literacy. In R. L. Venezky, D. A. Wagner, & B. S. Ciliberti (Eds.), *Toward defining literacy* (pp. 2–16). Newark, DE: International Reading Association.

Chapter 8
The Influence of Assessment on Moderating Science Teachers' Beliefs and Practices

Deborah Corrigan and Rebecca Cooper

Introduction

This chapter will consider the moderating influence of assessment on secondary school science teachers' notions of what is important in science education and their practices. Six senior secondary science teachers from different states of Australia (Queensland and Victoria) were involved in this research as these two states represent very different policy approaches to assessment at senior years of schooling. This research highlights the influence such policies can have on moderating teachers' practices and the differences between what they think is important in science education and the reality of their classrooms.

Assessment Policies in Victoria and Queensland

In Victoria, teachers undertake assessment of student learning within their schools in all years with the exception of the final year. In this final year (Year 12, 17–18-year-old students), science students are assessed through two externally set and marked examinations (66 % of the weighting) as well as School Assessed Coursework (SACs) (34 % of the weighting) set by the classroom teacher. SACs can be in the form of a report on an experiment, a student-designed extended practical investigation, an

D. Corrigan (✉)
Centre for Science, Mathematics and Technology Education, Faculty of Education, Monash University, Building 6, Wellington Road, Clayton, VIC 3800, Australia
e-mail: debbie.corrigan@monash.edu

R. Cooper
Faculty of Education, Monash University, Australia

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_8,

analysis and response to media items, a PowerPoint presentation, an oral presentation or an analysis and response to a given set of data. SACs should be part of the regular classwork.

In Queensland an internal assessment system similar to the Victorian one operates up until the last three years of schooling. During these final 3 years, students are assessed on work that is set and assessed by their classroom teacher (100 % of the weighting). These school-based assessments fall into two main categories: externally moderated and quality assured. Externally moderated assessment is set by the teacher but is then taken to a panel of teachers of that subject for moderation of the task. The quality assured assessment is set and moderated only within the school and is similar in nature to the SACs in Victoria. The tasks set for either category of assessment could take the form of school-based exams, observation of practical performances, projects, assignments or fieldwork and often include a student-designed extended experimental investigation (EEI). As all of the assessment is school-based, it is often integrated into the classroom activities and used for both formative and summative purposes.

While there are many similarities in these two assessment contexts, Queensland's lack of an externally set examination and the emphasis on internally set assessment that is moderated by teachers has meant that this moderation process is often viewed as important, highly valued professional development for the teachers. The different demands made on Queensland and Victorian teachers in senior secondary school in terms of meeting assessment requirements place the teachers from these two states in different teaching environments that can often challenge their views of what is important for their students to learn in science.

This chapter will focus on looking at what science teachers see as important in science (their ideal view of science education) and the reality they experience in their classrooms. This reality includes the structural frames they work in, which importantly for senior secondary science teachers include the assessment of their students at exit levels of schooling, their interpretation and implementation of science curriculum as well as the nature of the specific students they teach. In this sense, this chapter will focus on an empirical study where researchers were working with teachers who use formative assessment in very different summative assessment systems as outlined above.

Values of Science

In exploring these teachers' ideal view of science education, the study attempted to establish what was important to the teachers through the lens of values that underpin the discipline of science. The examination of values is appropriate, as according to Halstead (1996), values are the

> … principles, fundamental convictions, ideas, standards or life stances which act as general guides to behaviour or as points of reference in decision making or the evaluation of beliefs or actions and which are closely connected to personal integrity and personal identity. (p. 5)

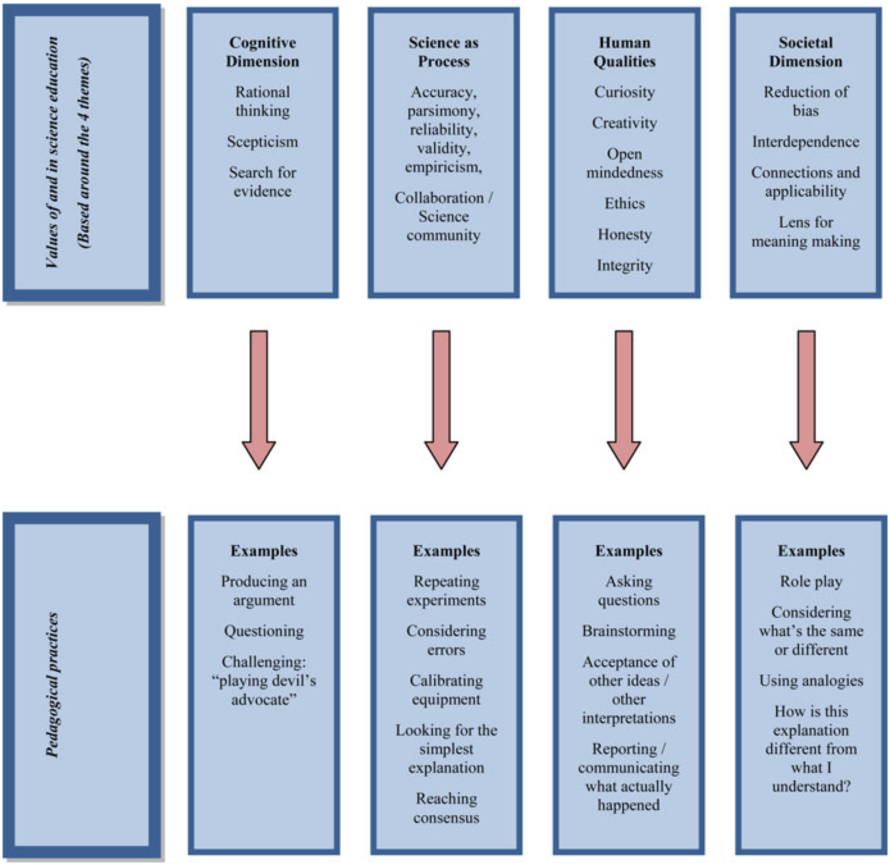


Fig. 8.1 Conceptual framework for the manifestation of values of science in teachers' practice (Adapted from Corrigan and Cooper 2011)

Values of science then should represent the nature of science including the epistemic and sociological values of science. Epistemic values of science include cognitive aspects, such as rational thinking and scepticism, as well as how such knowledge is generated, such as the following: is the data accurate, reliable, valid and empirical in nature? Sociological values of science include both external values such as reduction of bias and the lens chosen for meaning making, and internal values such as curiosity, creativity and open-mindedness. In studying science teachers in Victoria, Corrigan and Gunstone (2007) explored values of science under four broad themes—science as a process, human qualities, cognitive dimensions and societal dimensions—which are represented below in Fig. 8.1. The values represented here promote science as a way of thinking and acting and are consistent with the notion of values as "guides to behaviour" as defined by Halstead (1996, p. 5).

In Fig. 8.1 we have also included pedagogical practices as ways of observing the manifestations of the values within science classes. This is largely because in reality it is the pedagogical practices that can be observed, while the values need to be inferred or specifically sought through means such as interviews, as in the project discussed below.

It is interesting to consider what the arrows linking these two dimensions might represent. Perhaps they are events or things that are experienced in some way that alters a person's beliefs or attitudes, which may eventually accumulate to alter the values held. In this instance, the arrows may represent the interviews both before and after the observation of the pedagogical practices, the interviews having allowed teachers to articulate their beliefs about teaching science and their views of the values underpinning science, providing greater validity to the inferences being made with respect to the observations.

In considering pedagogical practices as what is being observed, it can often be difficult to understand what inferences are being made from these observations. For example, values, attitudes and beliefs, terms that are often used erroneously as interchangeable, comprise intellectual (or cognitive) as well as affective dimensions. Pajares (1992), in reviewing the literature in this field, considers Rokeach's (1968) perspective that "all beliefs have a cognitive component representing knowledge, an affective component capable of arousing emotion, and a behavioral component activated when action is required" (cited in Pajares 1992, p. 314). According to Pajares, Rokeach sees knowledge as a component of belief and relates beliefs to attitudes and values in the following way:

> When clusters of beliefs are organized around an object or situation and predisposed to action, this holistic organization becomes an attitude. Beliefs may also become values, which house the evaluative, comparative, and judgmental functions of beliefs and replace predisposition with an imperative to action. Beliefs, attitudes, and values form an individual's belief system.
>
> Understanding beliefs, Rokeach cautioned, requires making inferences about individuals' underlying states, inferences fraught with difficulty because individuals are often unable or unwilling, for many reasons, to accurately represent their beliefs. For this reason, beliefs cannot be directly observed or measured but must be inferred from what people say, intend, and do—fundamental prerequisites that educational researchers have seldom followed. (p. 314)

In the next section, we will examine an example of a science teacher and learning project and consider how the research being reported in this chapter utilised aspects of that project along with the frame represented in Fig. 8.1 to collect and analyse data.

The Science Teacher and Learning Project

The Catholic Education Office (2005, 2006, Melbourne, Australia) in collaboration with researchers from the Centre for Science, Mathematics and Technology Education (CSTME), Monash University, has been developing ways of growing and supporting science teacher researchers for eight years (2004–2011). Science teacher research is catalysed through opportunities for science teachers to collaborate more closely with science education academic supporters through more systematic research into their practice and their students' learning and to reframe (Schön 1987) ideas about science education such that science teaching can be viewed as problematic and become a site for genuine inquiry. Reframing is, in part, facilitated through a consistent focus on the nature of quality in science teaching and learning. It creates a questioning stance in teachers and simultaneously creates new ways to inquire into the practice setting, which is critical in shaping the outcomes, and ultimate value, of science teacher research as a means for developing, articulating and sharing sophisticated knowledge of practice. With the clear articulation of such knowledge and with a documented evidence base underpinning such knowledge, science teacher researchers have the opportunity to close the gap between science education research and practice—and also science education policy—as the new knowledge becomes communicated and developed more broadly.

This chapter will utilise aspects of the project outlined above to investigate links between what the six teachers involved in this research articulate as being important in science education and the pedagogical practices that they use in their classrooms. In identifying what the teachers felt was important in science, it was essential to establish their personal stance on the view of science as a way of thinking and acting as well as their views on what their students need to learn. Many of these aspects were elicited through semi-structured in-depth interviews with teachers prior to and after observing their science classes. These insights into the teachers' beliefs about and attitudes towards teaching science allowed the researchers to observe the pedagogical practices with a clearer focus on what the teachers intended. This, in turn, allowed the researchers to match the practices and inferences with the values of science that the teachers were promoting.

Outlining the Research

Through peer recognition, six expert science teachers were identified with whom we explored their conceptions of science as a way of thinking and acting, what they felt it was important for their students to learn, what values of science underpinned these conceptions and how this matched with their teaching practice. The six teachers (three in Victoria and three in Queensland) teach both senior secondary (students aged 16–18 years) biology, chemistry and/or physics and middle secondary (students aged 12–15 years) science classes.

Initial exploration of their conceptions and values was conducted via a semi-structured interview (Robson 2002). This allowed the teachers to speak openly about their teaching experiences and to expand their thinking throughout the interview. The direction that the interviews took was therefore determined by the researchers and the teacher as the interview progressed. This approach provided rich, detailed stories about the teachers' practice that, when analysed, allowed us to identify examples of the values underpinning their teaching practice.

The six teachers were filmed teaching two lessons each: one senior science class and one class of Year 7–10 science. These videos again provided rich examples, this time of their actual practice in the classroom, and were an additional data source to the unstructured interviews where they verbalised their practices.

The next steps in the research included providing these teachers with an opportunity for exploring their own practices based on the frame outlined in Fig. 8.1. Teachers were able to use both their interview and video data to comment on the frame as they saw it in their practice. The teachers had the opportunity to discuss their beliefs about science and science education and the values of science that had been promoted in their practice. To conclude the process, teachers were asked to plan a lesson that would portray at least one specific value of science to their students.

The Ideal Versus the Reality

In their middle school science classes, the teachers from Victoria all value *human qualities* (see Fig. 8.1 above), paying particular attention to encouraging curiosity and open-mindedness in their students. They also value *science as a process*, emphasising accuracy, but with varying degrees of conviction. This suggests that in their middle school classes, the teachers are promoting values related to exploration, seeing what is possible and maintaining an appropriate level of accuracy.

In their senior school classes, the teachers from Victoria all emphasised *science as process* with their classes, along with *cognitive dimensions*. It appears that teachers are promoting values of science in their senior classes that are more related to responding to questions and understanding content. The teachers felt that the higher stakes of Year 12 (the final year of secondary school) meant that they could not take as many risks with their teaching and that the purpose of teaching science shifted to focus on the external examination as a major driver of what and how to teach. Mark, a secondary school science teacher in Victoria who teaches middle school science and senior school Biology, summed up this thinking:

> So for me at Year 10 level, the example for Year 10 level is driver safety and relating that to forces and motion. Fifty years ago no one thought about crash investigation, why crashes happen and the science involved, and now it's everywhere. So that's an example of how science has come into an area of society where it has relevance in a lot of different areas. How we build roads, how we educate drivers, laws that are made and how we treat people with injuries who experience these sorts of things. So I think that's an area of science that I think has become… relevant to this time in their lives at Year 10 because they are probably getting behind the wheel of a car for the first time. So that's one example that's probably been around since I've started teaching [and] I've made it a priority of teaching motion and forces at Year 10. I just couldn't do it without relating it to driving and driver safety.
>
> … It's Year 12, different views, the pressure of an exam at the end. I would like to think that I'm enthusing them to understand biology rather than just sitting the exam, but I can't avoid the content, I have to teach it. So for me, it [the exam] has to be a big part of where I need to get their understanding to… if I don't want parents ringing up or a principal or head of campus coming to me.

In an effort to emphasise some of the values of science that he believes are important, Mark created a Year 10 (16-year-old students) subject called Creative

Science that explored the links between art (visual, aural and dramatic) and science. The links considered were wide ranging and could include considering the work of Jackson Pollock and the effect of drip rate and hole size on the end result, looking for shapes that are naturally formed within the work of artists, or considering the inspiration that art can provide to science and vice versa. During the lesson that was filmed, the students were experimenting with dripping and flicking paint and analysing the patterns formed:

> Mark is quietly moving around the classroom having put some background music on. He goes over to assist a student who is investigating the difference that height makes when using the drip method for painting. He assists her to take measurements but then continues to move around the room. Mark moves towards a student who has stopped what she is doing as the music dramatically changes. She remarks that this music no longer suits her work and he engages in quite an in-depth discussion with her about the effect of music on his own work, particularly his work as a research scientist. (Researcher field notes)

Contrasted with this is Mark's Year 12 biology class that was filmed:

> The day before this class was filmed the students used a computer simulation to collect data to answer several questions set for homework. Mark begins the lesson with a brief class discussion about the similarities and differences between two of the questions. He then moves on to another activity that is explained as being a way of consolidating their understanding and use of the terms autosomal dominant, autosomal recessive, X-linked dominant and X-linked recessive. The students are given four pedigree charts to discuss in pairs. They are asked to decide which type of inheritance is affecting each pedigree and to justify their selection. The pairs then join up to make groups of four and the groups of four then have to come to consensus about the type of inheritance in each pedigree. Each group records their answer on the board and the group's selections are discussed. The class agreed on 3 out of the 4 pedigrees and Mark facilitates a long and in-depth discussion of the possibilities for the fourth pedigree that culminates in the class discovering that there are actually two correct answers. Mark concludes the discussion by asking the students to consider what they would do if this had been an examination question, which results in one answer being suggested as more correct. (Researcher field notes)

While it is understandable that Mark would certainly want to draw links to the examination at some stage, it is interesting to note that he does this at the end of the class, in a sense displaying it as the final pinnacle that the students are working towards.

In Mark's own words, taken from interview:

> The dilemma I have and you're right that content is a killer. I succumbed to it even though it's not the way I would necessarily like to teach a Year 12 year. But I've always felt that we're doing that [working with the content] now, and I've been fortunate that I've been able to introduce another subject in Year 10 called creative science, where I do exactly what you've just explained [work in a more exploratory and open style].... And I guess for me the key here is to develop that in the Year 10 and Year 11 classes so that in Year 12 it is already part of their thinking. And then I feel that there is an opportunity to be able to do what I like to do in Year 12 given the rigours of Year 12 and the situation.

This dilemma was found to be common among Victorian teachers. For Victorian teachers teaching Year 12, preparing students well for the external examinations is more important than the values of science they want to promote in their science classrooms.

This research process has allowed us to explore the degree of connectedness between what teachers say they believe is important in science and science education (teacher's ideal) and their practice in the classroom (the reality). The data indicates that for these Victorian teachers, with these classes, the level they are teaching makes a difference to the values of science that they promote with their students. The teachers from Victoria clearly articulate that they believe that certain values of science are important (i.e. human qualities) but feel that at the senior levels they are unable to emphasise these due to the pressures of external assessment and, in particular, the examinations. The teachers feel that they are not doing the right thing for the students unless they focus on the external assessments and provide the students with what they believe they need in order to be able to perform at their best.

In Queensland the teachers, within a set of guidelines, determine the assessment at all levels of secondary science education. Greg, a secondary school science teacher in Queensland who teaches middle school science and senior school biology, describes the assessment process he implements across his teaching in secondary school:

Greg	… So if they're honest with their experimenting—especially in Year 10 and Year 9—it transfers in Year 12 where they do work as if it was on the outside [of school]. Like a SAL[inity] test, you don't make up the results for a SAL[inity] test. You just do it and you sign off on it [be]cause it's your name you're putting to the results so to speak.
Interviewer 1	Yeah, building that sense of reputation… so people have faith.
Greg	Yeah, faith in your results.
Interviewer 2	So are there opportunities for you to assess that sort of thing as you go through the years?
Greg	The Year 12 kids now I had in Year 10. And it's good because the ones I did teach…it's like as they get older, the boys that I've had before know how I sort of work. So they're always looking for stuff on the net and always coming up to me and saying, "Hey Sir have you seen this?" and "What do you reckon?" You know, just discuss things, where the other boys…I've started trying to get them into that sort of thinking process, but it's hard when you have a different class every year. It changes. That's how I sort of gauge it.
Interviewer 1	So your Year 12 was pretty similar to Year 10…or Year 9. We saw your Year 9 class. Is that just because you think that's the important way to teach? I mean the fact that it's Year 12....
Greg	Yeah, it doesn't matter the age because they're still kids. They're still 17 and they're boys and they still want to do hands on stuff and they still want to do fun things but obviously to a bit of a higher level. So I'll try to still do the same things throughout all the years, like you know the same values throughout the Year 9s, just so if they had me in Year 12, they'd know what to expect.
Interviewer 1	So you don't seem to shift in your teaching and what you think is important just because it's Year 12… So that whole notion… I mean you want them to do well… But that's not the driver? It doesn't appear to be the driver for you.

Greg The driver for that is the parents and for them [the students] to do well. But I think more of the drive is.... Well for me, I think you learn a lot more of the science at the university level, and if you enjoy science then you're going to study it. So if you enjoy something, then you're going to pick up science at university and study it and that's when you sort of get the real nitty-gritty concepts. I was speaking to the careers counsellor years ago and science is going backwards. There's hardly anybody doing it now in the university level. So if you get kids to scientifically think, then it's fine…. Ah well, the parents don't know that I think like that [be]cause you know, it's all about the boys at the end of Year 12. And I try and tell the boys it's not the be-all [and] end-all of Grade 12 getting that O.P. [overall position, the final result for students completing secondary school in Queensland, Australia]. No one looks at an O.P. after Grade 12 anyhow.

Considering the differences between the responses from these two teachers who teach the same thing but in different states of Australia, it is clear that the different assessment procedures imposed are having an impact on the teachers. Where Mark feels compelled to work with his students to prepare them for the end of year exam, Greg is more concerned with promoting values of science related to enjoying science and encouraging his students to discuss issues related to science in order to get them thinking. This trend was observed between all the Victorian and Queensland teachers.

Promoting Values of Science and Links to Pedagogical Practices

A little while after Greg had participated in the interviews and classroom observations, the researchers asked him to try to plan a lesson that would promote one of the values of science that he felt was important. What he found when he sat down to plan was the set of values we presented him with (see Fig. 8.1) matched his thinking about this unit with respect to the progression of the learning that he wanted the students to experience. The results, presented in Table 8.1, include a selection of the values of science that Greg selected, Greg's lesson descriptions with a pedagogical practice highlighted and the further explanation of the lessons that he gave during a semi-structured interview with the researchers.

In focusing on Greg in this section, we are able to highlight the evaluations he has made about his pedagogical practices in his science class as he uses his beliefs about science teaching to promote the values of science that he sees as important in teaching and learning science. From his interview and also from his lessons that were filmed, it was clear Greg had a firm belief about the importance of the values of science as an intrinsic aspect of his teaching. He was able to evaluate how his lessons promoted the values of science, based on what he believed was important for students to learn. He did not seem compelled to use all of them or use them evenly, and when interviewed

Table 8.1 Greg's selected values, his lesson's focus and further explanation of the lessons from the interview

Values	Lesson description	Further explanation from interview
Human qualities *Curiosity*	Lesson 1: Urban myths. Class discussion about stories or movies about infectious disease, fact or fiction. Responses recorded o n the board *Pedagogical practice:* Using myths, stories, movies to generate a desire to know	"Yeah, the process in which I'd teach. You know, to value… to engage students initially is **curiosity**. Like for Year 10 science, they're doing diseases and stuff so we do have a discussion about urban myths and stories and movies about infectious diseases. We just talk about that. You know, what diseases they've heard of or what's in the media at the moment about diseases and stuff like that"
Cognitive dimension *Scepticism*	Lessons 2–4: Movie*, what do you believe?* Watch a popular movie (I Am Legend, Outbreak, etc.) and ask students to construct a list of fictional and factual references to infectious diseases throughout the movie *Pedagogical practice:* Generating what is not reasonable—fic-tional as opposed to factual	"But with **scepticism**, we watch a movie, either *I am Legend* or *Outbreak* or *Infectious Disease,* and then they ask…yeah *Outbreak* is great. And then I actually get kids… so **scepticism**, they construct a list of fictional and factual references they use in the movie… like that, and then they have to… whatever is fictional they write down, say why it was fictional and then turn it around to make it factual"

Cognitive dimension *Rational thinking*	Lesson 5: Write a letter *Pedagogical practice:* "Myth bust" the movie using research and logic to improve the scientific credibility for the movie producer/director. This culminates in the students writing a letter to the movie producer/director suggesting corrections for the scientific inaccuracies	"So they came in with their…the **rationality** one… so after the movie I got the kids to write a letter to the director or producer saying, "You need to change this to fix the movie"
Science as process *Accuracy, reliability* **Human qualities** *Integrity*	Lessons 6–7: Experiment *Pedagogical practice:* Engage students in the science process	"After that we do an experiment with the old agar and swab so that covers a lot I think. That covers **integrity**, and **searching for evidence**, and **accuracy**, and **reliability** when they do an experiment. And then they did an experiment, basically an experiment from the book"
Cognitive dimension *Search for evidence*	Experiment: growth and control of pathogens. Swab surfaces on agar plates. Variables for controlling pathogens include different household disinfectants added to the agar—observe colony growth, identify colonies, infer best/worst disinfectant for different pathogens	

(continued)

Table 8.1 (continued)

Values	Lesson description	Further explanation from interview
Societal dimension *Interdependence*	Lesson 9: Investigation *Pedagogical practice:* Applying what you know and making decisions Investigation topic: Global warming and the spread of disease are there links?	"Ah yeah, so **interdependence**…I had just an investigation topic. Sometimes I just throw this in, with nature and science and human beings and interactions in nature. I sometimes do that if the kids are an advanced group. I try and get them to write what's the link between global warming and the spread of diseases, like sort of what the humans do and stuff like that"
Human qualities *Creativity*	Lesson 10: Design *Pedagogical practice:* Make your own design Students design their own disease in a group. They should consider: transmission, symptoms, mortality, treatment/control, creative microscopic drawing and name of disease. Students are to present their disease to the class (gallery walk). This could be viewed as **valuing the science community**	"And then **creativity**—I get the kids to design their own disease. So they get a piece of paper. It's like a Venn diagram but sort of a little bit different. They've got to say the transmission, symptoms, mortality, treatment, and they have to draw it microscopically. Like say if it's a bacteria, like what type of bacterial shape it is. And then you put them around the class, and they do a gallery walk so they look at all the other diseases"

he commented that he used them when they fit (which was frequently) with what he believed it was important for his students to understand.

For Greg, the values of science in Fig. 8.1 have meaning because they represent what science means for him. His prior experience as an environmental scientist assists him to see science beyond his own classroom and he is able to help his students consider the implications of what happens if one works in science. For example, Greg commented: "Students should always measure and record accurately. They [the students] know they need three results for an average but they don't want to do it. What they don't realise is that in industry accuracy equals dollars". Greg's experience in industry has strongly influenced the way he thinks about and uses science and this has manifested in the values of science that he promotes with his students through his teaching; this he does in ways he believes represent science more authentically for his students.

Formative Assessment in a Summative System

Promoting particular values of science in their classrooms sees teachers assessing different things throughout their lessons. These assessments are often formative and take place as an integrated part of the teaching. They may involve the entire class, a small group of students or an individual student, and the feedback is usually provided immediately as a discussion. For example, we filmed Larry, a senior physics and middle school science teacher in Queensland, teaching a Year 9 class (15-year-old students) that had just completed a series of experiments to investigate chemical and physical change:

> Towards the conclusion of the lesson, Larry brought the class together and asked the students, "Why is it a chemical change? What's the clue?" This was followed by asking the students, "Can you get me back to where we started?" The students replied that they couldn't because they were chemical changes but Larry, clearly not satisfied with this response, pushes them for more and asks again, "What's the clue for that?" The students' hands shoot up and Larry is shown the actual test tubes, beakers and petri dishes containing the results of the several experiments. Larry moves around the room to each group and has them explain what they have observed as the clue for a chemical change. (Researcher field notes)

Larry's assessment of their explanation, although informal and brief, does assist the students to concisely explain what they have observed during the experiment. Through assessing their explanations of the chemical changes in the experiment, he is promoting cognitive aspects of science, challenging the students to produce a stronger argument by referring to evidence.

Larry also taught a Year 12 physics class that was filmed and that also involved a student investigation. In this case, Larry had set up several demonstrations that used magnetism, electricity or both and progressed through them to help students explore electromagnetism:

> Larry has worked through all of the demonstrations and a pair of students asks if they can have another look at the demonstrations and have a go themselves. Larry excitedly replies,

> "Go for it" and leaves them to explore on their own. This prompts other members of the class to do the same with the other demonstrations. While this is happening, Larry begins to tell the students about the Maglev train as an example of where an electro-magnet is used in real life and talks about the friction being reduced and the speed of the train increasing. A student then asks, "So, how does the train stop?" This takes the discussion in a whole different direction as Larry facilitates a brainstorming session on what you could do or need to know in order to stop the train. Larry concludes the discussion with a captivating thought about how science can solve a problem (make a train go faster) and create a problem (make the train difficult to stop). (Researcher field notes)

Larry is pushing his students to think well beyond remembering basic facts and to think more about ways they can use what they know to help them solve problems. Larry is assessing his students' abilities to make links between the work they are doing in class and what they see outside of the classroom (i.e. valuing the societal dimension), which is very important to him. He comments during an interview, "It's definitely about making links and… I don't know why, they don't seem to be able to see a connection between doing changing colours in a test tube and what's happening down in the creek".

As can be seen from these examples, Larry consistently promotes his values of science through his pedagogical practices, where the formative assessment practices during his classes build student competency for summative assessment practices. The opportunities he affords his Year 9 students to clarify their explanations of chemical change and his Year 12 students to utilise their knowledge and articulate their links will be useful, he believes, to them during summative assessment tasks.

Contrasted with Larry is a lesson we filmed of Shona, a senior chemistry and middle school science teacher from Victoria, teaching a Year 12 Chemistry class (18-year-old students). During the lesson Shona has been working through some questions from previous Year 12 chemistry exams to try to get students to consider the process of answering the question as opposed to just answering it.

> Shona has a question from a previous exam question projected on the board and she says, "Let's look at this question, what topic is this from?" The students respond and she continues, "Ok, this is an equilibrium question. Right, now, let's write down what are the main things we need to know about equilibrium." The students brainstorm ideas about equilibrium and write them down. "Ok, now let's look at this question, what is it trying to target here?" Shona refers her students back to the ideas they have just written down and some students begin circling a couple of the words on the page. In an interview, Shona explains that she does this in class "so they [the students] can make more of a link between this knowledge that they've got in their head and then see what part of equilibrium is it wanting you to demonstrate? This helps them make links, but it also helps them understand how a question is constructed, how they might go about solving it and encourages them to question the validity of the actual question before they even attempt it." (Researcher field notes)

In contrast to Larry's lesson, this lesson is focused on the summative assessment of this subject, the examination. Shona is making an effort to have her students look beyond getting the correct answer in that she trying to make them think more carefully about the process of answering the question. The formative assessment that is taking place during this class is focussed on the students' ability to structure a response to a question, rather the response itself. While there is certainly not a sense that Shona is promoting science as a set of facts to be recalled, there is a sense that her teaching during this lesson is driven by the impending examination and her responsibility to her students to prepare them for it.

Conclusion

There are multiple ways for teachers to utilise formative assessment in their classrooms while still giving consideration to summative assessment. The teachers in this research who teach in Victoria, while for the most part promoting the values of science that they felt were important, tended to allow these values of science to be pushed aside and somewhat overrun by the requirements of the summative assessment practices that are applied during the final year of secondary school. These teachers felt that unless they allowed the summative assessment practices to drive their pedagogical practices when teaching students in their final year of secondary school, they would not be doing the right thing by their students. Thus, when considering the pedagogical practices of the teachers from Victoria, there was, at times, a discrepancy between what they said were important values of science during the interview and what they actually did in the classroom.

In contrast, the teachers from Queensland had a great deal of connectivity between the values of science they espoused as important during the interview and their observed pedagogical practices, regardless of the level of secondary education. These teachers were able to embed their assessment practices far more fluidly within their pedagogical practices and therefore create opportunities for formative assessment during their classes that better matched the summative assessment tasks that the students were to undertake, regardless of the year level. In this sense, the formative assessment was seen as integral in building student competency for the summative assessment practices.

Teachers' articulation of the values of science that they feel students should know in conjunction with their articulation of their beliefs about teaching science provides a powerful notion of what the teachers hold as their ideal view of teaching and learning science. Unfortunately, the reality does not always align with the ideal. While the values of science may provide enduring principles and life stances for teachers' pedagogical practices, they can be challenged by the environment in which the teacher is working and even be sidelined in the face of enforced structures such summative assessment practices.

This chapter has attempted to provide some explanation around the links (arrows) between the different components within the frame provided in Fig. 8.1 above. These links need to be modified to show the interdependence between each of these elements rather than the one-directional representation in Fig. 8.1. Beliefs about teaching science are an important aspect of the links as they challenge the nature of the values of science held by teachers and at times override the promotion of such values and strongly influence the teacher's pedagogical practices. Examination of the assessment practices that teachers employ in a classroom, or are bound by due to the environment they work in—particularly the formative assessment practices and how they related to the summative assessment practices—can help teachers identify the values of science they are promoting in their science classrooms and consider the impact of both values and beliefs on their pedagogical practices.

References

Catholic Education Office & Monash University. (2005). *Science teaching and learning project*. Melbourne: Catholic Education Office.

Catholic Education Office & Monash University. (2006). *Science teaching and learning project* (Vol. II). Melbourne: Catholic Education Office.

Corrigan, D. J., & Cooper, R. (2011). A frame for linking values and scientific literacy. Paper presented at the 42nd annual Australasian Science Education Research Association, Adelaide.

Corrigan, D. J., & Gunstone, R. (2007). Values in school science and mathematics education: Issues and tensions. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The re-emergence of values in science education* (pp. 133–148). Rotterdam: Sense.

Halstead, M. J. (1996). Values and values education in schools. In M. J. Halstead & M. J. Taylor (Eds.), *Values in education and education in values* (pp. 3–14). London: Falmer.

Pajares, F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307–332.

Robson, C. (2002). *Real world research. A resource for social scientists and practitioner-researchers* (2nd ed.). Oxford, UK: Blackwell.

Rokeach, M. (1968). *Beliefs, attitudes, and values: A theory of organization and change*. San Francisco: Jossey-Bass.

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass.

Chapter 9
Issues in Teaching for and Assessment of Creativity in Mathematics and Science

Mike Askew

Introduction

Creativity is increasingly highlighted on education agendas. Internationally, mandated curricula explicitly include calls for creativity. For example, Scotland's 'Curriculum for Excellence' talks of 'successful learners' who are able to 'think creatively' (Scotland Curriculum Review Programme Board and Scotland Scottish Executive 2004) and Finland has 'competitiveness, creativity, and social justice' as central curriculum aims (Hargreaves et al. 2007). Even nations known for having traditional curricula, such as Japan, Singapore and Korea, are raising the profile of creativity (Park et al. 2006; Schwartz-Geschka 1994; Tan 2000).

The recently introduced Australian National Curriculum is typical, with one of the stated three overall aims being the development of 'confident and creative individuals' (Australian Curriculum Assessment and Reporting Authority [ACARA] 2011a, b). Drilling down into the curriculum details finds creativity repeatedly emphasised. The list of general capabilities for learners to develop as set out by ACARA reiterates the overall aim through the inclusion of 'critical and creative thinking'. At the detailed level of curriculum content, the specifications both for mathematics and for science each again explicitly stress creativity, calling for 'confident, creative users and communicators of mathematics' (ACARA, 2011a, p. 1), and the science curriculum reiterates science as providing the opportunity for learners to 'develop critical and creative thinking skills' (ACARA, 2011b, p. 1).

Calls for creativity within mathematics and science teaching and learning are not new, but having them enshrined in mandated curricula is relatively recent. Despite, however, such curriculum aims and claims from research for the importance of encouraging creativity in science and mathematics, evidence still points to teachers

M. Askew (✉)
Monash University, Building 6, Wellington Road, Clayton, VIC 3800, Australia
e-mail: mike.askew@monash.edu

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_9,

treating the curriculum, particularly for science and mathematics, as a body of fixed knowledge to be delivered and students relying on rote learning and recall (imitative) methods, particularly in mathematics (Hiebert et al. 2003). If creativity is an important learning outcome, then why is the focus in schools still largely on the transmission of facts and the drilling in procedures?

Craft (2000) sums up the tensions in teaching, in being caught between 'soft' skills such as creativity that are seen as necessary for dealing with change and uncertainty and 'hard' skills that are the core of centralised prescription for teaching and assessments:

> If we look at creativity in education, then, we see a need, on the one hand, for teachers to become increasingly experts in fostering creativity and, on the other, an attempt to crush all artistry from the profession and to reduce teaching to a technicist activity. (p. 146)

Jackson (1986) distinguishes between education that is *mimetic*, based on predetermined and measurable content, and the *transformative*, which attends to developing qualities such as values or attitudes. Transformative education means learners are more likely to use maths and science concepts in meaningful ways in their lives outside school (Boaler 1993; Pugh 2004). Despite this, teachers and educational systems are 'more likely to ask questions like, "Do students understand the concepts correctly?" than "Do the concepts make any difference in the students' everyday, out-of-school lives?"' (Pugh and Girod 2007, p. 10).

In this chapter, I examine why calls for creativity in science and mathematics learning and teaching seem so difficult to bring about and the role that assessment may play in promoting creativity. I begin by defining creativity in general and specific aspects of it in mathematics and science. I briefly consider current external assessments and whether or not they suppress or encourage teaching for creativity. I then argue why tests are not appropriate for assessing creativity and how performance tasks are a productive way forward, particularly if supported by holistic assessments and dynamic standards.

While the focus in the main is on looking at the arguments in relation to mathematics, I hope that readers from a science background will be prompted to consider how similar issues hold for science education.

Defining Creativity

Broadfoot (2002) argues that a difficulty in defining standards for attainment arises from 'the failure to locate the search for standards—an assessment challenge—within an appropriate conception of learning itself—a curriculum challenge' (p. 158). The assessment challenge—what assessing creativity might look like—thus cannot be examined until after addressing the curriculum challenge of how to conceive of creativity within mathematics or science classrooms.

One problem is that there are as many definitions of creativity as there are theorists writing about it: in a review of the literature, Treffinger (1996) identifies over

100 different definitions. However, one defines creativity, explicitly or implicitly, will impact upon teaching and learning. There is not the space here to examine the range of definitions and their implications. Instead, I start with this definition:

> Creativity involves intentional imaginative activity producing locally novel and valued outcomes. (Adapted from National Advisory Committee on Creative and Cultural Education [NACCCE] 1999)

I chose this definition because it encapsulates several core ideas. First, although creative activity may involve imagination in considering possibilities, creativity has to go beyond imagination and involve some sort of external output (Robinson 2001). Creative outcomes are usually thought of in terms of some sort of (semi)permanent product but can also include performances, in the broadest sense of that word: learners orally justifying a mathematical conjecture or speculating on a scientific explanation can be thought of as being creative. As Sawyer (2006) points out, the creativity behind performance is even less studied than that behind products.

Second, creativity arises from intentional activity. Without intention, accidental outcomes can come to be seen as creative output. Even outputs that are presented as though they were the result of accident, for example, Jackson Pollock's paintings, turn out to be carefully planned and intentionally executed.

Third, outcomes of this intentional activity need to be novel; otherwise, the activity is reproductive rather than creative. Including 'novel' as a criterion can lead to arguing that school learners cannot be truly creative (in science or mathematics at least) as they are unlikely to advance these disciplines by producing anything novel (Csikszentmihalyi 1996). Others take the position (as I do) that novel can be interpreted as novel to someone, somewhere at sometime. So although the outcomes of creativity in science or mathematics classes may not be novel to experienced scientists or mathematicians, from the perspectives of the learners, they can be judged as novel and hence as creative. This distinction between the creativity that moves a discipline forward and the creativity that produces locally novel outcomes is denoted by some as the difference between Creativity and creativity (Craft 2000). Big C Creativity is associated with 'great works' by experts in contrast to the more everyday small c creativity that might arise when a student creates a solution to a novel problem or connects together two seemingly disparate ideas. Small c creativity moves creativity away from being something that only the few are capable of to being something all learners can engage in (although a substantial amount of the writing on creativity in education is still located within the 'gifted and talented' literature; see, e.g. Treffinger et al. 2002). Herein lies a first difficulty in assessing whether or not learners' outcomes are creative: how to decide whether they have produced things which, although possibly familiar to the teacher, are novel to the learner, or whether learners are reproducing ideas that they have previously encountered. I return to this question of 'novel to whom?' later.

The fourth aspect of creativity is that there has to be some value to the outcome. The value of a solution to a mathematical problem is largely down to whether or not the solution is correct, while considering the difference between two solutions in terms of creativity brings into play such considerations as 'elegance' or 'economy'.

Similarly in science an explanation needs testing out and would also be looked at in terms of the breadth of phenomena for which it can account. The inclusion of value in the definition makes creativity 'fundamentally and unavoidably social' (Sawyer 2006, p. 122). This is important in challenging the view, still popularly held, that creativity is an attribute of individuals. As most of the literature on creativity comes from psychology, the individualist view of creativity still tends to be dominant with sociocultural explanations still in the minority. Vygotsky (1971) acknowledged this many years ago (although sadly only with regard to art!): 'Art is the social within us, even if its action is performed by a single individual, it does not mean that its essence is individual' (p. 249).

One obvious implication of the definition for creativity provided above is that the outcomes of creative activity cannot be fully predicted in advance. This may be one reason why teachers shy away from lessons that involve creative activity as they cannot predict and control the outcomes. And if teachers expect to be able to produce assessment criteria in advance of student activity, then this adds to the view that creativity cannot be assessed. As I shall argue later, this need not be a deterrent to assessing creativity.

Creativity in Science and Mathematics

While the shift from big C Creativity to little c creativity allows teachers to accept that even young children can be creative, a barrier to teaching mathematics or science for creativity and consequently with assessing creativity is the view that these disciplines do not lend themselves to creative endeavour. Robinson's (2001) four-phase model of the creative process provides a framework for thinking about creativity in these disciplines:

– The importance of the medium
– The need to be in control of the medium
– The need to play and take risks
– The need for critical judgment (p. 111)

Importance of the Medium: This draws attention to the fact that creativity is context bound. Psychologists no longer hold to the idea of creative individuals in the sense of possessing a general talent or disposition that they can apply to many contexts. If we want students to be creative in mathematics or science, then the opportunities for that to happen must be made within the mathematics and science lessons. Ways to assess this creative activity must also be developed so that teachers can help learners become more creative within these subjects.

Being in Control of the Medium: This could be taken as support for the argument raised earlier that since young learners cannot be considered to be in control of the medium (be it mathematics or science), then they cannot be capable of creativity activity. One could also argue, however, that the breadth of the disciplines is such that no mathematician or scientist now could ever consider themselves to be in (total) control of the medium and so creative activity cannot be engaged in by anyone. I prefer to think of control here in the sense of control of parts of the medium. In other words, students can demonstrate creativity in the parts of the curriculum of which they have gained control. Young learners have control in this sense: once confident in adding pairs of numbers, they can look for patterns in working with odds and evens and create and test conjectures about patterns they have noticed.

Assessing such creativity requires teachers to disentangle the content that is being played with from the creative processes. Teachers are often most interested in assessing content that they have recently taught, yet research shows that this may not be the content that learners can use creatively. Students in the later years of secondary school, for example, displayed a 2-year gap between being taught some mathematics and being able to apply it (Bell et al. 1983). If we assume that learners are in control of the medium too soon after meeting particular content, then they are unlikely to display creative activity, but this should not be taken as an indication that they cannot be creative.

The Need to Play and Take Risks: Being creative often involves failing or getting things wrong. These are qualities that are often absent from mathematics or science lessons. The measures of success in mathematics that learners pick up on are speed and correctness, neither of which is conducive to playing and taking risks. As Edwards and Mercer (1987) showed in science lessons, students' observations or explanations that did not fit with expected outcomes were often reconstructed to fit with teachers' desired outcomes.

The Need for Critical Judgement: This is linked to the point that creative outputs need to have value. It is often assumed that it is the teacher's role to decide on the value of learners' creative output but opening up the validation to the class could benefit all learners. The work of Cathy Fosnot and colleagues in the 'Young Mathematicians at Work' programme (e.g. Fosnot and Dolk 2001) has many examples of even young learners engaged in dialogue about each other's solutions to problems and going beyond simply judging whether or not these are correct, as they develop their critical judgement. In assessing creativity, self and peer evaluations may be core.

Spaces for Creativity in Science and Mathematics

In science, Newton and Newton (2009) offer suggestions for (at least) four types of creative activity. The first two are encapsulated within the theme of making sense of the world scientifically and in learners constructing either descriptions of or explanations for phenomena. The second two sources of creative activity arise from collecting and evaluating evidence and then constructing either means of gathering descriptive data or ways to test explanations. Newton (2010), drawing on the work of Klahr and Dunbar (1988), further describes generating explanations as working in 'the hypothesis space' and testing these as working in 'the experiment space' (p. 188). As an exemplar of these, Newton offers the task of exploring what happens when dropping a wooden metre rule onto the floor. The question raised, 'What does the ruler bounce?' engages learners in creative activity in the hypothesis space, with a move into the experiment space in testing out their hypotheses.

I suggest that parallels to the hypothesis space and experiment space in mathematics education are 'the conjecturing space' and 'the justifying space'. Children exploring number patterns may notice that the sum of two odd numbers is always even and, rather than taking this as a mathematical 'fact' established on the basis of only a few examples, be encouraged to frame this as a conjecture: 'the sum of any two odd numbers will always be even'. Children enter the justifying space in creating convincing arguments for whether or not they consider their conjecture to always hold true. Different levels of 'justifying' in this example could include checking the conjecture with several more examples; taking an extreme case of two very large odd numbers and showing their sum is even; proving the conjecture by arguing that any odd number is made up of an even number plus one, so adding two odd numbers involves adding two even numbers plus the two 'ones' which must be even.

As pointed out earlier, the content of the learners' creations in these hypothesis/experiment or conjecture/justify spaces has to be considered independently of whether or not it is correct, but in terms of whether or not it might be judged as locally novel and creative. Boesen (2006) sets out a model elaborating mathematical reasoning, making a distinction between 'creative reasoning' and 'imitative reasoning'. Creative reasoning displays elements of being novel, flexible and plausible (which does not mean that it is necessarily correct) with a mathematical foundation. In contrast imitative reasoning involves either recall of reasons, when a complete answer is remembered, or algorithmic reasoning, when a solution procedure is recalled and applied. Helpful though this distinction is, it does not get us round the insider-outsider issue (Newton 2010). As an outsider to the children's world, how can a teacher assess whether the learner's output is new and novel to the (insider) child and not a reproduction of something from elsewhere? Is the learner presenting an argument for why the sum of two odd numbers must be even displaying mathematical creativity or sharing something learnt from elsewhere? In science, Newton (2010) found that pre-service teachers were inclined to rate a student's answer as more creative when the answer fitted with the correct explanation, even when it was clear that the student was reproducing something they had previously learned rather than constructing an explanation.

Assessment and Creativity

I have argued so far that calls for more creativity in mathematics and science lessons involve being clear what creative activity in lessons in these disciplines might look like so that we can then begin to think about ways of assessing learners' creative outputs. But can this argument be turned on its head and assessment itself used as a lever for bringing about changes in classroom practices? So before discussing what assessing creativity might look like, I briefly explore whether current assessment practices, particularly external assessments support or inhibit creativity. I look first at the argument that assessments, and in particular, external assessments (i.e. national or local assessments not chosen or devised by teachers themselves) have a narrowing effect on teaching and learning and drive out opportunities for creativity in science and mathematics. I then look at research that raises questions about whether assessments that do value creativity can in and of themselves encourage teaching for creativity.

Assessment as a Barrier to Creativity

Those arguing that assessment has a narrowing effect base this on the claim that current assessment techniques favour certain learning outcomes, in particular those outcomes that are easier to assess and which tend to be based around recall and application of procedures. If this is the case, then teachers in the knowledge of what will be assessed focus their energies on 'teaching to the test', and students' awareness of the sort of assessments they are going to encounter impacts both on what and how they learn (Gipps 1994; Sadler 2002). Recall and procedures come to dominate lessons as these are what are going to be valorised through assessment, effectively sidelining other, possibly more valuable, learning outcomes. Even if the intended curriculum includes statements about the importance of creativity, the implemented curriculum comes to focus on outcomes that are more easily assessed, even if these are less educationally valuable.

Nevertheless, it is important also to ask whether it actually is the case that national assessments have a focus on recall and imitative reasoning. The item in Fig. 9.1 is adapted from England's national mathematics test for 11-year-olds and demonstrates how assessments that may look, on the surface, as assessing recall or procedures can have more to them.

At first glance this assessment item appears to simply assess recognition of fractions, but it is more challenging than that. It is unlikely that children will have met the noncanonical representation of a third in the first diagram. In reasoning out an answer, children have to coordinate the information presented in the diagram with the worded direction as to what the unit is and conclude that, despite two of the 'thirds' represented in the first diagram being recreated in the second, the fraction shaded is still 1/3. Teaching to such test items is not simply a matter of practising old papers in the knowledge that similar questions regularly appear over the years. The national tests in England have a preponderance of such items that are not easily answered by recall or application of memorised procedures. Yet despite this, there is still much talk of teachers 'teaching to the test' and how the national tests prevent teaching focusing on reasoning or inquiry.

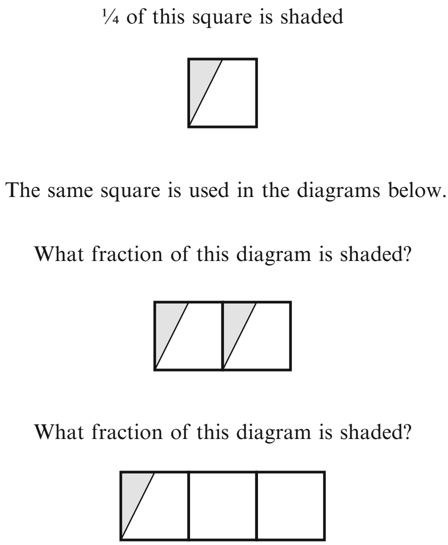


Fig. 9.1 A test item that assesses more than recall or procedures (Adapted from England's national mathematics test for 11-year-olds)

Assessment as a Lever for Change

Is it possible that introducing new forms of external assessment that explicitly attend to the spaces of hypothesis/testing and conjecture/justify could act as a lever to encourage teaching for creativity? While shifting assessments in that direction is no doubt a good idea in and of itself, as Boesen's (2006) research shows the evidence that this will be a strong force for change to teaching is less clear.

Boesen (2006) was one of the team of national assessment developers in Sweden devising assessments with an emphasis on 'reasoning, modelling, generalising, communicating and the ability to critically examine things' (p. ii). In subsequent research he examined teachers' construction of assessments to see if the teachers had included items that might require reasoning. In line with other findings, the majority of teacher assessments focused on tasks that required only 'imitative reasoning'. Teachers' exclusion of tasks requiring higher-order thinking was not, as might be expected, due only to teachers' lack of awareness of the need for such tasks, but was also because of their deliberate intent not to include them. The teachers made this decision to exclude tasks requiring non-imitative, creative reasoning tasks on the grounds that they believed such tasks to be too difficult for most students to deal with. Assessment tasks were chosen to get as many students gaining pass grades as possible and, particularly for lower attaining students, the teachers thought this to be more easily achieved with items requiring recall rather than reasoning. This points to the importance of teachers' beliefs about the nature of learning and learners in mediating whether or not teaching provides space for creative endeavour. It also suggests that teachers' views on creativity fit with the popularly held perception of creative activity being something that only a small group of learners are capable of engaging with.

Boesen (ibid.) suggests that another possible reason for this mismatch between the intent of assessments and teachers' perceptions of the nature of the tests may be a consequence of teachers not being privy to the thinking involved in the test development. Thus, although tests may be designed with the intent of promoting reasoning and creative thinking, without professional development that helps teachers come to appreciate it, the reform intentions may not succeed.

Assessing Creativity

Treffinger and colleagues (2002) suggest four ways to assess creativity: behaviour or performance data, self-report data, rating scales and tests. I think there is a category error in including rating scales in this list, as this is a way of grading the results of assessment not a form of assessment itself. I shall argue that self-report data is best considered within making judgements and therefore look at tests and performance data.

Tests

As indicated in the discussion above, test items can indeed have an element of assessing creativity built into them, but rather like the argument that because dogs can be taught to walk on their hind legs, they should be taught do so, we have to ask whether, however, well-designed tests are the best means of assessing mathematical or scientific creativity. A particular issue from the literature on creativity suggesting a drawback to the use of tests is the time-bounded nature of traditional testing.

The 'eureka' moment of quick and extraordinary insight is a popular view of creativity, but creative outcomes are more likely to arise from deep, flexible knowledge in specific content areas and extended periods of work and reflection (Silver 1997). Reflection time leading to creative outputs may be intentional, but it does not always involve self-conscious activity. Craft addresses the issue of 'insight' in the creative act, which has an element of the nonconscious to it, defining insight as 'the ability to build sense making bridges between different experiences and stimuli, and to be able to reflect on these' (Craft 2000, p. 120). Nonconscious aspects of insight, Craft argues, must not be underestimated and given the emphasis on teaching and learning about how things 'should' be, there is a danger of unwittingly blocking the 'non-conscious creative insights of children, given their relative powerlessness in claims on time, space, knowledge and experience' (p. 121). With regard to most testing practices, learners are powerless over time and space as, generally, tests are mandated to be carried out in particular spaces at particular times.

The growing empirical evidence for the power of 'sleeping on it' to promote insight suggests that assessing the insightful aspects of creativity requires a rethinking of current teaching and 'testing' practices. For example, adults participated in what they thought was a test of memory: they were taught a rule for generating a numerical sequence and asked to return a day later and report on whether they could recall the rule. What the participants were not told was that there was a much simpler rule for generating the sequence than the rule that they had been taught. When asked the following day if they could remember the rule, a significant number of the participants spontaneously reported that there was a simpler rule—an insight that they had reached without being prompted to try and find (Stickgold and Ellenbogen 2008).

Performance Tasks

As Treffinger and colleagues (2002) note, there are two potential sources of performance data: from learning in 'everyday' settings (in other words, occasions when creativity might arise spontaneously) and in tasks specifically set up for their potential to promote creativity. As learners creating within the hypothesis/test or conjecture/justify spaces are most likely to encounter these within specific classroom activities, I restrict the discussion here to this second type of data. While there is no shortage of suggestions for the sort of tasks in mathematics and science that might encourage learners to enter these creative spaces, how the teacher sets these up is a key determinant of whether or not the outcomes are creative. An example from my own research illustrates this.

As part of a 5-year longitudinal study of learning numeracy (defined in this case as the number aspects of the mathematics curriculum) in English primary schools, a team from King's College London devised two forms of assessment. The first was a fairly 'traditional' assessment (although unusual in that the majority of questions were not presented on the student test papers—the teacher had to orally administer the assessment) and the second a performance assessment aimed at exploring learners' extended problem solving and reasoning. We supplied the teachers with details of how to administer each assessment. In the case of the orally administered assessment, the students' papers and subsequent interviews with the teachers both indicated that these assessments had been administered appropriately. But the data returned on the performance tasks was so varied as to be unusable for the research. Responses ranged from students' scripts that showed so little evidence of productive activity as to suggest that the learners had been set off to do the task with virtually no help in becoming engaged with it (despite advice to the teachers on how to do this). At the other extreme, scripts returned showed every child's response was almost identical, suggesting heavy direction from the teacher. These assessments thus revealed much about how the teachers had set the assessment up but little about learners' reasoning.

Besides the difficulties in setting up tasks, the research shows the need to help teachers develop the range of things that they look for when assessing performance tasks. In particular teachers need to set aside expectations of 'correctness' in order to consider creativity. As noted earlier, Newton's (2010) research shows that this is challenging, for pre-service teachers at least. While it may be that serving teachers are more able to bracket out their knowledge of correct answers, research still needs to be done to investigate whether this is the case. If creativity is to be considered in the light of what is creative from the perspective of the learner doing the creating, then practices of normative assessment will make assessing creative performance difficult. The question of what sort of criterion-based assessment might be helpful here is discussed in the next section.

Other commentators have argued that variation in teacher assessment is too broad and that other factors include teacher assessment being unreliable, both between teachers and by the same teacher over time; the impact of the order of assessments—the effect of judgements from other assessments, either at a different time or in the sequencing of items; the 'halo effect'—the impact of general views of a learner; a teacher's overall leaning to harshness or not; and the effect of non-relevant factors such as neat handwriting (from Sadler 2002).

Such limitations would seem to support the case against teacher assessments and the need for externally validated assessments. Such objections, however, seem largely to be grounded in the primary purpose of assessment as summative, for comparing individuals, and assessment done in order to communicate 'standards' to others and the consequent drive for grades. In assessing creativity, the emphasis needs to be more formative and supportive of helping learners to become more creative rather than on assigning a 'creativity grade'.

Judging Creativity

> Assessing creative and cultural development is more difficult than testing factual knowledge … We noted earlier that creative outcomes have to be both original and of value. But there are different types and degrees of originality. Moreover, judging value depends on a sense of clear and relevant criteria. Teachers are often unclear about the criterion to apply to children's creative work and lack confidence in their own judgment. (NACCCE 1999, p. 127)

Although it may not be helpful to 'grade' learners on creativity, making judgements is necessary if teachers are to help learners improve. In doing so the specification 'of clear and relevant criteria' is no simple matter. Two approaches do appear to hold promise: dynamic standards and holistic judgements. Newton (2000) argues that despite the apparent difficulty in assessing creativity, teachers can make holistic assessments of creativity in scientific explanations and this is easy and the assessments are reliable. This builds on Amabile's (1983) pioneering work in holistic judgements. In a review of subsequent research, Hennessey (1994) makes a strong case for such an approach, noting that the bulk of the research draws on experts in the field to make the judgements. As many primary school teachers would not consider themselves to be experts in mathematics or science, this raises the question of how effective holistic judgement might be in the elementary years of schooling. In fact, in research with pre-service teachers, Newton found little agreement in their holistic assessment of learners' work in science (Newton 2010). Further research into this with experienced teachers is needed.

Newton's research demonstrated better agreement when the student teachers assessed learners' outputs against three attributes of creativity: novelty, scientific accuracy and elegance. In mathematics, in judging the work of learners in the conjecture/justify space, arguments could be considered against criteria such as strength of justification or elegance. These could be further refined. For example, Mason et al. (1985) argue for judging whether a justification would convince yourself, convince a friend, or convince an enemy (that is a mathematician!), and Sullivan categorises the language of arguments into 'naïve' (empirical and based on checking specific examples), 'crucial experiment' (considering extreme cases) or 'conceptual' (based on analytic or deductive reasoning).

It may be that teachers coming together and drawing out criteria for assessing creativity on the basis of initially holistic judgements—developing dynamic standards—would be the most productive way forward. Arguing for dynamic over arbitrary (external) standards and, drawing on the work of Moss and Schutz (2001), Broadfoot (2002) advocates the process of generating standards as 'the essential dynamic of educational quality and innovation' (p. 158).

In other words, teachers working together on dynamic standards to assess creative activity in science and mathematics are likely to lead not only to developments in assessing creativity but also to innovation in pedagogy. Engaging teachers in the processes of defining assessments is more important than providing them with an assessment product against which to judge learner outcomes. From this perspective of dynamic standards, the 'difficulties' associated with assessing creative activity become transformed into resources to work with.

> Where the emphasis is on generating, discussing and using 'dynamic standards' in a formative way, assessment is a key tool for system improvement. Where, however, the emphasis is on the imposition of 'arbitrary standards', not only does this represent a misguided belief in the power of numbers and words to contain the wealth of human creativity, the coercion and exclusion, the 'teaching to the test' to which it so often leads represents a tragic loss of opportunity for genuine progress and real learning. (Broadfoot 2002, p. 158)

Conclusion

It seems clear if we are to value and promote creativity in mathematics and science classrooms, then shifts are needed in both classroom cultures and assessment practices.

A shift is needed in the classroom culture, not simply in teachers' practices. For creativity to be encouraged, mathematics and science lessons need to have an element of playfulness and be safe places where learners can take risks. A shift is needed in assessment practices to help teachers adopt 'insiders' perspectives'—both the learner perspective and the discipline perspective—and so enlarge their repertoire for making judgements of creativity. Research is needed into the nature and support of both these types of shifts.

References

Amabile, T. M. (1983). *The social psychology of creativity*. New York: Springer.

Australian Curriculum Assessment and Reporting Authority (ACARA). (2011a). *The Australian curriculum. Mathematics F to Y10*. Canberra: ACARA.

Australian Curriculum Assessment and Reporting Authority (ACARA). (2011b). *The Australian curriculum. Science F to Y10*. Canberra: ACARA.

Bell, A. W., Costello, J., & Kuchemann, D. (Eds.). (1983). *Research on learning and teaching. A review of research in mathematical education, Part A*. Windsor: NFER-Nelson.

Boaler, J. (1993). Encouraging the transfer of "school" mathematics to the "real world" through the integration of process and content, context and culture. *Educational Studies in Mathematics, 25*, 341–373.

Boesen, J. (2006). *Assessing mathematical creativity—Comparing national and teacher-made tests, explaining differences and examining impact*. Unpublished Ph.D. thesis, Department of Mathematics and Mathematical Statistics, Umeå university, Umeå.

Broadfoot, P. (2002). Editorial. Dynamic versus arbitrary standards: Recognising the human factor in assessment. *Assessment in Education, 9*(2), 157–159.

Craft, A. (2000). *Creativity across the primary curriculum: Framing and developing practice*. London/New York: Routledge.

Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention*. New York: Harper Collins.

Edwards, D., & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom*. London: Methuen.

Fosnot, C. T., & Dolk, M. (2001). *Young mathematicians at work: Constructing number sense, addition, and subtraction*. Portsmouth: Heinemann.

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer.

Hargreaves, A., Halasz, G. Z., & Pont, B. (2007). School leadership for systemic improvements in Finland: A case study report for the OECD activity 'improving school leadership'.

Hennessy, B. A. (1994). The consensual assessment technique: An examination of the relationship between ratings of product and process creativity. *Creativity Research Journal, 7*(2), 193–208.

Hiebert, J., Gallimore, R., Garnier, H., Givin, K. B., Hollingsworth, H., Jacobs, J., Chui, A. M., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeck, W., Manaster, C., Gonzales, P., & Stigler, J. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Jackson, P. W. (1986). *The practice of teaching*. New York: Teachers College.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1–48.

Mason, J., Burton, L., & Stacey, K. (1985). *Thinking mathematically*. Wokingham: Addison-Wesley.

Moss, P., & Schutz, A. (2001). Educational standards, assessment and the search for consensus. *American Education Research Journal, 38*(1), 37–70.

National Advisory Committee on Creative and Cultural Education (NACCCE). (1999). *All our futures: Creativity, culture and education*. London: Department for Education and Employment (DfEE).

Newton, D. P. (2010). Assessing the creativity of scientific explanations in elementary science: An insider-outsider view of intuitive assessment in the hypothesis space. *Research in Science & Technological Education, 28*(3), 187–201.

Newton, D. P., & Newton, L. D. (2009). Some student teachers' conceptions of creativity in school science. *Research in Science & Technological Education, 27*(1), 45–60.

Park, S., Lee, S. Y., et al. (2006). Changes in Korean science teachers' perceptions of creativity and science teaching after participating in an overseas professional development program. *Journal of Science Teacher Education, 17*(1), 37–64.

Pugh, K. J. (2004). Newton's laws beyond the classroom walls. *Science Education, 88*, 182–196.
Pugh, K. J., & Girod, M. (2007). Science, art and experience: Constructing a science pedagogy from Dewey's aesthetics. *Journal of Science Teacher Education, 18*, 9–27.
Robinson, K. (2001). *Out of our minds: Learning to be creative*. Chichester: Capstone.
Sadler, D. R. (2002). Learning dispositions: Can we really assess them? Assessment in Education: Principles. *Policy & Practice, 9*(1), 45–51.
Sawyer, R. K. (2006). *Explaining creativity: The science of human innovation*. Oxford: Oxford University.
Schwartz-Geschka, M. (1994). Creativity in Japanese society. *Creativity and Innovation Management, 3*(4), 229–232.
Scotland Curriculum Review Programme Board and Scotland Scottish Executive. (2004). *A curriculum for excellence. Progress and proposals: A paper from the curriculum review programme board*. Edinburgh: Scottish Executive.
Silver, E. A. (1997). Fostering creativity through instruction rich in mathematical problem solving and problem posing. *Zentralblatt fuer Didaktik der Mathematik, 29*(3), 75–80.
Stickgold, R., & Ellenbogen, J. M. (2008). Sleep on it: How snoozing makes you smarter. *Scientific American Mind*. http://www.scientificamerican.com/article.cfm?id=how-snoozing-makes-you-smarter. Accessed 25 Oct 2012.
Tan, A. G. (2000). A review on the study of creativity in Singapore. *The Journal of Creative Behavior, 34*(4), 259–284.
Treffinger, D. J. (1996). *Creativity, creative thinking, and critical thinking: In search of definitions*. Saratosa: Center for Creative Learning.
Treffinger, D. J., Young, G. C., Selby, E. C., & Shepardson, C. (2002). *Assessing creativity: A guide for educators*. Storrs: The National Research Center on the Gifted and Talented.
Vygotsky, L. (1971). *The psychology of art*. Cambridge: Cambridge University.

Chapter 10
Science Teachers' Understanding of the College Entrance Examination in a Climate of National Curriculum Reform in China

Hongming Ma

Introduction

This chapter examines Chinese secondary school science teachers' understanding of the College Entrance Examination (CEE) in a climate of national curriculum reform. Currently, general education (from Year 1 to Year 12) in China is undergoing curriculum reform. The role of teachers in the implementation of curriculum has long been recognised (Olson 1980). From a pedagogical view of curriculum (Grundy 1987), individual teachers' beliefs and knowledge determine how teachers play their role in interpreting intended curriculum, designing lessons and making relevant decisions. What shapes teachers' beliefs and knowledge has been nicely captured by Fensham (2006):

> What is assessed in these systems determines what teachers and students recognize as knowledge of worth.
> Teachers in general are conscientious in doing their best to ensure that their students will learn this knowledge of worth well. (p. ix)

Given this understanding, assessment plays an influential role in shaping teachers' pedagogical practice.

The most important assessment in China's general education system is CEE. A student takes CEE after finishing Year 12. The student's CEE results affect whether the student can receive higher education and which university or tertiary college they can go to. CEE is a nationwide examination and is usually conducted on the same days across the nation. As a high-stakes test, it is often likened to a conductor's baton in that it directs and controls almost all decision-making in the educational system. The decisions include how teachers bring the intended curriculum to life—their pedagogical practice. Lingard et al. (2003) argue, 'in order to achieve improved outcomes for all students, it is necessary to align curriculum,

H. Ma (✉)
Monash University, Building 6, Wellington Road, Clayton, VIC 3800, Australia
e-mail: hongming.ma@monash.edu

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_10,

pedagogies and assessment' (p. 4). CEE is supposed to be in line with the curriculum standards so that it provides guidance for the development of pedagogies. Until now, there have been few studies that go into detail about what values promoted by CEE are recognised by teachers as important indicators for pedagogical practice.

This chapter analyses articles about CEE written by secondary school science teachers and particularly considers teaching strategies suggested by the teachers based on their understanding of what promotes by CEE. It also examines the new science curriculum standards in order to identify the alignment (or lack of alignment) between strategies suggested by the teachers and the values advocated by the science curriculum standards. After giving some background information to locate the issue in the cultural context of present-day China, this chapter first describes the method used in this study and then presents the findings of the document analysis. The discussion focuses on what the alignment or the lack of alignment among curriculum, pedagogies and assessment implies for the improvement of outcomes for all students.

Background

The Curriculum Reform

The Ministry of Education (MOE) issued its 'Synopsis of General Education Curriculum Reform' in 2001. This document identifies several disadvantages of conventional education in China, contending that it:

- Attaches too much importance to knowledge teaching rather than all-round education
- Lacks integration in curriculum design
- Attaches too much importance to book-learning
- Consists of contents that are too complicated and difficult, with some out-of-date and not commonly used content
- Attaches too much importance to passive learning, rote learning and rigid training in the implementation of the curriculum
- Involves assessment that attaches too much importance to the function of selection
- Is too centralised in the management of curriculum

In contrast to these perceived faults in conventional education, the synopsis states that the purpose of the curriculum reform is to change the status quo of general education in China from 'education for examination' to 'quality education', which attaches particular importance to the development of the overall quality of all students. This is seen to include intellectual, affective and ethical and physical development. The synopsis comprises reform objectives, new curriculum structure, curriculum standards, teaching suggestions, development and management of teaching materials, curriculum assessment, curriculum management and teacher education. It has become the document underpinning general education reform.

The pilot of the new curriculum was carried out in 2001, starting with compulsory education (Years 1–9). In 2003, the MOE issued 'Curriculum Design for General Senior High School (Years 10–12) (Draft)' and 15 curriculum standards including the standards for physics, chemistry and biology.

The draft of the curriculum design states that senior high school education is based on compulsory education and it should be regarded as fundamental education for the general public, aiming at further enhancing the quality of all citizens (MOE 2003). Specifically, the draft puts forward five reform objectives (MOE 2003), arguing that the educational system should:

1. Select basic knowledge and skills that students need for lifelong learning, enhance the connection between curriculum contents and students' life experience and the development of science and technology in contemporary society and promote innovation and practice.
2. Adapt the curriculum structure to diverse social needs and the need for students' individual development. The structure should be basic, diverse and integrated, and with clear layers.
3. In the implementation of the curriculum, encourage active learning, autonomous learning, teamwork and the development of skills in analysing and solving problems.
4. Establish a formative assessment system, improve school-based internal assessment, develop comprehensive assessment methods which combine students' academic records and development portfolios and establish a monitoring mechanism for quality education.
5. Give reasonable autonomy to schools, empowering schools to creatively implement the national curriculum and develop a school curriculum suited to local conditions.

The new curriculum document promotes quality-oriented education, echoing the trend in many English-speaking countries in that it values education for not only knowledge but also process and attitudes. The pilot of the new senior high school curriculum started in 2004 in three provinces and one autonomous region. Inevitably, the curriculum reform has been accompanied by assessment reform.

The Assessment Reform

In 2002, the MOE issued its 'Notice on Promoting Evaluation and Examination System Reform of Primary and Secondary Schools'. The notice states that the purpose of the assessment system reform is to improve students' overall quality and teachers' teaching standard and to ensure quality education (MOE 2002). According to the notice, the assessment system should pay equal attention to students' academic achievements and their values education, the assessment should promote the development of students' innovation and practical ability, the assessment should have requirements for all students while considering students' individual differences, and the assessment system should have diverse methods—in addition to standardised examinations, the development of scientific, easy-to-use assessment methods is encouraged.

Particularly, the notice sets benchmarks for CEE reform: the CEE should contribute to both the selection of students for higher education and the implementation of quality education at the secondary level; the assessment should ensure all-round evaluation including the ethical, intellectual and physical development of students; and the assessment should embody the principles of fair competition and fair selection. Further, universities and tertiary colleges are encouraged to explore selection methods which do not consist only of students' academic performance. Senior high schools are encouraged to establish a comprehensive assessment system to provide more information about students' development for the purpose of selection. The school-based evaluation records include the performance of students in inquiry learning and their involvement in community activities. It is expected that the school-based evaluation records will gradually become one of the most important references for the selection for higher education.

The CEE reform is a part of the national curriculum reform. Districts that have implemented new curricula are given the authority to administrate the CEE independently, including developing their own examination papers. In 2007, the first four districts that had implemented the new curriculum carried out the new CEE. It is expected that by 2013, all districts throughout the nation will carry out CEE based on the new curriculum. It is worth noting that although the notice sets high expectations for CEE reform, the main assessment method of the new CEE is still an external paper-and-pencil test with a focus on students' academic performance only.

The Method

The study reported in this chapter included three steps of analysis, all of which were conducted by the author alone. The first step was analysis of science curriculum standards. The senior high school science curriculum standards in China consist of physics, chemistry and biology curriculum standards. Therefore, the curriculum documents analysed were the following: 'General Education Senior High School Physics Curriculum Standard (Trial Edition)' (MOE 2004a), 'General Education Senior High School Chemistry Curriculum Standard (Trial Edition)' (MOE 2004b) and 'General Education Senior High School Biology Curriculum Standard (Trial Edition)' (MOE 2004c). Except for subject-specific contents, the three curriculum standards have many features in common. As a result, the analysis focused on identifying the common themes shared by all three documents, which included underpinning values, curriculum objectives, suggestions for teaching, and suggestions for assessment.

The second step was the analysis of secondary teachers' articles, in which they analysed the CEE examination papers taken in recent years and gave teaching suggestions based on their understanding of the values promoted by these CEE papers. The articles were searched and collected from 'China academic journals full-text database'. As part of the 'China National Knowledge Infrastructure (CNKI)', this database covers more than 6,000 academic journals published in China and is currently the most comprehensive and largest source of China-based information resources in the world. The basic criteria for the selection of the articles were that they were:

1. Written by secondary school teachers. As acknowledged earlier in this chapter, the implementation of intended curriculum depends on individual teachers' beliefs and knowledge. Although CEE may influence teachers at all year levels, the immediate pressure for implementation has drawn secondary school teachers to the forefront of the effort. Therefore, secondary school teachers' views are most important for understanding how the new curriculum reform influences teaching and learning in school settings. It is worth noting that the articles written by teachers are second-hand data—not directly from interviews or questionnaires or classroom observation. However, these articles reflect teachers' understandings of the values promoted by CEE, and the strategies suggested by the teachers are most likely to be practised in classroom settings. What is more, as the readers of these articles are schoolteachers, ideas and suggestions in these articles may widely influence teaching practice among schoolteachers. Therefore, these articles constitute valuable data to study.
2. Related CEE examination papers based on the new science curricula. So far, because not all administrative districts have taken CEE based on the new curricula, some articles still discuss the CEE based on the old curriculum. As this chapter concerns the progress of the current curriculum reform, only the articles that discuss the CEE based on the new curriculum were selected.
3. Published from 2010 to 2011. CEE based on the new curriculum started in 2007. Because only four districts were involved in the examination reform and the examination was in the pilot phase, the influence of 2007s new CEE was limited. By 2010, about half of the districts had taken the new CEE and the styles of the new CEE tended to be stabilised. Because articles published in 2010 and 2011 usually discuss the new CEE taken in 2009 and 2010, they were considered more representative of established views.

By 'fuzzy' matching, the search targeted articles published from 2010 to 2011, which contained the Chinese characters 'Gaokao Wuli' (CEE physics), or 'Gaokao Huaxue' (CEE chemistry), or 'Gaokao Shengwu' (CEE biology) in their titles. The search results showed 736 articles in total, among which 304 were physics-related, 300 chemistry-related and 132 biology-related. Articles were then selected against the criteria above. Some articles were not written by schoolteachers but by academics from universities or tertiary colleges, and were thus excluded from the analysis. Articles discussing the CEE based on the old curriculum were also excluded.

The selection of articles was also combined with the analysis of articles. The analysis was conducted in search of three categories of information: values underpinning the new CEE identified by teachers, skills promoted by the new CEE identified by teachers and teaching strategies suggested by teachers based on their interpretation of the new CEE. A preliminary analysis was conducted with ten articles discussing

CEE physics examination papers and themes under each category were identified. Next, articles discussing CEE chemistry and CEE biology were analysed to identify shared themes. A further search and analysis was subsequently conducted to identify additional themes to add to the preliminary findings. When there seemed to be no new theme to add, the search stopped. As a result, 42 articles were analysed among which 18 were related to CEE physics, 12 to chemistry and 12 to biology (see Appendix 1 for the list of articles studied).

Based on the findings from the first two steps, the third step was a multiple alignment analysis of curriculum, assessment and pedagogy. First, the new science curriculum standards were analysed and some common themes identified, which included underpinning values, curriculum objectives, suggestions for teaching and suggestions for assessing. The alignment analysis examined the alignment between curriculum objectives, suggested assessment and suggested pedagogy within the intended curriculum standards. Finally, consideration was given to the alignment between curriculum objectives proposed by the science curriculum standards and teaching strategies suggested by the teachers as a result of their interpretation of CEE examination papers.

Findings

From the Analysis of Science Curriculum Standards

The first step focuses on identifying the common themes shared by the three science curriculum standards. Findings from the first step are presented in terms of the underpinning values, curriculum objectives, suggestions for teaching and suggestions for assessment.

1. Values underpinning the new standards
   Across the three science curriculum standards, the following themes were identified:

   - Valuing basic knowledge and skills
   - Valuing the process and methods of scientific inquiry
   - Valuing creativity and practice ability
   - Valuing development of interest in science and formation of scientific world view
   - Emphasis on the common basis for all students but, at the same time, considers of individual students' needs
   - Valuing the need for the development of society and also give consideration to the lifelong development of students
   - Valuing the humanistic connotation of science and address issues related to STS (science, technology and society)
   - Valuing autonomous learning
   - Valuing diverse teaching methods
   - Valuing multifunction, multi-agent and multimodal assessment

2. Curriculum objectives

   The curriculum objectives involve the development of students in terms of three perspectives: knowledge and skills, processes and methods, and attitudes and values. More specifically, the objectives are presented as follows (as the alignment analysis was conducted against the curriculum objectives, each objective item was assigned a code such as 1-1, 1-2, 2-1, 3-2, etc. for the purpose of reporting):

   - Knowledge and skills

     1-1: Understand basic disciplinary knowledge.
     1-2: Obtain basic experimental knowledge and skills.
     1-3: Integrate knowledge across disciplines.
     1-4: Apply knowledge and skills to solve problems in real-life contexts.

   - Processes and methods

     2-1: Develop ability of scientific inquiry.
     2-2: Develop ability of collecting and processing information.
     2-3: Develop ability of autonomous learning (reflective, self-regulated, solving problems independently).

   - Attitudes and values

     3-1: Appreciate the wonder and harmony of the natural world.
     3-2: Develop curiosity about science, desire for knowledge and interest in scientific exploration.
     3-3: Form scientific attitude: think rigorously, critically and creatively; search for truth from facts.
     3-4: Nurture a sense of social responsibility to serve the community with science.
     3-5: Understand STS (science, technology and society) issues.
     3-6: Develop a consciousness of sustainability.

Compared with the old curriculum, the new curriculum attaches particular importance to developing students' inquiry ability. When defining inquiry ability, priority is given to experimental skills which include the following:

- Identifying a problem (including describing the problem)
- Forming a hypothesis (including making predictions)
- Designing experiments (procedure, equipment, variation control)
- Conducting experiments (operating instruments safely, recording data)
- Analysing data (describing and interpreting findings, applying mathematics knowledge and skills)
- Evaluating (analysing the differences between hypothesis and findings, paying attention to unsolved problems, identifying new problems, learning from the experience, improving the design)
- Communicating and cooperating (writing experiment reports, valuing teamwork)

The inquiry ability is also emphasised in suggestions for teaching and assessing.

3. Suggestions for teaching
   When teachers design teaching contents and activities, the standards suggest that consideration should be given to the three perspectives: knowledge and skills, processes and methods, and attitudes and values. Teachers should create opportunities to promote autonomous learning, inquiry learning and collaborative learning. More specific strategies include exhortations for teachers to:

   - Make good use of experiments. Try to enhance students' experimental knowledge and skills, using both hands-on and minds-on strategies.
   - Encourage students to participate in discussions to enhance their ability to evaluate, communicate and express themselves.
   - Choose teaching materials that are closely related to real life.
   - Use multimedia in teaching.
   - Use history of science in teaching.
   - Conduct cross-disciplinary research activities.

4. Suggestions for assessment
   The basic principle of assessment in the standards is that assessment should be based on the curriculum standards and be conducted in terms of three perspectives: knowledge and skills; processes and methods, and attitudes and values. More specifically, assessment should:

   - Pay attention to the functions of assessment both in promoting the development of students and in improving standards of teaching.
   - Equally value formative and summative assessments—pay attention to assessment that reflects the learning process; do not use examination results as the sole basis of assessment.
   - Promote multimodal assessment including paper-and-pencil tests, portfolios and observation records of students' performance and attitudes during their learning process.
   - Promote multi-agent assessment—involving teachers, parents, students themselves and peers.

From the Analysis of Articles Written by Teachers

In the second step, articles written by teachers were analysed in terms of values underpinning the new CEE identified by teachers, skills promoted by the new CEE identified by teachers and teaching strategies suggested by teachers based on their interpretation of the new CEE examination papers.

1. Values underpinning the new CEE identified by teachers:

   - CEE is for selecting students for higher education (other functions of assessment were not mentioned by the teachers).
   - CEE is a standardised test which requires standardised answers.

- Turning knowledge into marks is essential.
- Inquiry abilities are assessed through questions about experiments.
- Values and attitudes are promoted by making connection to hot topics (new scientific achievements) and/or history of science, everyday life practice and industrial production (using these as background materials for solving problems).
- CEE focuses on basic knowledge (concepts, theories) and basic skills (especially mathematical skills).
- Textbooks, examination outlines and examination guidance material are the most important resources.

2. Skills promoted by the new CEE identified by teachers
When teachers discussed the CEE papers, they did not talk about the skills explicitly in terms of the three perspectives promoted by the new curriculum standards. As a result, the findings are presented under the following six categories: general skills, language skills, information skills, mathematics skills, experimental skills and examination skills.

General skills:

- Apply knowledge in a flexible way (e.g. across contexts).
- Reason in a logical way.
- Memorise.

Language skills:

- Correctly use language in general.
- Correctly use terminologies.
- Correctly describe experiment procedures.
- Correctly describe diagrams.

Information skills:

- Extract useful information (ignoring interference and irrelevant information) from given texts, tables, (mathematical) diagrams, flow charts, pictures, photos, etc.
- Analyse and synthesise information.
- Make connection between information obtained and basic disciplinary knowledge to solve set problems.

Mathematical skills:

- Solve problems using mathematics methods and models.
- Calculate.

Experimental skills:

- Design or improve an experiment (fair test and variables control, identify problem, put forward hypothesis, test hypothesis).
- Conduct an experiment.
- Observe.

- Record (tables, graphs).
- Be aware of safety issues.
- Read and draw equipment diagrams.
- Analyse findings and draw conclusions.
- Evaluate experiment designs and identify mistakes and weaknesses (theories, process, methods, text descriptions, equipment drawings, experiment flow charts, etc.).

Examination skills:

- Keep a stable state of mind.
- Use standardised terminology.
- Weigh marks gained against effort needed to decide whether to answer certain questions.
- Use tips to guess answers (especially in multiple-choice questions).

3. Teaching strategies suggested to other teachers based on their interpretation of the new CEE examination papers.
The strategies suggested by the teachers are summarised in three categories: examination question-centred, teacher-centred and student-centred strategies.
Examination question-centred:

- Compile a collection of typical exam questions.
- Analyse and summarise authentic CEE examination questions and answers over the years.
- Arrange time effectively.
- Offer mental training for examination.
- Figure out the intention of examiners.

Teacher-centred:

- Strengthen basic knowledge and skills (basic knowledge is seen as the basis for analysis and synthesis).
- Guide students to use basic skills to solve problems in new contexts.
- Strengthen language skills (experimental and disciplinary terminologies).
- Strengthen calculating and mathematics skills (modelling, diagrams).
- Encourage students to read widely about ‘hot topics’, history of science, new sciences, application of science in everyday life (for a variety of purposes, e.g. motivation, to memorise concepts, to become familiar with background information).
- Link knowledge points into network structure.
- Analyse students’ mistakes from their homework and examination papers, identify the mistakes that appear to result from specific teaching and correct the mistakes.
- Use experiments wisely in teaching, which includes designing experiments that are simple and easy to operate and that advance by steps, keeping the experiments interesting, analysing experiment designs and designing experiments on paper.

Student-centred:

- Encourage students to design and complete experiments independently.
- Encourage students to form and internalise knowledge networks.
- Encourage students to reflect on their own strategies for answering examination questions.
- Individualise teaching (instruction, selection of exercises and examination questions, volume of training)—teaching students according to their abilities and cognitive development.
- Guide students to diagnose their own mistakes.

From the Multiple Alignment Analysis

The multiple alignment analysis is based on the first two steps and has two layers. The first layer is 'within the new curriculum standards' and the second layer is 'between the curriculum standards and teachers' interpretation of CEE examination papers.

Alignment Within the New Curriculum Standards

As described above in the method section, each new science curriculum standard has sections for curriculum objectives and suggestions for assessing and teaching. Therefore the analysis was first conducted to examine how the suggested assessment and teaching strategies match the new curriculum objectives. Except for subject-specific content (which are beyond the concern of this chapter), the three curriculum standards (physics, chemistry and biology) share many common features. Consequently, the findings are reported here in an integrated way and the curriculum documents are referred to as 'the new science curriculum standards' or 'the standards' for short. Table 10.1 shows the summary of the alignment analysis of curriculum objectives, suggestions for teaching and suggestions for assessment within the intended curriculum standards.

It is worth noting that the suggestions for both assessing and teaching are guidelines rather than practical measures. It is difficult to match suggestions with individual curriculum objectives. Therefore, the analysis was more general than specific. To some extent, the suggested assessment and teaching strategies are in alignment with the objectives set out by the new science curriculum standards. Both objectives and suggestions emphasise the three perspectives: knowledge and skills, processes and methods, and attitudes and values.

In the new curriculum, for the first time, formative assessment is seen as equally important as summative assessment. Compared with summative assessment, formative assessment is more flexible in monitoring student development and makes it possible to cater for the second and third dimensions of curriculum

Table 10.1 Alignment analysis of curriculum objectives, suggestions for teaching and suggestions for assessment within the intended curriculum standards

Curriculum objectives		Suggestions for teaching	Suggestions for assessment
Knowledge and skills	1-1: Understand basic disciplinary knowledge	Make good use of experiments	Equally value formative and summative assessments
	1-2: Obtain basic experimental knowledge and skills	Conduct cross-disciplinary research activities	Promote multimodal assessment
	1-3: Integrate knowledge across disciplines	Choose teaching materials that are closely related to real life	Promote multi-agent assessment
	1-4: Apply knowledge and skills to solve problems in real-life contexts		
Process and methods	2-1: Develop ability of scientific inquiry	Encourage students to participate in discussions	
	2-2: Develop ability of collecting and processing information	Make good use of experiments	
	2-3: Develop ability of autonomous learning		
Attitudes and values	3-1: Appreciate the wonder and harmony of the natural world	Use history of science in teaching	
	3-2: Develop curiosity about science, desire for knowledge and, interest in scientific exploration	Choose teaching materials that are closely related to real life	
	3-3: Form scientific attitude: think rigorously, critically and creatively; search for truth from facts		
	3-4: Nurture a sense of social responsibility to serve the community with science		
	3-5: Understand STS (science, technology and society) issues		
	3-6: Develop a consciousness of sustainability		

objectives—processes and methods (coded 2-1 to 2-3) and attitudes and values (coded 3-1 to 3-6). The suggestions of using multimodal and multi-agent assessment techniques also offer a wider range of solutions to evaluating the process of student learning and the development of students' attitudes and values.

The suggestions for teaching reflect the curriculum developers' intentions in seeking to balance curriculum objectives across three dimensions. In addition to teaching knowledge and skills (1-1 to 1-4), the suggestions also attach particular importance to strategies that promote the process of student learning (2-1 to 2-3) and the development of students' attitudes and values (3-1 to 3-6). Learning through experimenting is seen as an ideal way of promoting scientific inquiry including understanding knowledge, developing skills and promoting autonomous and collaborative learning. The standards suggest that teachers choose materials related to real life. This, to some extent, can help to motivate students and raise awareness of integrating knowledge across disciplines (1-3), as well as helping students to apply knowledge and skills to real-life situations (1-4) and guiding students to pay attention to social issues (3-1 to 3-6). Conducting cross-disciplinary research should contribute not only to integrating knowledge but to enhancing disciplinary knowledge and inquiry skills as well. The purpose of involving the history of science is mainly to help foster positive attitudes and values (3-1 to 3-6).

Both the curriculum objectives and the suggestions are part of the intended curriculum, which reflects the intentions of the curriculum developers. Whether the suggestions are feasible in practice and how they are enacted depends on how they are made meaningful through the interaction of teachers and students. It is expected that the CEE reform should be consistent with the new curriculum standards. Compared with the curriculum standards themselves, CEE examination papers have more direct influence on teachers and students. The following analysis focuses on whether there is alignment between the curriculum objectives and teaching strategies suggested by teachers based on their interpretation of CEE examination papers.

Alignment Between the Curriculum Standards and Teachers' Interpretation of CEE Examination Papers

The analysis of teachers' articles shows that most teachers noticed that the new CEE examination papers focus on basic knowledge and skills (including mathematical skills) (1-1). They also noticed that the new CEE examination papers try to promote inquiry learning by attaching particular importance to questions related to experiments (1-2, 2-1). Many teachers compared the questions in new CEE examination papers with those in old CEE examination papers and found that instead of asking questions based on abstract models, there was a trend in the new papers towards providing students with various forms of information and requiring students to process and make use of valid information (2-2). Further, teachers found that the new exam questions are, for the purpose of promoting values and attitudes (3-1 to 3-6), often put into contexts connected to real-life issues or history of science. Table 10.2 is a summary of the alignment analysis between curriculum objectives proposed by the science curriculum standards and teaching strategies suggested by the teachers.

In line with these understandings, teachers made suggestions for teaching. Most teachers emphasised that teachers should guide students to enhance their basic knowledge and skills (including logical thinking, mathematical skills, language skills

Table 10.2 Alignment analysis between curriculum objectives proposed by the science curriculum standards and teaching strategies suggested by the teachers

Curriculum objectives		Teaching strategies suggested by the teachers
Knowledge and skills	1-1: Understand basic disciplinary knowledge	Strengthen basic knowledge and skills Guide students to use basic skill to solve problems in new contexts Strengthen calculating and mathematics skills Link knowledge points into network structure Identify students' mistakes Use experiments wisely in teaching Encourage students to form and internalise knowledge networks
	1-2: Obtain basic experimental knowledge and skills	Strengthen calculating and mathematics skills Use experiments wisely in teaching
	1-3: Integrate knowledge across disciplines	
	1-4: Apply knowledge and skills to solve problems in real-life contexts	
Process and methods	2-1: Develop ability of scientific inquiry	Strengthen language skills Strengthen calculating and mathematics skills Use experiments wisely in teaching
	2-2: Develop ability of collecting and processing information	Use experiments wisely in teaching
	2-3: Develop ability of autonomous learning	Encourage students to design and complete experiments independently Encourage students to reflect on their own strategies for answering examination questions Guide students to diagnose their own mistakes
Attitudes and values	3-1: Appreciate the wonder and harmony of the natural world	Encourage students to read widely about hot topics, history of science, new sciences and application of science in everyday life
	3-2: Develop curiosity about science, desire for knowledge and interest in scientific exploration	
	3-3: Form scientific attitude: think rigorously, critically and creatively; search for truth from facts	
	3-4: Nurture a sense of social responsibility to serve the community with science	
	3-5: Understand STS (science, technology and society) issues	
	3-6: Develop a consciousness of sustainability	
		Teaching strategies for examination

and calculating skills) (1-1). For various reasons, teachers suggested that experiments should be given more attention. Some thought that experiments would help the students to answer inquiry-oriented examination questions (1-2, 2-1); some suggested that it can help students to memorise basic knowledge (1-1). Also for various reasons (motivation (3-2), memorising science concepts (1-1) and getting familiar with the background information for possible examination questions), teachers suggested that students read widely about hot issues, history of science, new science and applications of science in everyday life.

Except for the above-mentioned aspects identified by teachers from CEE examination papers, which are in alignment with curriculum objectives to some extent, many aspects mentioned by teachers are not in alignment with curriculum objectives. Moreover, various aspects set as curriculum objectives were not valued by the teachers. Basically, those valued by teachers but not emphasised in the curriculum objectives are examination skills and examination-oriented teaching strategies, while those emphasised as part of curriculum objectives but not valued by the teachers are integrating knowledge (1-3), application of knowledge and skills in real-life contexts (1-4) and attitudes and values (3-1 to 3-6). It is worth noting that although teachers noticed that the new CEE seeks to promote attitudes and values education, the strategies they suggested are not values-education-oriented.

The implications of the alignment and the lack of alignment of the intended curriculum and the teaching practice suggested by the teachers are discussed in the next section.

Discussion

Findings from the alignment analysis show that teaching strategies suggested by the teachers are in line with some aspects of the curriculum objectives. However, in terms of the changes expected to be made by the implementation of the new national curriculum, the analysis result does not show a very positive trend.

The major positive influence the new CEE has on teaching practice is related to the importance of experiments. As identified in the first step analysis, the new science curriculum standards promote students' inquiry ability by emphasising experimental skills. There used to be an inclination among teachers: 'talking about experiments is better than conducting experiments'. This is because conducting experiments takes more time to organise and did not contribute much to the marks of conventional CEE. The new CEE places particular importance on hands-on abilities. Although the examination questions are still in a paper-and-pencil format, at least some questions can be answered only when the students have conducted the experiment personally. Consequently, the teachers suggested that the importance of having students actually conduct experiments should not be ignored in teaching.

Another purpose of making good use of experiments as expressed by the teachers is related to developing students' information skills (2-2). In conventional CEE, examination questions are usually based on abstract formulae and models. Recognising that students need to develop more comprehensive abilities to collect information, identify and analyse valid information and make decisions based on new information obtained, the new CEE tends to require students to process more complicated information including texts, pictures and diagrams. In response to this requirement, some teachers suggested using experiments to develop students' abilities to collect and process information.

It is worth noting that although inquiry ability is the aspect that the new CEE highlights, to what extent this ability can be evaluated by paper-and-pencil tests is in doubt. Teachers' ultimate purpose of organising experiments is for the students to better answer CEE examination questions, not for the sake of developing inquiry ability per se. Also, in their articles, teachers focused more on those experiments outlined in the textbooks or in the examination guidance. The experiments referred to are laboratory-based. Although some inquiry skills can be developed by designing and conducting laboratory-based experiments, holistic inquiry activities involve broader authentic contexts and process skills. As paper-and-pencil tests such as CEE cannot reflect these features of inquiry, teachers did not show much intention of organising activities beyond textbooks or outside of the laboratory.

Teaching strategies suggested by the teachers are most in line with curriculum objective 1-1 (understand basic disciplinary knowledge). However, this should not be seen as evidence of positive changes brought by the CEE reform. Because conventional CEE is content-knowledge oriented and textbook bound, traditional ways of teaching usually stick to textbooks and focus on helping students to build their own knowledge network so that students can apply knowledge in a flexible way. The slight difference that the new CEE has made may lie in the degree of difficulty of the knowledge being taught. Knowledge and skills involved in the new CEE are supposed to be more basic than those of conventional CEE. Likewise, the positive link between teaching strategies and curriculum objective 2-3 (develop ability in autonomous learning) is not necessarily a result of CEE reform. Since the strategies focus on dealing with examination questions, the old CEE can have the same influence on student learning.

On the other hand, some major changes promoted by the new curriculum standards cannot be assessed by CEE. These changes are reflected by curriculum objectives 1-3, 1-4, and 3-1 to 3-6. As the new CEE is still based on disciplinary knowledge and administered in a paper-and-pencil format, it is difficult to assess students' ability to apply knowledge across disciplines and in authentic contexts. Likewise, as a high-stakes test, CEE is unable to monitor the development of students' attitudes and values. Although some new CEE examination questions have real-life contexts with the purpose of promoting the development of affective objectives, these objectives are not actually assessed. Consequently, teachers' responses to this are more examination-oriented. They suggested that the students should be encouraged to read widely about hot topics, history of science, cutting-edge sciences and applications of science in everyday life. Other than the purpose of motivation, which is part of nurturing attitudes and values (3-2), the main purpose expressed by the teachers is to have students become familiar with background information so that they will feel comfortable when encountering these issues in the examination.

As the findings show, although there is some degree of alignment between the objectives of the new science curriculum standards and teachers' suggested pedagogical strategies, there are more aspects in which teachers' understanding is not in line with the values underpinning the new curriculum standards. The new curriculum standards and the CEE place China's general education reform in several dilemmas. Without acting responsibly in confronting these dilemmas, the reform may cause more problems than solutions.

The first dilemma and perhaps the most fundamental one is that of catering to elite education versus mass education. The new science curriculum standards represent a transformation from elite education to mass education in contemporary China. Mass education is different from elite education in many aspects. Conventional elite education relies on examinations to select students for higher education. This has led to a trend of 'education for examination'. Mass education aims at 'education for all' and quality education that emphasises students' intellectual, affective and physical development. Education for examination values students of high performance, while quality education for all is committed to help every individual student reach his or her full potential to achieve both academic success and personal fulfilment. The new science curriculum standards make an effort to seek a balance between the needs of society and individual differences. However, as Hu (2005) argues, there is conflict between collective-oriented values and the needs related to individual development. Assessment is an area that embodies the conflict.

The differences between elite education and mass education mean that they must have very different assessment mechanisms. Instead of stressing the function of selection, the new science curriculum standards suggest that the purposes of assessment are to promote the development of all students and to improve teaching standards. The CEE is expected to be in line with this purpose. However, the CEE adopts a unified form of assessment, promotes a uniform learning style rather than a diversity of styles and represents the expectations of society. In essence, the one-size-fits-all techniques of CEE ignore individual differences and reflect the requirements of society for individuals. In their articles, teachers clearly recognised that the function of the new CEE still focuses on selecting students according to the criteria expected by society. As a result, the teachers' suggestions for teaching are inevitably focused on 'teaching to the tests' and turning knowledge and skills into marks.

The second dilemma is the formative nature of learning and student development versus the summative high-stakes nature of CEE. The new science curriculum standards attach significant importance to the process of learning. One dimension of the three major curriculum objectives is 'processes and methods'. According to the requirements of this dimension, it is expected that teaching should help students to develop abilities in scientific inquiry, collecting and processing information and autonomous learning (curriculum objectives 2-1, 2-2 and 2-3). In line with these requirements, the standards suggest that teachers choose diverse formative assessment methods and encourage multi-agent assessment which involves the teacher, students and peers. The purposes of assessment, as stated in the standards, are to promote teaching and learning.

In contrast to the formative nature of learning, CEE is a summative test. Often referred to as 'high-stakes testing', the CEE is a chance in a lifetime for most of the students to determine their future direction. Although it is possible for the students who have failed one CEE to take it again, teachers, schools, parents and students themselves are under tremendous pressure. It is not a surprise that in their articles, many teachers suggested that students should learn how to adjust their mental state in preparing and taking the examination and almost no teacher mentioned strategies for evaluating learning processes as they are not testable.

It is evident that the policy makers have realised the problem and as an effort to effect change, one of the CEE reform measures is to introduce an 'evaluation of overall quality' as one of the selection criteria for higher education. It is expected that this measure will overcome the malady of the conventional CEE which implies that the 'once-in-a-lifetime' mark is everything. However, this measure has been questioned in practice (Chen and Fu 2010). For one thing, at the preliminary stage of the reform, although the 'evaluation of overall quality' is combined in the selection mechanism in many districts, there is no detailed operational regulation. This makes it difficult to implement this measure in practice; for another, many candidates are concerned that this measure will cause corruption and cheating with the absence of a well-developed monitoring mechanism.

The third dilemma is that of holistic quality education versus the manageability of reductionist tests such as the CEE. The new science curriculum standards promote a wide range of competencies from three dimensions: knowledge and skills, processes and methods, and attitudes and values. Although described from three dimensions, the qualities they embody are an organic whole, that is, the amalgam of cognitive abilities, process skills and the affective domain in the development of individuals. However, the CEE is a standardised paper-and-pencil test. To minimise ambiguity and ensure that the examination is manageable, CEE has to break the holistic competency down into analytical and measurable pieces. The intrinsic conflict between holistic quality and reductionist assessment becomes inevitable. As Huber and Moore (2002) observe, 'Tests typically emphasize the wrong content because all too often that which is easy to assess is not that which is important to learn, especially in the sciences' (p. 19).

Most skills identified by teachers from the CEE examination papers are cognitive skills, and the strategies they suggested are related to instruction on the content and skills being tested (e.g. logical thinking and mathematical skills). Although the curriculum standards place the three objectives in equal positions, the CEE is dominated by content knowledge. Even with content knowledge, it is still doubtful whether correctly answering examination questions reflects real understanding. As Yager (1991) argues, standardised tests often fail to detect students' misconceptions. Because CEE science examination questions are divided into physics, chemistry and biology, there is no real chance to evaluate students' ability to integrate different disciplinary knowledge (1-3). The delivery of CEE in a paper-and-pencil format also makes it difficult to evaluate skills required in fulfilling authentic tasks such as hands-on ability (2-1), interpersonal competencies (2-1) and problem-solving in real-life contexts (1-4). The 'attitudes and values' (3-1 to 3-6) is another dimension that has been marginalised in CEE. A study of 2009s CEE physics examination papers found that although objectives of 'attitudes and values' were touched on by some CEE examination papers, there was no item that could assess affective objectives (Cheng 2010). The absence of assessment of affective objectives was also noticed by teachers in their articles. Consequently, in teachers' suggestions for teaching, they did not pay much attention to nurturing students' affective abilities.

There is one set of skills that is not mentioned in the new science curriculum standards, though it is central in teachers' suggestions for teaching. These skills are examination skills. Although in the analysis of teachers' articles, the strategies suggested by teachers are presented under three categories—teacher-centred, student-centred and examination question-centred—they are all examination-oriented in nature. Even though some teachers pointed out that teaching should aim at promoting real understanding and nurturing inquiry abilities, their final aim is still to 'turn knowledge and skills into marks'.

The fourth dilemma is the imported underpinning theories versus the national context of China. There are still many debates about the origin of the fundamental rationale of the new curriculum reform. Theories that are mentioned in the debates include Marxist theory, Dewey's pragmatism, constructivism, multiple-intelligence theory and postmodern curriculum theories (Zhong 2005; Wu and Ning 2008). Most of the theories mentioned are rather new to Chinese educators, especially teachers. This raises the issue of the localisation of imported theoretical foundations.

Wang (2010) points out that the theoretical foundations of the new curriculum reform such as constructivism have their roots in Western culture and reflect Western individualist values. Further, he argues that individualism sits in contrast with Chinese traditional values. Individualism values individual rights and choice. Every individual learner has his or her own experience and personal needs. Teaching should acknowledge the need of individual learners and advocate autonomous learning and teamwork. In terms of the relationship between teachers and students, teachers are not seen as authorities of knowledge; rather, they are seen as facilitators and even the peers of students in the course of learning.

Chinese traditional culture, on the contrary, is collectivist in respect of its ethics and values. This cultural tradition emphasises that individual interests are subordinate to collective interests. In terms of education, uniform content and method are widely accepted. Students are expected to respect and obey teachers. Although some advantages of constructivism are acknowledged, in practice, the influence of traditional values is deeply ingrained. The current curriculum reform is implemented in a top-down manner. For the majority of teachers, they face not only the purely cognitive dimension of updating knowledge but a strong cultural conflict. The involvement of two distinct sets of cultural values will inevitably lead to teachers' cognitive and behavioural confusion. Given this understanding, the localising of imported theoretical foundations should not conceptualised solely in academia and with curriculum developers. This localising has to involve schoolteachers to help them overcome the difficulties of cultural adaption and to gather feedback on the effectiveness of the adapted theory.

These dilemmas reflect the conflict between curriculum developers and curriculum implementers, as well as the conflict between theory and practice. On the one hand, curriculum standards show good will to promote quality education for all students; on the other hand, CEE continues to exert pressure on teachers to devote all their efforts to preparing students for standardised high-stakes tests. For the teachers, although the curriculum standards set the objectives, it is the CEE that indicates the way to go. Without fundamental reform of the examination mechanism, the curriculum reform will only have a superficial effect. Without dealing with the aforementioned dilemmas effectively, the curriculum reform is at the risk of eventually losing its way.

Conclusion

The analysis of the curriculum documents and teachers' articles shows that there is a substantial gap between the changes proposed by the new curriculum standards and those deemed important by the teachers. This is mainly because the high-stakes assessment, the CEE, is inadequate for evaluating the synthesised objectives of the new curriculum. Four dilemmas are identified to describe the current situation of the education reform. As CEE plays a strong part in shaping implemented curriculum, it is concluded that the new curriculum reform will incur the risk of achieving only superficial effects if it fails to successfully address these dilemmas. More studies with Chinese teachers and students should help to identify the support teachers need and to develop more localised theory and practice.

This study analyses articles written by secondary school teachers about their interpretation of the new CEE science examination papers. These are second-hand data that reflect teachers' intentions regarding their teaching of science in classrooms. It is worth noting that although teachers' intentions can be seen as part of the implemented curriculum compared with the intentions of curriculum developers (the policy makers), the curriculum is not really implemented until lessons are delivered. Moreover, to understand how the new CEE influences science learning, the learned curriculum also needs to be explored. The reform effort would benefit from further research using more naturalistic approaches such as classroom observation, interviews with teachers and students and collection and analysis of students' work samples.

Appendix 1: List of Articles Studied (Written in Chinese)

Ai, J. Y. (2010). Suggestions on the preparation of physics CEE. *Xinkecheng (Zhongxueban) [New Curriculum (Secondary School Edition)], 1*, 21.

Bai, R. (2010). Mastering physics concepts and laws—Strategies for the preparation of physics CEE. *Jiaoyu Jiaoxue Luntan [Education Teaching Forum], 19*, 244.

Bao, G. J. (2010). Guide on how to answer information questions in physics CEE. *Kaoshi Zhoukan [Examination Weekly], 12*, 1.

Cai, S. (2010). Analysing difficult CEE questions and reflecting on the shortcomings of teaching—The analysis of 2010's CEE chemistry paper of Jiangsu province. *Xinkecheng Yanjiu (Jichu Jiaoyu) [New Curriculum Study (Basic Education)], 12*, 168–169.

Cao, J. T. (2010). Valuing two bases and integrated application, improving quality of chemistry teaching—Related issues on the marking of 2010's CEE chemistry paper. *Zhongxiaoxue Shiyan yu Shebei [Experiment and Equipment for Primary and Secondary schools], 5*, 10–11.

Chai, H. L. (2011). Strategies for the preparation of new physics CEE—Comments on 2010's CEE paper of Zhejiang province. *Hunan Zhongxue Wuli [Hunan Middle School Physics], 1*, 49–51.

Chen, J. (2011a). Preparation for the new chemistry CEE. *Xinkecheng Yanjiu [New Curriculum Study], 1*, 185–187.

Chen, J. Q. (2011b). Starting from small things—Comments on the new chemistry CEE. *Jiaoxue Yuekan (zhongxue Ban) [Teaching Monthly (Secondary School Edition)], 3*, 58–62.

Chen, L. (2011). Innovative teaching preparation of chemistry CEE. *Wenli Daohang [Navigation for Arts and Science], 1*, 35.

Chen, L. H., & Qiu, H. M. (2011). The analysis of 2010's physics CEE of Jiangsu province. *Zhongxue Jiaoxue Cankao [Reference for Middle School Education], 8*, 33.

Chen, P. X., & Jin, J. (2011). The examination of innovation illustrated by 2010's physics CEE. *Wuli Jiaoxue Tantao [Journal of physics Teaching], 2*, 47–51.

Cheng, Y. A. (2011). Physics teaching in the context of new curriculum and CEE reform. *Zhongguo Jiaoyu Jishu Zhuangbei [China Educational Technology & Equipment], 5*, 92.

Fang, S. Z. (2010). Strategies for answering questions about experiment inquiry in biology CEE. *Xinkecheng Xuexi [Journal of New Curriculum Learning], 11*, 102.

Feng, W. Q. (2011). Teaching strategies for thinking training in inquiry experiment—The analysis of question No. 32 in 2009's biology CEE of Jiangsu province. *Kaoshi Zhoukan [Examination Weekly], 14*, 6–7.

Gao, S. L. (2011). Effective preparation strategies for chemistry experiment in CEE. *Xinkecheng (Jiaoyu Xueshu) [New Curriculum (Educational Scholarship)], 1*, 54–55.

Huang, G. L. (2011). Construct effective question system in the preparation of physics CEE. *Jiaoxue Yuekan (Zhongxueban) [Teaching Monthly (Secondary School Edition)], 3*, 47–57.

Huang, J. L. (2010). Strategies for the preparation of CEE—Comments on 2009's science CEE biology part of Jiangxi province. *Jilin Jiaoyu [Jilin Education], 2*, 56.

Huang, L. K. (2010b). The change of thinking in answering questions in the preparation of biology CEE. *Shengming Shijie [Life World], 5*, 103–105.

Jia, Y. C. (2010). Reflections on examination paper analysis lesson in the context of new physics CEE. *Kaoshi Zhoukan [Examination Weekly], 4*, 4–7.

Jiang, W. (2010a). Value teaching experiment design for biology CEE. *Shengming Shijie [Life World], 5*, 106–107.

Jiang, X. M. (2010b). Strategies for the preparation of biology CEE—The analysis of 2010's biology CEE of Jiangsu province. *Zhongxue Shengwu Jiaoxue [Biology Teaching in Middle Schools], 9*, 49–51.

Li, C. Z. (2011). Strategies for the preparation of biology CEE. *Zhongxue Jiaoxue Cankao [Reference for Middle School Education], 3*, 124.

Li, J. H. (2010). Develop the ability of physics modelling—Comments on the latest three years of physics CEE of Guangdong province. *Zhujiang Jiaoyu Luntan [Zhujiang Education Forum], 2*, 4–6.

Lin, J. (2010). Effective strategies for the preparation of physics experiment in CEE. *Fujian Jichu Jiaoyu Yanjiu [Fujian Basic Education Study], 2*, 89–90.

Liu, L. J., & Lang, J. H. (2010). The analysis of 2010's chemistry CEE papers. *Gaokao Jinkan [Journal of College Entrance Examination], 9*, 58–59.

Liu, Y. D. (2010). Enhance the ability of answering questions according to standards—Implications of the marking of biology CEE papers. *Yanbian Jiaoyu Xueyuan Xuebao [Journal of Yanbian Institute of Education], 24*(1), 118–120.

Shen, S. Y., & Xia, H. H. (2011). Strategies for the preparation of chemistry experiment in CEE—On 2010's science CEE of Hunan province. *Hunan Jiaoyu [Hunan Education], 2*, 55–56.

Shen, Z. R. (2011). Strategies for the preparation of physics CEE—Comments on physics CEE of Jiangsu province. *Xinkecheng [New Curriculum], 2*, 109.

Shi, Y. H. (2011). The three years of physics CEE in the context of new curriculum. *Wuli Jiaoshi [Physics Teacher], 32*(3), 65–66.

Song, J. (2011). The analysis of the questions on organic chemistry in CEE. *Kejiao Xinbao (Jiaoyu Keyan) [New Journal of Science Education (Educational Research)]*, *1*, 157.

Tang, J. H. (2011). The examination of ability in 2010's physics CEE of Jiangsu province. *Wuli Jiaoshi [Physics Teacher], 32*(1), 65–66.

Tao, H. B. (2010). Suggestions on the preparation of 2010's physics CEE. *Zhongxue Wuli [Secondary school Physics], 28*(9), 12–15.

Tian, B. X. (2010). Regulate thinking for the biology CEE. *Hubei Zhaosheng Kaoshi [Hubei Examination], 4*(1), 54–56.

Wang, H. Q. (2010). Preparation for 2011 physics CEE—The analysis of 2010's physics CEE of Jiangsu province. *Xinkecheng (Zhongxueban) [New Curriculum (Secondary School Edition)], 11*, 44.

Wang, L. P. (2011). Strategies for enhancing the accuracy of answering questions—A mental training lesson for the preparation of biology CEE. *Zhongxue Jiaoxue Cankao [Reference for Middle School Education], 1*, 104–106.

Xu, Y. (2011). The analysis of new examination forms in the new chemistry CEE. *Zhishichuang (Jiaoshiban) [Knowledge World (Teacher Edition)], 2*, 74–75.

Yang, C. Y. (2010a). Strategies for the preparation of biology CEE. *Shanxi Shifan Daxue Xuebao (Ziran Kexue Ban) [Journal of Shanxi Normal University (Natural Science Edition)], 24*, 47–48.

Yang, X. L. (2010b). Standards and skills—Implications of marking CEE chemistry paper for teaching. *Jiaoyu Gexin [Chemistry Teaching and Learning], 12*, 72–73.

Yang, Z. S. (2011). Teaching strategies and analysis of the status quo of biology CEE. *Xinkecheng [New Curriculum], 4*, 181–182.

Yuan, Z. C. (2010). Valuing basis, practice, and ability—Comments on 2010's CEE chemistry paper of Xiangsu province. *Zhongxue Huaxue Jiaoxue Cankao [Teaching and Learning References for Middle School Chemistry], 9*, 49–50.

Zhang, Z. (2011). Return to the basic, value ability, pay attention to curriculum reform—Comments on electrical experiments in 2010's physics CEE. *Zhongxiaoxue Shiyan yu Zhuangbei [Experiment and Equipment for Primary and Secondary schools], 2*, 29–31.

Zhou, H. W., & Tao, J. J. (2010). Teaching senior high school physics based on new curriculum—Study on the latest two years of physics CEE of Zhejiang province. *Jiaoshi [Teacher], 6*, 102–103.

References

Chen, H. L., & Fu, X. Q. (2010). Problems and solutions of College Entrance Examination reform in the context of curriculum reform. *Teaching and Educating, 36*, 10–11.

Cheng, L. (2010). Affective objectives in 2009's College Entrance Examination physics papers. *Journal of Education, 1*, 80–82.

Fensham, P. (2006). Forward. In B. H. W. Yung (Ed.), *Assessment reform in science—Fairness and fear* (p. xi). Dordrecht: Springer.

Grundy, S. (1987). *Curriculum: Product or praxis?* London: Falmer.

Hu, D. R. (2005). *Cultural studies on curriculum reform*. Beijing: Jiaoyu Kexue Chubanshe.

Huber, R. A., & Moore, C. J. (2002). High stakes testing and science learning assessment. *Science Educator, 11*(1), 18–23.

Lingard, B., Hayes, D., Mills, M., & Christie, P. (2003). *Leading learning*. Maidenhead: Open University.

Ministry of Education. (2001). *Synopsis of general education curriculum reform*. Retrieved from http://www.gov.cn/gongbao/content/2002/content_61386.htm

Ministry of Education. (2002). *Notice on promoting evaluation and examination system reform of primary and secondary schools*. Retrieved from http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe_441/201001/xxgk_78509.html

Ministry of Education. (2003). *Curriculum design for general senior high school (Years 10 to 12) (Draft)*. Retrieved from http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_711/201001/78378.html
Ministry of Education. (2004a). *General education senior high school physics curriculum standard* (Trialth ed.). Beijing: Renmin Jiaoyu Chubanshe.
Ministry of Education. (2004b). *General education senior high school chemistry curriculum standard* (Trialth ed.). Beijing: Renmin Jiaoyu Chubanshe.
Ministry of Education. (2004c). *General education senior high school biology curriculum standard* (Trialth ed.). Beijing: Renmin Jiaoyu Chubanshe.
Olson, J. (1980). Teacher constructs and curriculum change. *Journal of Curriculum Studies, 12*(1), 1–11.
Wang, P. (2010). The exploration of the inherent cultural values in constructivist theory on cultural adaptation in the curriculum reform. *The Modern Education Journal, 2*, 29–32.
Wu, Y. J., & Ning, T. T. (2008). Comments on the theoretical basis of the new curriculum reform of general education in China. *Theory and Practice of Education, 12*, 25–27.
Yager, R. (1991). The constructivist learning model: Towards real reform in science education. *The Science Teacher, 58*(6), 52–56.
Zhong, Q. Q. (2005). Curriculum reform in China: Challenge and reform. *Comparative Education Review, 187*(12), 18–23.

Chapter 11
Pedagogy in Theory and in Practice: Formative and Summative Assessments in Classrooms and in Systems

Paul Black

Introduction

The last few years have seen the growth of interest in formative assessments by teachers. At the same time, summative assessments have continued to arouse debate, both about their use as instruments of accountability and about the variety of ways, within different national systems, that affect teaching and learning. These two main purposes are intertwined in the everyday practices of teachers and schools, a relationship which varies, from one system to another, over a spectrum ranging from the harmonious to the destructive. The aim of this chapter is to explore this relationship, and in doing so to argue for its importance, and to map out some of the ways to promote its positive potentialities.

The starting point, in section 'A Model of Assessment in the Context of Models of Pedagogy', is to explore the relationship between these two purposes by framing both within an overarching model of pedagogy. The next two sections then examine, in the light of this model, some of the principles and the practices in the two arenas of formative (section 'Formative Assessment: Principles, Practices and Quality') and summative (section 'High-Stakes Summative Assessments') assessments. For the account of formative assessment, the focus is on the conditions needed for teachers' formative work to realise to the full its potential to improve pupils' learning. For the discussion of summative systems, the starting point is the argument that high-stakes assessments can only be fully valid and therefore supportive of effective learning, if they give significant weight to teachers' assessment judgements. The discussion surveys the wide variety of national and state systems, with particular attention to the various ways in which they support and make use of teachers' own summative assessments.

P. Black (✉)
King's College London
e-mail: paul.black@kcl.ac.uk

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_11,

Section 'Developing Teachers' Assessment Practices' develops the theme further in two ways. It starts with an argument that the quality of the day-to-day work of teachers depends on the quality of their summative assessments and on their positive link with the formative aspects of their teaching. To explore the micro-level, that is, the types of innovation needed to improve the daily work in classrooms and schools, the section then gives an account of the strategy and the main findings of a project which aimed to explore such innovation. This leads to further discussion of how a state's assessment policies might support, or at least avoid undermining, such work at a macro, that is, system, level. The closing section 'Conclusions' summarises the ideas developed in this chapter in the light of its initial aim.

A Model of Assessment in the Context of Models of Pedagogy

While assessment is a central component of teaching and learning, it is arguable that the neglect of assessment in studies of instruction has led to its being regarded as a marginal element. It follows that to reinstate it, it is necessary to propose a model of instruction in which the specific roles of assessment are made clear. To explain this argument, this section will first discuss the minor role that assessment has played in theories of pedagogy. It will then discuss some of the effects of this neglect of the central role of assessment in teaching and learning. Finally, it will propose a model which will serve as a basis for the arguments developed in the rest of this chapter.

Pedagogy and/or Instruction

Studies of teaching and learning in schools use the terms *Pedagogy* and *Instruction*. Many do not give, or adhere to, precise definitions: in particular, pedagogy is used by most authors as an inclusive term to cover all aspects of teaching and learning. Writing in the late 1990s, Watkins and Mortimore (1999) pointed out that a focus on pedagogy had been missing hitherto from most of the literature in England; they summarised the ways in which it was then beginning to emerge, each approach reflecting a different emphasis among the many aspects of teaching and learning.

A notable example has been the work of Alexander (2008), who states that:

> Pedagogy is the act of teaching together with its attendant discourse of educational theories, values, evidence and justifications. It is what one needs to know, and the skills one needs to command, in order to make and justify the many different kinds of decision of which teaching is constituted. Curriculum is just one of its domains, albeit a central one. (p. 47)

In Alexander's treatment, the definition above includes considerations of the exercise of power on, and thereby through, curriculum and teaching. Examples would be the study by Paolo Freire (1992) and similar works which add such adjectives as critical, conflict, liberatory and gender, all of which highlight the political function of pedagogy.

Where such a broad definition is adopted, it follows that the term *instruction* must represent some realm which is one component of *pedagogy*. Shulman (1999) adopted a similar scheme, as did Hallam and Ireson (1999) in describing pedagogy as 'those factors affecting the processes of teaching and learning and the inter-relationships between them' (p. 78). By contrast, Bruner (1966) inverted the terminology, arguing that a theory of *instruction* should serve as a guide to *pedagogy*, the latter being a collection of maxims. In what follows, the term *instruction* will be used to represent a subset of the broader realm of *pedagogy*.

What is notable in most of this literature is that assessment receives scant attention. Alexander (2008), for example, lists the core acts of teaching as task, activity, interaction and assessment (p. 49), but in the subsequent discussion, the last of these is given scant attention. The focus here is to consider assessment in the context of classroom teaching and learning. It does not follow that the literature on pedagogy and instruction is irrelevant to this task.

Neglect and Confusion

It is hard to imagine that classroom teachers would think that assessment plays only a marginal role in instruction, particularly those working in systems in which accountability is dominant. There are several ways in which such dominance has negative effects. These include constraining the quality of learning and teaching, lowering the professional status of teachers, undermining the quality of in-school assessments and the confounding of the formative and summative functions of assessment. These features interact in several ways and it is important to try to identify the underlying determinants which lead to these effects. I suggest that the most important of these are (a) confusion about the relationship between the formative and the summative, (b) misunderstanding of the criteria for the quality of any assessment and (c) mistrust of teachers—justified in part by the profession's poor grasp of assessment principles. I shall discuss issue (a) in this section; the other two will be considered in later sections.

In relation to point (a), it is helpful to emphasise that to speak of formative and summative assessments as separate entities may be one source of confusion. There are only assessments—instruments and practices for evoking information about the knowledge, understanding and attitudes of learners. The information so produced can be interpreted and used for formative purposes, or for summative purposes, or for both. An instrument can be so designed and used that it is more useful for one purpose than for the other—it is this feature that can lead to confusion. In many state systems, the formative feedback to learners to tell them why they failed the state's summative high-stakes test would be seen as extraordinary, but as a good employer the same state might be telling employees why they had failed in (say) an appraisal for promotion.

Assessment with a Comprehensive Model of Instruction

The model which is proposed here (from Black et al. 2011b) describes the design and implementation of instruction as a sequence of five main steps as follows:

A. *Clear aims* The task here is to consider different priorities and to decide how to achieve balance between them. For example, priority may be given *either* to promoting the understanding of the concepts and methods of a particular school subject *or* to developing students' skills in reasoning, listening and collaborating with others, and overall in becoming independent and confident learners. While neither of these two is disposable, they are in competition for priority and the teacher has to strike a balance between them.
B. *Planning activities* This involves work at several levels, from the overall plan of a school's subject department for several years' work down to one teacher's preparation for the next lesson. In composing the activities for a particular lesson, one important criterion is the potential of any activity to elicit responses which clarify or challenge and then extend students' understanding. Also relevant are the level of cognitive demand in a task, the relation of the task to previous learning experiences, and its potential to promote interest and engagement.
C. *Implementation* The implementation of a plan is crucial. If all students are engaged, then the teacher can elicit responses and work with these to help advance students' learning. This may be described as 'interactive regulation', which, as will be argued below, is a necessary condition for promoting successful learning. Such implementation can be hard to achieve because a class may respond in unexpected ways.
D. *Review* At the end of any learning episode, there should be review to check understanding before moving on. It is here that tests, with formative use of their outcomes, can play a useful part. The results may be used to revisit some issues with the class as a whole, so that this step may lead to revisiting Step C, or even Step B. Yet it also looks forward to Step E.[1]
E. *Summing up* This is a more formal version of the *review*. Here, the results may be used to make decisions about a pupil's future work or career; to report progress to other teachers, school managements and parents; and to report the overall achievements more widely to satisfy the need for accountability. This may be done by assessments by teachers, by external testing or by a combination of the two. The synergy and the overlap between this step and the four previous steps is the central feature in the design of assessment strategies, at classroom, school and state levels.

This scheme will be used as a framework of reference for the discussions in the subsequent sections. While it is presented here as a linear sequence, it will become clear as this account is developed that interactions, in both directions, between these steps must be a feature of any analysis of the practical business of instruction.

[1]Effectiveness of the learning is assessed in both Steps C and D; in this respect, the model differs from Steps C and D in the model of effectiveness described by Abrahams and Millar (2008).

It should also be noted that it is a model which serves to locate assessment within instruction: thus Step A is about the aims of the instruction and B about the teacher's skills in translating planning to help students achieve these. Assessment as such arises only in Steps C, D and E. Step E raises larger issues because state politics are driven by international league tables and demands for accountability, so that whether national and state systems support or undermine the achievements hoped for in Step A becomes questionable (see Ma, this volume).

Formative Assessment: Principles, Practices and Quality

As its title indicates, this section discusses this issue in three parts, assuming from the outset that the quality of a pupil's learning can be enhanced by improvements in the formative assessment aspect of the teacher's work. The first part examines how the aims of formative assessment can be understood in the light of principles of learning. The second then gives a short summary of the various practices in classroom learning through which these aims are realised, while the third considers the overall criteria for quality of formative work. This will serve to link to the model of section 'A Model of Assessment in the Context of Models of Pedagogy' and to the exploration in sections 'High-Stakes Summative Assessments and Developing Teachers' Assessment Practices' of the quality of summative assessments.

Formative Assessment and Principles of Learning

The definition of formative assessment which I shall use here is as follows:

> An assessment activity can help learning if it provides information to be used as feedback, by teachers, and by their students, in assessing themselves and each other, to modify the teaching and learning activities in which they are engaged.

This is the definition proposed in the Black and Wiliam's (1998) review of evidence. In a commentary on that review, Perrenoud (1998) said:

> This [feedback] no longer seems to me, however, to be central to the issue. It would seem more important to concentrate on the theoretical models of learning and its regulation and their implementation. These constitute the real systems of thought and action, in which feedback is only one element. (p. 86)

The model proposed in the previous section is a response to that challenge. To develop this, it is necessary to see interactive feedback as a way of describing the art of promoting dialogue, the central importance of which has been emphasised by Alexander (2006) as follows:

> Children, we now know, need to talk, and to experience a rich diet of spoken language, in order to think and to learn. Reading, writing and number may be acknowledged curriculum 'basics', but talk is arguably the true foundation of learning. (p. 9)

Two issues follow. The first is the need to ensure that such dialogue is encouraged. As explained above, Step B is precondition—it is in the planning of activities that the conditions favourable to dialogue are created. As Perrenoud (1998) expressed it:

> I would like to suggest several ways forward, based on distinguishing two levels of the management of situations which favour the interactive regulation of learning processes:
>
> - the first relates to the setting up of such situations through much larger mechanisms and classroom management;
> - the second relates to interactive regulation which takes place through didactic situations. (p. 92)

Step B is about his first level; his second level becomes important in Step C. Alexander's emphasis on the importance of dialogue is reflected in this focus on interactive dialogue, an emphasis which is lent additional weight by Wood (1998) in the following quotation:

> Vygotsky, as we have already seen, argues that such external and social activities are gradually internalized by the child as he comes to regulate his own internal activity. Such encounters are the source of experiences which eventually create the 'inner dialogues' that form the process of mental self-regulation. Viewed in this way, learning is taking place on at least two levels: the child is learning about the task, developing 'local expertise'; and he is also learning how to structure his own learning and reasoning. (p. 98)

Perrenoud (1998) adds a further perspective pointing out two related difficulties. One is to interpret any contribution from a student, the accuracy of which is a precondition for meeting the second, which is to frame a response in the light of such interpretation. He highlights this problem in an intriguing analogy:

> In other words, part of the feedback given to pupils in class is like so many bottles thrown out to sea. No one can be sure that the message they contain will one day find a receiver. (p. 87)

The Practice of Formative Assessment

In their accounts of the outcomes of work with six schools and about 40 teachers of science, mathematics and English (Black et al. 2003; Wiliam et al. 2004), the King's team reported in detail the developments of formative practices and their evaluation. Four main types of activity were developed: oral feedback, written feedback—mainly in relation to homework, peer and self-assessment and the formative use of tests. Each of these will now be considered in turn.

Teachers found the first, which required the development of interactive *oral dialogue* in their classrooms, the most difficult of the four. The potential of the question or task was a first requirement. A good example at primary level is as follows:

> Which is the "odd-one-out"—bird, cat, fish, elephant? Why? (Harrison and Howard 2009, p. 10)

However, this formative potential can only be realised by involving the pupils, as one science teacher explained:

> When a question is asked or a problem posed who is thinking of the answer? Is anybody thinking about the problem apart from the teacher? How many pupils are actively engaged in thinking about the problem? Is it just a few well motivated pupils or worse is it just the one the teacher picks out to answer the question? The pupil whose initial reaction is like that of a startled rabbit 'Who me sir?'
>
> Tom, Riverside School. (Black and Harrison 2001a, p. 56)

A frequent comment from teachers was 'It's pretty scary', reflecting the fear of losing control as they encouraged more involvement of their students.

Alexander (2006) describes the evidence of the predominance of limited forms of dialogue, such as 'recitation' (see also Applebee et al. 2003). The task of facilitating dialogue is a delicate one in that, given effective preparation in Step B, the teacher must, in Step C, carefully 'steer' the discussion, avoiding that close control which will inhibit participation, trying to respect, draw relevance from and make use of any pupil contribution even when it might seem unintelligent or bizarre, yet keeping the dialogue 'on track' in relation to the broader aim of the lesson. Study of the fine-grained detail of such dialogue can be revealing, particularly if results can be shared with the teacher(s) involved (see Nilsson and Loughran, this volume). A detailed account of this aspect is given in Black and Wiliam (2009). In particular, that paper emphasises the contingent nature of the implementation task of Step C and draws on the literature on self-regulated learning to explore Wood's point, as quoted above, that the pupil will be 'learning how to structure his own learning and reasoning'.

A different aspect of difficulty here is that in the close interaction needed to engage pupils in learning dialogue, the teachers' planning (Step B) must aim to draw on the diversity of cultural resources and expectations that pupils will bring, or can be encouraged to bring, to the classroom (see Cowie, this volume). This aspect is particularly important where cultural norms are such that pupils will not participate interactively in oral dialogue (see Lee Hang and Bell, this volume). Confucian cultures are a notable example (Song, this volume): in these cultures, dialogic interaction will have to depend on its development in writing.

The *feedback* which teachers give *on written work* may also be seen as a form of dialogue which can promote formative interaction and self-regulation, albeit within a different mode and a longer time scale. Dialogue in writing can become particularly productive when teachers compose feedback comments individually tailored to suggest to each student how his/her work could be improved, and expect the student to do further work in response to the feedback, as in the following example from Black and Harrison (2004):

> You are mixing up the terms power, energy and force. First check your glossary for explanations of these terms, then read pages 27–30 to see how your textbook uses power and energy correctly. Finally see if you can rewrite this using the terms correctly. (p. 13)

An additional lesson emerges from the findings of Kluger and De Nisi (1996), whose analysis of 131 research studies of feedback showed that the average effect size of feedback intervention was +0.4 but also that these effect sizes ranged from −0.6 to +1.4. Negative effects arise where learners are given the result but no help, while the most effective feedback gives specific comments on errors plus suggestions for strategies to improve. However, an extra feature was their conclusion that learners must believe that success is due to internal factors that they can change, and not due to factors such as ability or being liked by the teacher. This reinforces the findings of Butler (1988) and Dweck (2000) that the choice between feedback given as marks, and feedback given only as comments, can make a profound difference to the way in which pupils view themselves as learners: confidence and independence in learning is best developed by the second choice, that is, by feedback which gives advice for improvement and avoids judgement.

An alternative to having the teacher compose feedback on written work is for students to evaluate one another's work by *peer assessment*. This has several advantages in that all students can be involved, all can start to talk the language of the subject, and each can see his/her own work through the eyes of peers, which can help develop objective self-assessment (Black and Harrison 2001b). However, such advantages do not follow automatically. The review of studies of group work by Johnson et al. (2000) showed that group work only secures significant learning advantages if the groups are genuinely collaborative and that similar advantages are seen when collaborative group work is compared with individual study. However, such advantages do not accrue if the group interactions are competitive. Similar findings were reported in the study by Blatchford et al. (2006), while Mercer et al. (2004) reported the positive outcomes of an intervention study designed to train pupil groups in effective collaboration. A particular feature of Mercer's work was that one pupil in each group had to ensure that every suggestion, assertion or contradiction was justified by arguments that included reasons; in consequence, after the training, such reasoning words as 'think', 'because', 'would' and 'should' were used three times more frequently than before.

The fourth area of practice is the use of—so-called—*summative tests for formative purposes*. This is the area described as Step D in section 'Formative Assessment: Principles, Practices and Quality', where it was pointed out that a test at the end of any learning episode could be designed not only to summarise, perhaps in the form of individual student marks, before moving on, but also to serve as an opportunity for a review in which test results could be interpreted in the light of the strengths and weaknesses of the learning achieved. The Black et al. study (2003) showed how teachers had used this opportunity. In their work preparing for such a test, students could be helped to engage in their own reflective review of the work they had done in order to enable them to plan their revision effectively. It was also found useful for students to attempt in groups to set questions of the type that they judged would be fair tests of their learning, and to mark answers in groups as a form of peer assessment. The main point of such work was to encourage them, through their peer and self-assessments, to apply criteria to help them understand how their work might be improved and to help deepen their understanding of the criteria by relating them to their own specific examples. In general, the formative uses involved using the opportunity of review to consolidate students' overviews of their learning and where necessary to go back over topics for which the test had revealed there were unresolved difficulties. Thus they could see many tests as a useful part of their learning, and not as a terminal judgement of their achievement. It followed, of course, that to make such formative uses possible, a test should take place a short time before the end of the time planned for a topic in order to allow for the results to be best used.

The Quality of Formative Assessment

The work described in the previous paragraph was clearly part of Step D, although, if particular difficulties were exposed, it might lead back to further work in Step C and even, if the need for a fresh approach were indicated, to Step B. Likewise, the work in Step D, because it serves both formative and summative purposes, anticipates and could overlap with the work in Step E. Looking ahead to discussion of Step E in section 'Developing Teachers' Assessment Practices', I consider here the relevance of the criteria of *reliability*, *validity* and *comparability* to the practices of formative assessment.

The reliability of assessments used for formative purposes is a less problematic issue than it is for the summative use. Any misinterpretation of information can usually be detected very quickly by follow-up discussions. In the close interactions involved in oral and written dialogue, the focus is on the specific issue in question as it arises in a defined context—so there is no problem due to aggregation based on limited sampling over many topics and over many possible occasions or contexts. However, such limitations do begin to obtrude in end-of-topic tests, but here again are not serious given the interaction that might follow in the use of the results and the lower level of any 'high-stakes' pressures.

For validity, insofar as the formative interpretations bear upon identifying the optimum choice of the next steps to take in supporting the learning, predictive validity is important but errors therein can be identified and corrected within a limited context and short time scales. For construct validity, the question is whether or not the work in Step B, as developed in Steps C and D, expresses and develops the aims and values of the teacher, given that these aims and values are those that justify the priority given in the curriculum to the topics to be learnt. The consequential validity in formative feedback encompasses the effects both on the learners' immediate learning tasks and on their motivation and self-esteem. As discussed above, positive effects in both aspects can be secured. Work on all of these aspects is made easier in that the constraints of classroom and associated work are far less severe than those required for formal public tests.

It is not clear that the issue of the comparability of assessment outcomes is relevant for the formative uses. Given that a teacher's formative work with any class ought to be attuned to the particular needs of that group of students and will have to explored and used with reference to that teacher's plans, it seems neither possible nor indeed desirable that there should be comparability between the evidence, and the interpretations thereof, of different teachers in their different classrooms. This is not to argue that teachers cannot learn from collaboration with one another in auditing and enhancing the quality of the formative aspects of their work but that comparability between their practices and their judgements may not have to be given high priority. Indeed, flexibility in adapting to each context may be a criterion of quality practices.

As presented here, these criteria apply directly in the contexts of the development of classroom dialogue, dialogue with written work, and the fostering of collaboration and of peer and self-assessment. Their application may change in emphasis as attention moves to Steps D and E, that is, to schools' uses, both informal and formal, of their assessment results, and similarities and differences will emerge as this area of practice is discussed in the next two sections.

High-Stakes Summative Assessments

This section starts by arguing for the central importance of the contribution that teachers' own judgements should make to high-stakes summative assessments. It then supports and develops this argument by a discussion of the criteria against which the quality of all summative assessments should be judged. A third section uses an account of a variety of national systems to illustrate how they support the quality of their high-stakes summative assessments, particularly the contribution to these by teachers.

The Importance of Teachers' Summative Assessments

A recent report (Mansell et al. 2009) has argued strongly for the advantages of teachers' own assessments:

> [T]eachers can sample the range of a pupil's work more fully than can any assessment instruments devised by an agency external to the school. This enhances both reliability (because it provides more evidence than is available through externally devised assessment instruments) and validity (it provides a wider range of evidence). (p. 12)

A more fundamental argument is proposed by recent documentations and debates in the EU Council of Education Ministers (2010), who have identified the importance of developing 'key competences' and are exploring the problems of reflecting these in national assessment systems:

> Key competences are a complex construct to assess: they combine knowledge, skill and attitudes and are underpinned by creativity, problem solving, risk assessment and decision-taking. These dimensions are difficult to capture and yet it is crucial that they are all learned equally. Moreover, in order to respond effectively to the challenges of the modern world, people almost need to deploy key competences in combination. (p. 35)

This same argument was made more incisively by Stanley et al. (2009):

> [T]he teacher is increasingly being seen as the primary assessor in the most important aspects of assessment. The broadening of assessment is based on a view that there are aspects of learning that are important but cannot be adequately assessed by formal external tests. These aspects require human judgment to integrate the many elements of performance behaviours that are required in dealing with authentic assessment tasks. (p. 31)

Such general arguments apply to many different aspects of learning aims. The obvious example in science is the engagement of pupils in practical investigations. Another is the case made by Askew (this volume) that both the development and the assessment of pupils' creative reasoning must be entrusted to their teachers.

Reliability, Comparability and Validity

These issues were discussed, in the context of formative assessments, in section 'Formative Assessment: Principles, Practices and Quality'. In the literature about the reliability of teachers' summative assessments, some have claimed that high reliability can be achieved, quoting evidence from research studies to spell out the conditions required for such assessments to achieve their potential quality (ARG 2006; Harlen 2005). A more recent survey (Johnson 2011) of the literature is more cautious, emphasising the complexity of the task and concluding that the evidence about current procedures is inadequate. To take, for example, one issue, the process of moderation, the finding is as follows:

> There is no systematic form of moderator training in the GCE/GCSE world, as there is for written test markers, and little is known about inter-moderator consistency at the present time. It would be costly and logistically complex to organise inter-moderator reliability studies on a regular basis for every subject that has a coursework element, but some further research does surely need to be carried out. (p. 42)[2]

Comparison with the reliability of high-stakes public examinations is inevitable, with some policy makers convinced that teachers' assessments cannot be trusted. However, the reliability of formal tests is both misunderstood and often overestimated. In England, public concern focuses on the few cases of marker and administrative errors which are bound to arise regularly in large-scale systems, even where these are carefully controlled. This focus on a minor source goes with ignorance of the main source, which is the error that is bound to arise when any assessment is based on a limited sample of the totality of a candidate's achievements. In the UK there has been strong focus in the past 3 years on this source: the various analyses which have been conducted (Baird et al. 2011) seem to show that between 10 % and 20 % of candidates in public examinations will have the 'wrong' grade. There is disagreement on whether and how the public should be made aware of this; the term 'error' can mislead many users insofar as it carries implications of incompetence.

For large-scale high-stakes assessment, *comparability* is secured by ensuring that all candidates are assessed in the same way, such as by common test papers, and by procedures and checks which ensure comparability in the interpretations of their

[2]GCSE: General Certificate of Secondary Education—the assessment taken at age 16 by most secondary students. GCE: General Certificate of Education—the assessment taken at age 18 by pupils following specialised studies and aiming for university entrance. Both operate in England, Wales and Northern Ireland, but not in Scotland.

work. If the instruments and practices were more diverse, and if the assessment were in part in the hands of candidates' own teachers, then it would be harder to assure all users that comparability had been secured.

Validity is a more complex concept in which the other two are subsumed (Crooks et al. 1996). The interpretations in terms of predictive, consequential and concept validity are the significant features here. For the predictive, formal tests may serve well if it is success in similar summative tests in future that is at issue, but if a broader set of futures is relevant, they might be weak because they assess only a narrow range of achievements. For consequential validity, the issue relates to the inferences that will be drawn by the users: in school uses, predictive and consequential obviously overlap, for the summative can both affect teachers in orienting their future practices (e.g. teaching to the test) and learners in taking decisions (e.g. about future choices of subjects). Construct validity is central, for if a result indicates that a pupil is good at (say) English or mathematics, the question that arises is about the meaning of 'being good at' the subject assessed. This is usually the main focus of criticisms of the limitations of formal testing: the criticisms imply that at least some part of high-stakes assessments should reflect broader measures, which will require assessment by teachers.

International Comparisons

The relationship between teachers' formative and their summative assessment practices is inevitably dependent on the state systems in which their schools operate. The nature of such dependence varies widely between different countries and states. This section discusses only a few examples to illustrate this diversity (see Berry and Adamson 2011; Black and Wiliam 2005, 2007 for more comprehensive accounts).

The system in France represents one end of this spectrum. The national test, the *baccalauréat*, is set and marked externally: it is the sole source of school-leavers' qualifications and the main criterion for selection for further education. The test is composed entirely of written tasks, except in the case of foreign languages. Each examination lasts for 3–3.5 h. Science papers, for example, include structured open response questions, with most requiring extended rather than short responses, and (in biology) an essay question. Multiple-choice questions are not used and all questions have to be answered. The strategy is to test students at some depth in a limited number of topics.

The case of the USA is more complex. External tests, often commercially generated, are dominant at state levels, with attempts being made at federal level to enhance national testing by linking it to funding of the states. Various stakeholders also make use of grades awarded by schools and assembled in pupil records, while higher education institutions supplement this last source with results of tests produced by private agencies. Multiple-choice items are the only, or main, instruments. Attempts to broaden the scope of assessments by promoting the use of pupil portfolios assessed by teachers have been unsuccessful: in three states, the use of non-standardised tasks, weak guidelines for the inclusion of evidence in portfolios and inadequate training in marking were all identified as causes of difficulty (Koretz 1998). Similar findings were reported in a smaller-scale study in another state by Shapley and Bush (1999), while a broader review by Brookhart (2011) also concluded that in the USA, teacher judgement for summative assessment had often been 'found wanting'.

The state of Queensland in Australia is at the other end of the spectrum. It bases students' school leaving certificates wholly on school-based assessments. New South Wales places 50 % of the weight of final assessments in the hands of school, the other 50 % being based on the state's formal tests. However, the latter play a dual role, being used as calibrators to audit the school-based results[3]—leading in a few cases to scaling overall—and to detect or correct anomalies in the results of particular schools. There is pressure from the federal government's desire to impose a national test.

What is significant in these last two states is that there is a coherent state system of training teachers as assessors and as participants in interschool alignment procedures to ensure that the assessment instruments and the procedures used to interpret them are comparable across schools. The aim is seen to require interschool collaborations in local groups in order to secure interschool consistency in judging the final portfolios produced by individual students; the systems of moderation which have been developed serve to ensure this comparability in the final results, but also support this outcome through collaborative professional development at earlier stages in the school year (Fensham and Rennie, this volume). Another significant difference between these systems and that in (say) England is that the state has responsibility for all aspects of the assessments—the external test administration and the cross-school moderations—and also for the curriculum and for all aspects of teacher education, including the funding of professional development work. In consequence, the state assessment system is planned as one aspect of an overall system to provide training and support to teachers in their summative assessment work so that they contribute effectively to the overall framework. Such a framework would seem to provide a way to establish positive and coherent relationships between Steps D and E of the model in section 'A Model of Assessment in the Context of Models of Pedagogy'.

Stanley et al. (2009) describe how the state systems in Queensland and New South Wales provide this training and support. A detailed account of the Queensland system is given in Colbert et al. (2011). A comparison (Corrigan and Cooper, this volume) of the Queensland system with that in Victoria, where the external examination at the end of compulsory schooling dominates teachers' work in the years leading up to it, shows how, in the latter system, teachers feel that their own values in science education must be overridden by the need to help their pupils achieve their best possible test scores.

The examples of Scotland and Wales differ from the two Australian examples—Queensland and New South Wales—in that the system in both is in a process of change. For Scotland, the issue of interest concerns the use of testing to satisfy

[3] External tests are also used calibrate the overall results of a school's assessment in Sweden.

demands for school accountability at all levels of schooling. A study by Boyd and Hayward (2010) of this changing scene provides evidence of the problems that arise when national changes do not take account of the contrasting needs of, and limited perceptions of, the different purposes of assessment. For over 10 years, the implementation of formative assessment had been effectively developed by the Scottish government and is widely regarded as a very successful and positive initiative (Hutchinson and Hayward 2005). For their summative assessments, schools were provided with tests on a national database, the expectation being that they should use the findings from such tests to inform, not to replace, their own assessments. However, this had not happened: teachers and schools, relying on results of these national tests as evidence of their school's success, were training pupils to do well in them rather than using the test results to inform their own judgement of a pupil's competence at a particular level. One reason for this focus on national assessment information rather than on teachers' professional judgement was the need to provide information to the local education authorities, whose strong culture of performance management was a response to a national requirement to demonstrate continuous improvement in their schools. When the Minister for Education announced that national assessment data would no longer be collected nationally, it was hoped that this negative influence on classroom practice would disappear. It did not: local authorities continued to collect the data from schools and these schools continued to depend on the external test instruments.

At a later stage, the government introduced its new Curriculum for Excellence, the aims of which expanded the range of the curriculum's intended achievements. It followed that schools should have been exploring changes to achieve validity of their assessments in relation to these aims. However, the need for support for teachers had not been considered—yet they were now expected to develop the quality of their assessments and to earn local and national trust in their assessments. This meant that their summative work ought to have changed to focus on improving learning, which could lead to and benefit from the establishment of positive and fruitful links between formative and summative assessment practices. No coherent system to provide guidelines, training and support was set up to help secure these changes. The collection of purely quantitative data on limited areas of activity continued as before. Boyd and Hayward (2010) concluded:

> there is an urgent need to tackle issues of assessment literacy in Scotland. Until individuals and groups have a better understanding of the uses and abuses of assessment data, the tensions emerging in this study are likely to persist. (p. 21)

A study of similar transition problems (Gardner et al. 2011) in the changes in Wales from reliance on external testing to use of teachers' assessments to serve accountability tells a story with several similar features. It pointed to a lack of any infrastructure to support the development, implementation and ongoing operation of a new assessment system, and highlighted the following two weaknesses:

> The first related to the lack of comprehensive planning ('under-designing') of many of the initiatives, whilst the second related to perceptions of what constituted quality assessment practice. (p. 112)

Overall, the examples of this section show different ways in which the problem of the formative-summative relationship, linked to the balance between external and teachers' assessment, has been resolved. These include ignoring teachers' assessments, complete reliance on teachers' assessments or combinations of the two sources of evidence in a variety of ways. What emerges is that, if a solution to the problem is to be effective, a coherent system is needed, built with close attention to the problems of providing national support and enhancing teachers' professional development, backed by a clear programme for enhancing teachers' assessment literacy.

Developing Teachers' Assessment Practices

The Importance of Summative Assessments with Schools

The focus in this section is on the ways in which the links between formative and summative assessment are implemented at the level of the everyday practices of teachers and their schools. These are important in their own right because in between the rapid and informal feedback envisaged above as Step C, and the slow and (often) uninformative feedback of high-stakes external tests, lies the whole range of assessments which teachers and their schools conduct for themselves, and which have to serve both formative and summative purposes. Of course, the extent to which, and the ways in which, teachers have responsibility for the high-stakes assessments of their students in Step E varies widely between different state systems. However, discussion in this section will focus on those summative assessments over which schools have full control and which are produced mainly or wholly for internal purposes. These purposes comprise guiding individual pupils, reporting to parents, making decisions about teaching sets, providing information for the next teacher of the class and reporting to the school's senior management.

All of these can have significant effects on the careers of teachers and of their pupils. So while they might not be classified as 'high stakes', it is still important that these assessments be as valid and reliable as possible. The 'high-stakes' implications can also vary. With very young children, teachers may build up the summative assessment that good guidance requires by observation of their involvement over a period of time in such a way that the children themselves are not aware of the process (Fleer and Quinones, this volume). However, many such assessments are implemented in a way that makes them stressful to students, but, as argued below, this need not be the case. The question which then arises is whether these summative purposes are well served by the data and the practices which are used by teachers and schools for this type of assessment. Writing in the context of the UK, Gardner (2007) has argued that teaching about assessment skills and values is the weakest aspect of the professional development of teachers, one consequence of which is that many teachers lack test construction skills (Fensham, this volume).

Thus, there are strong arguments that the quality of teachers' summative assessments is of prime importance, not only, as argued in section 'High-Stakes Summative Assessments', because the validity of high-stakes tests, whether external or internal or both, is going to be more rigorously challenged in future but also because of their effects within schools. The account of the formative use of summative tests in section 'Formative Assessment: Principles, Practices and Quality' illustrated how, for summative work within the control of the school, positive links can be built between the dual-purpose use of some assessments, both in respect of informal reviews (Step D) and the more formal assessments such as those used at the end of a school year (Step E).

To take this further, teachers' summative assessments must be developed so that both their validity and their comparability within and across schools can be trusted. To illustrate what such development may involve, this section presents the following account of a recent exploratory study.

Exploring the Prospects for Positive Development

This study, the King's-Oxfordshire Summative Assessment Project (KOSAP: Black et al. 2010, 2011a) was designed both to explore the issues with secondary phase teachers of English and of mathematics and to attempt, through a small-scale collaborative intervention, to improve the quality of the summative judgements of these teachers. In what follows, I shall give a brief account of the findings.

A *first* finding emerged from an initial audit of the internal assessment practices in the three schools involved. Although these three had been chosen because of their good reputation, in particular in their development of formative assessment practices, it was clear that their summative work left much to be desired. While teachers were regularly assessing their students and using the results both for reporting and for decisions about setting,[4] they had not considered whether the means they were using served these purposes in a dependable way. The accountability pressures exerted by England's national assessment systems made it hard for them to do otherwise. This auditing phase of the work evolved into a validity-oriented intervention which explored teachers' development of their own critique of 'off-the-shelf' test questions which they used. As one of the teachers put it:

> The maths department of [School C] has in place a series of end of unit tests which, at the time, generated a series of numbers without any relation to areas of the national curriculum. The first thing that [colleague] and I did was to rewrite the tests in order that they generate useful information about how well students are doing in each area against NC levels.
>
> Mathematics teacher. (Black et al. 2010, p. 221)

This critical auditing served as a stimulus for teachers to produce for themselves better student tasks more attuned to their own teaching aims and values.

[4] Otherwise known as streaming or tracking.

The *second* finding arose because it was a condition of the project's work that it help the schools produce results for which it could be claimed were comparable both within and across the three schools. This aim called for a system which could produce evidence of work done by pupils in several tasks, including a portfolio, with some degree of uniformity so that cross-moderation of the products, both within and between schools, would be feasible. This meant that it would be necessary to attend to five main features, in each of which threats to both validity and reliability can arise. These were as follows:

(a) Validity of each component of a portfolio: each task or test had to be justified in relation to the aims that it was said to assess
(b) Agreed conditions for the presentation and guidance under which students would work in producing the various components of a portfolio
(c) Guidelines about the ways in which each domain was sampled within a portfolio, both by the separate components and by the collection as a whole
(d) Clear specification of the criteria to which all assessors have to work
(e) A requirement for comparability of results within and between schools, which would require moderation procedures at both intra- and interschool levels

It was expected that formal, timed tests might play a part but that to have them as the only component would limit the validity of any portfolio. I shall expand here on the work involved in features (a) and (e) above; the outcomes for the other three are not presented here.

As the audit triggered teachers' concerns about the quality of their own summative procedures, it led to debate about validity, notably starting from the question, 'What does it mean to be good at (English or maths)?' Through their engagement in this debate, the teachers began to see that they had hitherto neglected to critique their assessment practices in the light of their beliefs and values concerning the purpose of learning in their subject/s. The uncertain understanding of validity was more marked among the mathematics teachers than among the English teachers, probably reflecting the differences between the pedagogic cultures of the two (Hodgen and Marshall 2005). Teachers came to realise the importance of the validity issue. As one teacher put it:

> The project made me think more critically about what exactly I was assessing. The first question I remember being asked ('what does it mean to be good at English?') gave me a different perspective on assessment. I find myself continually returning to this question.
>
> English teacher. (Black et al. 2010, p. 222)

The initial audit also showed that to ensure comparability both within and between schools would be a formidable task and that it might create an unwelcome extra workload for teachers. However, this concern was not borne out by the final reflections of the project teachers. One positive feature was that the moderation processes were seen to have valuable effects on many aspects to their work:

> …that the moderation and standardisation process was incredibly valuable in ensuring rigour, consistency and confidence with our approach to assessment; that teachers in school were highly motivated by being involved in the process that would impact on the achievement of students in their classes (like the moderation and standardisation at GCSE).
>
> English teacher. (Black et al. 2010, p. 225)

Teachers in one school, having experienced the benefit of the end-of-year moderation meetings, planned to have such meetings three times a year; they wished to explore the value of meetings to 'moderate formatively', that is, to explore ongoing progress in developing shared judgements and criteria in advance of the occasions where decisions would have to be taken. Another positive feature was illustrated by the following statement from a teacher, reflecting on pupils' involvement in an investigation of the standard paper sizes:

> I think one of the reasons is that it's a good task is that it's a real task. It's all based on A4, A5 and A6 you know, which is in real life, kids know that. …. they can physically hold up an A5 sheet against an A4 … a lot of kids are engaged straight away … it's a piece of work that every kid could achieve from.
>
> Mathematics teacher. (Black et al. 2011a, p. 455)

These statements illustrate the general finding that pupils could be positively involved in tasks even when they knew that the outcomes would form part of their individual summative assessment. They also show that the results revealed to their teachers new possibilities for improving such engagement and motivation, mainly because the broader focus of their assessment work gave more opportunity for pupils to perform.

The Links Between Formative and Summative Assessments

The evidence and arguments of this section serve two of the aims of this chapter. One is to show that teachers are able to take responsibility for summative assessments or at least to play an active part in meeting this responsibility, in such a way that the practices used would still be supportive of learning. The other is to replace the assumption that a sharp discontinuity between the formative and the summative is the norm, by a realisation that there can and ought to be synergy between them. This last issue, already discussed in sections 'Formative Assessment: Principles, Practices and Quality and High-Stakes Summative Assessments' above, can now be taken further. In that discussion, it was emphasised that informal summative assessments could be used for formative purposes, perhaps in a last phase of the learning of a topic, designed to come after the 'terminal' summary assessment in order to enable use of that assessment to improve learning. Here, links between Steps C and D were the focus

Positive links between the formal summative assessments of Step E and the formative and informal summative functions of Step D could develop because teachers, being involved in the formulation of the tasks and the operational procedures for their use, had confidence in implementing them, so breaking down the gaps, in ownership and agency, between the formative and summative aspects of their work. Commenting on the use of school's summative results as guidance to parents, one teacher commented:

> But I think if all the teachers had more, possibly more ownership of what we are actually doing in terms of summative assessment then you would have more confidence in saying to parents, which I think is one of the biggest things I find with lower school.
>
> Mathematics teacher. (Black et al. 2011a, p. 460)

This was in sharp contrast to the initial opinions of many teachers, to the effect that parents would be satisfied if they merely reported to them results of formal tests which closely resembled the external national tests, preferably being composed by using versions of these from previous years.

There were also benefits in the engagement of pupils in the summative assessments. They began to see this summative assessment work as a shared enterprise. As one teacher put it:

> They feel that the pressure to succeed in tests is being replaced by the need to understand the work that has been covered and the test is just an assessment along the way of what needs more work and what seems to be fine.
>
> Mathematics teacher. (Black et al. 2003, pp. 56–7)

For evidence to take home, the pupils' 'working-at-grade' reports were replaced by their portfolios which they regarded as theirs and which they could describe to their parents. A comparable example, where pupils summed up their learning by preparation and presentation of posters, is reported in the paper by Fitzgerald and Gunstone (this volume).

These arguments tell against any assumption that there has to be a sharp discontinuity between the formative and the summative. In fact there can be and ought to be synergy between them, and such synergy should be seen as the natural healthy state.

Accountability Pressures: Must They Be Malign?

Where teachers experience the accountability pressures of externally imposed tests, achieving helpful links between formative and summative practices may well seem impossible: my assessment diagnosis for England is that it has a disease that has developed from a flawed concept of accountability, as Black and Wiliam (2007) explain:

> Where summative becomes the hand-maiden of accountability matters can get even worse, for accountability is always in danger of ignoring its effects on learning and thereby of undermining the very aim of improving schooling that it claims to serve. Accountability can only avoid shooting itself in the foot if, in the priorities of assessment design, it comes after learning. (p. 10)

This phenomenon can be demonstrated in two ways. It was recognised in the national curriculum in England that to meet the requirements for validity in the assessment of science learning, extended practical investigations should play some part in the assessment programme. Such work has to be assessed by teachers in the classroom. However, limited to external inspection of samples and by examiner visits, moderation operated in the shadow of accountability pressures, leading to the following statement by a science teacher, which was characteristic of the views of groups of teachers consulted in the study of Black et al. (2004):

> My department and myself have very strong views about the coursework component of GCSE. Unfortunately, because these investigations are mainly not assessing what they are supposed to, we were unable to find anything positive to say about them apart from 'in theory'. (p. 16)

A similar pathology is described in the New Zealand context by Jones and Bunting (this volume). Such examples reinforce the need to focus on moderation procedures, as set out above.

A second example is Fairbrother's (2008) analysis of the questions and marking schemes of the national assessment test papers for age 14 pupils used in England in the years 2003–2006, which showed that these paid very little attention either to the skills of scientific enquiry or to the use in science of the mathematics tested at those ages, and almost no attention to the higher order objectives of analysis, synthesis and evaluation. A telling feature was that in the marking schemes, 80 % of the marks were allotted to components for each of which the marking alternatives were just one or zero. One source of this syndrome is that if test agencies have to work to curriculum aims described in only very general terms, they have to interpret these in the concrete terms of explicit questions, and if they have to frame these within constraints, of cost, and of testing to 'cover' several years of work within a few hours, concern for validity and for effects on learning are set aside. The process produces tests with the qualities described in Fairbrother's analysis. Similar consequences were reported in the more general 2004 survey by Black et al.

These failures of a system of state education arise because of lack of alignment between the design and implementation of the three components of curriculum, instruction and assessment. Division of responsibility between state agencies responsible for these different components can lead to interagency tensions (Kuiper et al., this volume). Fensham (this volume) shows how lack of such alignment, between some of the stated aims and the actual tests, is evident in both the TIMSS and the PISA tests, while Ma (this volume) describes a similar lack of alignment in the state system in China. Again, the account of the history of the national assessment surveys in the USA (Champagne, this volume) shows how the disparity between declared aims and the validity actually achieved in the test has been an enduring feature over the last 40 years.

Ways of overcoming this problem of tension between formative and summative functions of assessment requires attention to three components. The first is that those who formulate curriculum specifications should be required to produce, at the same time, assessment instruments which implement a valid interpretation of the stated, or implied, learning aims, so designed that teaching guided by those instruments will be supportive of those aims (Millar, this volume). The second is that teachers' should take responsibility for serving the summative purposes or at least play an active part in meeting this responsibility. The third is that teachers and schools should be accountable only for those determinants of their work over which they have control—which requires that some form of (contextual) value-added system be used. This issue will not be discussed here (see Leckie and Goldstein 2011).

Conclusions

The model of section 'A Model of Assessment in the Context of Models of Pedagogy' has been used here to emphasise the view that assessment plays several central roles in any comprehensive discussion of teaching and learning, a view which has not been understood by the public, by politicians or by many academics. It is also argued that an understanding of the roles of the formative and summative assessments can only be achieved by identifying these roles and their interdependence on other determinants of learning quality. It is hoped that this model has at least helped to open up the appropriate debates, as attempted in sections 'Formative Assessment: Principles, Practices and Quality, High-Stakes Summative Assessments and Developing Teachers' Assessment Practices', on these roles. In particular, it serves to show that the tensions between the formative and the summative can be reduced, if not resolved, where teachers have ownership and agency in respect of summative assessments.

The accounts in sections 'High-Stakes Summative Assessments and Developing Teachers' Assessment Practices' serve to underline two aspects of the responsibilities of teachers and schools for the summative assessments of their students. One is that the status of teachers' judgements in the accountability debate ought to be enhanced. The other is that the quality of judgements will not earn such status unless there is sustained investment to improve that quality. Such investment must promote the professional development of teachers through programmes designed in the light of the evidence about the conditions required to make such work effective (Harrison, this volume).

The accounts in these sections serve to emphasise the various aspects which such sustained support should consider, notably assessment literacy, understanding of criteria of validity in each school subject and the disciplines and benefits of rigorous procedures for intra- and interschool moderation.

The international spectrum surveyed in section 'High-Stakes Summative Assessments' describes the wide range of state systems. Some pay little or no attention to teachers' summative work, others alternately recognise and then belittle it, some give it very constrained status, some are giving it strong support in principle but leave it adrift in practice, and a few have both accepted the needs and have provided for them. The direction of change that is needed is fairly clear. The challenge is to discern how arguments can be framed and targeted to influence those who might be able to make such change. This brings the argument full circle, back to section 'A Model of Assessment in the Context of Models of Pedagogy' and so to the cultural, social and political dimensions of pedagogy within which instruction by schools has to function.

References

Abrahams, I., & Millar, R. (2008). Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education, 30*(14), 1945–1969.

Alexander, R. (2006). *Towards dialogic thinking: Rethinking classroom talk*. York: Dialogos.

Alexander, R. (2008). *Essays in pedagogy*. Abingdon: Routledge.

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*(3), 685–730.

ARG: Assessment Reform Group. (2006). *The role of teachers in assessment of learning*. London: Institute of Education, University of London.

Baird, J-A., Black, P., Beguin, A., Pollitt, A., & Stanley, G. (2011). *The reliability programme: Final report of the technical advisory group*. Coventry: Office of Qualifications and Examinations Regulation (OFQUAL). Retrieved from www.ofqual.gov.uk/downloads/category/193-reliability-compendium

Berry, R., & Adamson, B. (Eds.). (2011). *Assessment Reform in Education*. New York: Springer.

Black, P., & Harrison, C. (2001a). Feedback in questioning and marking: The science teacher's role in formative assessment. *School Science Review, 82*(301), 55–56.

Black, P., & Harrison, C. (2001b). Self- and peer-assessment and taking responsibility: The science student's role in formative assessment. *School Science Review, 83*(302), 43–49.

Black, P., & Harrison, C. (2004). *Science inside the black box*. London: GL Assessment.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Black, P., & Wiliam, D. (2005). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *Curriculum Journal, 16*(2), 249–261.

Black, P., & Wiliam, D. (2007). Large-scale assessment systems: Design principles drawn from international comparisons. *Measurement, 1*(5), 1–53.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

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

Black, P., Harrison, C., Osborne, J., & Duschl, R. (2004). *Assessment of science learning 14–19*. London: The Royal Society.

Black, P., Harrison, C., Hodgen, J., Marshall, M., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education, 17*(2), 215–232.

Black, P., Harrison, C., Hodgen, J., Marshall, M., & Serret, N. (2011a). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education, 18*(4), 450–469.

Black, P., Wilson, M., & Yao, S.-Y. (2011b). Roadmaps for learning: A guide to the navigation of learning progressions. *Measurement: Interdisciplinary Research and Perspectives, 9*(2–3), 71–123.

Blatchford, P., Baines, E., Rubie-Davies, C., Bassett, P., & Chowne, A. (2006). The effect of a new approach to group-work on pupil-pupil and teacher-pupil interaction. *Journal of Educational Psychology, 98*, 750–765.

Boyd, B., & Hayward, L. (2010). *Exploring assessment for accountability*. Glasgow: Universities of Strathclyde and Glasgow.

Brookhart, S. M. (2011). *The use of teacher judgment for summative assessment in the United States: Weighed in the balance and (often) found wanting.* Oxford, UK: Oxford Centre for Educational Assessment. Retrieved from http://oucea.education.ox.ac.uk/events/teacher-judgment-seminar/papers-2/

Bruner, J. (1966). *Toward a theory of instruction*. New York: Norton for Harvard University.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *The British Journal of Educational Psychology, 58*(1), 1–14.

Colbert, P., Wyatt-Smith, C., & Klenowski. V. (2011). *A systems level approach to building sustainable assessment cultures: Moderation, quality task design and dependability of judgment.* Oxford, UK: Oxford Centre for Educational Assessment. Retrieved from http://oucea.education.ox.ac.uk/events/teacher-judgment-seminar/papers-2/

Crooks, T. J., Kane, M. Y., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education, 3*(3), 265–285.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality and development*. Philadelphia: Psychology.

E.U. Council of Ministers. (2010). *Assessment of key competences: Draft Background Paper for the Belgian Presidency meeting for Directors-General for school education.* Brussels: E.U.

Fairbrother, R. (2008). The validity of key stage 2 science tests. *School Science Review, 89*(329), 107–114.

Freire, P. (1992). *Pedagogy of hope*. New York: Continuum.

Gardner, J. (2007). Is teaching a 'partial' profession? Making the grade. *Journal of the Institute of Educational Assessors,* (Summer), 18–21.

Gardner, J., Harlen, W., Hayward, L., & Stobart, G. (2011). Engaging and empowering teachers in innovative assessment practice. In R. Berry & B. Adamson (Eds.), *Assessment reform in education* (pp. 105–120). New York: Springer.

Hallam, S., & Ireson, J. (1999). Pedagogy in the secondary school. In P. Mortimore (Ed.), *Understanding pedagogy and its impact on learning* (pp. 68–97). London: Paul Chapman.

Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education, 20*(3), 245–270.

Harrison, C., & Howard, S. (2009). *Inside the primary black box*. London: GL Assessment.

Hodgen, J., & Marshall, B. (2005). Assessment for learning in mathematics and English: Contrasts and resemblances. *Curriculum Journal, 16*(2), 153–176.

Hutchinson, C., & Hayward, L. (2005). The journey so far: Assessment for learning in Scotland. *Curriculum Journal, 16*(2), 225–248.

Johnson, S. (2011). *A focus on teacher assessment reliability in GCSE and GCE.* Coventry: Office of Qualifications and Examinations Regulation. Retrieved from www.ofqual.gov.uk/downloads/category/193-reliability-compendium

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis*. Retrieved from http://www.clcrc.com/pages/cl-methods.html

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

Koretz, D. (1998). Large scale portfolio assessments in the US: Evidence pertaining to the quality of measurement. *Assessment in Education, 5*(3), 309–334.

Leckie, G. B., & Goldstein, H. (2011). A note on 'The limitations of using school league tables to inform school choice'. *Journal of the Royal Statistical Society: Series A, 174*(3), 833–836.

Mansell, W., James, M., & Assessment Reform Group. (2009). *Assessment in schools: Fit for purpose? A commentary by the Teaching and Learning Research Programme*. London: Economic and Social Research Council.

Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learn science. *British Educational Research Journal, 30*(3), 359–377.

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes. Towards a wider conceptual field. *Assessment in Education: Principles, Policy and Practice, 5*(1), 85–102.

Shapley, K. S., & Bush, M. J. (1999). Developing a valid and reliable portfolio assessment in the primary grades: Building on practical experience. *Applied Measurement in Education, 12*(2), 111–132.

Shulman, L. S. (1999). Knowledge and teaching: Foundation of the new reform. In J. Leach & B. Moon (Eds.), *Learners and pedagogy* (pp. 61–71). London: Chapman.

Stanley, G., MacCann, R., Gardner, J., Reynolds, L., & Wild, I. (2009). *Review of teacher assessment: What works best and issues for development*. Oxford, UK: Oxford University Centre for Educational Development.

Watkins, P., & Mortimore, P. (1999). Pedagogy: What do we know? In P. Mortimore (Ed.), *Understanding pedagogy and its impact on learning* (pp. 1–19). London: Paul Chapman.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*(1), 49–65.

Wood, D. (1998). *How children think and learn*. Oxford, UK: Blackwell.

Chapter 12
An Assessment *Perezhivanie:* Building an Assessment Pedagogy *for, with* and *of* Early Childhood Science Learning

Marilyn Fleer and Gloria Quiñones

Introduction

Assessment *for, with* and *of* learning within the field of *early childhood*[1] science education brings to the fore the child within that child's social and emotional context. In contrast to subject-specific areas, the underlying assessment traditions in early childhood seek to capture the whole child in the dynamics of children's everyday interactions within the range of learning settings—playgroups, family day care, long day care, kindergarten and the early years of school. Discipline-specific areas, such as science, are assessed within the context of close observations of children as they go about their day-to-day activity within these early childhood settings—often across the borders of home and the early childhood setting which they attend. The tradition has been to capture the wholeness of the situation alongside how the child is engaging scientifically and how she/he is coming to understand scientific concepts at their site of use or construction. That is, the child and the context are not separated but are dialectically related within the assessment context (see Fleer and Jane 2007). Early childhood education assessment practices therefore do not seek to reduce or dissociate the learner from the learning context. As Vygotsky (1997) cautions, we often see 'an enormous mosaic of mental life developed comprised of separate pieces of experience, a grandiose atomistic picture of the dismembered human mind' (p. 4).

Many assessment techniques for older children which rely upon the spoken word or a response in a written form are simply not possible or very difficult to use reliably in early childhood education. The child cannot yet read, is still developing her/his language and does not yet have a concept of performing one's best for an assessment context. Importantly, what is understood about a child at one point in time can

[1] This term is also known as early years education, early childhood education and development, or early child care and development.

M. Fleer (✉) • G. Quiñones
Monash University, Building 6, Wellington Road, Clayton, VIC 3800, Australia
e-mail: marilyn.fleer@monash.edu

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_12,

change within seconds as a child flips back and forth in her/his thinking about the social and material world. The challenge is how to capture the dynamic and myriad scientific interactions that a child encounters (Carr 1998) in ways that give confidence that a reliable judgement has been made about a child's understandings in science.

In this chapter we seek to explore the *wholeness of the assessment context* where the learner and the social context are inextricably fused, in order to discuss the complexity of assessment in early childhood science education. We draw specifically on cultural-historical theory to present our arguments. We begin with a broad overview of general assessment in education and later in the context of early childhood education, followed by the historical and contemporary assessment practices that inform how science content knowledge is being determined. We conclude with a framework which seeks to make visible an assessment approach capturing the dynamic relations between development and learning within the early childhood period, and through this, we provide an assessment approach which is centred on illuminating children's science content knowledge and capability as located and understood within meaningful contexts.

Assessment in Education

In the field of science education, the nature of the relationship between the teacher and the students is a key characteristic of assessment (Black 1993). Formative assessment recognises the relationship the pupil has with the teacher and her/his knowledge that the information being assessed is important and makes a significant difference to pupil outcomes (Black 1993; Wiliam et al. 2004). For instance, this can be seen when '…the assessment information is used, by both teacher and pupils, to modify their work in order to make it more effective' (Black 1993, p. 49). Formative assessment includes assessment *for* learning (Wiliam et al. 2004). Investigating how teachers undertake formative assessment is clearly important (Black 1993).

In their research of assessment in classrooms, Black and Wiliam (1998) investigated formative assessment. These researchers looked at primary school and college classroom context and examined student perspectives and the teacher's role in formative assessment. They found examples of 'classroom experience' (p. 10) which centred on assessment for learning, where a lot of assumptions were made regarding the nature of learning and how this related to the assessment of the students. Black and Wiliam revealed that assessment between peers was beneficial and argued that formative assessment had a major impact on improvements to student's learning. For this chapter it is significant that a diverse range of strategies were used by teachers for formative assessment. Moreover, Black and Wiliam (1998, 2003) found that there is a close relationship between assessment, teaching practice and the tools used by teachers to assess. One strategy proposed by Black and Wiliam (1998) that is particularly relevant for early childhood assessment is *questioning*. They noted that when teachers have good teaching strategies, such as being able to use questioning as an effective teaching tool, then these teachers could also use these techniques for effective assessment.

Research undertaken by Wiliam et al. (2004) in primary and secondary science classes found that teachers' formative assessment techniques tended to focus mostly on questioning as the main way of assessing *for* student's learning. They noted that teacher support in their assessment practices was found to be important for the quality of children's learning. This strategy involved both oral and written tasks. Teachers questioned students, which provoked pupil thinking and in turn resulted in students asking questions. Questioning also proved to be important for knowing more about students' prior knowledge. Another aspect the researchers considered in the use of formative assessment was the quality of feedback. When this was investigated, it was found that formative assessment was often implemented as a 'static process' (p. 52) where a student's knowledge or knowledge of the concept was simply fed back to the student. And in extreme cases, some teachers gave solutions to a mathematical problem to students for them to gain knowledge of the answer to the problem.

Some issues found in relation to formative/summative assessments were how questioning design is relevant or not for students' learning and how students can be given a more active and stronger voice in their own learning (Black 2001). Deep consideration of the design of assessment tools is important. Further, how questioning provides links with a student's zone of proximal development and how it links with teacher scaffolding need to be further researched (Black 2001).

Summative assessments also provide a problematic pedagogical challenge for teachers. In their longitudinal study with English and mathematics teachers, Black et al. (2011) showed how summative assessment practices are beneficial to students because they are able to see their progress. They also note that teachers see summative assessment as a 'shared enterprise' (p. 460).

The next section discusses how these assessment concepts are dynamically practised by early childhood professionals within science learning contexts. We begin with a presentation of the assessment traditions found in early childhood education and discuss contemporary approaches to early childhood assessment.

Foundational Framing of Assessment Practices Within the Field of Early Childhood Education

Due to the complexity and challenges associated with the assessment of very young children, there has been a long history of assessment theory dating back to G. Stanley Hall's (1844–1924) original child study movement. Aspects of this work are still foundational in much of what occurs in practice within the field of early childhood education in countries that have supported the establishment of kindergartens. In essence, child study practices saw early childhood professionals trained to make detailed observations of young children, paying close attention to the reliability and validity of their observations. Central to the child study movement and how it was originally taken up in early childhood education has been a maturational view of development, a focus on stages and the strict positioning of the teacher as an objective observer of young children's development. Assessment practices within early childhood education were conceptualised through the child study movement as a *method of scientific observation*.

The basic approach to assessment practice in early childhood education therefore requires close observations of children in the learning context independent of the assessor. That is, the observer must be conceptualised as objective and distant from the child—despite the realities of the teaching-learning context where child activity is almost completely reliant upon adult support. Mainstream practices in early childhood assessment still overwhelmingly require children's development to be recorded in individual portfolios, often under the headings of cognitive, social-emotional, physical and language development, despite these documents being quite dated (see Tasmania Department of Education 2008). In these portfolios, subject-specific content is recorded, but usually not foregrounded as much as children's overall development (e.g. NAEYC and NAECS/SDE 2003). Being aware of this assessment practice is important because science as a field of knowledge has not been seen as needing to be assessed within early childhood education. Rather, an assessment of children's overall development has been central to the work of early childhood professionals.

In summary, conceptualising how scientific knowledge and scientific capability are to be assessed is challenging because the field has concentrated on the assessment of children's overall development, rather than being oriented towards discipline-specific content knowledge. Further, the scope and sequence of school science content as presented in curriculum documents has traditionally not been foregrounded in early childhood curriculum (see Commonwealth of Australia 2009; Hedges 2002; Nuttall 2005). Internationally, science content in early childhood curriculum has had a minor part to play in the planning of children's cognitive development. In addition, the assessment of children's conceptions in science has a short history. However, the field of early childhood education does have a long-standing and mature approach to assessment practice. How this has been transformed over time is discussed in the following section.

Contemporary Assessment Practices Within the Field of Early Childhood Education

Two major changes in early childhood education have had a significant impact on assessment practice affecting early childhood science. The first has been the recent introduction of discipline content knowledge into many early childhood curricula around the globe (Commonwealth of Australia 2010). This has provided the space and impetus for early childhood teachers to engage in thinking about children's development beyond the broad traditions of cognitive, social-emotional, physical and language development (e.g. Cullen 1994, 2003; Fleer and Richardson 2009). Second, the profession has moved beyond maturational theories of child development, having been influenced by critical theories (see Fleer et al. 2008), poststructuralist theories (MacNaughton 2009) and cultural-historical theories (e.g. Jordan 2009). The former two have invited the profession to ask questions about whose development was captured in the original child study movement literature, to problematise the image of the child as being *innocent and not capable* as mooted by Stanley Hall himself and to call into question measurement as predefined lockstep stages of development (e.g. Dahlberg et al. 1999).

Alternative theories of child development have been put forward, and early childhood educators have sought to try and capture the dynamics and complexity of the assessment context as they document children's learning across a much broader range of areas, including science content knowledge (e.g. Cowie 2000; Podmore 2009). These influences have allowed a new view of children and childhood to emerge, which also takes into account the *scientific and technological child* (Fleer 2011). Through this we have seen major changes in assessment theory and practice for the field of early childhood education (e.g. Carr 2000, 2001; NAEYC and NAECS/SDE 2003; NZ Ministry of Education 2011).

Cultural-Historical View of Assessment

Groundbreaking work by Carr (2000, 2001) has turned assessment practice away from a deficit view (i.e. identification of developmental problems) to a credit model (what children can do), leading to a more ethical approach to assessment practice in early childhood education. A credit model is captured as a principle of ethical assessment practice which now underpins contemporary child observations and which forms a central dimension of what is foregrounded in many countries, including Australia. For instance, the Department of Education, Employment and Workplace Relations (2010) published an educators' guide advocating an assessment pedagogy which is ethical, dynamic, forward-oriented and child-focused, as is summarised below in Fig. 12.1 (Fleer et al. 2008, cited in Commonwealth of Australia 2010, p. 39).

How an ethical, dynamic, forward-oriented and child-focused assessment pedagogy is realised in practice can be seen in Fig. 12.2 (assessment example) where we see a series of five photographs which document a toddler's investigation of shadows. This represents an example of how educators capture the demonstrated abilities of even the youngest members of the early childhood community. Alongside of the photographs is a narrative associated with the images. In addition to this data is usually an analysis of the learning, along with how this information will be used for planning for the infant's future learning. Links are usually made directly to the curriculum framework, highlighting important outcomes. In this particular case, the educator stated that

> Ryder was curious to discover the functions and attributes of the shadow as he used his senses to explore it. He watched movements and patterns created and then became confident to lean in and test it and touch it with gentle strokes at first and then by tapping. (Commonwealth of Australia 2010, p. 123)

Links to the curriculum outcomes of problem solving, inquiry, experimentation, hypothesising, researching and investigating were made in this assessment example.

Unlike previous approaches to assessment, such as seen in the child studies movement discussed in the previous section where the observer is always absent

Assessment principle	Assessment practice (what this means in practice)
Assessment pedagogy	
Ethical assessment	Systems and staff select assessment methods and tools which provide children with opportunities to confidently demonstrate their capabilites.
Dynamic assessment	Systems and staff select assessment methods and tools where interactions between the assessor and the assessed are in the context of meaningful, supportive and respectful interactions.
Forward measuring assessment	Systems and staff select assessment methods and tools which include the ability to assess children's potential, rather than just their actual development/learning.
Child oriented assessment	Systems and staff include assessment methods and tools where children can assess themselves.
Assessment content	
Assessment should match the curriculum	*Criteria for checking the content and method of assessment:* • What is the *emphasis*? Is there an *over-emphasis* on things that can be easily measured? • Are the *important understandings* able to be assessed? • Is *community specific* valued content well represented? • Are the assessment pedagogies not too *time consuming* leading to assessing skills or understandings which are not rich, complex or integrated? • Are they the *most appropriate* or useful tools or methods? Sometimes, tools and *strategies that are 'easy'* to use are not always the most appropriate. • *Do they encourage staff to teach to the assessment task, rather than planning a broader curriculum?* Remember: important learning for children may involve learning outside of set assessment tasks. Fleer et al (2008)

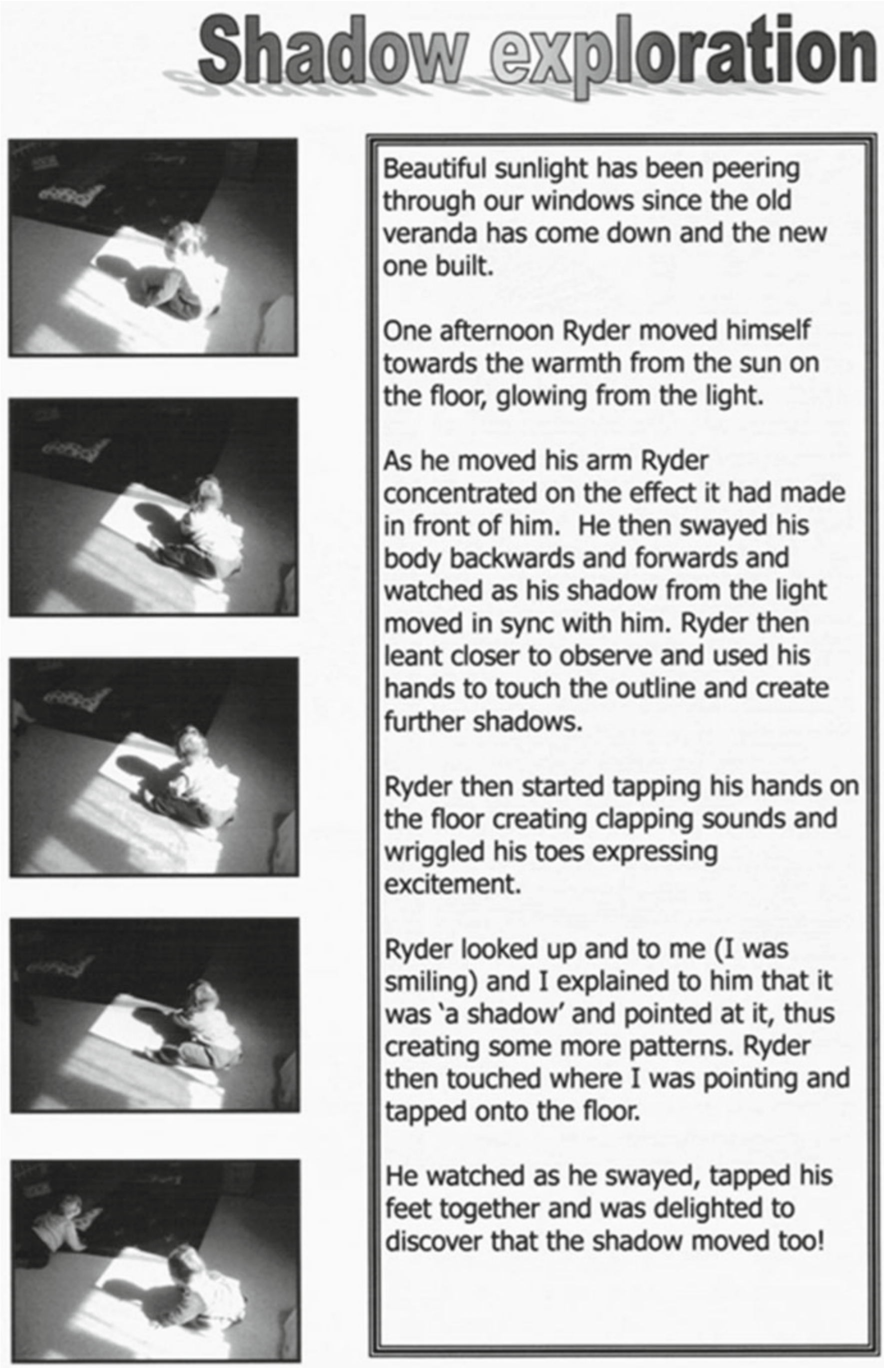
Fig. 12.1 Assessment tools for early childhood (Commonwealth of Australia 2010, p. 39)

from the observation, this educator specifically takes a more active stance in her cultural-historical analysis and states, for example:

> Ryder expressed great satisfaction with his discovery as he clapped and wriggled his toes with excitement. He was also happy to share his exploration with me as he looked towards me and watched to find further sights and movements as I pointed with my finger. (Commonwealth of Australia 2010, p. 123)

Including the adult in the observations is now more common and is more reflective of the dynamic assessment contexts that make up early childhood assessment practice.

Dynamic assessment (Lunt 1993) seeks to bring together the assessor and the assessed into a dynamic assessment situation with a focus not on a child's past experience but rather their capacity for learning. Much of the pioneering work has focused on older children and often in controlled clinical situations (e.g. see Feuerstein et al. 1979, 1980; Tzuriel 2001). Its appeal is summed up well by Elliott (2003), who states that 'the strength of dynamic measures is that these may yield



Shadow exploration

Beautiful sunlight has been peering through our windows since the old veranda has come down and the new one built.

One afternoon Ryder moved himself towards the warmth from the sun on the floor, glowing from the light.

As he moved his arm Ryder concentrated on the effect it had made in front of him. He then swayed his body backwards and forwards and watched as his shadow from the light moved in sync with him. Ryder then leant closer to observe and used his hands to touch the outline and create further shadows.

Ryder then started tapping his hands on the floor creating clapping sounds and wriggled his toes expressing excitement.

Ryder looked up and to me (I was smiling) and I explained to him that it was 'a shadow' and pointed at it, thus creating some more patterns. Ryder then touched where I was pointing and tapped onto the floor.

He watched as he swayed, tapped his feet together and was delighted to discover that the shadow moved too!

Fig. 12.2 Assessment example: learning stories assessment in practice (Commonwealth of Australia 2010, p. 122) better predictions of subsequent educational performance than traditional static cognitive tests…' since in dynamic assessment people are 'now less likely to ask 'how can we most appropriately sort and classify children?' but rather 'how do we teach this child" (p. 20).

Underpinning a dynamic view of assessment that speaks to the early childhood context is the idea of adult mediation, which has allowed the assessor to move beyond a static and individual construction of the assessment context. That is, rather than measuring what a child can do on her/his own, dynamic assessment seeks to assess the child and the adult working together at a higher cognitive level, where the extent of the mediation is measured alongside what is achieved. For example, in an early childhood context, the assessor takes into account the number and type of prompts needed for a child to work with a collaborator to achieve at a higher plane of thinking than when the child works alone. This can be demonstrated in the assessment example (Fig. 12.2) where the educator takes note of how many times and in what ways he/she needed to point to the changing shape of the shadow or deliberately emphasise the shadows with additional hand gestures, before the infant responded. The child's exploration of the shadows and the educator's pointing gesture as a tool for mediation are included in the assessment documentation.

Dynamic assessment in early childhood contexts goes beyond capturing a static moment in time, seeking rather to position assessment as being in motion and mediated through interaction in social situations. This is evident in the assessment example because the adult and the child *together expanded the learning situation* through their ongoing interaction. This is a more robust approach for gaining insights into children's scientific thinking because it captures the motion or movement of children's thinking in relation to the context in which the thinking is occurring. It is therefore a more reliable and valid approach to assessment in early childhood than assessment which is conceptualised as gaining knowledge about some 'end point of learning'. *Dynamic assessment* captures the embedded and fluid nature of assessment practices undertaken by early childhood professionals.

Dynamic assessment draws on Vygotsky's (1987, 1997) concept of the zone of proximal development (ZPD), and this concept has had traction because it helps explain the significance of the educator in the assessment pedagogy (Fleer and Richardson 2003, 2006, 2009). Much of the theorisation and empirical work has centred on either measuring a learner's ZPD within a dynamic assessment context or framing it as an interactive formative assessment that is integrated into classroom instruction (see Allal and Ducrey 2000). This chapter focuses on the latter perspective. Assessment practices which seek to document what a child can do on their own relate directly to the *actual zone of a child's development,* while the *proximal zone* captures the dynamic relations between the assessor and the assessed. What is new for the field of early childhood education is the concept of the *zone of potential development* (Kravtsova 2008), which takes into account future possibilities which orient the child to particular events or concepts, but does not yet expect the child to perform or achieve in these. For instance, in the example of the infant playing with his shadow, the infant does not yet have an understanding that white light travels in straight lines and that it is possible to change its path or its speed in specific ways. Creating the conditions around the infant for future conceptual possibilities for science learning falls into the concept of the *zone of potential development*. For example, orienting the infant to how he can redirect light would fall within the realms of potential development, such as when the infant is provided with a hand-held puppet in the same context described in the assessment example, and together with the educator, they create a shadow puppet show. Introducing a prism and refracting light may also be something that the educator and the infant could do together. However, it is unlikely that the infant would gain scientific understandings about how white light travels in straight lines, that it can be absorbed, reflected or refracted. An infant is unlikely to gain the scientific understandings that light changes speed and direction when it passes through one medium and to another or that when a light reaches a surface, it may be partially or completely transmitted, absorbed or scattered back from the surface. Rather, the interaction between the infant and the educator constitutes the zone of potential development because the child is not psychologically able to engage in the experience with scientific conceptual understanding. Yet this experience is necessary for later learning because these kinds of social experiences in early childhood lay important foundations for scientific thinking about things to notice about how light behaves.

This idea is captured in the concept of *potentive assessment* (Fleer 2010), which sits within the suite of formative and summative assessment practices (MacDonald 2007). *Potentive assessment* seeks to capture an important interactional sequence between the educator and the infant, where the infant is oriented to scientific playfulness that helps her/him to further investigate everyday life with a more scientific lens in the future. Capturing these experiences as part of the assessment interaction is necessary for determining the kinds of experiences and interactions in early childhood that build scientific concepts for future learning. Why this is important in early childhood education is because traditionally early childhood educators have interacted and assessed in the 'here and now', looking back on what has been achieved or what is currently being enacted, rather than examining how interaction sequences can orient learners in particular ways to work and think scientifically in the future. This practice is assessment *for* future learning. Naming this practice as *potentive scientific assessment* helps the early childhood educator pay close attention to the significance of their role in scientifically orienting children towards their natural and built environment.

An Assessment Perezhivanie

Although assessment *for* learning now underpins assessment practices within most countries, early childhood educators still find the concept of assessment challenging, as noted by Fleet and Patterson (2011): 'The term itself makes some early childhood educators nervous, as it may imply a formal testing regime, which is associated often with bleak personal memories' (p. 2), and 'there is a hesitancy about tackling the concept of "assessment" in prior-to-school settings' (p. 1). How educators feel about the practice of assessment, combined with how they feel about their science knowledge, makes an interesting emotional cocktail.

There is a vast and long-standing body of empirically based literature on teacher knowledge and confidence to teach science concepts to young children (Appleton 2006). However, little attention has been directed towards acknowledging the relations between how a teacher feels in the moment of teaching and how the children take up these emotionally charged scientific interactions. We argue that we need to capture this emotionally charged dynamic in the assessment pedagogy because we believe that all learning of concepts is not only connected to how one feels at the point of learning the concept, but this emotionality frames the child's future attitudes towards science. We draw upon Vygotsky's (1994) concept of *perezhivanie* to name the relations between cognition and emotion when assessing science and scientific activity of very young children.

Vygotsky (1994) showed in his writing the importance of the bringing together both individual and social relationships in the concept of *perezhivanie*: 'The emotional experience [*perezhivanie*] arising from any situation or from any aspect of his [sic] environment determines what kind of influence this situation or this environment will have in the child' (p. 339, emphasis in original). According to Gonzalez Rey (2009), the concept of *perezhivanie* remains unfinished. Perezhivanie encapsulates the unity of not only cognitive factors in a learning situation but also emotional ones. In the translations into English, this concept is defined as 'emotional experience'. We argue that this concept is useful for the assessment of scientific thinking and capability of young children because it puts cognition in unity with emotions.

How a teacher feels about teaching science or how a child feels about their learning in science has always been deemed important in science education research (Cowie 2000). In the assessment context, this should be taken into account because it goes beyond cognitive capabilities alone. We argue that children's thinking should not be the only aspect that is assessed but that we need to capture how they are emotionally experiencing the learning or assessment situation. This is what is central to the concept of *perezhivanie*.

Documentation of children experiencing their social and material environments, such as classroom and home, alongside how children relate to each other and to the educator, makes up the assessment context of science learning, giving a broader reading of the child 'doing and learning science'. Vygotsky (1994) emphasised the *attitudes* children have to situations and noted how different children will experience the same *situation and conditions* differently. Bozhovich (2009) captured this central idea when stating that 'Children have different attitudes towards one and the same reality and experience it differently' (p. 68). In assessment, these emotionally charged situations should be studied *in unity with cognition.* Through the different 'living experiences' and attitudes a child holds of a specific subject, such as science, we build a picture of not only the child's thinking of the experience but their *emotional commitment* to the experience. We argue that in assessment, educators become aware of, and interpret the event from, the child's emotional experience and analyse the attitude to learning and to assessment that is evident. As Vygotsky (1994) points out, 'It ought to be able to find the relationship which exists between the child's emotional experience [perezhivanie], in other words how a child becomes aware of, interprets, [and] emotionally relates to a certain event' (p. 341).

Experiences are not only one moment in time but are a sequence of moments and events for the child. Planning different experiences and their sequence from moment to moment becomes important. This is because the child in the different experiences relates not only to the physical environment but also to the social environment, which offers different conditions and determines experiences and social relations between peers and teachers. It becomes important for educators to assess these moments.

Perezhivanie—living experiencing—is a complex system that is difficult to make visible. The commitment of the teacher to make this visible is essential for showing not only how the child is thinking but also the child's 'affective relationship' (Bozhovich 2009, p. 66) to science. Significantly, assessment must be related to how the environment affects the course of the child's development and how the physical and social environment is understood by the child. In drawing on Bozhovich's (2009) work to re-theorise assessment, we argue that the assessment context should examine:

– What is the 'nature of the experience' (Bozhovich 2009, p. 67)?
– How do children understand (perceive and are able to conceptualise) the circumstances affecting them?
– How do children make meaning?
– What kind of 'affective state' (Bozhovich 2009, p. 68) is the child in when in the pedagogical experience or assessment context?

In determining the child's emotional attitude towards the science context and content, we can begin to make better judgements about the child's scientific learning. This can be exemplified through the assessment example, where it would be possible to build understandings about light through darkening a room and controlling a single light source to beam into the darkened room, thus making it easier for young children to begin to understand the nature of how light travels. However, we know that because light surrounds young children all of the time, it is probable that they do not think about the 'sea of light that surrounds them' but rather would be more focused on its absence. Darkening a room in an early childhood centre is helpful for supporting children's learning. However, darkness can also be associated with a significant amount of fear, and being aware of this when planning for teaching and assessment is highly significant.

> If we want to understand exactly how the environment affects children and what influence it exerts in their mental development, then we must analyze the relationship between the environment and the child's needs, the extent to which it is capable of satisfying them or, in some cases, hinders their satisfaction. (Bozhovich 2009, p. 70)

It is extremely important, as suggested by Bozhovich (2009), that pedagogical experiences (and we argue assessment) take into account the circumstances and conditions that 'affect' children. As such, it is important to document the *affective state* of the environment that children find themselves in during the assessment moments. Some of these moments can have a profound and long-lasting effect on a child. For instance, when teaching children the concept of light, focusing only on the cognitive side and not the emotional dimensions could potentially build negative associations with learning this particular scientific concept. Bozhovich has suggested that some children might develop a negative relationship towards learning, and school in general, when their early learning environments are not positive experiences. We argue that the nature of these emotional connections to learning concepts such as light is related to how they may behave towards school science later in secondary school or even when studying early childhood teacher education later in life. Analysing the conditions in which emotional attitude is formed is important in influencing children's affective relation to the experience, the nature of which is a 'complex system of feelings, affects, and moods' (p. 74). Its significance at that time or even in the future when these same children as adults may be teaching science content to children is often not considered.

We believe that the educator in the assessment moment needs to be aware of what kinds of tools are available to analyse both cognitive and affective development and give thought to finding the balance between these dimensions in assessment. Vygotsky (1987) conceptualised this mental and affective unity as 'motive forces, [where] we identify, needs, interest, incentives' of children (p. 50). In thinking and affect, feelings and emotions are interrelated with the child's needs in a learning experience. As children relate differently to their experience of the learning, they also have different needs, interests and motives for learning. In order to assess specific moments in the learning experience authentically, the educator should note how she/he affects children with regard to scientific interest and motives. Consequently, it is important to examine the complex system of feelings, affects and moods held by the child in the process of assessing learning. The pleasure that was being experienced by the infant in the example and that was written into the assessment (see Fig. 12.2) begins to capture the unity of affect and cognition. These moments in assessment are special and require deliberate planning by teachers to involve not only children's thinking but also their feelings. Affect relates to the child's emotions and feelings towards the nature and content of the experience and how it is transformed as the child is able to understand and make meaning of these experiences. It can be seen how this concept of *perezhivanie* encapsulates both the affective relations to the child's social environment and their thinking about it (through making meaning and understanding) and the emotional and affective state of the child at that moment in time and what they bring to that particular environment.

The concept of perezhivanie that we discuss in this chapter is therefore related to the living experiences of children generally and where they are exposed specifically to science concepts. The need for children to have science experiences becomes important in their everyday life and academic learning. In this chapter we also argue for deliberate attempts to capture, in the assessment moment, future possibilities that are being enacted by the educator by their orientation of the child to their scientific world. Science as a human invention created over time cannot be simply imparted and assessed, but requires an orientation to new ways of seeing everyday life for a child. The infant who notices the shadow, and the educator who interprets the infant's actions and then creates the conditions which help build further investigations, is working to capture an emotional engagement with science and a motive for children to work scientifically to further explore their world. We capture the concept of perezhivanie and potentive assessment in relation to the assessment example (Fig. 12.2) in Table 12.1 below.

Table 12.1 An assessment perezhivanie in science—related to the assessment example (Infant playing with his shadow)

Perezhivanie—the unity of cognition and emotion	Dynamic assessment	Potentive assessment
What is the nature of the experience that 'affects' children?	*What kind of living learning experiences do teachers plan that make 'affect' visible in their assessment of science?*	
	In the moment: through close observation of an emotionally charged situation, such as when the infant interacted with his shadow, the educator can join the assessment moment and extend the exploration of shadows through playful gesturing, resulting in the documentation of affective dimensions of the noticing and the exploring by the infant	*Beyond the moment:* through introducing the absence of light (darkness) the infant is placed into his potential zone of development for later learning about light as something which exists, while at the same time allowing him to control the light source so that he can control 'the level of light and darkness' in relation to how he feels
How do children perceive and conceptualise the circumstances affecting them?	*How do children perceive the science learning experiences?*	
	In the moment: through close observation of the interaction between the infant and the assessor when engaged in playfulness with the shadow, the educator can make judgments about the *infant's intentionality* regarding his developing understanding of blocking light to create a shadow	*Beyond the moment:* children's perceptions of the learning experience deepen over time, such as when an educator deliberately draws attention to a range of light sources, such as a torch, the sunlight or a candle. Assessment of *emotional attitude towards exploration* in partnership with the educator gives insights into future activity for working scientifically
What is the 'affective state' of the assessment context?	*What is the educator's and the child's 'mood' for experiencing science?*	
	In the moment: the raw emotions that are expressed as the infant moves about and blocks the light are an indication of the current affective state of the infant	*Beyond the moment:* the feelings that children and teachers promote in the learning experience can be assessed in relation to how children 'feel' (safe and supported) when 'thinking' and working with science concepts
How do children make meaning? How do emotions contribute to that meaning?	*How is science meaning making captured in the moment?*	
	In the moment: extending the learning experience to include assessment of how children are making meaning when 'in' the experience. The educator documents both thinking in the moment and emotions in the moment	*Beyond the moment:* extending the learning experience to include assessment of how children are making meaning after the experience. The educator documents a tapestry of both scientific thinking and emotions at different points in time related to the science concept

Conclusion: Assessment Pedagogy for Early Childhood Science Education

It has been argued in this chapter that the field of early childhood education is a biologically and culturally unique period requiring a specific form of assessment pedagogy for determining understandings of, and capabilities in, early childhood science. We conclude this chapter by bringing together all of the assessment principles discussed above into a specific form of assessment pedagogy suitable for early childhood science education.

In the conceptualisation captured in Table 12.2, the principles include ethical, dynamic and forward-measuring child assessment. The pedagogies needed to enact these principles include credit models, assessable moments and proximal and potentive assessment framing. The practices observed should result in positive experiences for children when in the assessment moment(s), feeling safe and supported when being assessed, being oriented to future learning contexts and having agency and control in the assessment situation. The concept of an *assessment perezhivanie* captures the importance of the unity of cognition and emotions as the child experiences science in everyday life at home and in the early childhood setting. The assessment approach takes account of cognitive capabilities and thinking alongside of what the children are emotionally experiencing in the learning or assessment situation in collaboration with others. This in turn allows children to have an 'affective relation' with, and develop a motive to learn, science concepts as they interact with competent others who are already using the science concept in their daily life.

Table 12.2 Assessment pedagogy of early childhood science

Assessment principle	Assessment pedagogy	Assessment practice (what this means in practice)	Assessment perezhivanie
Ethical assessment	*Credit models*	Assessment practices determine what children can do and know, rather than highlighting what they cannot yet do	Assessment context is positive, resulting in a positive attitude towards the assessment experience
Dynamic assessment	*Assessable moment*	Interactions between the assessor and the assessed support performance at a higher level than when the assessed is working alone	Children feel supported and safe as they collaborate with others to achieve a given assessment task
Forward measuring assessment	*Proximal and potentive assessment*	Assessment approaches allow for the documentation of children's supported and potential levels, rather than just their actual level	Eavesdropping on what matters or being supported or engaged in science activities with more highly competent others develops a motive orientation to science

Capturing moments deliberately through assessment means that a teacher supports the whole performance of science learning and documents both the development and the support given in these organised interactional situations. Through science being collectively experienced and assessed, where thinking and the felt experience of the science concept are documented, it becomes possible to make judgements about children's engagement with the context and concepts of science.

We believe these principles and practices captured as assessment pedagogy for early childhood science education make an important contribution to supporting early childhood professionals as they reconceptualise their assessment approaches to take account of children's science content learning. Our assessment pedagogy *for*, *with* and *of* early childhood science content knowledge and capability will provide one way forward for early childhood teachers who wish to engage more intentionally with science content knowledge.

References

Allal, L., & Ducrey, G. P. (2000). Assessment *of*-or-in the zone of proximal development. *Learning and Instruction, 10*, 137–152.

Appleton, K. (2006). *Elementary science teacher education. International perspectives on contemporary issues and practice*. New Jersey: Lawrence Erlbaum Associates.

Black, P. J. (1993). Formative and summative assessment by teachers. *Studies in Science Education, 21*(1), 49–97.

Black, P. (2001). Dreams, strategies and systems: Portraits of assessment past, present and future. *Assessment in Education: Principles, Policy & Practice, 81*(1), 65–85.

Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Black, P. J., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal, 29*(5), 623–637.

Black, P., Harrison, C., Hodgen, J., Marshall, B. M., & Street, N. (2011). Can teachers' summative assessments produce results and also enhance classroom learning? *Assessment in Education: Principles Policy & Practice, 18*(4), 451–469.

Bozhovich, L. I. (2009). The social situation of child development. *Journal of Russian and East Europeon Psychology, 47*(4), 59–86.

Carr, M. (1998). *Assessing children's learning in early childhood settings: A series of three video tapes*. Auckland: Video Campus and Wellington/NZCER.

Carr, M. (2000). Technological affordances, social practice and learning narratives in an early childhood setting. *International Journal of Technology and Design Education, 10*, 61–79.

Carr, M. (2001). *Assessment in early childhood settings. Learning stories*. London: Paul Chapman.

Commonwealth of Australia. (2009). *Belonging, being and becoming. The early years learning framework for Australia*. Canberra: Australian Government Department of Education, Employment and Workplace Relations.

Commonwealth of Australia. (2010). *Educators belonging, being and becoming. Educators' guide to the early years learning framework for Australia*. Canberra: Australian Government Department of Education, Employment and Workplace Relations.

Cowie, B. M. (2000). *Formative assessment in science classrooms*. Unpublished doctoral dissertation, University of Waikato, New Zealand.

Cullen, J. (1994). Why retain a development focus in early education? In E. Mellor & E. Combe (Eds.), *Issues in early childhood services: Australian perspectives* (pp. 53–64). Dubuque: Wm C. Brown.
Cullen, J. (2003). The challenge of Te Whaariki: Catalyst for change. In J. Nuttall (Ed.), *Weaving Te Whaariki: New Zealand's early childhood curriculum document in theory and practice* (pp. 269–396). Wellington: New Zealand Council for Educational Research.
Dahlberg, G., Moss, P., & Pence, A. (1999). *Beyond quality in early childhood education and care: Postmodern perspectives*. Philadelphia: Falmer.
Elliott, J. (2003). Dynamic assessment in educational settings: Realizing potential. *Educational Review, 55*(1), 15–28.
Feuerstein, R., Rand, Y., & Hoffman, M. B. (1979). *Dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments and techniques*. Baltimore: University Park.
Feuerstein, R., Rand, Y., Hoffman, M. B., & Miller, R. (1980). *Instrumental enrichment: An intervention program for cognitive modifiability*. Baltimore: University Park.
Fleer, M. (2010). *Early learning and development: Cultural-historical concepts in play*. Cambridge: Cambridge University.
Fleer, M. (2011). Technologically constructed childhoods: Moving beyond a reproductive and critical view of curriculum development. *Australasian Journal of Early Childhood, 36*(1), 16–24.
Fleer, M., & Jane, B. (2007). *Science for children* (3rd ed.). Australia: Pearson Education.
Fleer, M., & Richardson, C. (2003). Collective mediated assessment: Moving towards a sociocultural approach to assessing children's learning. *Journal of Australian Research in Early Childhood Education, 10*(1), 41–55.
Fleer, M., & Richardson, C. (2006). *Observing children in early childhood environments: Using a sociocultural approach to assessment*. Canberra: Early Childhood Australia.
Fleer, M., & Richardson, C. (2009). Cultural-historical assessment: Mapping the transformation of understanding. In A. Anning, J. Cullen, & M. Fleer (Eds.), *Early childhood education: Society and culture* (pp. 130–144). UK: Sage.
Fleer, M., Agbenyega, J., Blaise, M., & Peers, C. (2008). *Assessment within the early years: Assumptions, beliefs and practices. A literature review for DECS*. Melbourne: Research node for furthering early childhood research and learning, Faculty of Education, Monash University.
Fleet, A., & Patterson, C. (2011). Seeing assessment as a stepping stone: Thinking the context of the EYLF. *Research in Practice Series, 18*(4).
González Rey, F. (2009). La Significación de Vygotski para la Consideración de lo Afectivo en la Educación: Las Bases para la Cuestión de la Subjetividad. Vygotsky's significance for the consideration of the affective processes in education: The basis for the matter of subjectivity. *Actualidades Investigativas en la Educacion, 9*, 1–24.
Hedges, H. D. (2002). *Subject content knowledge in early childhood curriculum and pedagogy*. Unpublished M. Ed. thesis, Massey University, New Zealand.
Jordan, B. (2009). Scaffolding learning and co-constructing understandings. In A. Anning, J. Cullen, & M. Fleer (Eds.), *Early childhood education: Society and culture* (pp. 39–52). UK: Sage.
Kravtsova, E. E. (2008). *Zone of potential development and subject positioning*. Paper presented at Vygotsky symposium, Monash University, Melbourne.
Lunt, I. (1993). The practice of assessment. In H. Daniels (Ed.), *Charting the agenda. Educational activity after Vygotsky* (pp. 145–170). London: Routledge.
MacDonald, M. (2007). Toward formative assessment: The use of pedagogical documentation in early elementary classrooms. *Early Childhood Research Quarterly, 22*, 232–242.
MacNaughton, G. (2009). Exploring critical constructivist perspectives on children's learning. In A. Anning, J. Cullen, & M. Fleer (Eds.), *Early childhood education: Society and culture* (pp. 53–64). UK: Sage.
NAEYC & NAECS/SDE (National Association for the Education of Young Children & National Association of Early Childhood Specialists in State Departments of Education). (2003). *Joint position statement. Early childhood curriculum, assessment, and program evaluation: Building*

an effective, accountable system in programs for children birth through age eight. Washington, DC: National Association of Early Childhood Specialists in State Departments of Education.

Nuttall, J. (2005). Educators and children learning together: Reflections on the early childhood assessment exemplars. *Early Education, 38*(Spring/Summer), 63–71.

NZ Ministry of Education. (2011). *Assessment for learning*. http://www.educate.ece.govt.nz/learning/curriculumAndLearning/Assessmentforlearning/KeiTuaotePae/Book2.aspx. Accessed 3 July 2011.

Podmore, V. N. (2009). Questioning evaluation quality in early childhood. In A. Anning, J. Cullen, & M. Fleer (Eds.), *Early childhood education: Society and culture* (pp. 158–168). UK: Sage.

Tasmania Department of Education. (2008). *Kindergarten development check*. Hobart: Tasmanian Department of Education.

Tzuriel, D. (2001). *Dynamic assessment of young children*. New York: Kluwer.

Vygotsky, L. S. (1987). *The collected works of L.S. Vygotsky: Vol. I. Problems of general psychology*. R. Rieber & A. Carton (Eds.) (trans: Minick, N.). New York: Plenum.

Vygotsky, L. S. (1994). The problem of environment. In R. Van der Veer & J. Valsiner (Eds.), *The Vygotsky reader* (pp. 338–354). Oxford: Blackwell.

Vygotsky, L. S. (1997). *The collected works of L.S. Vygotsky. Vol. 4. The history of the development of higher mental functions*. New York: Plenum.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*(1), 49–65.

Chapter 13
Classroom Assessment: Making Space for Diversity

Bronwen Cowie

Introduction

Current recommendations to address the decline in student engagement and differentials in student achievement, especially that of students from indigenous, minority and lower-socioeconomic groups, are that school science needs to make a better connection with students' lives out of school so that instruction emerges out of their everyday experiences (Fensham 2009). Classrooms nowadays are characterised by diversity and so this poses a number of challenges for teachers. Students not only come from different ethnic, language and socioeconomic backgrounds but also differ in their familiarity with scientific ways of speaking, thinking and acting and in their orientation to and experiences of classroom learning and interaction (Dweck 2006; Hung and Bell, this volume; Lee and Buxton 2010; Pea and Collins 2008). One of the challenges teachers face is how to access and understand this diversity. Another is how to work with it as a resource that supports all their students to develop the science knowledge and skills they need to participate in society and, simultaneously, to enrich their wider cultural and social identities. In this chapter I suggest that classroom assessment has the potential to help teachers address these challenges.

In this chapter I illustrate the formative potential of assessment (Black, this volume) and assume that classroom assessment encompasses teacher and student decisions and actions about ways forward in teaching and learning that are based on evidence of student learning (Black and Wiliam 2009). Seen this way assessment is a ubiquitous and often invisible aspect of classroom life (Moss 2008). Despite this the pervasiveness of classroom assessment means that it is a crucial influence on student

B. Cowie (✉)
University of Waikato, Gate 1 Knighton Road, Hamilton 3240, New Zealand
e-mail: bcowie@waikato.ac.nz

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_13,

motivation and on how students come to conceptualise what it means to learn and know (Moss et al. 2008). In the science classroom, everyday teacher assessment practices shape what students come to understand science is about and what it means to learn, know and do science. They are a key influence on how students come to conceptualise their own relationship with science—whether or not they are a successful knower, learner and user of science for school purposes and as part of their everyday lives. It also shapes whether or not they see a role for science in their future.

Researchers working to an agenda of 'science for all' (Aikenhead 1996; Barton and Tan 2009; Carlone et al. 2011), along with those working to develop culturally relevant/responsive pedagogies (Gay 2000; Ladson-Billings 1995), have endorsed a similar list of suggestions. They are united in advocating a credit-based view of students and their communities, arguing that diversity is 'a fundamental quality of all human interaction and, crucially, a source of creativity, insight, and learning for teachers, students and researchers alike' (Warren et al. 2001, p. 45). Seen this way the diversity in student experiences, languages and orientations is a resource for rather than an impediment to individual and collective learning. These scholars are insistent that classroom assessment practices need to contribute to student science-related identities *and* affirm and enhance the identities students have forged as part of their involvement in communities outside the classroom.

Researchers interested in the equity and social justice aspects and impacts of assessment have pointed out that whose knowledge is included in the curriculum and legitimated via assessment and what opportunities students have to learn and demonstrate their learning are important and need to be taken into account (Gipps and Murphy 1994; Stobart 2005). For all of these reasons, it is important that classroom assessment encompasses the full range of science learning outcomes, including conceptual, epistemological and social aspects (Duschl 2008) as well as supporting students in learning how to learn and how to use science (Aikenhead et al. 2011). To foster these outcomes it is also important that assessment provides opportunities for all students to demonstrate the breadth of what they know in ways that are meaningful and accessible to them and that also help them move their learning forward.

In the sections that follow, I trace the evolution of research out of the University of Waikato, New Zealand, as a means of reflecting on shifts in thinking about the nature and value of the diversity of knowledges, beliefs and experiences students bring to their learning of science. Starting with the first Learning in Science Project [LISP] in the 1980s, Waikato science education researchers have had a strong tradition of giving careful attention to children's ideas about and explanations for the natural world. In construing these ideas as 'children's science', this early work took a non-deficit view of student ideas and experiences (Osborne and Freyberg 1985). Subsequently, a number of studies have sought to promote science teaching and assessment that is responsive to the breadth in student learning orientations and preferences, cultural knowledges and practices, and everyday and home experiences. Across these studies I illustrate that, over time, science educators at

Waikato have moved from a focus on the diversity in student explanations of natural phenomena to a focus on the breadth of knowledge and expertise students and their communities have to contribute to science learning. At the same time, I note and consider the implication of a shift in the goals for student science learning from the accumulation of knowledge to the development of student identities as learners and users of science. Through an analysis of these shifts, I aim to reconceptualise classroom assessment as a responsive practice that is clearly and inextricably integrated with a culturally responsive approach to pedagogy and interactions in the classroom and with families. While using research from Waikato researchers may seem somewhat restrictive, shifts in the Waikato research are broadly consistent with shifts elsewhere (Bell 2005). But first I provide some contextual detail, which I hope will allow the reader to draw parallels with their own context and interests.

The New Zealand Classroom as a Context

In New Zealand it is fundamental that education honours the Treaty of Waitangi (the founding document for New Zealand as a bicultural nation) and supports Māori students to live as Māori and to participate actively as citizens of the world (Durie 2004). The *New Zealand Curriculum* (Ministry of Education 2007) reiterates this and adds that the classroom curriculum needs to reflect New Zealand's cultural diversity and 'value the histories and traditions of all its people' (p. 9). For New Zealand science teachers, this means that they need to ensure that as students develop their expertise and affiliation with science, their cultural and social identities are affirmed and enriched in both the short *and* the long term. In providing advice for how this might be achieved, the Ministry of Education (2008) document *Ka Hikitia–Managing for Success: The Māori Education Strategy 2008–2012* reiterates the need for schools and teachers to take greater responsibility for learning approaches that engage Māori students more effectively with their own families, communities and other cultural institutions.

Recommendations by Bishop and Glynn (1999) about how to support the learning of Māori students, although firmly grounded in the New Zealand context, resonate with those by scholars of culturally relevant pedagogy and researchers working to an agenda of 'science for all' discussed earlier. Bishop and Glynn report that teachers who show respect for their Māori students, who care about what is happening in their students' lives outside the classroom and who respect the knowledge, experience and expertise students bring into the classroom will be respected in turn by their students. Bishop and Glynn contend that in this environment students learn to make responsible choices, exercise agency and contribute from their own knowledge and expertise. Significant for the discussion in this chapter about how assessment might make a space for diversity, Bishop and Glynn note that these strategies benefit all students.

A Diversity of Explanations for Phenomena

The early LISP work indicated, as has been found around the world, that students come to class with a diversity of ideas about and explanations for the natural world. The assumption underpinning the LISP work was that student explanations 'provide a sensible and coherent understanding of the world from the child's point of view' (Osborne and Gilbert 1980, p. 376). However, the research was framed from within a canonical science perspective, and the conceptions sought were those that relate to canonical science such as energy and force. The research identified children's thinking about Western science only, with views that did fit within a Western science framework labelled as alternative (Fleer 1999). Other worldviews were excluded and positioned in ways that demonstrate that they were not equally valued. The early LISP research was valuable, however, in directing attention to the diversity of science ideas and experiences students bring to class. Pedagogically, it led to recognition that students' existing ideas are tightly held and difficult to change, especially when science explanations conflict with everyday experiences and language use, and that they influence how students make sense of teaching activities so teachers need to take these into account (Bell 2005).

The variations in student interview responses to questions grounded in different instances and events highlighted the need for teachers to attend to the role of context in shaping student responses. In terms of assessment this work has been important in New Zealand in directing attention to the need for ongoing classroom assessment to monitor the sense students are making of and taking from teaching and learning activities (Bell and Cowie 2001). The following vignette illustrates how Grant, a teacher in the LISP (Assessment) project, drew on his knowledge of earlier LISP work Cosgrove (1995) to surface and challenge student ideas about current as part of a unit on electricity.

Vignette 1: Assessment that targets student alternative conceptions
One of the first challenges Grant posed in the class was for students to 'make a bulb glow'. He provided them with a bulb, battery and wires and, as he expected, many students struggled with this task. Then when he asked them why their circuit worked they offered the range of explanations the earlier LISP researchers had identified. Grant then showed the class the four explanations the LISP team had developed and asked them to choose the one that best reflected their view and to explain why they thought this. Next he used the ammeter demonstration detailed in the LISP unit of work to introduce and provide an evidential basis for the science explanation of current flow.

The contribution of the LISP work in helping Grant set up a situation to surface and understand student ideas is illustrated in this example. The example also highlights the important role well-designed assessment tasks can play in encouraging students to talk through their thinking—it is not enough to simply observe the product of activity. The LISP unit on electric current also included suggestions for contexts to help students test out and consolidate new thinking, such as the design of Christmas tree lights that would not fail when one bulb blew. This sequence allowed Grant to encourage students to consider whether their reasoning fitted with the evidence rather than them needing to defer to his authority as a basis for accepting the science view, something that is important when the goal is to develop student learning autonomy (James and Pedder 2006). However, it needs to be acknowledged that such an in-depth approach is not always possible, and nor are students' ideas necessarily amenable to change through the presentation of contradictory empirical evidence—changing one's mind has a social dimension (Driver et al. 1994).

Diversity in the Social Processes of Assessment

The Learning in Science Project (Assessment) (Bell and Cowie 2001), and later the Classroom Interactions in Science and Technology Classroom [InSiTE] study (Cowie et al. 2008), focused directly on formative assessment. Working with 10 teachers of Year 7–10 students (ages 11–14 years) over 2 years and 12 teachers of Year 1–8 students (ages 5–12 years) over 3 years, respectively, these studies sought to understand and enhance the classroom assessment practices involved in teachers recognising and responding to student learning throughout the learning process. In terms of diversity, these studies highlighted the social nature of the assessment process and the variations in student goals and preferences for demonstrating and gaining feedback on what they knew and could do.

Accommodating Diversity in Expressing Understanding and Accessing Feedback

Teachers and students in the LISP and InSiTE studies identified that from their perspective, the most effective formative assessment is accomplished through informal one-on-one or small group teacher-student interactions when students are working through activities. These interactions were said to have the advantage that students were more prepared to reveal what they were unsure of and that feedback was more likely to be co-constructed through conversation. To optimise the possibility of constructive dialogue and feedback, the LISP students stated it was important that teachers and students respected and trusted each other. They were more prepared to disclose their thinking when they trusted their teacher to respond to their questions and concerns in a respectful manner and in language they understood.

Nevertheless, while the majority preference was for dialogue, some students said they liked to be able to sit quietly and read and think about teacher feedback. While students were aware of the challenge for teachers of coming to know the background knowledge and expertise of each and every student in a class, they were emphatic that teacher feedback tailored to take account of this diversity was not only more useful but also affirmed that the teacher understood and cared for *them* and *their* learning and this was important to them, or at least to those students who spoke to me (Cowie 2000). This said, on occasions, the LISP students were quite clear that understanding was not their goal. Sometimes they simply wanted to complete the teacher-assigned task, and on these occasions they wanted feedback that assisted with no more than this. Layered over the variations in student conceptual understandings, this diversity in student goals and preferences adds considerably to the complexity of the challenge teachers face within interactive formative assessment practices.

Accommodating Diversity Through Multiple Opportunities, Modes and Media

The InSiTE study demonstrated the value of providing students with multiple and multimodal opportunities to demonstrate and access feedback on what they know and can do and what they are coming to know. In orchestrating this, teachers accommodated student diversity by catering for differences in student background knowledge and expertise in the ways students were able to express their ideas and by allowing for differences in the pace of student learning. This helped maintain the focus on students' strengths and what they could do, rather than what they could not. If a student could not articulate their learning, they were often able to demonstrate it in other ways as happened with Roger, whose actions spoke louder than words.

Vignette 2: The value of assessment in situated purposeful activity
Roger was a five year old student with English as a second language. He struggled to express his ideas in English and his written expression and ability to draw diagrams appeared very restricted. On the other hand, after carefully observing how his teacher constructed a kite he was able to construct one himself. He diagnosed the problem with his second kite by comparing it with his first kite, which had successfully flown, and with a commercial kite. The next day he brought a book from home that included a sequence of diagrammatic steps on how to construct a box kite. With teacher encouragement he proceeded to make and fly a box kite.

This example illustrates two important aspects to do with providing space for diversity. Firstly, the teacher's talking through and modelling how to build a kite not only signalled the value of careful observation and analysis but also provided multiple entry points into the ideas she was seeking to highlight. It attests to the value of teachers consciously seeding the classroom environment with resources (books, pictures, model kites) that the students can access as aids and stimuli to their thinking. Secondly, as Kelly and Brown (2003) have pointed out, student dialogue and actions in situated, purposeful activity represent a more democratic and productive way of understanding student learning because they evidence what is being constructed by students for their local purposes rather than for a more structured, teacher-directed response. While the teacher had set up the environment, it was Roger who activated the resources as sources of information and feedback. By considering Roger's actions his teacher was able to gauge his developing understanding as well as the extent of his initiative and problem solving skills.

The LISP (Assessment) and InSiTE studies also illustrated the value of activating students as instructional resources for one another (Wiliam 2007). Janet's orchestration of the *Odd One Out* game to develop students' understanding of the categorisation of animal groups in the InSiTE study provides an example of this.

Vignette 3: Making student ideas public and discussable with peers
Janet, an InSiTE teacher who knew about the LISP research on student ideas about living/non-living and animals, began a unit on cicadas with a whole class discussion on the distinction between living and non-living and followed this up with discussions about the distinction between plants and animals and the different animal groups. This aspect of the unit culminated in a sorting activity, which required students to select pictures of four different animals and explain why one of them was 'the odd one out'. Initially Janet introduced the *Odd One Out* game to the class using a set of pictures that were attached by Velcro onto the classroom wall. After this the students played the game before and after school and when they had free time.

The Odd One Out game provided a context in which students were obliged to explain their reasoning and could expect to have it challenged. At the same time it allowed Janet to cater for a range of student interests in, and knowledge of, particular animals and to accommodate student preference to talk through and access feedback on their ideas with some peers and not others and at times of their choosing. The game supported student learning autonomy and at the same time provided Janet with unobtrusive access to student thinking, and she was able to intervene if required.

As can be seen in the previous examples, considerable variation is possible in what can come to serve as productive opportunities for assessment. Even quick quizzes can have a formative function if the teacher uses these as an opportunity to discuss the responses with students. However, a caveat is needed here. If teacher assessment converges too often and too tightly on the ideas she or he intends students to learn, this can limit her or his ability to make sense of and respond to student comments and actions (Torrance and Pryor 1998). This was sometimes the case in Grant's class during the electricity unit described earlier.

Vignette 4: The limitations of a convergent formative assessment focus
When Grant asked, "What does a fuse do?" as part of a quick quiz a student replied, "A fuse is used to light a bomb". Grant did not comment on this reply but instead sought additional responses and when a student suggested fuses are used to break an electrical circuit he outlined how fuses are used to prevent house electrical fires. After the lesson Grant commented that he had noticed that the first student had written that fuses are used to make bombs go when he was circulating the class and that he had been puzzled by this. When I suggested that the student might have been thinking about a bomb from an action movie he acknowledged this was a viable explanation but that his frame of reference had been on science and so, in the moment, he had not been able to make sense of, nor build on, the student's answer:

Oh, of course, fuse. Oh, yes. I never of thought of that. Yes, it's a different sort of fuse, isn't it? A bomb fuse or a firework fuse. I had a mind set on electrical fuses although I suppose I used to use fuses and fuse boxes to set off flashes for stage shows years ago. (Cowie 2000, p. 131)

On this occasion Grant missed an opportunity to help students make a conceptual link between the functions of fuses in different contexts. On the other hand, his nomination of the first student was based on his close attention to student ideas as they were recorded in response to the quiz questions he had posed. His follow up to the suggestion that a fuse is used to break an electrical circuit by explaining the value of this function was representative of his concern to help students link what they learned in science class to their everyday lives. Indeed, of all the teachers I have worked with, he spent the most time sourcing and developing resources that illustrated ideas in everyday contexts. Moreover, when questioned, Grant immediately made further links with this alternative version of a fuse. Unfortunately his frequent convergent focus and apparent lack of immediate responsive action was interpreted by a number of students as a lack of interest in them.

Although Grant put considerable effort into planning and preparation, his efforts were only really appreciated by students with a strong background in electronics. These students completed the assigned tasks very quickly and then engaged Grant in exploratory conversations about possible next steps. The four girls I spoke with who had limited relevant prior experience commented that most of their interactions with Grant were around getting tasks done and this often led to their feeling like 'robots' in the sense of not being expected or required to think. As a result of this experience, they were revisiting their decision to continue with science. Here, we have an instance of short-term gain in terms of task completion at the expense of longer-term engagement, which is not desirable given the current goals of science education.

In sum, a focus on assessment as a social process highlighted the need for flexibility in teacher assessment practices if they were to both accommodate and leverage the possibilities within the various ways students engaged with tasks, the diversity of learning goals they pursued and the kinds of interaction they preferred.

Diversity in Student Funds of Knowledge

The University of Waikato's Quality Teaching Research and Development [QTR&D] (Glynn et al. 2008, 2010) and Culturally Responsive Pedagogy and Assessment [CRP&A] (Cowie and Otrel-Cass 2011; Cowie et al. 2011; Parkinson et al. 2011) studies represent a shift to consider more directly the breadth of cultural and everyday knowledges and skills students bring to the classroom and to the way science learning intersects with and can either enrich or undermine students' sense of who they are and who they want to be and become. As noted earlier, it is fundamental to education in New Zealand that it support Māori students to succeed as Māori and as citizens of the world, which in the context of this chapter I am interpreting as helping Māori students to succeed in science without requiring that they leave much of who they are and what they value at the classroom door.

In relation to assessment, both the above studies were underpinned by a view of learning as the transformation of identity (Wenger 1998) and premised on the understanding that part of making learning visible and attributing value assessment 'makes up people' (Stobart 2008, p. 6). Or as Moss (2008, p. 239) explains it, 'assessment offers learners identities and positions and presupposes aspects of their identities in the situations in which they are assessed'. In this statement Moss acknowledges the role of assessment in shaping identity and also that identities are not fixed and unitary but fluid, situated and multiple. In thinking about the implications of this, Shepard (2001) notes that a commitment to equal opportunity for diverse learners means providing ways into the curriculum that are consistent with the language and interaction patterns students have developed in their homes and communities. In this section I provide evidence of how teachers in the QTR&D and CRP&A projects invited in discussed and built on the many and varied ideas and experiences students had to contribute to the science curriculum. The examples aim to illustrate the nuanced interactions and impacts of classroom assessment with student funds of knowledge and science-related classroom identities.

Valuing Student and Community Funds of Cultural and Everyday Knowledge

Gonzalez and Moll (2002) coined the phrase 'funds of knowledge' [FOK] to describe the historical and cultural knowledge, strategies and resources that families and communities have acquired through their life experiences and cultural practices. In science, Barton and Tan (2009) have illustrated the value of teachers drawing on the student family, community, peer group and popular culture funds of knowledge. Their study also highlighted the need for teachers to co-create with students a classroom culture where students felt safe to share their funds of knowledge, confident in the expectation they would be respected even if their ideas were discussed and debated. In the QTR&D and CRP&A studies, the teachers used a range of strategies to seek out and invite student and community funds of knowledge into the classroom science curriculum as integral to classroom assessment (Fleer and Quinones 2009).

In the CRP&A study, Asri chose to study adaptation in the context of sea life in part because she knew most of her students and their families would have been to the seashore. She invited their FOK into the classroom using a 'home learning' task. In Asri's experience, asking the students what their families knew about the sea and having family members come to class to talk about their experiences 'shows that what they already know is valued at school' (Parkinson et al. 2011, p. 2), which was important to her. She noted that students are more confident when sharing home learning because they have had support from someone at home.

Vignette 5: Deliberately seeking out and inviting FOK into the curriculum
To complement and extend an initial brainstorm activity Asri assigned a 'home learning task' that required students to talk with their families about their own experiences with the sea and to invite their family members to class on the next Friday to share a myth, legend or experience, and to support them with their own sharing. On the next Friday the children responded very positively to a father's story about being a fisherman and a mother sharing about holidays spent exploring the seaside.

As a consequence of this task, Asri found out that most of her students' experiences were to do with boogie boarding or swimming. Very few had really examined sea creatures. To accommodate this she downloaded a number of additional videos and images of different jellyfish and flatfish and accessed more books about sea creatures. She used these resources to ground and focus whole class discussions on adaptations. Knowing her students often went to the beach over the weekend, she also provided time for students to share any experiences of the sea at the beginning of her Monday lessons.

Teachers in both studies reported they struggled to 'let go' and move towards a more culturally responsive approach that sought to incorporate student cultural knowledges and that allowed students to contribute to and share some of the responsibility for their own learning. However, they soon found that when they handed over some responsibility, their students readily made meaningful connections between their home, their whānau (nuclear and extended family) and their school learning. When students realised these ideas and experiences would be welcomed and treated with respect, they proved more than willing to seek out and contribute them. In one example, when Jane, QTR&D teacher, wondered which native plants had medicinal properties, a student said, 'We could ask Tare's nana since he said that they used rongoa [medicinal] plants in his family. We could read books or ask people to come and talk to us'. In Tina's class (Tina was another teacher in the QTR&D study), a student responded to her question about where the class could get information about local landscape features by saying, 'We could go on the net or we could ask our kaumatua [elder]'.

These student comments are typical of the responses the teachers experienced and indicate the breadth of sources of knowledge students are aware of and that teachers might employ to input new information and to provide students with feedback. Significantly, in Tina's class, her students found that different hapu (subtribes) had different cultural stories about the same maunga (mountain). This paved the way for Tina to introduce that there were various ways to explain phenomena and for her to articulate the students' knowledge alongside that of the 'English science story'. As part of this process, she took the time to discuss when and where different explanations might be more appropriate and when and why scientific ways of thinking and working are useful and how they came to be valued highly (Moje 2007).

Knowing when and why science explanations are powerful is important for all students, but when students bring strongly held and very different home/community understandings to their science learning, this focus allows space for them to think about when and why other ways of thinking might be useful and valued. McKinley and Stewart (2009), writing about the New Zealand context, advise that this requires teachers to hold indigenous knowledge in tension with science knowledge and use this tension to catalyse insights into the nature of science. As Driver et al. (1994) point out, scientists understand perfectly well what is meant when they are told 'Shut the door and keep the cold out' and that they use different conceptions of 'solid' depending on the task at hand. In terms of student identity development, holding multiple knowledges fits well with Wenger's (1998) assertion that 'To be able to have effects on the world, students must learn to find ways of coordinating multiple perspectives' (p. 274). He goes on to describe that a central role of education is to help students learn to straddle boundaries and find ways of being in the world that can encompass multiple, conflicting perspectives in the course of addressing significant issues. This goal is embedded in the Māori aspirations that their children develop as Māori and as citizens of the world and in the aims of culturally relevant/responsive pedagogy elsewhere (see Ladson-Billings 1995).

The following vignette details the flow on consequences of Taylor, one of Asri's students, sharing the knowledge she had gained from interviewing her uncle as part of a home learning task.

Vignette 6: Funds of knowledge enriching the science curriculum
Taylor reported that her uncle was a kaitiaki [guardian] for the local harbour. She explained that "his job is important because we need to protect and preserve our seafood, our beaches and our wild life for the future generations". In subsequent lessons Asri described her own commitment to conservation and affirmed students who contributed ideas to do with conservation. She encouraged discussion of a newspaper article on people taking undersized fish and whether or not a paua shell a student brought to class was under size. When a student shared the Māori cultural practice of returning the first fish caught to the sea a number of students from first home countries other than New Zealand shared the cultural practices from these countries. Asri asked students to bring to class cultural and community myths about animals that might explain their adaptive features.

Asri's action to include a focus on conservation in the science curriculum enriched the unit by further linking it to students' experiences and served to reinforce that students sharing what they knew was valuable because others could learn from it. The student stories of what to do with the first fish caught indicated a shared concern with aspects of conservation, which Asri was able to pursue within a science frame. By prompting the children to consider the adaptive feature that was the focus of a particular myth, Asri drew her students' attention to myths as explanatory activities, albeit without the basis in observation and systematic investigation that characterise science knowledge. Just as importantly, sharing and building on their funds of knowledge allowed Asri and her students to get to know each other as people who are enmeshed in multiple spheres of activity, not just as a teacher or as a student, thereby helping to build bridges between science learning and the other aspects of students' lives.

The teachers in both studies sought to reflect and affirm their Māori students' identities by using Māori legends to introduce Māori cultural knowledge and by using Māori words, names and/or images in classroom displays, lesson plans and learning materials. In some instances students learned and were assessed on their knowledge of science-related Māori words. This was the case in Jude's class where the worksheet students used to plan their weather forecast to the school included both English and Māori words. Jude encouraged the students to use Māori in their presentation, and a number of groups began with a Māori greeting. In Asri's class Tama, a Māori boy who was fluent in *te reo* (the Māori language), offered to read the Māori version of a story about whales alongside Asri's reading of the story in English. Tama's action and Asri's acceptance of his expertise is congruent with the principle of *ako* (to learn, teach, instruct), which construes the roles of teacher and learner as fluid and interchangeable so that both parties benefit and learn from each other.

The other Māori boys in the class were overtly supportive of Tama during the reading and more visibly engaged after Asri's affirmation of Tama's expertise in *te reo*. These examples, although they do not relate directly to science ideas and practices, illustrate the value of teachers taking a holistic view of what students have to offer and what is taken to be of value within science lessons as a means of engaging the students with science. Later in the unit, these boys enthusiastically contributed their experiences of floundering and took a lead role in the discussion of how flounder are adapted to bottom feeding as part of the explicit science focus for the unit.

Across the QTR&D and CRP&A classes, we were reminded of the need to refrain from stereotyping individuals and groups. Māori students do not always want to work in groups. In addition, parents are just as likely to talk with their children about Western science views as Māori cultural views. For example, prior to the teacher introducing the scientific explanation of the life cycle of a plant, a student asked her whānau about the Māori explanation for how forests are created. Her mother was unaware of any Māori legends and suggested the student ask her uncle who worked in a native tree nursery (Glynn et al. 2008). His reply to her was, 'I don't know the Māori story, but I can tell you the scientific way'. This further highlights the need for teachers to know each of their students, taking into account the whole of their lives and their individual strengths, interests, needs and preferences.

Assessment to Engage and Empower Families Within the Curriculum

The QTR&D and CRP&A studies highlighted, as Upadhyay (2009) has pointed out, that teaching science to empower students is not only about fostering student classroom engagement. There is value in providing a space for community and family engagement in the curriculum and assessment. Indeed, this is an imperative under *Ka Hikitia*. Engaging with families in this way was important to all of the QTR&D and CRP&A teachers, but it was a particular focus for Marion, who was a new entrant teacher (all of her students were aged 5 years).

Vignette 7: Assessment empowering students and their families and whānau
In Marion's class parents played an important role in supporting student learning about tuatara (a reptile endemic to New Zealand). They took the children to a tuatara sanctuary and guided their observations and sketching of live tuatara while there. They helped the children prepare a *PowerPoint* presentation about what they had learned for a whole school assembly and assisted with the construction of a clay model of a tuatara. Each of these activities was an occasion for rich dialogue about tuatara features and habitat but the clay modeling proved to be a particularly effective forum for parental curricular involvement. A number of parents offered their help and one dad, who had not been in to help before, was just as keen as his son to take the finished product home. With parent-helper guidance that encouraged them to observe carefully photographs of tuatara each student produced a credible model of a tuatara. Student comments indicated they valued the involvement of their siblings and family members.

In this vignette, student opportunities to learn and to gain feedback on their learning were embedded in their relationships with families, with other children's families, with their peers and with the conservator at the sanctuary. They were not restricted to teacher-student or student-student interactions alone. Involvement with a wide range of adults with whom students can interact to test out and gain feedback on their ideas is of value in classroom assessment that seeks to provide a space for student diversity because it increases the likelihood of interaction that is timely and responsive to students' needs and interests. Informal conversations with parents indicated that an early positive experience with their children's science learning had established the beginnings of a productive form of participation for adults in their children's science learning—a relationship that, hopefully, will help sustain student interest in science as they move up the school year levels.

In another example, Frank's Year two students (6-year-olds) were learning about the classification of animals. Frank was aware of the LISP work on living/nonliving and animal classification and encouraged the students to find out what they could about an animal of their choosing from their koro (grandfather figure), kuia (grandmother figure) and other whānau members, as well as from books and other resources. In this way students came to see that their whānau knowledge was valued and taken seriously by the teacher. Parent comments as they dropped off or picked up their children indicated that in many cases the whole family had become interested in the animal their child was studying. Diane's mother commented:

> Since you have started this study, Diane has not stopped talking about the kiwi. She has hounded me every day to go to the library. Last night I took her into the town library and we were there 45 minutes while she chose books to bring home to read to her father. Then this morning she wanted me to come and show you a book about the kiwi. (Glynn et al., 2008, p. 32)

Social Purposes and Consequences for the Demonstration of Knowledge

The different social purposes and consequences for the demonstration of knowledge are worthy of deliberate consideration when the goal is to open up a space for meaningful assessment. With no prompting from the research team, each of the ten teachers in the QTR&D project designed a summative task that involved students sharing what they had learned with others. Knowing that they would be sharing what they learned with an audience beyond the teacher added purpose and meaning to the learning and its demonstration. The students prepared displays for local community groups, visited more junior classes to share what they had learned and presented an overview of their learning at whole school assemblies. The students in Jude's class who presented a weather report to the school were able to represent their knowledge in a culturally appropriate manner, through korero (talk). Jude was careful to provide them with a number of opportunities to practise gathering, analysing and interpreting temperature, rainfall and wind direction and speed data and to gain feedback on their presentation. This process was culturally appropriate for the Māori students because it was important that their oral report was informative and polished. Parents responded very positively to videos of the presentations at a parent-teacher-student report evening.

The positive reinforcement students received from these different audiences added to students' sense that their science learning was of value and to their sense of being someone who could know and do science. Students from a number of classes reported that their class had gained a school-wide reputation for being interested in and knowledgeable in science. One student was affronted when the class found that the principal had an unused digital microscope in his office. The student couldn't understand how this could be the case when his class was known to be actively engaged in science investigations. Alongside this, a number of students took obvious pleasure in being able to contribute to the science learning of others.

Concluding Thoughts

In this chapter I have illustrated how teacher classroom pedagogy and assessment practice might open up a space for diversity. The diversity in student explanations for natural phenomena is an aspect of diversity that all teachers need to consider. By illustrating the use of the LISP work on children's science/alternative conceptions, I hope I have made a case for its efficacy in helping teachers take account of this aspect of diversity in their classrooms. Science teachers need to be aware of and take into account differences in student explanations for natural phenomena and the sense they are making of and taking from tasks if they are to guide them towards more science-oriented understandings. Similarly, teachers need to be aware of differences in students' preferences for interaction and feedback as well as their abilities and inclinations to represent or express what they know and can do if they are to access what the students know and provide useful feedback.

It is fundamental to my argument, however, that culturally responsive classroom assessment is inextricably integrated with (culturally) responsive pedagogy and that it seeks out, invites in and incorporates student and community funds of knowledge into the curriculum. I concur with Bang and Medin (2010) who state:

> We believe that central to the future of science and science education is to understand, support, and leverage the ways in which diversity—of people, practices, languages, meaning, knowing, epistemologies, goals, values, and the like…in learning environments and professional practice are an asset and expand the possibilities for human knowing and meaning. (p. 1009)

At the same time, the nature and focus of classroom assessment that supports students' successful participation in science needs to enrich and empower students as members of the many other cultural and community groups they encounter over the course of the day. In the early years of students' science education, I believe that it is particularly important that assessment establish productive forms of participation for family, whānau and community members so that they are able to encourage student participation over the long term.

In thinking about the wider implications for classroom assessment, it is important to remember that diversity also plays out in relation to the diverse funds of knowledge and relationships that teachers have to bring to bear on the conduct of assessment as a social and cultural practice that shapes identities. Teacher identities are at stake, and in the process of ongoing production, during classroom assessment. Teachers implementing pedagogy and assessment that makes space for diversity requires them to know their students and their students' communities well and to have a breadth and a depth of science content and pedagogical content knowledge. They need to appreciate the critical intersections between science, school science, and the cultural, community and home backgrounds of their students (Southerland et al. 2007). This is a very demanding task, one that has increased in complexity as classes have become more diverse. Dutro et al. (2008) make the point that there is a difference between culturally responsive teaching when the students in the class are relatively homogeneous and teaching where the class is highly diverse and 'many identities come into contact and conversation' (p. 271). The examples in this chapter have only touched on this aspect and its wider implications. This is an important area for further investigation in the ongoing development of classroom assessment that makes a space for the various forms of diversity that teachers encounter in contemporary classrooms.

Acknowledgements The InSiTE and CRP&A projects were funded by the Teaching and Learning Research Initiative. All examples were drawn from the reports and papers listed when the different projects were introduced.

References

Aikenhead, G. S. (1996). Science education: Border crossing into the subculture of science. *Studies in Science Education, 27*, 1–52.

Aikenhead, G., Orpwood, G., & Fensham, P. (2011). Scientific literacy for a knowledge society. In C. Linder, L. Ostman, D. Roberts, P. Wickman, G. Erickson, & A. MacKinnon (Eds.), *Exploring the landscape of scientific literacy* (pp. 28–44). New York: Routledge.

Bang, M., & Medin, D. (2010). Cultural processes in science education: Supporting the navigation of multiple epistemologies. *Science Education, 94*(6), 1008–1026.

Barton, A., & Tan, E. (2009). Funds of knowledge and discourses and hybrid space. *Journal of Research in Science Teaching, 46*(1), 50–73.

Bell, B. (2005). *Learning in science: The Waikato research experience*. London: Routledge.

Bell, B., & Cowie, B. (2001). *Formative assessment and science education*. Dordrecht: Kluwer.

Bishop, R., & Glynn, T. (1999). *Culture counts: Changing power relations in education*. Palmerston North: Dunmore Press.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5–31.

Carlone, H., Haun-Frank, J., & Webb, A. (2011). Assessing equity beyond knowledge- and skills-based outcomes: A comparative ethnography of two fourth-grade reform-based science classrooms. *Journal of Research in Science Teaching, 48*(5), 459–485.

Cosgrove, M. (1995). A study in science-in-the-making as students generate an analogy for electricity. *International Journal of Science Education, 17*(3), 295–310.

Cowie, B. (2000). *Formative assessment in science classrooms*. Unpublished doctoral thesis, The University of Waikato.

Cowie, B., & Otrel-Cass, K. (2011). Exploring the value of 'horizontal' learning in early years science classrooms. *Early Years, 31*(3), 285–295.

Cowie, B., Moreland, J., Jones, A., & Otrel-Cass, K. (2008). *The Classroom InSiTE Project: Understanding classroom interactions to enhance teaching and learning in science and technology in Years 1–8*. Wellington: Teaching and Learning Research Initiative.

Cowie, B., Otrel-Cass, K., Glynn, T., Kara, H., Anderson, M., Doyle, J., Parkinson, A., & Te Kiri C. (2011). *Culturally responsive pedagogy and assessment in primary science classrooms: Whakamana tamariki*. Summary report. Wellington: Teaching and Learning Research Initiative.

Driver, D., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*(7), 5–12.

Durie, M. (2004). *Mauri Ora*. Melbourne: Oxford University.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education, 32*, 268–291.

Dutro, E., Kazemi, E., Balf, R., & Lin, Y. (2008). "What are you and where are you from?" Race, identity, and the vicissitudes of cultural relevance. *Urban Education, 43*(3), 269–298.

Dweck, C. (2006). *The mindset: The new psychology of success*. New York: Random House.

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

Fleer, M. (1999). Children's alternative views: Alternative to what? *International Journal of Science Education, 21*(2), 119–135.

Fleer, M., & Quinones, G. (2009). Assessment of children's technological funds of knowledge as embedded community practices. In A. Jones & M. de Vries (Eds.), *International handbook of research and development in technology education* (pp. 579–597). Rotterdam, The Netherlands: Sense.

Gay, G. (2000). *Culturally responsive teaching: Theory, research, and practice*. New York: Teachers College.

Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckingham: Open University.

Glynn, T., Cowie, B., & Otrel-Cass, K. (2008). *Quality Teaching Research and Development Science Hub (Waikato): Connecting New Zealand teachers of science with their Maori students. Report to Ministry of Education*. Hamilton: Wilf Malcolm Institute of Educational Research, University of Waikato.

Glynn, T., Cowie, B., Otrel-Cass, K., & Macfarlane, A. (2010). Culturally responsive pedagogy: Connecting New Zealand teachers of science with their Māori students. *Australian Journal of Indigenous Education, 39*, 118–127.

González, N., & Moll, L. (2002). Cruzando el puente: Building bridges to funds of knowledge. *Journal of Educational Policy, 16*(4), 623–641.

James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *Curriculum Journal, 17*(2), 109–138.

Kelly, G., & Brown, C. (2003). Communicative demands of learning science through technological design: Third grade students' construction of solar energy devices. *Linguistics and Education, 13*(4), 483–532.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Education Research Journal, 35*, 465–491.

Lee, O., & Buxton, C. (2010). *Diversity and equity in science education: Research, policy and practice*. New York: Teachers College.

McKinley, L., & Stewart, G. (2009). Falling into place. Indigenous science education research in the Pacific. In S. Ritchie (Ed.), *The world of science education: Handbook of research in Australasia* (pp. 49–65). Rotterdam: Sense.

Ministry of Education. (2007). *New Zealand curriculum*. Wellington: Learning Media.

Ministry of Education. (2008). *Ka Hikitia—Managing for success: The Mäori education strategy 2008–2012*. Wellington: Ministry of Education.

Moje, E. (2007). Developing socially just subject-matter instruction: A review of the literature on disciplinary literacy teaching. *Review of Research in Education, 31*, 1–44.

Moss, P. (2008). Sociocultural implications for assessment I: Classroom assessment. In P. Moss, D. Pullin, E. Haertel, J. Gee, & L. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 222–258). New York: Cambridge University.
Moss, P., Pullin, D., Haertel, E., Gee, J., & Young, L. (Eds.). (2008). *Assessment, equity, and opportunity to learn*. New York: Cambridge University.
Osborne, R., & Freyberg, P. (1985). *Learning in science: The implications of children's science*. Auckland: Heinemann.
Osborne, R., & Gilbert, J. (1980). A technique for exploring students' world. *Physics Education, 15*, 376–379.
Parkinson, A., Doyle, J., Cowie, B., Otrel-Cass, K., & Glynn, T. (2011). Engaging whānau with children's science learning. *Set: Research Information for Educational Research, 1*, 3–9.
Pea, R., & Collins, A. (2008). Learning how to do science education: Four waves of reform. In Y. Kali, M. Linn, & J. Roseman (Eds.), *Designing coherent science education: Implications for curriculum, instruction, and policy* (pp. 3–12). New York: Teachers College.
Shepard, L. (2001). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.
Southerland, S., Smith, L., Sowell, S., & Kittleson, J. (2007). Resisting unlearning: Understanding science education's response to the United States' national accountability movement. *Review of Research in Education, 31*, 45–72.
Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education, 12*(3), 275–287.
Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. London: Routledge.
Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Maidenhead: Open University.
Upadhyay, B. (2009). Teaching science for empowerment in an urban classroom: A case study of a Hmong teacher. *Equity & Excellence in Education, 42*(2), 217–232.
Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A., & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The logic of everyday sense-making. *Journal of Research in Science Teaching, 38*(5), 529–552.
Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. New York: Cambridge University.
Wiliam, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1051–1098). Greenwich: Information Age.

Chapter 14
Formative Assessment as a Cultural Practice: The Use of Written Formative Assessment in Samoan Science Classrooms

Desmond Lee Hang and Beverley Bell

Introduction

As formative assessment is known to improve learning outcomes (Black and Wiliam 1998; Hattie and Timperley 2007), it is a form of assessment that is increasingly being used internationally and included in the curriculum and assessment policies of governments, for example, in Samoa (Department of Education 1995) and New Zealand (New Zealand Ministry of Education 2007). However, there has been slow progress in the use of formative assessment in Samoan secondary schools due to the predominance of summative assessment for ranking in Samoan schools which reflects the hierarchical nature of Samoan society (Pereira 2005), the popularity of high-stakes examinations in Samoan society (Pereira 2005; Pongi 2004), the general lack of knowledge of teachers with respect to the international literature on formative assessment due to access issues (M. Matau, pers. comm.) and the lack of professional development in this area (Lee Hang 2011). Another reason for the slow progress in Samoan secondary of teachers using formative assessment is that of *le-tautala*, or silence, in the classroom. The cultural practice of *le-tautala* by students in Samoan classrooms is documented in detail in Lee Hang (2011). In this chapter, we will report on the research exploring written worksheets as a way to undertake formative assessment in Samoan secondary science classrooms, given *le-tautala,* and hence a culturally appropriate formative assessment practice for Samoan classrooms.

Samoa (formerly known as Western Samoa) is an independent Pacific island nation state situated near the equator and 2,900 km north-east of New Zealand, with a total land area of 2,934 sq km and a population of 180,741 in the 2006 census. The

D. Lee Hang (✉)
Faculty of Science, National University of Samoa, P.O. Box 1622, Apia SAMOA
e-mail: d.leehang@nus.edu.ws

B. Bell
University of Waikato, Gate 1 Knighton Road, Hamilton 3240, New Zealand
e-mail: beebell@waikato.ac.nz

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_14,

majority of the population speak Samoan as their first language, with Samoan and English being the official languages of government, education, trade and commerce. In the twentieth century, Samoa was a German protectorate until the outbreak of World War I. Subsequently, New Zealand administered Western Samoa under the auspices of the League of Nations and then as a UN trusteeship until independence was regained in 1962.

In the last New Zealand (2006) census, Samoans were the largest Pacific ethnic group in New Zealand, making up 131,100 or 49 % of New Zealand's Pacific population (265,974) in a total New Zealand population of 4,027,947 (New Zealand Statistics Department 2011). Samoans also live and work in Australia, with 39,992 recorded in the 2006 census (Australian Department of Foreign Affairs and Trade 2011). There are also Samoans living on the West Coast of the USA. Hence, there are as many Samoans living away from Samoa as in Samoa. Although this research was undertaken in Samoa, with Samoan teachers and students, it is of interest to New Zealand, Australian and US teachers of Samoan students.

This chapter is co-authored by two people with different cultural backgrounds. Des Lee Hang is Samoan born and lives in Samoa, working in initial teacher education of science secondary teachers who are also Samoan and will teach Samoan students in Samoa. He has lived in New Zealand for extended periods of time for educational purposes. Beverley Bell is sixth-generation New Zealand born, with a British ancestry, and works in initial teacher education of secondary teachers of many cultures in New Zealand: Pākeha (European New Zealanders), Māori (indigenous New Zealanders), Pasifika, Indian, Taiwanese, British, South African and others. To add to this complexity, the intended readership of this chapter is multiple: Samoan teachers teaching Samoan-born students in Samoa, New Zealand teachers teaching students in New Zealand who identify as Samoan (they may be New Zealand or Samoan born) and teachers in other countries, for example, the USA and Australia, who teach students who identify as Samoan.

The main part of this chapter is written by Des based on the views expressed to him by the Samoan teacher participants during the course of his doctoral research, with interpretation and further elaboration based on Samoan cultural practices he has learned from elders. In doing so, Samoan culture is described, documented in written form and communicated to Samoan teachers within the Samoan culture. Des' writing is about teaching as a cultural practice within his culture, that is, the Samoan culture. Other parts are written by Beverley to link the research findings to the concept of culturally responsive teaching, which is a cross-cultural concept. The purpose is to communicate the findings to non-Samoan teachers of Samoan children, for example, Pākeha teachers in mainstream New Zealand schools and classrooms, teaching students who identify as Samoan.

Le-Tautala and Formative Assessment

In the Samoan culture, *le-tautala* is based on the cultural value that young people do not orally question or answer their elders in public or in front of the whole class, as to do so would be considered disrespectful (Lee Hang 2011; Moli 1993). Whereas this practice of questioning of elders and replying to elders in public may be more acceptable in Western culture, it is not practised in many Pacific cultures, for example, in Samoa. In Samoan culture, the person who is valued is someone who speaks little, and when they do speak, it is in thoughtful and measured terms. As Des (the first author) noted:

> My current concern came about as a result of what I experienced while teaching pre-service science teacher classes at the National University of Samoa. One of the main frustrations that I found while teaching pre-service science teachers was not being able to solicit verbal replies from teacher education students during classes. My own questions were often met with silence. It has been frustrating not being able to have the student and teacher formative interactions in classes which enhance student learning. (Lee Hang 2011, p. 15)

Using the cultural practice of *le-tautala* to explain Des' comment, we can theorise that the students were practising their culture of silence with a respected older person in the university and community. At the tertiary level, there is also a silence due to the social cultural impact of making a mistake in English at this level, which would last for generations to come (it would be the subject of social ridicule, *pona*, and mockery—*tausuaga*). So, there is silence by the students in terms of addressing the teacher's question, but you can still hear murmurings of students to each other.

Le-tautala has many nuances to its meaning, with 16 having been reported (Lee Hang 2011). One meaning is that of *va-fealoa'i,* sacred relational space. *Va-fealoa'i* refers to the protocols of respect that are generally observed to maintain the *va* (space) between Samoan people:

> Va is the space between, the between-ness, not empty space, not space that separates but space that relates, that holds separate entities and things together in the Unity-that-is-All, the space giving meaning to things. The meanings change as the relationships/ the contexts change. (Wendt 1996, p. 18)

This space between people is occupied by a shared understanding of culture and relationships. It is governed by the values of respect, loyalty and love for familial ties and the meanings that each relationship allocates to such a space. The following quote from one of Des' research participants highlights the importance of *va-fealoa'i* and its potential to assist teachers in classroom teaching:

> *Ia o Samoa, because e mo'i a i tu ma aga fa'aSamoa, ia ae e tatau ona tatou faatinoina le va-fealoa'i ma tamaiti. Ia e fa'aaoga ai le va fealoa'i o matua ma fanau. Aua e leai se eseesega o le ta'uina o faiaoga o matua faaleaoaoga lea o le fanau. So, e tatau a la ona iai pea se taua o le va fealoa'i o faiaoga ma tamaiti ina ia mafai ai fo'i ona draw le attention i se lesona a?*
>
> [In Samoa because of our culture or way of life, we need to observe the cultural practice of va-fealoa'i between elders or parents and children. Because there is not much difference between parents' role at home and that of the teachers in schools, therefore emphasis should be placed on this cultural relationship to enable teachers to draw the pupils' attention to a lesson]. (Pre-service secondary science teacher)

Within this *va*, or sacred relational space, is practised *faaaloalo* (respect for elders), *faalogo ma usita'i i le matua* (having to listen to elders/teacher) and *le taliupua* (not answering back). As a science teacher educator commented to Des:

> Yeah I think … because in our culture the children are taught not to answer back to the elders. So in the Samoan classroom we see a lot of students not asking questions …, even though if they don't understand or don't know (anything) of what is discussed. It's like they don't want to ask questions, I mean they are not used to that questioning process because of the way they have been brought up at home. (Science teacher educator)

The resultant silence in the classroom, *le-tautala*, may also communicate *matamuli* (shyness), *leiloa le tali* (not knowing the answer), *fefe i sese* (fear of making mistakes and of being mocked), *ma* (shame and embarrassment), *amanai'a pe'a pasi* (wanting to gain respect that comes when you are doing well at school). *tamali'iaga* (family and personal pride), *mana'o e tali fa'atasi mai le vasega* (preferring the collective choral response rather than individual responses), *mana'o e fa'aaoga le gagana Samoa* (preferring to use the Samoan language rather than the English of the classroom), the *eseesega o le fale ma le aoga* (home and school differences), *le nu'u e sau ai* (your village of origin, whether it is rural or urban), *tuatuagia i feau ma tiute i le fale* (whether you have been doing many chores at home and are tired), *that o fanau matutua e vaaia latou tei laiti i le fale* (the older children looking after their younger siblings at home and therefore not having time to do the homework) and *tuatuagia i fa'alavelave* (disrupted by or being encumbered in family obligations) (Lee Hang 2011). These cultural practices for communication tend to be more pronounced in rural than urban schools in Samoa.

Cultural values also underlie the practices of formative assessment in that there needs to be a trusting relationship between teacher and students before students will disclose what they know and don't know in order for teacher feedback and feedforward to be given (Cowie 2000). Within this, what counts as trustworthy is culturally determined. In the Samoan classroom, a trusting teacher-student relationship is one in which both teachers and students understand that there is no undue pressure within the teacher-student relationship to talk (question or answer) in front of the whole class. To be invited to talk might be considered unexpected, culturally inappropriate and disrespectful to the status of the teacher as a knowledgeable elder. The Samoan students' cultural practice in the classroom, when the teacher is talking to the whole class, is to be silent as a way of communicating respect. Hence, Samoan classrooms are typified by the teacher talking to the whole class, for most of the lesson time, with Samoan children being largely nonverbal (Day 1981), unquestioning (Moli 1993) and silent (Tanielu 2001) towards the teacher in the classroom. Teacher-directed and teacher-dominated teaching is the norm. There is little oral interaction between teacher and student, making it nearly impossible to do oral planned formative assessment (Bell and Cowie 2001), also called formal formative assessment (Furtak and Ruiz-Primo 2008), or interactive formative assessment (Bell and Cowie 2001), also called informal formative assessment (Furtak and Ruiz-Primo 2008).

Much of the recent literature on formative assessment is contextualised in 'Western' classroom settings, and the teacher-student relationship tends to be based on Western values. Oral formative assessment is a culturally appropriate form of formative assessment in, for example, New Zealand mainstream classrooms. But students may not disclose what they do not know for the reasons of not wishing to be seen as wrong or lacking in learning and knowledge, especially if they are then put down or embarrassed by the teacher or other students. Their silence is seen as reticence. For formative assessment to be possible, this reticence needs to be minimised if oral feedback and feedforward is to be given by the teacher. In addition, the giving of feedback and feedforward as a part of formative assessment can build or destroy a sense of trust in the teacher-student relationship. Whether the feedback is given in front of class peers in a whole class discussion, given within a conversation between teacher and student or given as written feedback to each student individually may impact on the levels of trust in the relationship. Without the trust, the eliciting of student thinking is difficult if not impossible.

In doing formative assessment, teachers wish to be able to trust the validity of the thinking being elicited from the students. Is the thinking elicited a valid picture of the student's thinking, especially with respect to whether the student has learned the intended goals? There is debate over how we as teachers elicit students' thinking and respond with feedback and feedforward to improve student learning, not just in terms of the questions asked but also in the mode of the responses: written or oral, in planned formative assessment or interactive formative assessment (Bell and Cowie 2001), and done individually with students or in whole class discussion.

The use of the written mode to elicit student thinking was researched by Furtak and Ruiz-Primo (2008), who concluded that 'students' below-level conceptions are more likely to be shared in writing as compared to discussion' (p. 820). It is important to elicit the below-level conceptions so that feedback and feedforward can be given to improve learning to the intended level or goal. But giving written feedback and feedforward is seen as time consuming and delayed beyond its useful time frame (Furtak and Ruiz-Primo 2008).

Students wish to be able to trust the teacher when they disclose what they know and don't know (Cowie 2000). Trustworthy feedback by the teacher is that which is perceived by the student to be fair and just (Carless 2006). We argue that formative assessment must also be culturally appropriate to be trustworthy to students. In summary, what constitutes trust in the teacher-student relationship, on which formative assessment depends, from both the teacher's and student's perspective is often based on what cultural values they bring to the relationship (Bell 2011; Bishop et al. 2009).

However, these cultural values are not necessarily static as cultures and cultural values can change at the level of society and the classroom (Marlina 2009). The notion of culture used here is described as:

> the ever changing values, traditions, social and political relationships, and worldview created, shared, and transformed by a group of people bound together by a combination of factors that can include a common history, geographic location, language, social class, and religion. (Nieto 2000, p. 139)

Hence, culture is not a 'characteristic of individuals, and as such a set of stable practices that can be described and taught' (May and Sleeter 2010, p. 4). Instead 'culture and identity are understood here as multilayered, fluid, complex, and encompassing multiple social categories, and at the same time as being continually reconstructed through participation in social situations' (p. 10). This view of culture holds that culture is socially constructed and therefore influenced by social, economic and political discourses (Brown et al. 2007; Nieto 2010) with the dominant discourses at any time determining what will be valued or not.

In addition, culture is seen as learned, like world views, from interactions between members of family, communities and schools (Nieto 2010). However, the culture of power and privilege (Delpit 1998), for example, in schools and classrooms in New Zealand, may enable a Eurocentric-dominant culture to maintain oppression and hegemony over cultures, such as Samoan knowledge, language and ways of knowing. Viewing Eurocentric knowledge and practices as 'mainstream' in multicultural educational situations (e.g. New Zealand classrooms) continues to position it at the top of a hierarchy and minimise the contribution and value of a wider spectrum of knowledges and practices.

Lastly, culture is seen as dialectical (Nieto 2010), meaning that there are tensions within the complexity of culture, and identifying with a culture does not require accepting all aspects, good and bad, of that culture. Dialectical thinking is characterised by a tolerance of contradictory beliefs. For example, Tongan tertiary students in New Zealand may choose to not practise aspects of their culture, for example, attending church meetings each night during the week, in order to succeed at their study (Kalavite 2010). In another example, Samoan culture today may contain many practices taught by the nineteenth- and twentieth-century missionaries and colonisers (Lee Hang 2011). The silence in the classroom may be the result of these twin colonising efforts to maintain power imbalances, as much as from indigenous Samoan practices. Western culture in New Zealand is also changing, with interaction between Pākeha, Māori, Pasifika and Asian cultures, for example, in practices at funerals, and in New Zealand classrooms with respect to language, knowledge that is valued and values.

One way to address the student silence in Samoan classrooms might be that of expecting Samoan teachers and students to adapt their practices in the classroom to become more Western, with oral responses. But for Samoan educators, the situation is more complex. Their dilemma is how to promote learning, for example, by undertaking formative assessment strategies so that Samoan students are educated to be participating members of a global community, for higher education and employment, while still teaching and practising Samoan cultural practices and values in classrooms. In Samoa, an educated and wise person is one who has both ways: both Western and Samoan knowledge and practices. A schooling that only teaches and promotes Western knowledge, practices and values is not necessarily seen as an education. Perhaps there is another way to enable formative assessment in Samoan classrooms: the use of written feedback and feedforward. While there is some research on written feedback on assignments, for example, Carless (2006), the research reported in this chapter addressed the question: Does written formative assessment constitute a culturally appropriate form of formative assessment in Samoan science classrooms?

Worksheet 4 **Yrs 12-13**

Formative Assessment on common student misconceptions in some biology terms

Introduction : **Student Name**:________________________________

Some pupilstudents studying biology find it difficult to distinguish between the following terms: ***respiration (cellular), breathing*** and ***gas exchange***

Instruction: Write down what you know about these terms based on how they differ from each other. After writing your answers, please take a few minutes to assess how well you have answered by placing any of the letters G, Y or R from the traffic lights criteria in the self-assessment column based on how you feel about your response.

Traffic Lights Criteria

Write G for Green, Y for Yellow or R for Red in the circles for the self assessment to indicate the following.

- Green for good understanding
- Yellow for partial understanding
- Red for little understanding

Task	Student responses	Student Self-Assessment	For Teacher's Use Only	
			Feedback on current student learning	Feedforward for future student learning
What is respiration? (cellular)		○		
What is breathing?		○		
What is gas exchange?		○		

Fig. 14.1 A sample of one of the worksheets trialled and evaluated at in-service workshops

The Formative Assessment Worksheets

Des' experience as a science teacher and science teacher educator in Samoa was used to develop written formative assessment worksheets with sufficient spaces for students' responses and for teacher's feedback and feedforward comments. The worksheets consisted of questions and student answers, with spaces for teachers' feedback and feedforward comments to improve students' learning. There was also a column for students to self-assess, thereby giving further information to the teacher doing feedback and feedforward. Each worksheet was designed to fit one A4-sized page and to be photocopy friendly. Figure 14.1 shows a sample of the worksheets that were developed as an alternative strategy to elicit written student responses in class as a means of overcoming any cultural factors such as *le-tautala* that may be impacting students' verbal responsiveness in class. The worksheets were seen as a way to engage the students' within the 'safety' of the written text and writer-reader interaction via feedback and feedforward.

The worksheets were developed to cover a variety of topics in the general science curriculum: osmosis in a potato chip; using a light microscope; preparing a wet mount; biological terms of respiration, gas exchange and breathing; diffusion in a solid, liquid and gas; protein synthesis; distance-time graphs; speed-time graphs; heat energy transfer; sound waves; chemical formulae and symbols; test for presence of some gases; atomic theory; elements and atomic numbers; electron arrangements in atoms; and balancing equations.

Of the total of 16 formative assessment worksheets developed in this study, eight were developed by Des prior to the commencement of the research, while another eight were developed, again by Des, based on teacher ideas from the workshop component of the study.

The Workshops

The purpose of the workshops was to assist in the science teacher participants' professional development and to generate research data on formative assessment. The workshops also promoted the cultural practice of *talanoa, fetalai, fetalaa'i, fesuaa'i or fefa'asoaa'i,* which all mean 'the sharing of information for mutual benefit'. Des had heard elders use these words to persuade others to speak their minds and share their wealth of experience to inform the consensual decision-making process. This is so that everyone contributes in the process until a satisfactory final decision is reached. This fits here, in the sense that the research participants were clear in their wish to share any useful knowledge generated from this study for the benefit of all students and teachers in Samoa.

The workshop was a crucial part of this study from the beginning not only because it was an opportunity to raise teachers' awareness about formative assessment but also because it was an opportunity to get the teachers to evaluate for themselves the usefulness of the developed worksheets. In addition, the workshop was also seen as a culturally appropriate form of reciprocity with the participants for their time and effort in this study. Rather than just asking them to contribute to the study, the workshop offered them an opportunity to enhance their own professional knowledge.

The objectives of the workshop were to:

(i) Raise the participants' awareness of and knowledge about formative assessment as described in this chapter.
(ii) Develop and explore formative assessment (i.e. feedback, feedforward) strategies.
(iii) Give participants the opportunity to practise using formative assessment strategies.
(iv) Solicit and share culturally appropriate strategies for formative assessment in Samoan classrooms.
(v) Meet the researcher's obligation to the cultural concept of *fa'ataualofa* (reciprocity) by contributing to the participants' professional development.
(vi) Generate research data for (ii), (iii) and (iv).

The workshop was planned for 2 days, allowing time for the participants to learn about and reflect on the ideas discussed in the workshop and how these could be incorporated or viewed in light of their own classroom practices. Six sessions were planned on various aspects of formative assessment, and within those six sessions, a total of ten activities were prepared to engage the participants and solicit their ideas about doing formative assessment in the Samoan classroom setting.

The workshop was run twice: in August 2006 for five associate teachers (teachers who mentor a student teacher on practicum) (T6–T10) and five teacher educators (T1–T5) and then again in April 2007 for three pre-service (T11–T13) and three in-service student teachers (T14–T16). Hence, 16 participants were involved in the study. A detailed summary of the topics involved and the activities that the participants were engaged in during the 2-day workshop in 2006 and 2007 is given in Lee Hang (2011). Topics included the purpose of assessment, a definition of formative assessment, how to do formative assessment, cultural challenges to formative assessment, developing written formative assessment worksheets and exploring a model of formative assessment in the Samoan context. The same workshop programme was used in 2006 and 2007.

Evaluation of the Worksheets Before the Workshops

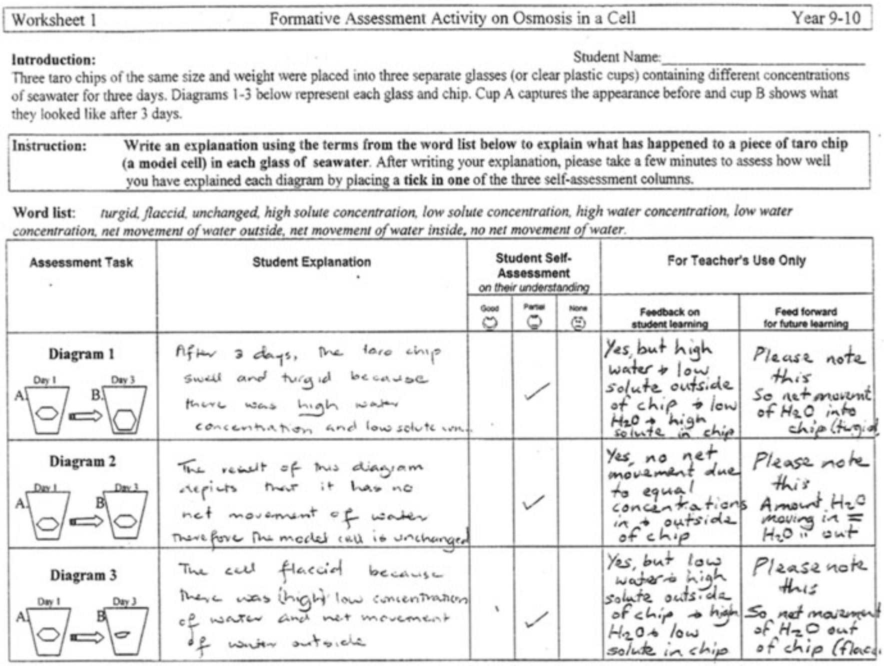
Within the time available, seven of the eight worksheets developed by Des prior to the first workshop were trialled at School A by Teacher 6 before the workshop. The aim of these trials was to see if the students found the instructions clear and easy to follow. From the returned worksheets, it was apparent that the students were able to follow the instructions, and all worksheets had been completed by the students (Lee Hang 2011).

Evaluation of the Worksheets During the Workshops

Copies of the student-completed worksheets (see Lee Hang 2011, for further details) without feedback and feedforward comments were given to the participants during the workshop component of this study, who wrote down their own feedback and feedforward comments. They also received copies of some worksheets, which the researcher had completed, as examples of what was required. An example is shown in Fig. 14.2.

The participants were also asked to evaluate the worksheets with feedback and feedforward comments written by the first author, based on the questions:

1. What do you think of it?
2. Is it useful?
3. In what way?
4. Would you use it?
5. How would you use it?
6. Any other comment?



Worksheet 1	Formative Assessment Activity on Osmosis in a Cell	Year 9-10

Introduction: Student Name:____________

Three taro chips of the same size and weight were placed into three separate glasses (or clear plastic cups) containing different concentrations of seawater for three days. Diagrams 1-3 below represent each glass and chip. Cup A captures the appearance before and cup B shows what they looked like after 3 days.

Instruction: **Write an explanation using the terms from the word list below to explain what has happened to a piece of taro chip (a model cell) in each glass of seawater**. After writing your explanation, please take a few minutes to assess how well you have explained each diagram by placing **a tick in one** of the three self-assessment columns.

Word list: *turgid, flaccid, unchanged, high solute concentration, low solute concentration, high water concentration, low water concentration, net movement of water outside, net movement of water inside, no net movement of water.*

Assessment Task	Student Explanation	Student Self-Assessment *on their understanding*			For Teacher's Use Only	
		Good	Partial	None	Feedback on student learning	Feed forward for future learning
Diagram 1 (Day 1, Day 3; A, B)	After 3 days, the taro chip swell and turgid because there was high water concentration and low solute con[…]		✓		Yes, but high water + low solute outside of chip + low H2O + high solute in chip	Please note this. So net movement of H2O into chip (turgid)
Diagram 2 (Day 1, Day 3; A, B)	The result of this diagram depicts that it has no net movement of water therefore the model cell is unchanged		✓		Yes, no net movement due to equal concentrations in + outside of chip	Please note this. Amount H2O moving in = H2O " out
Diagram 3 (Day 1, Day 3; A, B)	The cell flaccid because there was (high) low concentration of water and net movement of water outside		✓		Yes, but low water + high solute outside of chip → high H2O + low solute in chip	Please note this. So net movement of H2O out of chip (flacc[…]

Fig. 14.2 Worksheet 1, with feedback and feedforward comments written by Des

The participants were deliberately not asked to rate the worksheets on a scale because it seemed to contradict the formative practice of not giving out marks or grades but rather to give comments in the form of feedback and feedforward.

Participants who filled out and returned their evaluation forms were positive in their first impressions of worksheets. When asked about the usefulness of the worksheets, participants thought that the worksheets would be useful in identifying students' areas of weakness in terms of improving learning and catering for different student levels of understanding. Participants stated that they would use the worksheets in their lessons, even though some had concerns about large class numbers and the time required in the planning and preparation associated with large class sizes (in Samoa, classes in secondary schools may be as high as 35–60 students).

Evaluation of the Worksheets After the Workshops

At the completion of the workshops, participants were given sets of the eight worksheets (see Lee Hang 2011, for further details) as well as set of the other eight worksheets developed during the workshop (see Lee Hang 2011) to practise using in their schools. The teachers were keen to try out these worksheets in their classrooms. The pre-service teachers were encouraged to try them out in their lessons during their teaching practicum. However, only four participants (three associate teachers and

FNT7/L2/23.08.06		
Teacher: T7	Lesson: L2	Date: 23/08/2006
Venue: School B	Time: 8.10—9.50am	Class: Y10
Subject: Science	Topic: Heat Energy	Size: 32 pupils

After the students had settled down, the teacher explained that their lesson today was a follow-up of their previous lesson on Heat Energy. Teacher asked questions in Samoan and pupils answered them chorally (in a group) in a mixture of Samoan and English. Choral answers seemed to provide a "safe" way or opportunity during class time for pupils to verbalise their answers as a group—and to avoid the embarrassing practice of being singled out or placed on the spot—when answering individually.

The teacher asked what the terms 'conduction' and 'convection' mean.He asked:*O le a le uiga o le upu conduction?* (What is the meaning of conduction?)*Ae faafefea le upu convection?* (What about the term convection?)

One student replied: "*O le conduction o le ability lea o se metal for example e conduct ai electrons*,"to which the teacher replied: "*Feololo*" (an average answer).

Another student said: "*O le convection o le direction lea e move iai air particles*". The teacher replied: "*Sa'o*" (correct).

The teacher continued, revising the concepts and drawing diagrams onthe board as illustrations. The teacher asked questions to solicit choral responses from the students, which they did. At one stage an individual student answered and the class burst out in laughter. This socially-tolerated practice of *aamu* (mockery) is very disheartening and discouraging for pupils who genuinely want to learn in classrooms. I saw it in the misty eyes of the pupil who gave a response and was ridiculed for it.

The teacher then asked the pupils if they had any questions, and asked them to write the notes on the blackboard into their books.

The teacher moved around the class to see whether the pupils were copying notes into their notebooks. The teacher then moved in front of the class, gave out the formative assessment worksheets and instructed the class that they had ten minutes to work on the worksheets before they will be collected. Students settled down and started reading and working on worksheets.

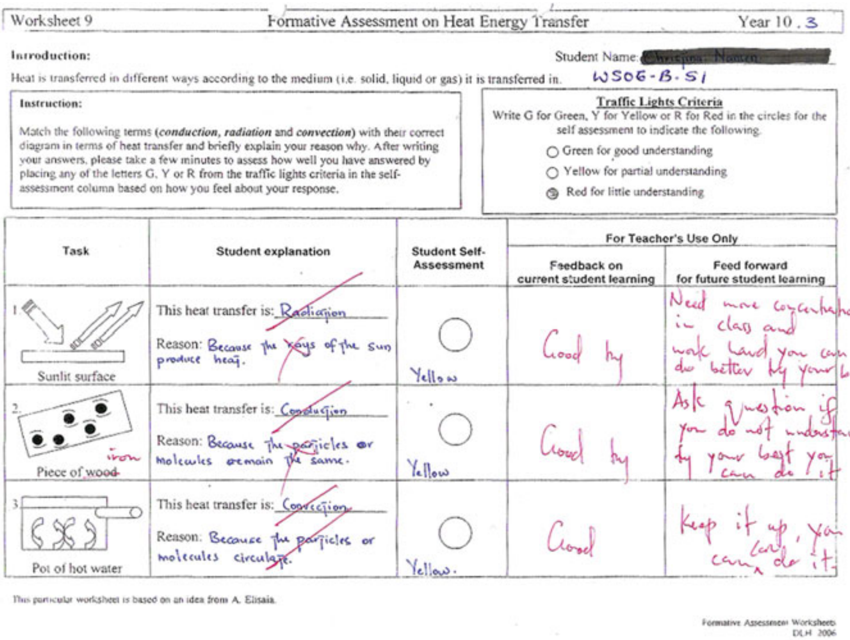
Bell went, teacher collected worksheets. The teacher told the pupils that he will give these back out to them the next lesson.

Fig. 14.3 Fieldnotes for a lesson

one pre-service teacher) returned their worksheets. While Des attempted to contact the remaining teachers before he returned to New Zealand to carry out data analysis, this was unfortunately not successful as telephone and email access is not reliable in Samoa.

Worksheets from an Associate Teacher

One worksheet was given to a Year 10 science class that an associate teacher had taught on the topic of heat energy. Of the 37 worksheets that were distributed to the class, 30 were returned by the teacher to Des. Des also observed the lesson and the following summary is from his fieldnotes (Fig. 14.3). The worksheet was used as an exercise for students at the end of a teacher-directed lesson revising a previous lesson's topic.


Worksheet 9 | Formative Assessment on Heat Energy Transfer | Year 10.3

Introduction:

Student Name: [redacted]

Heat is transferred in different ways according to the medium (i.e. solid, liquid or gas) it is transferred in.

WS06-B-S1

Instruction:

Match the following terms (*conduction, radiation* and *convection*) with their correct diagram in terms of heat transfer and briefly explain your reason why. After writing your answers, please take a few minutes to assess how well you have answered by placing any of the letters G, Y or R from the traffic lights criteria in the self-assessment column based on how you feel about your response.

Traffic Lights Criteria

Write G for Green, Y for Yellow or R for Red in the circles for the self assessment to indicate the following.

- Green for good understanding
- Yellow for partial understanding
- Red for little understanding

Task	Student explanation	Student Self-Assessment	For Teacher's Use Only: Feedback on current student learning	For Teacher's Use Only: Feed forward for future student learning
1. Sunlit surface	This heat transfer is: Radiation Reason: Because the rays of the sun produce heat.	Yellow	Good try	Need more concentration in class and work hard you can do better try your best.
2. Piece of wood / iron	This heat transfer is: Conduction Reason: Because the particles or molecules remain the same.	Yellow	Good try	Ask question if you do not understand try your best you can do it
3. Pot of hot water	This heat transfer is: Convection Reason: Because the particles or molecules circulate.	Yellow.	Good	Keep it up, you can do it

This particular worksheet is based on an idea from A. Elisaia.

Formative Assessment Worksheets DLH 2006

Fig. 14.4 Sample student worksheet from the associate teacher's class

It is interesting to note the practice of 'choral response' and the form of bilingual talk known as 'code-switching' in student-teacher exchanges captured in this field-note account. In this particular lesson, the formative assessment worksheet was used as a post-teaching exercise.

Figure 14.4 shows one of the worksheets from the lesson that was described in the fieldnotes. As can be seen, the student seemed to find the self-assessment aspect of the worksheet difficult. For example, in Fig. 14.4, the student gave an appropriate self-assessment for 1 and 2 but not for 3. Perhaps the novelty of self-assessment in the classroom or the issuing of judgements about one's own learning (which from a cultural point of view is the prerogative of elders, that is, teachers in the school and parents at home) is a factor.

Nevertheless, it is the teacher's comments that are of particular interest to this chapter in terms of their ability to enact the offering of written feedback and feed-forward. While the teacher was willing to incorporate formative assessment work-sheets into the lesson, he needed more help with the writing of succinct and helpful comments for students learning. The teacher's feedback column is filled with praise such as 'good' or 'good try', while the feedforward column featured encouraging remarks such as 'need more concentration in class', 'work hard', 'you can do better', 'try your best', 'ask questions if you don't understand', 'keep it up' and 'you can do it'. The comments did not address the substantive content of the student's responses; hence, the teacher's feedback and feedforward was not responsive to the student's

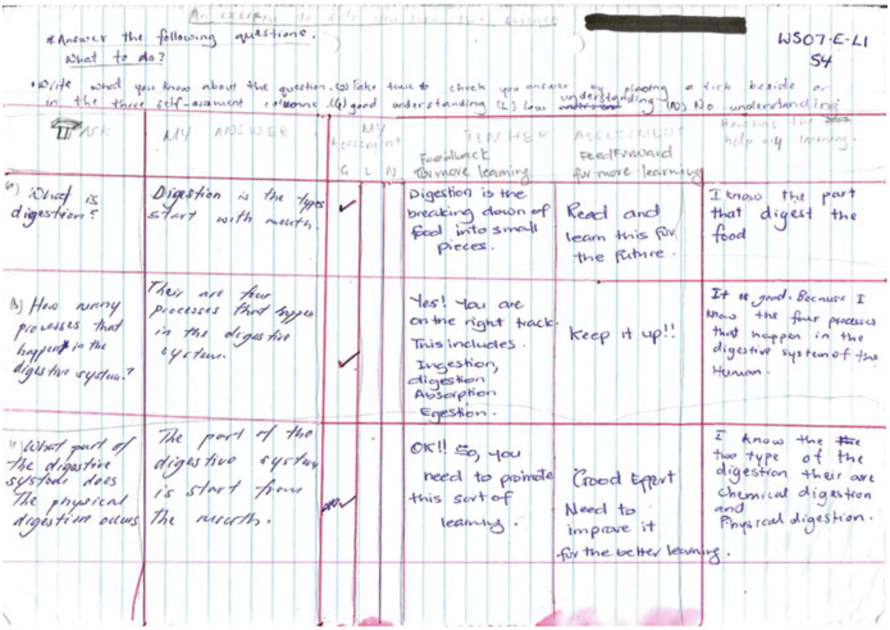


Fig. 14.5 Sample student worksheet from a pre-service teacher's class

thinking. It could be that the teacher lacked the science content knowledge to provide meaningful feedback and feedforward for this topic.

A Pre-service Teacher's Lesson

In 2007, student teacher T12 attended the workshop prior to her teaching practicum, and her placement was the only one that went according to the original research plan of placing pre-service students with an associate teacher who took part in the 2006 workshop. The teacher was enthusiastic about her use of the worksheets, and this was evident in how she created her own worksheets and incorporated them in the two lessons that were observed by Des.

In her first lesson, the student teacher taught the concept of digestion in animals to a Year 10 science class. She prepared notes and diagrams on newsprint, which were taped on the blackboard for students to copy into their notebooks. This gave the teacher some free time to walk around the class and interact with a few students or just simply walk around and check whether the students were 'on task'. In addition, she also prepared a worksheet (see Fig. 14.5) written on newsprint, which the students copied onto an A4-sized loose-leaf page before attempting it. The associate teacher only gave Des 11 worksheets at the end of the lesson.

It should be noted that this particular worksheet trialled four assessment columns. The first was for self-assessment, the second for feedback, the third for feedforward and the fourth was for the student's thoughts after reading the teacher's feedback and feedforward. The purpose of adding the extra column for students to write what they thought of the feedback and feedforward was to increase teacher-student formative interactions. The last column is like a 'thinking aloud' exercise but written.

With regard to the sample worksheet presented in Fig. 14.5, the student teacher could have commented in the feedback comment by stating specifically that the student's answer is incorrect and then providing the correct definition of digestion. Her feedforward could have been, 'You need to learn the correct meaning of digestion'. The last column, which was a new development of the worksheet, was a column for students to comment on how the teacher's feedback and feedforward has helped with their learning. This particular student seemed to have no change in his thinking, not having learned anything new. However, the fact that this student actually commented on his learning in the last column suggests that written feedback is opening up an avenue for interaction between students and teachers in Samoa. That the student teacher improvised with the use of newsprint to prepare her worksheet because she had no access to a photocopy machine shows her willingness to create opportunities for written formative assessment in her class. A significant drawback, unfortunately, was that the students spent most of their lesson copying down the worksheet onto a separate piece of paper.

Summary of Findings

Overall, the findings indicate that the worksheets or written formative assessments did enable some of the teachers to elicit written responses from students and to give written feedback and feedforward directly to each student in the Samoan setting. The four teachers who were able to return their questionnaires indicated their willingness to take a new assessment activity on board. This finding is consistent with the findings of another study on formative assessment and teachers in the New Zealand context, in which teachers of science developed their thinking and practices to undertake more meaningful formative assessment (Bell and Cowie 2001).

The second finding was that students were able to respond to the worksheets; of the worksheets returned, all had been attempted if not completed. Perhaps the fact that students were used to written summative assessments promoted by the high-stakes, examination-oriented education system in Samoa made them attempt the written formative assessment worksheets. However, that many students *did* return their worksheets was a significant step forward in the context of *le-tautala* in a Samoan classroom. It seems the worksheets provided an avenue for teachers to negotiate *le-tautala* to enable formative assessment. It may be a way to do formative assessment on a regular basis or an initial step towards oral formative assessment once a teacher and the students have gained confidence.

Thirdly, the four teachers' written responses indicated the need for further professional development to enable them to more competently undertake formative assessment. This finding is not unexpected. The teacher feedback and feedforward indicated a need for further learning and practice with respect to commenting on students' existing knowledge in relation to the lesson's learning objectives and how to get there. The teacher comments also reflected a need for deeper understanding of science so as to give more specific feedback and feedforward and be responsive to student responses. This is a reality arising from the issue of schools not being able to recruit or retain qualified science teachers in Samoan secondary schools due to poor employment benefits, such as salaries. A suggestion for improvement from the literature is to perhaps incorporate 'learning objectives' and 'success criteria' (Clarke et al. 2003) into the worksheets. This idea is basically about establishing the direction for students' learning so that when they attempt the worksheets, they will then be able to assess and see for themselves where they are in relation to where they want to be (Sadler 1989). To self-assess, one needs to know where one is, where one is going and the criteria used to judge when one has achieved the learning goals.

Fourth, the students did not include much science knowledge in their responses, reflecting either that there is a lack of understanding in the concepts of science or that the students did not perceive the need to use the science concepts in their answers. In addition, the students appeared to have difficulties with science language and English. The English language skills of students are known to influence their learning of science concepts (Muralidhar 1992). There is also the need for scientific literacy, that is, teachers and students being able to talk science and to use the language of science and science concepts in their thinking about the world.

The fifth key finding related to the students' self-assessment. The Samoan students in the research were not used to the learning strategy of self-assessment promoted in this study, and their responses to the exercise showed inconsistencies and difficulties, which reflects a natural response to an unfamiliar event or request. With specific teaching on how to self-assess, the use of learning intentions and success criteria, and more practice, the students may be better able to do this.

Lastly, resourcing issues like access to photocopying were a big constraint in the use of worksheets. Resourcing has the potential to significantly limit the use of worksheets when a photocopier is not available. Teachers may then not use the written formative assessments; ask the students to copy the worksheet from the board, which takes away time from learning; or use the written worksheets as an initial and temporary step on their learning journey to doing oral formative assessment.

Concluding Comments

This chapter has documented research on the use of written formative assessment worksheets as a culturally appropriate form of formative assessment in Samoan science classrooms. The research findings indicate that some Samoan science teachers and students were able to engage in formative assessment using the written format, when oral interaction between teacher and student is not a cultural norm in the actual teaching in the classroom, respecting the relationship *va* between teacher and student. However, as mentioned earlier, cultural practices, such as those in a classroom, may change, with resulting changes in teaching and assessment practices. Moreover, teaching, of which formative assessment is a part, may be viewed as a cultural practice (Bell 2011), meaning that the culture of the teacher influences his or her thinking and practice of teaching.

It is hoped that the findings of this research may be useful to Samoan teachers of science in Samoa, such as Des, who are teaching students of the same culture as themselves. The written formative assessment worksheets have potential as a classroom formative assessment tool and activity and also as a preliminary formative assessment activity until the teacher feels confident to use oral formative assessment interactions in giving feedback and feedforward. Also, the written formative assessment sheets can be used to bridge the gap between students who practise *le-tautala* and those who do not, for example, students in some urban high schools.

If the teacher is of a culture different to that of the students, the term 'culturally responsive teaching' is often used to acknowledge the mindfulness of the teacher to be responsive to the students' culture (not just their own) in informing their teaching decisions. It is hoped that the findings will be useful to non-Samoan teachers, such as Beverley, who are teaching Samoan students in science classrooms in New Zealand, Australia and the West Coast of the USA. In being a culturally responsive teacher, a non-Samoan science teacher of Samoan students is mindful of the following points (Bell 2011):

Culturally responsive teaching is not ethnicity blind and takes into account, rather than ignores, the culture and ethnicity of the students. In doing formative assessment, a non-Samoan teacher would be aware of and responsive to the Samoan cultural practices involved in *le-tautala* (silence) in the classroom in the teacher-student relationship.

Culturally responsive teaching does not use deficit theorising to explain differences in the achievement of students of different ethnicities. In doing formative assessment, a non-Samoan teacher would understand that silence in response to a teacher question does not necessarily mean the Samoan student does not know the content of the lesson.

Culturally responsive teaching includes having high expectations of students, not expectations based on stereotypes. In doing formative assessment, a non-Samoan teacher would expect her or his students to be able to understand feedback and respond to feedforward when done in culturally appropriate ways and use both to improve their learning.

Culturally responsive teaching involves forming relationships with students for professional caring and a commitment that students will achieve academically. In doing formative assessment, a non-Samoan teacher would form trusting relationships with Samoan students based on Samoan cultural values and not just those of her or his own culture. The teacher would care that the Samoan students achieve and change her or his pedagogy if need be.

Culturally responsive teaching includes teachers knowing and relating to their students as culturally located human beings. In doing formative assessment, a non-Samoan teacher would view the Samoan students as bringing with them into the classroom their cultural practices, knowledge and values and not leaving them at the classroom door.

Culturally responsive teaching includes building relationships and communication with the families and communities of students. In doing formative assessment, a non-Samoan teacher would communicate with families and the community about the purposes of formative assessment, as distinct from assessment for summative purposes—why it is important and how it will be done.

Culturally responsive teaching involves using the cultural and ethnic knowledge, language, values and practices of the students as resources to inform teacher decision-making about curriculum and pedagogy. In doing formative assessment, a non-Samoan teacher of Samoan students would be using the students' knowledge and their practice of *le-tautala* (silence) to make decisions about how to do formative assessment in the classroom.

Culturally responsive teaching is emancipatory and transformative, and hence, it is political for social justice. In doing culturally appropriate formative assessment, a non-Samoan teacher would be working to improve the learning outcomes of Samoan students and hopefully their employment futures, with the students' learning as bicultural: both Western and Samoan. To achieve this, the non-Samoan teacher is becoming bicultural as well.

References

Australian Department of Foreign Affairs and Trade. (2011). *Samoa country brief*. http://www.dfat.gov.au/geo/samoa/samoa_brief.html. Accessed 9 Nov 2011.

Bell, B. (2011). *Theorising teaching in secondary classrooms: Understanding our practice from a sociocultural perspective*. London: Routledge.

Bell, B., & Cowie, B. (2001). *Formative assessment and science education*. Dordrecht: Kluwer.

Bishop, R., Berryman, M., Cavanagh, T., & Teddy, L. (2009). Te Kotahitanga: Addressing educational disparities facing Maori students in New Zealand. *Teaching and Teacher Education, 25*(5), 734–742.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Brown, T., Devine, N., Leslie, E., Paiti, M., Sila'ila'i, E., Umaki, S., et al. (2007). Reflective engagement in cultural history: A Lacanian perspective on Pasifika teachers in Aotearoa New Zealand. *Pedagogy, Culture and Society, 15*(1), 107–118.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*(2), 219–233.

Clarke, S., Timperley, H., & Hattie, J. (2003). *Unlocking formative assessment: Practical strategies for enhancing students' learning in the primary and intermediate classroom. New Zealand version*. Auckland: Hodder Moa Beckett.

Cowie, B. (2000). *Formative assessment in science classrooms.* Unpublished doctoral dissertation, University of Waikato, Hamilton.

Day, R. R. (1981). Silence and the ESL child. *TESOL Quarterly, 15*(1), 35–39.

Delpit, L. D. (1998). The silenced dialogue: Power and pedagogy in educating other people's children. *Harvard Educational Review, 58*, 280–298.

Department of Education. (1995). *Western Samoa education policies 1995–2005*. Apia: Department of Education Printery.

Furtak, E., & Ruiz-Primo, M. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education, 92*, 799–824.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Kalavite, T. (2010). *Fononga 'A Fakahalafononga: Tongan students' journey to academic achievement in New Zealand Tertiary Education.* Unpublished doctoral dissertation, University of Waikato, Hamilton.

Lee Hang, D. (2011). *Fa'afatāmanu talafeagai mo lesona fa'asaienisi: O le tu'ualalo mo a'oga a faia'oga saienisi fa'aōliōli. A culturally appropriate formative assessment in science lessons: Implications for initial science teacher education.* Unpublished doctoral dissertation, University of Waikato, Hamilton.

Marlina, R. (2009). "I don't talk or I decide not to talk? Is it my culture?" International students' experiences of tutorial participation. *International Journal of Educational Research, 48*, 235–244.

May, S., & Sleeter, C. (2010). Introduction: Critical multiculturalism. In S. May & C. Sleeter (Eds.), *Critical multiculturalism: Theory and praxis* (pp. 1–16). New York: Routledge.

Moli, S. (1993). *Towards a relevant science education for Western Samoa: A'oa'oga tau saienisi mo Samoa i Sisifo.* Unpublished MEd dissertation, University of Waikato, Hamilton.

Muralidhar, S. (1992). Learning science in a second language: Problems and prospects. *Directions, 14*(1), 14–28.

New Zealand Ministry of Education. (2007). *The New Zealand curriculum*. Wellington: Learning Media.

NewZealandStatisticsDepartment.(2011).http://www.stats.govt.nz/Census/2006CensusHomePage.aspx. Accessed 9 Nov 2011.

Nieto, S. (2000). *Affirming diversity: The sociopolitical context of multicultural education* (3rd ed.). White Plains: Longman.

Nieto, S. (2010). *The light in their eyes: Creating multicultural learning communities* (10 anniversaryth ed.). New York: Teachers College.

Pereira, J. A. (2005). *Aspects of primary education in Samoa: Exploring student, parent and teacher perspectives.* Unpublished doctoral dissertation, University of Otago, Dunedin.

Pongi, V. (2004). *The role of assessment in improving quality in education: The shift from assessment of learning towards assessment for learning*. Unpublished discussion paper prepared for the Forum Education Ministers' meeting held on January 28–29, 2004, Pacific Islands Forum Secretariat, Apai.

Sadler, R. (1989). Formative assessments and the design of instructional systems. *Instructional Science, 18*, 119–144.

Tanielu, L. (2001). *Pacific vision international conference – A society of learners*. Ministry of Pacific Island Affairs.

Wendt, A. (1996). Tatauing the post-colonial body. *Span, 42–43*, 15–29.

Chapter 15
The Disparity Between Achievement and Engagement in Students' Science Learning: A Case of East-Asian Regions

Jinwoong Song

Introduction

> Certain *inputs* from the outside—students, teachers, other resources, management rules and requirements, parental anxieties, standards, tests with high stakes, and so on—are fed into the box. Some *outputs* are supposed to follow: students who are more knowledgeable and competent, better test results, teachers who are reasonably satisfied, and so on. But what is happening inside the box? …. (Black and Wiliam 1998, p. 140, emphasis in original)

In this convincing metaphor, Black and Wiliam (1998) see the classroom as a 'black box'. The dictionary meaning of black box is 'a complex system or device whose internal workings are hidden or not readily understood' (http://oxforddictionaries.com/definition/black+box). Black and Wiliam use the metaphor to indicate a place where all substantial interactions among students, teachers and science are supposed to take place, but, as what is actually happening is rarely seen or understood, the system resembles a black box. They argued that to help teachers to bring about actual meaningful changes, formative assessment is essential since it gives necessary information and feedback that can rarely be obtained by the means of summative assessment. By introducing the metaphor of black box, they might have hoped to transform science classrooms into a kind of 'transparent box'.

In many comparative studies of education, one of the most typical black boxes is the case of East-Asian students' achievement, frequently noted as something deserving of admiration and less frequently (but still often) as something of a surprise and curiosity. For example, US President Barak Obama lauded Korean education with the intention of calling for reforms of the US education system:

> Our children spend over a month less in school than children in South Korea. That is no way to prepare them for a 21st century economy. That is why I'm calling for us not only to expand effective after-school programs, but to rethink the school day to incorporate more

J. Song (✉)
Seoul National University
e-mail: jwsong@snu.ac.kr

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_15,

> time—whether during the summer or through expanded-day programs for children who need it. … If they can do that in South Korea, we can do it right here in the United States of America. (March 10, 2009, Education Speech)

Obama's praise of Korean education included a focus on mathematics and science education. He also called for the USA to look at Korea in adopting not only longer days and after-school programmes but also more positive social attitudes towards teachers:

> Let's also remember that after parents, the biggest impact on a child's success comes from the man or woman at the front of the classroom … In South Korea, teachers are known as 'nation builders.' Here in America, it's time we treated the people who educate our children with the same level of respect. (The Korea Herald, 26 Jan. 2011)

President Obama's compliments to Korean education were received by many Koreans with a degree of surprise. This is because whenever educational issues in Korean are under debate, US education has usually been the very example suggested for Korea to follow to resolve the issues. At his keynote address on March 25, 2011, at the annual meeting of the Association for Education Finance and Policy in the USA, Mr. Ahn, the former Minister of Education in Korea, warned the USA against copying his nation's approach, which he said had grown too test centred and often detracted from students' love of learning. He went on:

> Although the pain of memorizing is unavoidable for young students to acquire new knowledge, they should also be motivated by the pleasure of creative expression. … However, we force the students to memorize so much that they experience pain rather than [the] pleasure [of] acquiring knowledge through the learning process.

He also commented on Korean parent's devotion to their children's education, praised much by Obama: 'Extreme parental pressure is not something to be envied… The Korean case illustrates it is possible to have too much of a good thing' (Cavanagh 2011, p. 18).

Given these very contrasting appraisals, Korean education (especially of mathematics and science) appears to be a 'black box' whose internal workings are hidden and not readily understood. However, this is by no means only the case in Korea, but is believed to be a common phenomenon across the regions of East Asia.

This chapter seeks to understand the inner workings of the larger black box of East Asia, a region whose science achievements are constantly praised by the outsider while at the same time its science education practice has been the object of continuous criticism by the insider. This contradiction has been most clearly illustrated by the disparity between the highest achievements and the lowest engagements of students in a series of international comparative assessment of science in PISA and TIMSS studies. Thus, this chapter considers the issue of the disparity between achievement and engagement with a focus on East-Asian regions, and especially data from Korea, in the following order: (a) reviewing the data about the disparity that appears in TIMSS and PISA studies, especially focusing on less widely known data showing low students' engagement; (b) considering personal opinions and studies on social and classroom cultures in East Asia, mostly from perspectives outside the field of science education; (c) reporting results of some recent studies illustrating aspects of the essential culture of Korean school science; (d) introducing a more comprehensive and systematic framework to enable consideration of science culture (the 'SCI'); and finally (e) providing some recommendations and implications drawn from this study.

Disparities of East-Asian Regions Appearing in TIMSS and PISA Studies

High Achievement

For over a decade of international comparative studies by PISA and TIMSS, the science performance of students of the East-Asian region has been truly remarkable, almost always being among the top tier of the ranking (see Table 15.1). Other countries beyond the region of East Asia also consistently belonging to this group are Finland and Singapore, of which the latter could also be considered a member of East Asia, not in terms of geography but having similar social and cultural bases. This tendency to high achievement in science seems to be so persistent that there is no sign of a decrease in the near future. Further, East-Asian regions have shown similar high achievements in other subjects, slightly less obvious in reading but certainly very strong in mathematics. For example, in the PISA 2009 study in which 65 countries or regions participated, Hong Kong, Japan, Korea and Taiwan were ranked 4th, 8th, 2nd and 22nd, respectively, in reading and 3rd, 9th, 4th and 5th, respectively, in mathematics (OECD 2010).

Low Engagement

One of most striking features of East-Asian students' science learning in the international comparative studies is that, despite their highest achievement among the participating countries, they consistently show unexpectedly low levels of engagement, something that has received relatively little attention from most people within and outside the region. What is happening inside the black box?

Table 15.1 Ranks and scores of science in PISA and TIMSS studies

	PISA				TIMSS			
Country	2009	2006	2003	2000	2007	2003	1999	1995
Chinese Taipei	11 (520)	4 (532)	–	–	2 (561)	2 (571)	1 (569)	–
Hong Kong SAR	2 (549)	2 (542)	3 (539)	–	9 (530)	4 (556)	15 (530)	16 (510)
Japan	4 (539)	5 (531)	2 (548)	2 (550)	3 (554)	5 (552)	4 (550)	3 (554)
Korea, Rep. of	5 (538)	11 (522)	4 (538)	1 (552)	4 (553)	3 (558)	5 (549)	4 (546)
International avg.	**(501)**	**(488)**	**(497)**	**(491)**	**(500)**	**(491)**	**(521)**	**(518)**

Table 15.2 TIMSS 2007 data on index of students' positive affect toward science (PATS) (grade 8) (Martin et al. 2008, pp. 174–175)

Country	High PATS		Medium PATS		Low PATS	
	Percent of students	Average achievement	Percent of students	Average achievement	Percent of students	Average achievement
Chinese Taipei	40	597	24	552	35	527
Hong Kong SAR	60	549	22	508	19	498
Japan	47	574	28	545	25	529
Korea, Rep. of	38	586	27	544	36	526
International avg.	**65**	**476**	**19**	**442**	**16**	**436**

Index based on students' responses to three statements about science: (1) I enjoy learning science, (2) science is boring (reversed) and (3) I like science. Average is computed across the three items based on a 4-point scale: (1) agree a lot, (2) agree a little, (3) disagree a little and (4) disagree a lot. Students agreeing a lot or a little on average across the three statements are assigned to the high level. Students disagreeing a little or a lot on average across the three statements are assigned to the low level. All other students are assigned to the middle level.

Table 15.3 TIMSS 2007 data on index of students' self-confidence in learning science (SCS) (grade 8) (Martin et al. 2008, p. 187)

Country	High SCS		Medium SCS		Low SCS	
	Percent of students	Average achievement	Percent of students	Average achievement	Percent of students	Average achievement
Chinese Taipei	23	619	36	552	41	536
Hong Kong SAR	33	561	49	516	18	515
Japan	20	601	44	554	36	529
Korea, Rep. of	24	603	40	556	36	516
International avg.	**48**	**492**	**38**	**439**	**13**	**427**

Index based on students' responses to four statements about science: (1) I usually do well in science, (2) science is more difficult for me than for many of classmates (reversed), (3) science is not one of my strengths (reversed) and (4) I learn things quickly in science. Average is computed across the four items based on a 4-point scale: (1) agree a lot, (2) agree a little, (3) disagree a little and (4) disagree a lot. Students agreeing a lot or a little on average across the four statements are assigned to the high level. Students disagreeing a little or a lot on average are assigned to the low level. All other students are assigned to the middle level.

Among the 29 TIMSS countries, East-Asian regions showed the lowest levels of students' positive affect toward science, something which was assessed by students' responses to three statements ('I enjoy learning science', 'science is boring' and 'I like science'). Japan, Chinese Taipei and Korea were ranked the third, second and first lowest, respectively (see Table 15.2). It is really surprising that the students with the highest scores show the most negative affect towards science learning.

East-Asian regions again showed the lowest levels of students' self-confidence in learning science, as assessed by students' responses to four statements ('I usually do well in science', 'science is harder for me than many of my classmates', 'I am just not good at science' and 'I learn things quickly in science'). Hong Kong, Korea, Chinese Taipei and Japan were ranked the fifth, third, second and first lowest, respectively (see Table 15.3).

TIMSS studies provide some valuable data indicating what is going on during science lessons. When 'doing science investigation' was considered, according to students' responses, Taiwan and Korea among East-Asian regions were particularly low compared to OECD members, while most East-Asian regions were below the OECD average on self-perceptions of being able to 'give explanations about what is being studied', 'watch the teacher demonstrate an experiment or investigation', 'design or plan an experiment or investigation' and 'relate what is being learned in science to our daily lives' (see Table 15.4). It is also interesting to see that there is a great disparity between the reports of students and those of their teachers, not only in East-Asian countries but also across the OECD members, especially with higher recognition by students in being able to 'make observations and describe what was seen', 'watch the teacher demonstrate an experiment or investigation' and 'design or plan an experiment or investigation', with higher responses by teachers on the ability to 'relate what is being learned in science to our daily lives'. Why the two groups see the same object—doing science investigation—differently is not yet clear, but establishing why this is would provide insights into significant aspects of what is really going on inside the black box.

With respect to the time that students report spending on various activities in science lessons, East-Asian regions generally showed a much higher percentage in 'listening to lecture-style presentations', while percentages were lower in 'working problems with teacher's guidance', 'working problems on their own without teacher's guidance' and 'taking tests or quizzes', illustrating how science teaching in general in East-Asian regions tends to focus on lecture-style presentation (see Table 15.5).

Although, as seen above, East-Asian regions share many common features, there seems a noticeable difference between two subgroups: Hong Kong and Taiwan versus Japan and Korea. This illustrates another subtle difference within the East-Asian culture, between Chinese and non-Chinese, possibly reflecting different historical and cultural orientations towards science and science-related careers.

PISA 2006 encompassed 57 countries and economies. Students' (15-year-olds') enjoyment of science was evaluated by five items related to (school) science learning. Japan and Korea belonged to the lowest group of students regarding their 'enjoyment of science', ranked 56th and 51st, respectively, whereas Hong Kong and Taiwan showed higher scores than the OECD average (see Table 15.6).

A similar pattern can be found in students' instrumental motivation to learn science: Japan and Korea ranked as the 57th and 53rd, respectively, out of 57 (OECD 2007), in particular, reflecting Japan's alarmingly low levels of instrumental motivation to learn science (see Table 15.7).

The same pattern was also found in students' future-oriented motivation to learn science: again, Japan and Korea ranked very low—54th and 55th, respectively, out of 57 (see Table 15.8).

Societal and Classroom Cultures in Korea and East Asia

Science education is not independent of its wider cultural backgrounds in terms of both the wider culture of the society and specific classroom cultures. The disparity found in East-Asian regions outlined above must also be deeply rooted in commonly

Table 15.4 TIMSS 2007 data on students'/teachers' reports on doing science investigations (grade 8) (Martin et al. 2008, pp. 299, 303)

Country	Make observation and describe what was seen	Give explanations about what is being studied	Watch the teacher demonstrate an experiment or investigation	Design or plan an experiment or investigation	Conduct an experiment or investigation	Work in small groups on an experiment or investigation	Relate what is being learned in science to our daily lives
Chinese Taipei	35/26	42/57	41/16	27/10	29/12	28/9	38/62
Hong Kong SAR	62/21	57/66	61/19	42/13	68/70	71/67	63/63
Japan	66/42	38/30	62/28	45/18	72/68	76/69	29/47
Korea, Rep. of	29/37	26/81	46/34	21/22	28/39	29/33	35/84
International avg.	**65/35**	**65/71**	**67/41**	**50/27**	**54/47**	**56/50**	**57/79**

The data are the percentages of students/teachers (?) who reported doing the activity about half of the lessons or more and the percentages of students whose teachers reported students doing the activity about half of the lessons or more

Table 15.5 TIMSS 2007 data on percentage of time in science lessons students spend on various activities in a typical week (grade 8) (Martin et al. 2008, pp. 320–321)

Country	Reviewing homework	Listening to lecture-style presentations	Working problems with teacher's guidance	Working problems on their own without teacher's guidance	Listening to teacher reteach and clarify content/procedures	Taking tests or quizzes	Participating in classroom management tasks not related to the lesson's content/purpose	Other student activities
Chinese Taipei	10	48	11	5	9	8	4	3
Hong Kong SAR	9	39	15	8	8	8	5	8
Japan	3	47	15	5	14	5	2	8
Korea, Rep. of	5	49	9	8	13	6	6	5
International avg.	**9**	**25**	**17**	**13**	**13**	**10**	**6**	**7**

Table 15.6 PISA 2006 data on students' enjoyment of science (From OECD 2007, p. 144)

Items	Chinese Taipei	Hong Kong-China	Japan	Korea	OECD average
I enjoy acquiring new knowledge in science	79	85	58	70	**67**
I generally have fun when I am learning science topics	65	81	51	56	**63**
I am interested in learning about science	64	77	50	47	**63**
I like reading about science	62	65	36	45	**50**
I am happy doing science problems	43	54	29	27	**43**
Average/total	62.6/313	72.4/362	44.8/224	49.0/245	**57.2/286**

Data represent the percentage of students agreeing or strongly agreeing with the statements of the items

shared traditions and cultures of the regions. Robust understanding of this disparity requires knowing more about the wider context of the phenomenon:

> On one of my first days of teaching an English class in Korea to tertiary Korean students … Moving from group to group I heard the following; "My name is… I'm studying … I come from…. My family…" I enquired about this and was told, "We don't know each other." Even though we had practiced introductions in other group activities there were now new group configurations and they "needed to introduce each other first." (Cronin 1995)

On his first encounter with a Korean classroom, an English teacher witnessed how important the Confucian tradition is to Koreans. In the Confucian tradition, as the teacher noted, it is extremely important to establish social status. Until formal introductions are made, the students are considered to be non-persons in terms of communicating with each other, and by the behaviours the teacher describes, they were accommodating both the teachers' instructions and adhering to their important Korean tradition (Cronin 1995).

It might be true that the distinctiveness of one culture could be better appreciated by an outsider from a different cultural background not yet accustomed to the new culture. As another example, Alex Case, an expert of TEFL (teaching English as a foreign language) who taught English for about 2 years in Korea and with ample other experience of teaching across the world, expressed his personal feelings and assessments about the cultures of Korean education as follows:

> Korean society is traditionally very stratified, with different vocabulary and grammar needed when speaking to a higher status person such as someone older, a teacher, boss, customer, or a man if you are a woman. The higher status person will also expect to initiate and dominate conversations. Koreans will therefore not be shy about asking each other and you about your ages, as it helps to put you all in your place on the social scale. (*Alex Case for TEFL.net*) (http://edition.tefl.net/articles/home-and-abroad/korean-cultural-differences/)

> [Korean] Students therefore enter class with an expectation that they will cover a lot, that they will be given lots of homework, and that the teacher will be strict about completing those things. However, apart from an initial surge of enthusiasm and preparation for tests,

Table 15.7 PISA 2006 data on students' instrumental motivation to learn science (From OECD 2007, p. 147)

Items	Chinese Taipei	Hong Kong-China	Japan	Korea	OECD average
I study school science because I know it is useful for me	83	72	42	55	**67**
Making an effort in my school subject(s) is worth it because this will help me in the work I want to do later on	76	73	47	57	**63**
Studying my school science(s) is worthwhile for me because what I learn will improve my career prospects	76	72	41	52	**61**
I will learn many things in my school science subject(s) that will help me get a job	73	64	39	46	**56**
What I learn in my school science subject(s) is important for me because I need this for what I want to study later on	65	63	42	45	**56**
Average/total	74.6/373	68.8/344	42.2/211	51.0/255	**60.6/303**

Data represent the percentage of students agreeing or strongly agreeing with the statements of the items

> the work rate of students is hardly likely to match those expectations. This could be because of tiredness, other demands on their time and energy …. (*Alex Case for TEFL.net*) (http://edition.tefl.net/articles/home-and-abroad/korean-cultural-differences/)

The differences between East-Asian cultures and the so-called Western cultures have been occasionally addressed by researchers in several fields of education (e.g. Chang 2008; Cho 2004; Kaiser et al. 2006; Lee 2001; Leung 2001). For example, the difference between Korean culture and that of the West is well illustrated in an article by a Korean English educator as follows:

> **Different Ways of Thinking**: In conwtrast to English students, Korean students tend to express themselves in general and indirect ways, even when asked to communicate their ideas. This is because they have been trained to think inclusively and express themselves indirectly in case they might offend others. Such a reserved attitude originates from Confucian thinking, in which moderation is considered the supreme virtue. (Cho 2004, p. 34)

> **Koreans' Group-oriented Thinking**: the Family and Society above the Individual. … Unlike English-speaking people, Koreans put their family name first, followed by their given name, showing the importance of the family. Such a strong sense of belonging is revealed most obviously in Korean students' use of 'our' for 'my.' Instead of saying "my teacher,"

Table 15.8 PISA 2006 data on students' future-oriented motivation to learn science (From OECD 2007, p. 149)

Items	Chinese Taipei	Hong Kong-China	Japan	Korea	OECD average
I would like to work in a career involving science	38	46	23	27	**37**
I would like to study science after secondary school	34	41	20	23	**31**
I would like to work on science projects as an adult	29	37	17	17	**27**
I would like to spend my life doing advanced science	22	25	23	12	**21**
Average/total	30.8/123	37.3/149	20.8/83	19.8/79	**29.0/116**

Data represent the percentage of students agreeing or strongly agreeing with the statements of the items

> "my father," they say "our teacher" even when the listener is not a classmate, "our house" even when talking to a stranger, and "our father" even when the listener is not a sibling. … This is because they do not call people by their names, but rather by their relationship or social/family roles, thus, instead of calling somebody "Tom," they call them "Jane's father" or "Bill's teacher." (Cho 2004, p. 34)

Leung (2001), in a study identifying the features of East-Asian mathematics education, claimed that the way of teaching mathematics in the region is deeply rooted in the traditional culture and values of the region and that a better understanding of the features of East-Asian mathematics education requires an understanding of the underlining traditions and values. These, Leung noted, emphasise the roles of teacher and subject matter:

> If a theory of mathematics education in East Asia is to be established, it must be able not only to organize the distinctive features into a coherent whole, but also to account for or justify these features in terms of fundamental East Asian views and values. … Student-centered education is the basic tenor in Western theories, while East Asian educators are affirming the importance of the teacher and the subject matter. (pp. 46–47)

What is seen as good instruction is also very much dependent on social culture. In a study to find out what may be counted as high-quality teaching and learning of mathematics in Korea, Pang (2009) argued that good mathematics instruction may be perceived differently with regard to underlying social and cultural norms by looking in detail at a sixth grade teacher's lessons. She also provided a cautionary remark on adopting new trends of teaching from outside, saying:

> a careless adoption of new trend such as seemingly student-oriented activities may result in the loss of strong content-oriented instruction, which contributes to the superior performance in international comparisons on mathematics. We need to reflect on the strengths and weaknesses of the existing teaching practice and analyze its origin with regard to the underlying cultural paradigms and societal needs. (p. 360)

As illustrated by these Korean cases, the culture of East Asia is more group oriented or collectivistic (Triandis 1995) and indirect in expression than that of the

West, and thus, the way of teaching (or of desirable teaching) would be quite different from the West. This kind of a wider context of the society would, directly or indirectly, influence many aspects of teaching and learning of science in schools.

Recent Studies on Latent Features of Science Learning in Korea

Having considered more general social and educational backgrounds of the East-Asian region and Korea in particular, I now turn to consideration of what is known about what is actually going on inside science classrooms. We will see results of selected recent studies focused on latent features of science learning, mostly carried out by a group of science educators at Seoul National University. This will provide some insights into the origins of the disparity in Korea and, by inference, East-Asian regions more generally.

Students' Perceived Purposes of Laboratory Work

In a study of students' views of the purposes of laboratory work (Kim and Song 2003), 7th grade students (n = 147) in a middle school in Seoul (Korea) were asked two questions: 'Why do you think scientists do laboratory work?' and 'why do you think you do laboratory work in science classes?' Students responded to the questions with short essays including the reasons for their answers and examples. It was found that students have very different views of their own school science laboratory work by comparison with their views of scientists' laboratory work. They tended to think that they do laboratory work mainly for understanding and memorising the contents of textbooks (83 % of students expressed this idea), while they saw scientists doing laboratory work to discover new facts or invent something (57 % expressed this idea).

In particular, their responses to the purpose of school laboratory work were much more diverse despite overwhelming support for understanding and memorising of textbook contents. For example, the students' responses belonging to this category could be further classified into the following subcategories: 'to facilitate the understanding of theories', 'to learn principles', 'to help remember facts and principles', 'to learn scientific knowledge', 'to satisfy curiosity', 'to learn contents in-detail', 'to verify scientific facts' and 'to find the results of the experiment'. Other students' responses to school laboratory work were also mostly related to educational functions, such as 'to get experience in lab-work procedures', 'to develop ability of thinking', 'to help apply scientific knowledge to the life' and 'to learn how to work in teams'. Two extracts from responses illustrate this:

> [The purpose of school laboratory work …] *Wouldn't it be to make the words of textbooks more easily understood? For example, let's have a case of force which we just learned. If the words or terms of force appear to be difficult to understand,* [The teacher…] *by doing*

> *experiments and by giving more explanations during the experiment, make students to have clearer understanding of the concept.* [It's …] *a kind of studying or learning methods, I think.* (Sunae)

> *Unlike scientists, we do experiments on the basis of already known facts. So there are not many facts we ourselves are going to find out. The reasons for doing experiments are not only to grasp the things in the textbook but also to feel them by your whole body through carrying out the experiments.* (Kyungsik)

Thus, there was a clear and distinctive difference in students' perceptions of purpose between scientists' laboratory work and school science laboratory work. They considered school laboratory work to be a kind of 'learning method', not as independent inquiry activity to be undertaken by them. In other words, for the students, laboratory work was a continuum of classroom learning of science, which would not easily stimulate students' active engagement.

Students' Images of 'Doing-Science-Well'

Starting from the assumption that students' images of learning would be a major factor affecting their ways of learning, Lee et al. (2008) investigated a group of high school students' (37 grade 11 science stream students) images of 'Doing-Science-Well'. The students were asked to write their personal opinions and experiences in response to the following three open-ended questions probing their general images—images from their personal experience with fellow classmates and images from their experience with somebody not confined to the classroom, respectively:

(Q1) What do you think about Doing-Science-Well? What would be Doing-Science-Well? Why do you think so?
(Q2) Who is a good example of Doing-Science-Well in your class? And why do you think so? What was the critical occasion to see that he or she was Doing-Science-Well? Describe the situation of the occasion in detail and explain why you were so much impressed.
(Q3) Throughout your life, who is a good example of Doing-Science-Well? And why do you think so? What was the critical occasion to see that he or she was Doing-Science-Well? Describe the situation of the occasion in detail and explain why you were so much impressed.

The students' responses to the questions were analysed by systemic networks (Bliss et al. 1983). This resulted in two major categories: 'according to social standards' and 'according to personal features'. Examples of 'according to social standard' were 'getting high grades in science', 'being experts like science teacher and scientists', 'having high social reputation' and 'winning a prize in contests'—all of which reflect the features of a person that are socially recognised rather than being based on his or her personal features.

More interesting data were those responses categorised as 'according to personal features' because these provided concrete examples of people who are typically perceived as Doing-Science-Well. The personal features were categorised into three aspects—cognitive, affective and psychomotor—and further divided into several

Table 15.9 Students' images on personal features of Doing-Science-Well (Lee et al. 2008)

Aspects	Sub-aspects	Q1	Q2	Q3	Total
Cognitive aspects	Understanding science easily	13	5	4	**22**
	Applying understanding and knowledge diversely	6	2	2	**10**
	Having abundance of scientific knowledge	4	2	3	**9**
	Being creative	2		1	**3**
	Subtotal	25	9	10	**44**
Affective aspects	Being interested, enjoying doing science	4	6	1	**11**
	Being curious	2	2		**4**
	Being open minded	2	1		**3**
	Being enthusiastic		1	1	**2**
	Concentrating on a task		1		**1**
	Having tenacity		1		**1**
	Subtotal	8	12	2	**22**
Psychomotor aspects	Expressing, explaining science concepts well	4	6	11	**21**
	Doing experiments well	3	6	1	**10**
	Discovering something new, finding reasons	5		1	**6**
	Using scientific knowledge in everyday life situations	5		1	**6**
	Inventing something new	3		1	**4**
	Participating many scientific activities	1	1		**2**
	Fixing electronic and mechanic devices well		2		**2**
	Solving math problems well	1		1	**2**
	Subtotal	22	15	17	**53**

sub-aspects (see Table 15.9). It was found that the two most frequently mentioned sub-aspects were 'understanding science easily' and 'expressing, explaining science concepts well'. This would seem to illustrate once again Korean students' strong tendency to highly value knowledge and understanding.

Based on the identified features and specific aspects, the students' responses to the questions were re-categorised into nine typical types of Doing-Science-Well. These were described as Einstein, Marie Curie, Socrates, Science-mania, MacGyver, Edison, Encyclopaedia, Pythagoras and Feynman. The use of these labels was intended to more easily convey and conceptualise the main characteristics of the images (see Table 15.10). Although the study did not investigate the distribution of the images possessed by students in terms of the identified Doing-Science-Well types, about a half of the types (Einstein, MacGyver, Encyclopaedia, Pythagoras and Feynman) are more concerned with knowledge and understanding of science, which would be more related to achievement than to engagement.

Fatigue in Learning Science

Kim (2011) identified factors causing the feeling of fatigue of learning science in school and science centres through a survey of 610 students comprising about 200 students from each of elementary, middle and high school. Following the idea of

Table 15.10 Types of personas who are Doing-Science-Well (Lee et al. 2008)

Type/persona	Features of the type of Doing-Science-Well
Einstein	A person who creates new knowledge with open mind
Marie Curie	A person who draws excellent findings and conclusions out of systematic experiments
Socrates	A person who keeps asking questions with curiosity
Science-mania	A person who enjoys doing various science activities (like, reading science books, visiting science centres) with great interest
MacGyver	A person who can find solutions to everyday life problems by applying scientific knowledge
Edison	A person who searches for the inconvenient around us and invents new things to ease them
Encyclopaedia	A person who gets very high scores in science with a full of scientific knowledge
Pythagoras	A person who is very good at formula, numbers and estimations
Feynman	A person who can explain difficult scientific concepts plainly

the contextual model of learning by Falk and Dierking (2000), three contexts (i.e. personal, sociocultural and physical-environmental) were introduced to classify the fatigue-causing factors. These were further divided into nine categories. Through two stages of the survey, Kim identified 50 fatigue factors for each of 'learning school science' and 'visiting science centres' and ranked the factors according to their relative frequencies. In particular, among the most popular factors causing students' feelings of fatigue in school science learning, many were related to learning knowledge and concepts and examinations, such as 'sciences are too difficult, and I have to think very hard'(1st); 'science lessons are theoretical, rather than practical'(4th); 'sometimes experiment results are different from what we learned'(6th); 'I feel the pressure of being asked questions by teachers and of making presentation'(9th); 'I have to just listen to teacher's explanations'(10th); 'there are too many things to remember'(11th); and 'I feel too much pressure of exams'(16th).

Value and Aspects of Science Lessons That Appeal to Students

Students spend much time in school studying science but what has been consistently shown by many previous studies is that students' interest in science is low and declines as they move through school (e.g. KSF 2004). To explore this disappointing phenomenon, Park and Song (2009) investigated the conditions of science lessons perceived by 16 secondary students (eight boys and eight girls, 13–15 years old) to be appealing. Each student was asked to write a short essay and then participated in a semi-structured interview. The questions given during the interview were intended to probe the students' views on the value of science lessons (i.e. reasons for learning science, the importance of science learning), appealing features of science lessons and other relevant and background information.

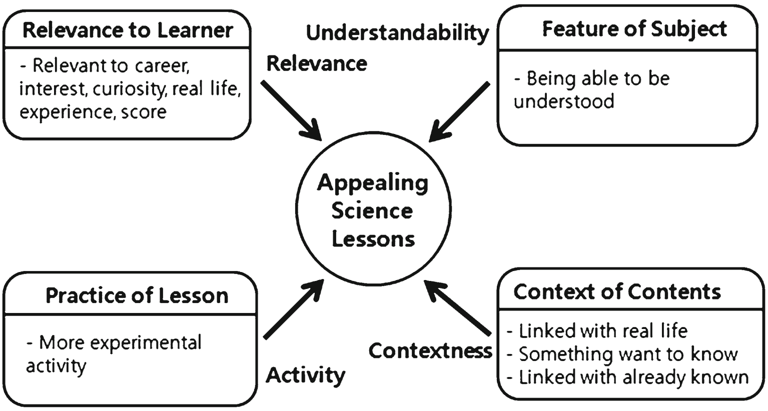


Fig. 15.1 The conditions of science lessons to be appealing to students (Park and Song 2009)

The valuing of science lessons for students was found to be either intrinsic, interest and curiosity/knowledge and common sense/relevance of science to life, or extrinsic, career choice/curriculum related/needs of and application to real life/importance of science and technology in society. Although most students agreed with the importance of science lessons, half of them thought that not all students needed to study science, while the other half supported all students learning science. The analysis of the appealing aspects of science lessons showed that students like and enjoy science lessons mainly for the following reasons: 'as long as I understand, science is easy to study', 'experimental and practical activities are interesting', 'it is good to know something in real life', 'if I learn what I want to know, I will be delighted to learn more' and 'I am happy when I can make a link between what I know already and what I don't know yet'. Based on this analysis, it was claimed that there are four main conditions for science lessons to be appealing to students: understandability of subject, relevance to learners, practical nature of the lessons and the degree of contextualisation of the content (see Fig. 15.1). Importantly, in practice, the last three conditions are the mostly needed ones in Korean science education as repeatedly discussed earlier.

A More Comprehensive Framework of Science Culture Indicators (SCI)

Despite providing a large volume of invaluable data and information about school science education, TIMSS and PISA studies, with their primary focus on cognitive and affective aspects of individual students' science learning in schools, cover only a limited scope of the current status of science education. Today's science education goes far beyond the boundaries of school and curriculum.

During the past several decades, under the names of scientific literacy or public understanding of science, various programmes, projects and surveys for the promotion

Table 15.11 Dimensions, modes and categories within SCI (Song et al. 2008)

	Mode	
Dimension	Potential mode	Practice mode
Individual dimension	Opinion	Learning
	Interest	Application
	Understanding	Participation
Society dimension	Human infrastructure	Event
	Physical infrastructure	Media
	Institutional infrastructure	Civil activity

of science in society have been developed around the world. The term 'science culture', which is being used frequently by science communicators, is also widely used with a meaning similar to that of scientific literacy or public understanding of science, but with a greater emphasis on viewing science as a culture (e.g. Godin and Gingras 2000; Song and Cho 2004). The idea of science culture (also known as scientific culture or scientific and technological culture) has become a main concern of both science educators and science communicators.

As efforts and investments in science culture increase, a need for effective monitoring systems has developed. Alongside this need, several science-related indicators have been developed and used to monitor the status quo of science and technology within nations around the world (e.g. the Science and Engineering Indicators [SEI] of the National Science Foundation in the USA, the European Commission's Eurobarometer in the EU and the Japanese Science and Technology Indicators [STI] in Japan) (see, e.g. EC 2005; KSF 2006; NSF 2006; NISTEP 2004). These indicators are largely concerned with statistical data on national situations related to science and technology in general and do not provide specific and systematic data for science culture.

To meet the need for developing a new indicator system specialised for science culture, Song et al. (2008) developed a comprehensive framework for the monitoring and assessment of science culture called SCI (Science Culture Indicators). Science culture is defined in the SCI as 'all modes and values of life which individuals and society share in relation to science and technology'. Based on this definition, the SCI is divided into two dimensions: personal literacy of science culture (individual dimension) and social foundation of science culture (society dimension). Further, each dimension is divided into two modes: 'potential' and 'practice'. Thus, the SCI has a 2×2 structure consisting of four areas (individual-potential, individual-practice, social-potential, social-practice). Each of these areas is divided into three area-specific categories. As a result, the SCI system has 12 categories (see Table 15.11). In addition, each category is further divided into subcategories and subcategories into indicators. Each indicator contains items which relate to data sources that can be checked or surveyed once or regularly. Thus, the SCI has a hierarchical structure: area → category → subcategory → indicator → item.

The individual dimension of SCI is called SCI-I (Science Culture Indicators for Individuals) and comprises two areas—individual-potential and individual-practice (see Table 15.12). The individual-potential area includes indicators concerning individual

Table 15.12 Areas, categories and subcategories within SCI-I (Science Culture Indicators for Individuals) (Song et al. 2008)

Areas	Category	Subcategory
Potential mode of individual dimension	Opinions about …	S&T and Research
		Formal Science Education
		Human Infrastructure
		Physical Infrastructure
		Institutional Infrastructure
	Interest in …	Learning of Science
		Application of Science in Everyday Life
		Participation of Science in Social Community
	Understanding of …	Scientific Knowledge
		Scientific Method
		Everyday Life and Science
		Science and Society
Practice mode of individual dimension	Learning for …	Formal Science Education
		Acquiring Science Information
	Application in …	Scientific Explanation of Phenomena
		Scientific Habits
		Pseudoscience
		Using High-technology Devices
	Participation in …	Lifelong Education
		Science Club
		Argumentation of Scientific Issues
		Civil Science Activities
		Occupations in S&T

persons' 'Opinion', 'Interest' and 'Understanding' related to science culture. 'Opinion' is a value judgement about social conditions or facts, for instance, a view on a science and technology (S&T) development or an attitude towards an S&T research project. The opinion category is further divided into five subcategories: 'S&T and Research', 'Formal Science Education', 'Human Infrastructure', 'Physical Infrastructure' and 'Institutional Infrastructure'. The interest category encompasses attitudes towards individual actions and practices, including 'Learning of Science', 'Application of Science in Everyday Life' and 'Participation of Science in Social Community'. Finally, 'Understanding' is composed of 'Scientific Knowledge', 'Scientific Method', 'Everyday Science' and 'Relationship between Science and Society'.

The individual-practice area is divided into 'Learning', 'Application' and 'Participation'. 'Learning' and 'Application' consist of indicators that represent individual learning of scientific knowledge and its application to his or her life. 'Participation' comprises subcategories that represent individual participation in scientific communication or in events or activities related to S&T. 'Participation' is a category of vital importance because it can directly monitor the output of various science culture programmes and projects. (Song et al. (2011) provide more details of SCI-S [for society] and an example of its application to the comparison of social foundations of science culture of cities.)

Table 15.13 Correlations among six categories of SCI individual dimension (Cheung 2008)

		Potential mode			Practice mode		
		Opinion	Interest	Understanding	Learning	Application	Participation
Potential mode	Opinion		.498**	−.054	.185**	.161**	.098
	Interest			.087*	.556**	.376**	.429**
	Understanding				.070	.120*	.031
Practice mode	Learning					.324**	.542**
	Application						.340**
	Participation						

*$p<0.05$, **$p<0.01$

Based on this framework of SCI-I, Cheung (2008) carried out a survey of the science culture literacy of middle school students in Seoul. A total of 1406 students were surveyed, drawn from four middle schools (Grades 7 and 9) of which each one half group took one of the two split-half survey items of the whole indicators.

Among the more important results from the statistical analysis of students' responses was that students with higher grades showed more negative opinions and less interest and practice than the lower-grade students, which was not unexpected. Another more interesting finding came from the correlation analysis among the six categories of the SCI individual dimension. While 'Interest' and 'Participation' categories had the strongest correlations with other categories, 'Understanding' had either no significant or very weak correlations with others, which, rather sadly, implies that 'more' knowledge and concepts do not improve other aspects of personal science culture (see Table 15.13).

Concluding Remarks

International comparative studies like PISA and TIMSS have been the subject of many academic discussions (e.g. Britton and Schneider 2007; Fensham 2007a, b; Guo 2007). In particular, TIMSS and PISA, prior to 2006, have been criticised as neglecting the affective aspects of science learning and interest in science (e.g. Fensham 2007a).

As we have seen from the results of TIMSS and PISA, countries in the East-Asian region (Hong Kong, Japan, Korea and Taiwan) show a considerable disparity between the achievements (high) and the engagements (low) in students' science learning. Students in the region always show performances among the very best, but very low and often the worst results in enjoyment, competence, interest in science and science study. This disparity is so persistent that it might be called the 'East-Asian Disparity'.

East-Asian regions are in general quite keen to import and implement the so-called Western approaches in education even though they have maintained such a long history of the Confucianism tradition, a tradition which is very different from the West. For example, as seen above, Confucianism significantly affects the communication between students and a teacher and even among students. East-Asian students also have a group-oriented culture, an attitude of expressing their opinion indirectly, and an expectation of covering large quantities of content in their science classrooms. Possibly due to these cultural features, East-Asian students often remain silent during science lessons, are less active in participating in classroom discussion and in practical work and are passive in expressing their opinions and feelings. These characteristics in turn might contribute to the low levels of their engagement in science learning.

It is, however, not yet clear whether the low status of student engagement is a consequence of the underlying cultures or just a reflection of the tendency to express rather passively and conservatively. For example, it has been suggested that in cross-national survey research, response styles (or biases) are often influenced by culture (e.g. Buckely 2009; Harzing 2006; Tellis and Chandrasekaran 2010). Nevertheless, as mentioned above, recent studies on Korean science lessons have repeatedly shown that science lessons and students in Korea, and likely in other East-Asian regions, have focused on the contents of textbooks, knowledge and concepts of science. Given that there appears to be no matching set of data from the so-called Western regions and cultures, it would be hard to make any definitive claim; however, it seems clear that this tendency is linked with the phenomenon of 'East-Asian Disparity'. What is not yet clear is whether or not this disparity, especially the extreme low levels of student engagement, would be a fair representation of the actual situation and whether or not this disparity can easily be resolved.

Whatever is the case, the disparity in East-Asian regions of high achievement but low engagement is an important and urgent issue to be tackled. Considering the long-term effect of motivation, attitude and experience on the future learning of science (e.g. Lau and Roeser 2002), this disparity could be even worse than its reverse (i.e. low achievement but high engagement). A great deal of effort by science education research and practice must be made to search for the causes, dynamics of the relationship and effective ways for overcoming it. In addition, the issue of the subdivision within East-Asian regions—Hong Kong and Taiwan versus Japan and Korea—should be the subject of further studies with cultural approaches.

As Fensham (2007b) has illustrated, the decisions about what science learning to test, what information relevant to science learning to collect and what mode of collecting data be used are inevitably reflections of the values and backgrounds of the evaluation projects (like TIMSS and PISA) and of the major players inside the projects.

Other researchers have criticised the narrow scope of assessment outcomes in science education, particularly in the case of international comparative studies. For example, Guo (2007), in his article on the international perspective of science education, acknowledged and stressed the recent developments in international comparative studies by saying:

> It is also important that besides achievement in the cognitive domain, student learning outcomes should include aspects such as motivation, self-concept, social-cultural and linguistic aspects, study skills, engagement learning how to learn, global awareness, and the effective use of ICT. (p. 240)

In their edited book focused on the international perspective of the influence of educational context on science education, Coll and Taylor (2008) rightly stressed the importance and vital role of the local context in curriculum and assessment of science education:

> The majority of the science curricula reforms reported in this book have drawn upon Western thinking, and new science curricula are inevitably based on some form of foreign or imported Western science curricula. … So, if not constructivism, certainly a learner-centered approach to teaching and learning is advocated in all curriculum documents. … A learner-centered approach to science education is the necessary yet not sufficient enabler of allowing for educational context in teaching and learning. … So despite this apparent beacon of hope, there has mostly been little recognition of the importance of educational context in learner-centered education. (pp. 356–357)

The significance of the relationship between assessment and cultural background therefore continues to be one of the important issues in science assessment:

> It could also be anticipated that students from cultural backgrounds in which it is unusual, even unacceptable, for the young to question or engage with argument with adult teachers, might find their beliefs about learning and about the role of the learner being challenged. All of these issues have practical import for the teachers and there is need for future research to guide them. (Black and Harrison 2010, pp. 208–9)

Although the disparity between achievement and engagement is an obvious feature of East-Asian regions, the issue of this disparity could be a general phenomenon which appears with varying degrees and directions across the world. In this sense, the current policy of international comparative studies like TIMSS and PISA announcing international rankings only in terms of student achievement needs to be reconsidered in order to also include ranks of students' engagement in science. In addition, more attention should be paid to finding ways of including additional data collection about social and cultural backgrounds.

When Black and Wiliam (1998) introduced the metaphor of black box, the black box was meant to represent the classroom as a place where unseen interactions among important participants must be understood. With the recognition of the existence and seriousness of the East-Asian disparity, we now might be in the position to make efforts to turn a bigger 'black box' (not a classroom but a set of cultural groups) into a 'transparent box'.

References

Black, P., & Harrison, C. (2010). Formative assessment in science. In J. Osborne & J. Dillon (Eds.), *Good practice in science teaching: What research has to say* (2nd ed., pp. 183–210). Berkshire: Open University.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, October.

Bliss, J., Ogborn, J., & Monk, M. (1983). *Qualitative data analysis for educational researchers: A guide to uses of systemic networks*. London: Routledge, Kegan & Paul.

Britton, E. D., & Schneider, S. A. (2007). Large-scale assessments in science education. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 1007–1040). New Jersey: LEA.

Buckely, J. (2009). *Cross-national response styles in international educational assessments: Evidence from PISA 2006*. https://edsurveys.rti.org/PISA/documents/Buckley_PISAresponsestyle.pdf. Accessed 20 Dec 2011.

Cavanagh, S. (2011). Excessive test hurts love of learning, officials says. *Education Week*, Wednesday, April 6, 2011, (Vol. 30, Issue 27, p. 18). http://www.edweek.org/ew/articles/2011/04/06/27korea.h30.html?r=2037452165. Accessed 20 Dec 2011.

Chang, S. J. (2008). A cultural and philosophical perspective on Korea's education reform: A critical way to maintain Korea's economic momentum. *Academic Paper Series* (Vol. 3, No. 2). Korea Economic Institute.

Cheung, M. (2008). *A survey of science culture literacy of middle school students in Seoul using the SCI (Science Culture Indicators)*. Unpublished master thesis of Seoul National University.

Cho, B.-E. (2004). Issues concerning Korean learners of English: English education in Korea and some common difficulties of Korean students. *The East Asian Learners, 1*(2), 31–36.

Coll, R. K., & Taylor, N. (Eds.). (2008). *Science education in context: An international examination of the influence of context on science curricula development and implementation*. Rotterdam: Sense.

Cronin, M. (1995). Considering the cultural context in teaching and learning for Korean tertiary students by western teachers. In L. Summers (Ed.), *A focus on learning* (pp. 53–56). Proceedings of the 4th Annual Teaching Learning Forum, Edith Cowan University, February 1995. Perth: Edith Cowan University. http://lsn.curtin.edu.au/tlf/tlf1995/cronin.html. Accessed 20 Dec 2011.

European Commission (EC). (2005). *Special Eurobarometer 225: Social values, science and technology*. Brussels: Author.

Falk, J. H., & Dierking, L. D. (2000). *Learning from museums: Visitor experiences and the making of meaning*. New York: AltaMira.

Fensham, P. (2007a). Interest in science: Lessons and non-lessons from TIMSS and PISA. In R. Pintó & D. Couso (Eds.), *Contributions from science education research* (pp. 3–10). Dordrecht: Springer.

Fensham, P. (2007b). Values in the measurement of students' science achievement in TIMSS and PISA. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The re-emergence of values in science education* (pp. 215–229). Rotterdam: Sense Pub.

Godin, B., & Gingras, Y. (2000). What is scientific and technological culture and how is it measured? A multidimensional model. *Public Understanding of Science, 9*(1), 43–58.

Guo, C.-J. (2007). Issues in science learning: An international perspective. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 227–256). New Jersey: LEA.

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management, 6*, 243–266.

Kaiser, G., Hino, K., & Knipping, C. (2006). Proposal for a framework to analyze mathematics education in Eastern and Western traditions. In F. K. S. Leung, K.-D. Graf, & F. J. Lopez-Real (Eds.), *Mathematics education in different cultural traditions—A comparative study of East Asia and the West* (pp. 319–351). New York: Springer.

Kim, M. (2011). *A framework development of and a comparative analysis of the factors causing science learning fatigue in science museums and schools*. Unpublished master's dissertation, Seoul National University

Kim, H., & Song, J. (2003). Middle school students' ideas about the purposes of laboratory work. *Journal of Korean Association for Science Education, 23*, 254–264.

Korea Science Foundation (KSF). (2004). *Study of the national survey on the students' recognition toward science & technology*. Seoul: Author.

Korea Science Foundation (KSF). (2006). *Report of public understanding of science and technology*. Seoul: Author.

Lau, S., & Roeser, R. W. (2002). Cognitive abilities and motivational processes in high school students' situational engagement and achievement in science. *Educational Assessment, 8*(2), 139–162.

Lee, J-K. (2001). Confucian thought affecting leadership and organizational culture of Korean higher education. *Radical Pedagogy*, *3*. http://radicalpedagogy.icaap.org/content/issue3_3/5-lee.html. Accessed 21 July 2011.

Lee, W., Kim, H., & Song, J. (2008). Secondary school students' images of Doing-Science-Well. *Journal of Korean Association for Science Education, 28*(1), 1–14.
Leung, F. K. S. (2001). In search of an East Asian identity in mathematics education. *Educational Studies in Mathematics, 47*, 35–51.
Martin, M. O., Mullis, I. V. S., Foy, R. (in collaboration with Olson, J. E., Erberber, E. & Galia, J. (2008). *TIMSS 2007 international science report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades.* Chestnut Hill: TIMSS & PIRIL International Study Center.
National Institute of Science and Technology Policy (NISTEP). (2004). *Science and technology indicators: 2004, A systematic analysis of science and technology activities in Japan*. Tokyo: Author.
National Science Foundation (NSF). (2006). *Science and engineering indicators 2006: Volumes 1&2*. Arlington: Author.
OECD. (2007). *PISA 2006 (Science competencies for tomorrow's world) Volume 1: Analysis*. Paris: Author.
OECD. (2010). *PISA 2009 Results: What students know and can do: Student performance in Reading, Mathematics and Science (Volume 1)*. Paris: Author.
Pang, J. S. (2009). Good mathematics instruction in South Korea. *ZDM Mathematics Education, 41*, 349–362.
Park, D., & Song, J. (2009). When do science lessons appeal to students? Secondary school students' views on the value of school science and the appealing aspects of science lessons to students. *Journal of Korean Association for Science Education, 29*, 593–610.
Song, J., & Cho, S. (2004). "Yet another paradigm shift?" From minds-on to hearts-on. *Journal of the Korean Association for Science Education, 24*, 129–145.
Song, J., Choi, J., Kim, H., Chung, M., Lim, J., & Cho, S.-K. (2008). Developing the indicator system for diagnosing the national status quo of science culture. *Journal of Korean Association for Science Education, 28*, 316–330.
Song, J., Chung, M., Choi, E., Kim, L., & Cho, S.-K. (2011). How to compare the social foundations of science culture: A trial with five cities in Korea. *Public Understanding of Science*. doi:10.1177/0963662511417351. Published online first.
Tellis, G. J., & Chandrasekaran, D. (2010). Extent and impact of response biases in cross-national survey research. *International Journal of Research in Marketing, 27*, 329–341.
Triandis, H. C. (1995). *Individualism and collectivism*. San Francisco: Westview.

Chapter 16
Embedding Assessment Within Primary School Science: A Case Study

Angela Fitzgerald and Richard Gunstone

The reluctance of primary school teachers to teach science in their classrooms is well documented with issues such as limited science content knowledge and low levels of confidence in teaching the subject matter cited as key deterrents. The barriers posed by these issues often result in primary school teachers implementing occasional science lessons that pique their students' interest, but not extending these lessons to promote the development of strong conceptual understandings or monitoring what learning has occurred. However, this is not the case in all primary school classrooms. This chapter documents the approach used by one teacher, Lisa (a pseudonym), to create a coherent set of science learning experiences to meet the learning needs of her students as well as piquing their interest. This case study focuses on her use of assessment. Through embedding assessment into her science teaching and learning approaches, Lisa was able to monitor the development of her students' science ideas, use evidence gathered from her students to inform her own practice and engage her students in assessment as part of their learning experience rather than the much more common approach of treating assessment as an additional or separate process. As Lisa's story unfolds over the chapter, the significant role that assessment can play in developing and strengthening science teaching and learning in primary classrooms is highlighted.

A. Fitzgerald (✉) • R. Gunstone
Centre for Science, Mathematics and Technology Education, Faculty of Education, Monash University, Building 6, Wellington Road, Clayton, VIC 3800, Australia
e-mail: richard.gunstone@monash.edu

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_16,

Assessing Science Learning as a Part of Science Teaching

Assessment is a tool for monitoring and evaluating student learning, but it is the ways in which teachers use this tool that are crucial. Teachers should monitor students' developing science understandings and, based on this, provide them with opportunities, experiences and feedback that will further enhance their learning.

Leach and Scott (2002) refer to the gap existing between students' everyday views or prior understandings of science and the accepted scientific view as 'learning demand'; the greater the difference between these two ways of thinking, the greater the learning demand faced by the student. In undergoing the conceptual growth involved in changing naive existing ideas to scientific ideas, students need to also take responsibility for their own learning through being aware of their existing understandings and how these understandings can be further developed (Baird 1990; Gunstone 1994). It is this process of students monitoring their learning, with the support of their teachers, which assists in bringing about this growth.

Feedback on learning assists students in the process of closing the gap by moving them towards scientifically recognised understandings for the phenomena they are experiencing and exploring (Black et al. 2003; Cowie 2002). Teachers need to consider both how this feedback is managed and offered and how it is integrated into their teaching and learning approaches. In providing feedback on the development of students' science understandings, teachers need to be aware of the purposes for which they use assessment and how they embed assessment within their practices to best support learning.

This chapter draws from a larger study, which examined the science teaching practices of primary school teachers (Fitzgerald 2010), and focuses on one of the participating teachers: Lisa. In this chapter, we explore the ways in which Lisa embedded assessment within the teaching approaches she used to deliver a science-focused unit of work. These practices are being documented to provide some insights into the assessment strategies being used in primary school classrooms. Capturing this is important because when assessment becomes embedded in teaching and learning, it often gets lost among the general classroom goings-on and those 'outside' the immediate classroom are unable to develop a clear sense of what is being assessed and why. This chapter will focus on un-embedding Lisa's use of assessment to enable her assessment practices and the learning consequences of these to be better understood.

Lisa: A Case Study

Lisa, the teacher whose teaching is the subject of this case study, used an instructional approach structured around the 5Es model (engage, explore, explain, elaborate and evaluate). The *Primary Connections* teaching and learning programme that Lisa used in her science teaching is based on this model (Australian Academy of

Science 2007). *Primary Connections,* widely implemented in primary schools across Australia, is a unit-based programme that has been designed to support teachers in teaching science and associated literacies (Hackling and Prain 2005). This case study draws on video footage collected from Lisa's classroom over the course of the astronomy-based *Spinning in Space* unit, consisting of nine weekly lessons averaging 90 min in length. The video data were supplemented with weekly interviews with Lisa, which further explored her objectives for the lesson and her thinking behind the strategies and approaches that she used. Work samples were collected from a small focus group of her students, a volunteer group of students based on Lisa's suggestions of students who would work well together and be willing to communicate their ideas with the researcher. In this chapter, little attention is given to this data source.

This case illustrates the ways in which Lisa enacted each of the 5Es phases of inquiry over the unit. The presentation of the case study is framed by the 5Es—engage, explore, explain, elaborate and evaluate—and the data presented serve to uncover the ways in which Lisa, as the teacher, embedded assessment within her practice to support student learning.

Engage

The engage phase occurs during the first few lessons of a unit and is designed to stimulate the students' interest in a topic. The role of the teacher in this phase is to provide students with learning experience(s) that actively makes connections with their past learning experiences and elicits the students' prior knowledge and understandings of the topic (Bybee 1997). In this case, Lisa used Lesson 1 to conduct the engage phase of the 5Es model.

The *Spinning in Space* science unit fitted within the broader whole-curriculum class theme for the term, Planet Earth. As Lisa explained to her Years 3 and 4 students (aged 7–9 years), in science they would focus specifically on the relationships between the Sun, Earth and Moon and what causes day and night. In this lesson, she encouraged students to think about their own understandings and experiences of the topic. She explained: 'I wanted them to start thinking about what causes day and night, and I wanted them to start make connections about things that they see in their own life and own experiences'.

Lisa generated interest and curiosity in the topic by showing two clips of footage from *YouTube*. The first clip was a series of satellite images of Earth and the second clip used time-lapse photography to show changes to the city of Melbourne over 24 h. She was selective about her choice of footage because she wanted to engage the students, not perpetuate or develop any alternative conceptions about the topic. 'You actually have to be very selective because a lot of the videos actually show the Sun moving across the sky because that is what people want to see.' Her choice of moving images connected with ways in which the students experienced these phenomena in their lives.

Lisa used whole-class discussion to tease out an observation made by one of the students, who noticed during the time-lapse imagery that '[the city] was going from day to night' (video footage). This short interaction led to a brainstorming session about the differences between day and night. Lisa used prompts such as 'How do we know when it is day or night?' and 'What might we see if it is day or night?' (video footage). This provided all students with the opportunity to contribute their ideas (video footage).

A TWLH chart (a strategy for recording what you *t*hink you know, what you *w*ant to know, what you have *l*earnt, and *h*ow you know what you have learnt) was used by Lisa as another way of eliciting the students' ideas about the new topic. In Lesson 1, the focus for contributing to the chart was limited to the sections of the chart headed T, what the students thought they knew about the topic, and W, what they wanted to know. Below is some dialogue between Lisa and her students documenting what they thought (T) they knew about the Sun, Earth and Moon (video footage).

Teacher	What are some of the things that we think we know already about the Sun, Earth and Moon? Ben, start us off.
Ben	When the Moon is a crescent, it's in the shade.
Teacher	In the shade? Can you explain that just a little bit more, Ben?
Ben	It's in the shade of the Earth and when the Sun comes around the Earth is blocking the Moon, so it only gets a little bit of Sun and you can see a shadow.
Teacher	OK. Thank you for clarifying that for me, Ben. Great. Simon, what else do we think we know?
Simon	That the Sun and Moon travel from one side of the world to the other each time.
Teacher	OK. Interesting. Leah?
Leah	The Sun shines at day and it goes away at night.
Teacher	Thank you very much for that Leah. Andrea?
Andrea	One side of the Earth has the Sun and the other side of the Earth has the Moon.
Teacher	Great. Rachel, last one.
Rachel	When the Sun goes away, it's the Sun having a rest.
Teacher	OK.

The students also had the opportunity to pose questions about what they wanted (W) to know about this science topic. Some questions included the following: How does the Sun disappear at night? What is the Moon made of? How does the Earth spin if there is no wind or air?

Lisa involved students in creating a word wall, a strategy for brainstorming words specific to a topic. The students contributed words related to the Sun, Earth and Moon, such as 'craters', 'star' and 'gravity', as well as words not directly linked to the topic, such as 'Venus', 'aliens' and 'calculus' (video footage). Lisa entered all the students' responses into a Word document, despite the focus drifting away from the key aspects of the topic. While this session took a different path to what she expected, Lisa believed it was important for students to express their ideas and experiences during this engage lesson.

> Every child should feel that what they're saying is valued even though it may simply not be relevant. You don't want to be cutting them off in the engage lessons when you really want them involved.

Lisa revisited the word wall in Lesson 2 as a way of refocusing students' attention. They brainstormed more words related to the Sun, Earth and Moon in small groups before sharing their ideas with the whole class (video footage). This resulted in words that better reflected the topic such as 'hot', 'light' and 'gas' for the Sun; 'oxygen', 'gravity' and 'round' for the Earth; and 'phases', 'crescent' and 'craters' for the Moon (video footage).

Lisa elicited each student's existing ideas about the topic through their individual responses to a worksheet about how day and night occur and their labelled scientific diagrams showing the relationship between the Sun, Earth and Moon. The students answered three questions about day and night on the worksheet.

While this lesson was structured to engage student interest in the new science topic, interlaced throughout were assessment strategies. In particular, diagnostic assessment strategies played a particularly dominant and important role in Lisa's teaching during the early stages of this unit. It is widely recognised that students enter into science lessons with prior knowledge about the concepts to be taught (Duit and Treagust 2003). While these understandings may seem sensible and coherent to the individual, it is not uncommon for a mismatch to exist between these ideas and the views that are universally accepted by the scientific community. Lisa's use of diagnostic assessment in this lesson enabled her to generate evidence about her students' existing ideas related to this astronomy-based topic and begin to identify the existence of alternative conceptions. This gathering of information was achieved both formally, in this case through the students' responses to the worksheet, their diagrammatical representations, the TWLH chart and word wall, and informally, as evidenced by the whole-class discussions that were instigated.

A significant next step is how this diagnostic information is then used to inform the development of the unit and the scientific understandings of the students. Consistent with constructivist teaching approaches, Lisa could have used the understandings that she had had gathered from these various modes of diagnostic assessment to assist in identifying the specific learning targets for each of her students and then adapted her teaching approach to best enable her students to reach these goals. However, in reality, this is not always easy to achieve. An existing unit of work, the *Primary Connections* module, was guiding Lisa's teaching approach and the identification of individualised learning targets for 27 students is a time-consuming task. In this case, Lisa used the evidence gathered through the diagnostic assessment tasks as points of reference to be revisited over the unit, such as the TWLH chart and word wall, or to be used by herself to gauge the development of students' understandings over the unit, such as the labelled diagram and worksheet responses.

Explore

The explore phase allows teachers to provide students with hands-on experiences of the science phenomena behind the topic. The role of the teacher in this phase is to provide students with a range of shared experiences to allow the concepts and processes relevant to the unit to be identified, explored and developed (Bybee 1997). This phase involved Lessons 2–4 of Lisa's teaching and learning sequence.

Lisa used several activities in Lesson 2 to enable students to explore the relative sizes of the Sun, Earth and Moon. A think-pair-share strategy was used to promote student thinking about the types of objects that are the same shape as the Sun, Earth and Moon. The students referred to objects such as an orange, balloon and various types of balls (e.g. basketball, netball, baseball) (video footage). These ideas led to a whole-class discussion about distinguishing spheres from circles. Lisa encouraged students to share their understandings of these concepts (video footage).

Teacher	It's really hard to draw these shapes on a piece of paper. But how is a sphere different to a circle? Simon, what do you think?
Simon	It's a three-dimensional shape.
Teacher	Fantastic. Simon, what do you mean by three-dimensional?
Simon	It's sort of like a cube, but round shaped.
Teacher	OK. Thank you, Simon. Georgia, what do you think?
Georgia	Well, two-dimensional shapes are flat, but then three-dimensional they're all different shapes. They're like two-dimensional shapes that have been pumped up.
Teacher	Fantastic. That's a good explanation, Georgia. Thank you. Naomi, what would you like to add?
Naomi	With a sphere, if you put something in the middle, it should be the same distance to each edge around it.
Teacher	Wow. Thanks Naomi. Andrea?
Andrea	Mine's a bit like Georgia's. A sphere is rounded, but a circle is just flat.
Teacher	Fantastic.

Lisa showed the students three spherical objects: a peppercorn, a marble and a basketball. The students agreed in a whole-class discussion that the basketball could be used to represent the Sun, the marble the Earth and the peppercorn the Moon. Lisa turned the classes' attention to the Sun and the Moon by explaining that there is a common misconception that the Sun and the Moon are the same size as they appear to be the same size in the sky when viewed from Earth. To explore this idea, the students moved outside to complete an activity in small groups. The activity required one student to hold a tennis ball (representing the Moon), while another student holding a basketball (representing the Sun) moved away from the first student until that (first) student perceived both balls to be the same size. Lisa led a discussion following the activity to assist the students in connecting their experiences of this activity with the idea that the apparent similarity in sizes of the Sun and Moon are due to the Sun being much further in distance from Earth than the Moon.

The following dialogue captures how Lisa used questioning to assist students in making connections between the activity and the concept (video footage).

Teacher	What did you notice? What did you see? Ella?
Ella	When we were taking the basketball back, when the basketball looked about the same size as the tennis ball we normally stopped around the start of the cricket pitch.
Teacher	OK. Fantastic. So which one was further away Ella, the tennis ball or the basketball?
Ella	Basketball.
Teacher	OK. Fantastic. How does this then relate to the Moon and the Sun? How does this help us understand how the Moon and Sun look about the same size?
Andrea	Because the Sun is further away than the Moon and because when we did [the activity] we held the Moon and said stop when [the Sun] looked about the same size. Even though [the Sun] was further away than the Moon, it looked the same size because it is bigger.
Teacher	Fantastic. So which one is bigger, Andrea?
Andrea	The Sun.
Teacher	Why did the Sun look about the same size [as the Moon]?
Andrea	Because it was further away.
Teacher	Fantastic.

The students created scale models of the Sun to further strengthen their understanding of the relative sizes of the Sun, Earth and Moon. Lisa elicited the students' personal experiences and understandings of model making before undertaking the activity. Lisa explained that 'a model is a representation of the real thing, so it's not the same size, it's a lot smaller, but it's a way to show what a car [for example] might look like' (video footage). Lisa created scale models of the Earth (1-cm-diameter circle) and the Moon (2.5-mm-diameter circle) for each group, while in small groups the students created 1-m-diameter models of the Sun. Lisa asked the students to predict how far they would need to stand apart for their Sun model to look the same size as the Moon model.

The students wrote about their experiences of these activities in their journals. Based on the students' journal entries, Lisa recognised the different connections that students made between the activities and the relative sizes of the Sun, Earth and Moon.

> There's a few in the class that sort of missed the idea, and the thing is that some of them can verbalise that to me, but they're not always able to write that. But when you look at the focus group, they've all got that idea that the basketball was further away or now they've transferred that to the Sun is much, much bigger than the Moon [due to] a much further distance.

The assessment focus shifted in this phase of the unit from determining what the students know about the science concepts as the topic is begun to developing an understanding of how their ideas are forming during the learning process. To assist in this process, Lisa engaged her students in a number of activities that continued to move their understandings towards the scientific explanations for the phenomena under study. Formative assessment practices, as they were used in this instance, are inherently part of the instructional process. Therefore, the ways in which these teaching and learning sequences were structured enabled Lisa to gather information about how her students' understandings were developing.

As seen with Lisa's use of diagnostic assessment, the gathering of formative information from students played out in formal and informal ways. This notion that the formative assessment used by teachers can be categorised as formal or informal has been explored and described by Cowie and Bell (1999) using the terms 'planned' or 'interactive' (p. 102). They argue that planned formative assessment occurs when the teacher decides what will happen before the lesson starts, whereas interactive formative assessment occurs when the teacher responds spontaneously as opportunities arise. Over the three 'explore' lessons, the planned formative assessment included students completing journal entries, participating in modelling activities and contributing to the TWLH chart. The products provided Lisa with concrete examples of her students' understandings. In developing these products, students were provided with the opportunity to reflect on and articulate their understandings.

An important aspect of formative assessment is that students are involved in this process to ensure that they are also informed about the development of their own understandings (Black et al. 2003). In this phase of the unit, this was probably most recognisable in the completion of the TWLH chart in Lesson 4, where the students were able to identify what they had learnt over the previous lessons and what evidence they had to support that they had learnt those ideas. Two students explained to the class what they had learnt about shadows and how they knew this (video footage).

Teacher	What is something that we have learnt (L)? Think back to the activities we have done. Rachel, what's something we have learnt?
Rachel	We learnt about day and night.
Teacher	What about day and night? You need to be more specific.
Rachel	How it's dark at night and light in the day.
Teacher	How do we know that Rachel?
Rachel	Umm.
Teacher	How (H) do you know when it's day and night? Which of our senses do we use?
Rachel	Because when it's night, we can't see many things because it is dark and in day, you can see lots of things.
Teacher	Well done. Excellent. Brilliant. And I like how people are matching up what they've learnt with some observation or some activity that we've done that helps them to know that.

Interactive formative assessment took place over the three lessons mainly through the opportunities that Lisa and her students had to engage in discussion, which allowed students to express their science understandings and opened up avenues for Lisa to recognise and respond to their ideas. With the goal of formative assessment essentially being to gain an understanding of what students know and do not know in order to adapt teaching and learning appropriately, techniques not commonly viewed as assessment tools such as teacher observations and classroom discussion have a particularly important place in this process (Boston 2002). While it is evident that Lisa did draw upon a range of formal/planned and informal/interactive procedures to inform the formative assessment process, what was not clear was exactly how she used this information to modify her practices over these lessons to assist and enhance student learning in science.

Explain

The explain phase requires students to discuss and develop explanations of the scientific phenomenon they are encountering to make sense of their observations and experiences. The role of the teacher in this phase is to provide students with opportunities to represent their conceptual understandings and ensure that they are aligned with current scientific understandings (Bybee 1997). Lisa used Lesson 5 to conduct the explain phase of the 5Es model, providing the students with learning experiences that introduced them to the current scientific views about what causes day and night and supporting them to represent their understanding through creating and performing a role-play (video footage). Her focus was for students to recognise that day and night were caused by the Earth rotating on its axis.

Lisa used five different demonstrations to represent how day and night occur. She believed that it was important that several examples were provided to support the development of students' understandings of the scientific explanation for how day and night occur:

> I really had to get across that idea that it was the Earth moving and not the Sun because a lot of them still had that idea that it was the Sun that was moving across the sky. And the Sun does appear to move across the sky, but that's because the Earth's rotating. I just wanted to make sure because we were in that explain phase that I was very clear that that was what was actually happening.

First, using a basketball to represent the Earth with a small wooden stick attached as an object on the Earth and a torch to represent the Sun, Lisa asked the students to share their observations of what happens to the shadow of the stick as the Earth rotates. The students noticed that the shadow was moving and Lisa reiterated that as the Earth moves, so do the shadows being formed on the Earth, while the Sun stays in the same position (video footage).

Second, Lisa asked a student to represent the spinning Earth by spinning around in front of the data projector, which represented the Sun. As the student rotated around, Lisa asked the class several questions related to what they observed happening (video footage).

Teacher Now pretend that Keisha is the Earth and the data projector is the Sun. As Keisha starts to rotate, what do you notice about Keisha as she is rotating slowly? What parts of her are in the light? What parts of her are in the dark? Georgia, tell me, what do you notice?

Georgia	The light is shining on her.
Teacher	Where exactly is light shining? Would someone else like to add to that? Andrea?
Andrea	When she turns around, the dark side is always opposite her because it's not facing the data projector. So if she was the Earth, one half would be like a shadow on the Earth.
Teacher	Excellent. As Keisha is standing now, which part of her is in the light? And you can all see this, so I should see all hands-up. Dana?
Dana	Her back.
Teacher	Which part of Keisha is in the shadow or hasn't got light shining on her? Leah?
Leah	Her face.
Teacher	Fantastic.

Third, Lisa added three more students to this model. The four students formed a circle and rotated around in front of the projector. Again, Lisa asked the rest of the class to respond to questions, such as 'When do the students start to come into or go out of the light?' (video footage). To create a more direct link to the occurrence of day and night, Lisa then connected this model to the Sun (data projector light) and the Earth (the ring of four students) by asking the students to identify which parts of the Earth were experiencing day and night. After repeating this line of questioning several times, Lisa asked the students to explain why they thought those parts of the Earth were experiencing day and night. Fourth, Lisa showed the students a clip from *YouTube* based on time-lapse footage from the space station Galileo showing the Earth rotating around its axis. After watching this clip, Lisa provided the students with the opportunity to share their observations with the class. She also used questioning to elicit what the students knew about how long it takes the Earth to rotate once on its axis (i.e. daily) and once around the Sun (i.e. yearly). Finally, Lisa used three student volunteers to demonstrate the movements of the Sun, Earth and the Moon. Lisa asked the student representing the Sun to remain still, while the student representing the Earth rotated around while moving around the Sun. She then added the student representing the Moon, who moved around the Earth.

Following this teacher-led modelling, Lisa provided students with the opportunity to explain their understandings of how day and night occurred by creating their own role-plays, which they performed for the class. Lisa felt that this activity was an effective way for the students to show their understandings of the phenomena being studied.

> It worked really well and [despite] the low literacy level of a lot of the kids actually doing it, the role-play was really good. I'd use it again, especially like I said [with] the low literacy levels in the classroom it's a good way, a different way for [the students] to explain their science without having to write it down.

However, as the students performed their role-plays, Lisa noted some confusion among the groups regarding the role of the Moon in causing day and night. For example, one group explained 'when the Moon is on one part of the Earth, it's night time [and] on the opposite side, the Sun is shining so it's daytime' (video footage).

Another group explained 'day is made by the Sun shining on the Earth, but when the Moon comes to this side and blocks the Sun's light on the Earth that makes night time' (video footage). Lisa addressed this issue by again modelling how day and night occur using a torch (Sun), globe (Earth) and a tennis ball (Moon). She did explain that sometimes the Moon does block the Sun's light from reaching Earth, which is known as an eclipse. Lisa believed, in hindsight, she should have left the Moon out of the role-play to lessen the conceptual confusion of the students.

> Unfortunately, I should have left the Moon right out of it because then they got that idea that the Moon was causing the day and the night. But I think by following that up at the end, talking about that idea of the eclipse rather than day and night really helped. [However] when I went around and was reading their responses to what causes day and night, [some of the students] still had [the notion] that the Moon causes day and night.

This phase of the unit also included the use of formative assessment to monitor and provide feedback on the ongoing development of the students' ideas. However, in contrast to the previous sequence of lessons, Lisa was now more intent on students' developing and representing strong conceptual understandings that were aligned with current scientific explanations of the astronomy-based phenomena being studied. Again, a structured sequence of activities was implemented; in this case, a set of demonstrations and a series of role-plays, which enabled Lisa to gather some planned formative information. Information through the more informal mode of interactive formative assessment was also gathered from the student group through discussions that took place around the set of demonstrations that Lisa orchestrated, the *YouTube* clip that was shown and the role-plays performed by small groups of the students. Resulting from this array of information was Lisa's realisation that an alternative conception had formed in the students' understandings regarding how day and night occur.

In effectively using formative assessment to improve student learning, it is important that feedback is given promptly to enable students to take account of it in their learning (Scottish Qualifications Authority 2009). By doing this quickly, students are motivated to make changes to their understanding while it still holds meaning to them. Lisa was able to react instantly as part of her response to the students' role-play performances through an additional teaching sequence. In this instance, Lisa's rich science pedagogical content knowledge enabled her to recognise students' stages of conceptual development and respond in ways that supported their conceptual growth and change.

Elaborate

The elaborate phase focuses on students planning and conducting an investigation as a way of applying and extending their conceptual understandings in a new context. The role of the teacher is to challenge the students' conceptual understandings by providing new experiences in which the students can develop a broader understanding of the science phenomena under examination (Bybee 1997). This phase involved

Lessons 6 and 7 of Lisa's sequence, during which Lisa provided students with the opportunity to conduct an investigation which examined two questions: 'What happens to the length and direction of shadows during the day?' and 'When are the shadows the longest and the shortest?' These questions were designed to assist students in applying and further developing their understandings of the ways in which the Earth moves in relation to the Sun and how this causes day and night.

Lisa worked with the students in a step-by-step approach to plan the investigation (video footage). She informed the students of the variables they would change (the time of the day) and measure (shadow length and direction). The students were provided with an investigation planner, one of the resources included in the *Spinning in Space* module, and used this to individually record a prediction about what they thought they would find out from the investigation, identify the variables that needed to be kept the same during the investigation and respond to the question, 'How are you going to keep it a fair test?' Lisa explained how the equipment associated with this investigation would be set up by sketching a diagram on the board. The class then moved outside onto the school oval where Lisa modelled how the equipment would be used and demonstrated how the students would record their findings over the day. In small groups, the students set up their shadow-stick investigation. They conducted their investigation by recording the length and position of the shadow cast by the stick at hourly intervals from 10 am to 3 pm.

The students shared their observations of the shadow-stick investigation with the class in the following lesson. Lisa used questioning to further elicit their observations and understandings of how the shadow was formed (video footage).

Teacher	What did you notice happening with the shadows during the day yesterday? What did you see? Ewan, what did you see?
Ewan	The shadows kept moving anti-clockwise.
Teacher	OK. What did you notice happening, Leah?
Leah	Well, when the Earth spins …
Teacher	No. I don't want explanations. I want to know what you saw.
Leah	Well, the shadows were moving around.
Teacher	What else did you notice about the shadows during the day? Joseph?
Joseph	They got smaller.
Teacher	The shadows got smaller. You saw them getting smaller. What else did people see? Ella?
Ella	I saw them getting bigger.
Teacher	You saw them getting bigger as well. Michael, what did you see?
Michael	I saw them move in a different direction to where the Sun was.
Teacher	Fantastic observation, Michael. Rachel?
Rachel	The shadows moved to the left.
Teacher	Fantastic. Why do we get those shadows? How were the shadows being formed? Imogen?
Imogen	The Sun was on this side and the pencil was in the middle and on the other side the shadow was formed.
Teacher	And why is that shadow formed Imogen? Can you explain that for me?

Imogen	Ah, because the pencil is in the way of the Sun.
Teacher	And so the pencil is doing what to the light to get a shadow?
Imogen	Blocking it.
Teacher	Good work, Imogen. Excellent answer.

The interpretation of the data and overall evaluation of the investigation was based on questions outlined in the investigation planner (e.g. What happened to the direction of the shadow during the day? What challenges did you experience doing this investigation?). Initially, the students participated in a whole-class discussion focused on assisting them in evaluating their investigation, particularly some of the challenges they faced, because Lisa found that the students experienced difficulty in this area. The students then worked in small groups to discuss their data in relation to six questions. The dialogue below captures how one group of students interpreted aspects of their data. During this dialogue, Lisa entered the discussion and used questioning to monitor the students' understandings of the key conceptual areas (video footage).

Teacher	Why did the shadow change?
Ella	Because the Earth rotates, so the Sun is pointing from a different direction to make the shadow.
Teacher	So what is moving? The Earth or the Sun?
All	The Earth.
Teacher	So why is the Sun in different positions in the sky?
David	It's not really in different positions in the sky, it just looks like it's in different positions.
Michael	The Earth is spinning.
Teacher	That's right. Because we are moving, the Sun appears in different positions in the sky. Great. It actually does look like that David because in the morning it's over here (pointing to the East), lunchtime it is up there (pointing to the zenith) and in the afternoon it's over there (pointing to the West). So it is in a different position because of the Earth's rotation. OK, go on. What's your next one? What happened to the direction of the shadow during the day?
Ella	It changed in an anti-clockwise direction. It moved in an anti-clockwise direction.
Teacher	OK, great.
David	And it changed by moving to the left.
Teacher	OK, good. What happened to the position of the Sun during the day?
Georgia	It's moved!
Ella	It looked like it moved, but it was actually our Earth rotating.
David	Well, let's just say, it didn't move, it just looks like it did.

This phase of the unit had two objectives: to extend the students' conceptual understandings of the topic being studied and to evaluate their investigative skills. In assessing student achievement across these dual-purpose lessons, Lisa drew on both formative and summative assessment practices. For example, as in previous lessons, Lisa integrated both planned and interactive modes of formative assessment into her teaching and learning approach. In this instance, the planned was made up of individuals' responses to a set of predetermined questions, and the interactive occurred through Lisa's monitoring of the small group discussions that were encouraged to tease out ideas linked to these questions. In contrast to the explain phase, Lisa's use of this information seemed more focused on further clarifying and reinforcing the students' conceptual understandings in preparation for the final evaluative phase of the unit than on determining how her future lessons needed to be modified to address gaps in learning. Therefore, her use of formative assessment was applied differently to the ways in which it is usually envisaged.

The use of summative assessment in this lesson connects to a more formal process of recognising the students' achievement of the investigating outcomes. This process resulted in the production of a final product from each student, which consisted of his or her responses on the investigation planner. While Lisa used this information to gain insights into the level of each individual's investigative skills, it is difficult to capture an overall sense of individual achievement of investigative outcomes through this type of product, which was completed as part of small group work and involved skills that were examined over a short period of time. It needs to be recognised that the students would have been developing their science investigating skills over the whole course of their primary education to this time (the previous 3–4 years). Therefore, Lisa's use of summative assessment in this case is an example of being provided with a snapshot of an individual's learning at a particular point in time (Garrison et al. 2009).

Evaluate

The evaluate phase provides students with the opportunity to reflect on their learning experiences over the unit by creating a product that represents their conceptual understandings. The role of the teacher is to encourage students to express their understandings and to create an appropriate opportunity from which to assess student progress over the unit (Bybee 1997). The evaluate phase involved Lessons 8 and 9 of Lisa's teaching and learning sequence, in which students created a poster (Lesson 8) and presented it to their peers (Lesson 9) (video footage). Lisa used an assessment rubric to provide scaffolding for what science information students would need to include on their posters. Lisa invited the students to use their experiences of creating posters to add other criteria to the rubric. Through a whole-class discussion, the students added Presentation, Titles and Spelling as additional areas to be assessed. As part of their poster, the students were required to include their understandings of the sizes, shapes, positions and movements of the Sun, Earth and Moon.

The TWLH chart was also revisited in Lesson 8 to enable students to reflect on what they had learnt (L) from the *Spinning in Space* topic so far and what evidence they had to demonstrate how (H) they developed this understanding. The following dialogue captures how students explained to the class what they had learnt about day and night and what evidence they had to support how they knew this (video footage).

Teacher	What can we add to our TWLH [chart]? What is something else that we have learnt? Ruby?
Ruby	When one side of the Earth is facing the Sun, it is day.
Teacher	So what would be the point that would lead from that? If one side is facing the Sun and that is daytime, what then goes with that? Ella?
Ella	The side that is not facing the Sun is called night time.
Teacher	Excellent. What evidence or what have we seen in the classroom to know that? We know from our own experiences, but what evidence have we seen in the classroom to help us understand that? We've done a couple of things to help us with that. Ben, what was one of those things?
Ben	The light through the windows.
Teacher	Yes, we can see that. But what activities have we done in the classroom to help us understand that day and night occurs? Leah?
Leah	When the people stood in front of the data projector and we could see them coming in and out of the light as they spun around.
Teacher	Excellent. One of our role-plays. Great what other role-plays did we do to help us understand about day and night? Michael?
Michael	We did a role-play where we had to explain how day and night occur.
Teacher	Fantastic. OK, is there anything else that we want to add? Actually, there was something that we did to help us with day and night. What other evidence have we looked at? We looked at it last week. Andrea?
Andrea	Images.
Teacher	That's right. We have also looked at images from space. And this leads back to what we talked about yesterday, different ways of learning. We've used the role-plays to help us, we've used pictures to help us.

Students presented their finished posters to their peers in Lesson 9 (video footage). The students were each given one minute to explain their posters to a small group of their peers. Lisa deliberately chose not to assess the students on their presentation skills, focusing instead on the conceptual understandings that were evident in the poster:

> I'm not going to have an assessment rubric on the presentation as such because we do, do a lot of assessing with their listening and speaking with their news. But because they're trying to explain their science, I don't want them actually worrying about anything else. I want them to concentrate on telling each other about the science.

This final phase of the unit provided students with the opportunity to represent their learning over the course of the unit and as a formal summative process enabled Lisa to identify what each student had achieved and what conceptual understandings had and had not been acquired. To begin this process, the TWLH chart was revisited. The completion of this activity provided an opportunity for students to share what they had collectively achieved in this science unit. This information also acted as a reminder of what had been covered over the unit and was used by students to inform the creation of their own individual summative piece of work. While the notion of individual summative assessment can provoke anxiety in some students (National Centre for Fair and Open Testing 1999), Lisa adopted several strategies that acted to keep adverse effects to a minimum. The production of a poster as a means of showing what had been learnt was considered by students to be an enjoyable task and also provided them with some flexibility and choice in the ways they presented their learning. The negotiation between Lisa and her students about what should be included in the marking rubric also engaged students in the summative process as their opinions about what mattered, in terms of how their posters would be assessed by her, were valued. Finally, the poster presentations, which were not assessed, provided the opportunity for students to showcase their work, and ultimately their learning across the unit, to their peers in ways that were nonthreatening and celebratory.

In the final lesson of the unit, it became evident that Lisa thought of summative assessment as a two-way process when she finished the lesson and the unit by asking students to reflect on their learning experiences using a PMI chart (i.e. a strategy for recording *P*ositives, *M*inuses and *I*nteresting things). She highlighted the importance of thinking carefully and identifying at least three points for each area. Lisa had found in the past that students had difficulty reflecting on their learning and, in particular, identifying minuses.

> The thing is [that] they always associate the minus [section] with bad and getting into trouble, and I think that's just a logical progression. Whereas, [I'm] trying to get them around to see that the minuses actually help us learn and help us do it better for next time.

The opportunity that she provided students through the PMI chart strategy was essentially summative feedback for her that took into account student opinions of the unit and how it supported their learning in science. This information was invaluable to Lisa and her teaching practice but also engaged students in the process of thinking critically about their own learning in science. Therefore, Lisa's approach to summative assessment not only provided her with information about student learning over the unit but also identified what helped, what did not and what improvements could be made to better support student learning in the future.

Using Assessment to Support Student Understanding in Science

Constructivist approaches to teaching and learning emphasise the influence of learners' prior experiences on the ways their understandings are constructed from new experiences or information (Fensham et al. 1994). Conceptual change models for teaching science seek to examine students' existing ideas about particular science phenomena before engaging them in different learning experiences. These approaches are focused on challenging existing ideas and developing understandings more closely aligned to currently accepted scientific views (e.g. Scott et al. 1992; for a perspective written specifically for intending and practising primary science teachers, see Skamp 2008). Assessment plays a key role in identifying these ideas and monitoring how understandings develop.

The level of conceptual change, or learning demand, required over the *Spinning in Space* unit of course varied from student to student. For example, the focus group students explained that many of the science phenomena introduced in the unit were not new to them (video footage). Therefore, the shifts required in their thinking may not have been significant. However, this clearly was not the case for all students. For example, Rebecca explained in Lesson 1 that day changed into night because 'the Sun [has] a rest' (video footage). While the amount of conceptual change required may have differed from student to student, each was individually supported by Lisa through the multiple opportunities she embedded within her teaching practice that enabled her to monitor students' learning and provide feedback on their progress.

Lisa's use of an inquiry-based approach allowed her to embed diagnostic, formative and summative assessment techniques into her ways of teaching as well as into her students' learning (Australian Academy of Science 2007). Evidence of these three types of assessment being used as part of Lisa's repertoire is highlighted within the case study. However, perhaps most notable was Lisa's use of formative assessment, particularly during the explore, explain and elaborate phases of the *Spinning in Space* unit, to monitor and provide feedback on the development of students' conceptual understandings.

This case study of Lisa's science teaching and learning approach with a particular focus on her embedded use of assessment is an example of the ways in which expert teachers use assessment as a pedagogical tool to foster student learning. By actively monitoring students' science understandings and providing appropriate feedback on the development of these understandings through embedding assessment within teaching approaches, teachers can better support their students' learning in science. A central issue in raising the quality of science learning and teaching, in all primary, secondary and tertiary classrooms, is finding ways to help many more teachers develop the knowledge, skills and confidence to use assessment in this highly engaging and positive way (see Nilsson and Loughran, this volume).

References

Australian Academy of Science. (2007). *Primary connections: Spinning in space*. Canberra: Australian Academy of Science.

Baird, J. R. (1990). Metacognition, purposeful inquiry and conceptual change. In E. Hegarty-Hazel (Ed.), *The student laboratory and the science curriculum* (pp. 183–200). London: Routledge.

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

Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation, 8*(9). http://PAREonline.net/getvn.asp?v=8&n=9. Accessed 19 Oct 2011.

Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices*. Portsmouth: Heinemann.

Cowie, B. (2002). Re-viewing conceptual change through a formative assessment lens. In R. K. Coll (Ed.), *Science and technology education research papers* (pp. 164–179). Hamilton: Centre of Science and Technology Education Research, University of Waikato.

Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education, 6*(1), 101–116.

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*(6), 671–688.

Fensham, P. J., Gunstone, R. F., & White, R. T. (Eds.). (1994). *The content of science: A constructivist approach to its teaching and learning*. London: Falmer.

Fitzgerald, A. (2010). *Beliefs, knowledge and practices of effective primary science teachers.* Unpublished thesis, Edith Cowan University, Perth.

Garrison, C., Chandler, D., & Ehringhaus, M. (2009). *Effective classroom assessment: Linking assessment with instruction*. Westerville: National Middle School Association.

Gunstone, R. F. (1994). The importance of specific science content in the enhancement of metacognition. In P. J. Fensham, R. F. Guinstone, & R. T. White (Eds.), *The content of science: A constructivist approach to its teaching and learning* (pp. 131–146). London: Falmer.

Hackling, M. W., & Prain, V. (2005). *Primary connections: Stage 2 research report*. Canberra: Australian Academy of Science.

Leach, J. T., & Scott, P. H. (2002). Designing and evaluating science teaching sequences: An approach drawing upon the concept of learning demand and a social constructivist perspective on learning. *Studies in Science Education, 38*(2), 115–142.

National Center for Fair and Open Testing. (1999). *The value of formative assessment.* http://www.fairtest.org/examarts/winter99/k-forma3.html. Accessed 19 Oct 2011.

Scott, P. H., Asoko, H. M., & Driver, R. (1992). Teaching for conceptual change: A review of strategies. In R. Duit, F. Goldberg, & H. Niedderer (Eds.), *Research in physics learning: Theoretical issues and empirical studies* (pp. 310–329). Kiel: IPN.

Scottish Qualifications Authority. (2009). *Guide to assessment*. http://www.sqa.org.uk/sqa/files_ccc/GuideToAssessment.pdf. Accessed 19 Oct 2011.

Skamp, K. (Ed.). (2008). *Teaching primary science constructively* (3rd ed.). South Melbourne: Cengage Learning.

Chapter 17
Formative Assessment in Learning to Teach Science

Pernilla Nilsson and John Loughran

Introduction

Worldwide, trends in education policy appear to be based on the assumption that testing is a relatively straightforward and undisputed way of assessing where and how students are placed in relation to the acquisition of knowledge in a range of given domains, such as age cohort and content area. We would argue that such a view of assessment is limited and that, by extension, assessment has often become synonymous with simply judging or ranking students according to test outcomes of ability and/or competence. Assessment then is typically understood in terms of grades or measures that can be applied in ways that assist in selecting for, or ordering, perceived ability.

Attempts to challenge superficial views of assessment are perhaps implicit in the differentiation of two forms of assessment: summative and formative. However, even though such a differentiation might be helpful in our thinking about assessment *of* learning versus assessment *for* learning, the relation between the two forms of assessment is not definite. Discussion of the explicit relationship between summative and formative assessment processes has been absent from much of the educational discussions (Taras 2008). While summative assessment tends to report how much students have learned at a particular point in time, formative assessment provides teachers and students with information they need during the learning process to make decisions that will bring about more learning.

Yorke (2003) highlighted that part of the problem of making explicit the relation between formative and summative assessment resides in the duality of meaning of the word 'assessment':

P. Nilsson (✉)
Halmstad University
e-mail: Pernilla.Nilsson@hh.se

J. Loughran
Monash University, Building 6, Wellington Road, Clayton, VIC 3800, Australia

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_17,

> On one hand an assessment is an outcome of the act of assessing: the grade and/or comment attached to a piece of work. On the other hand, it is a process that involves the assessor, the piece of work or behavior in question, and the student: formative assessment is quintessentially process-oriented. (p. 485)

Formative assessment requires an effective communication system in which the teacher provides regular information and frequent descriptive feedback to the learner (Stiggins and Chappuis 2005). As such, formative assessment can be viewed as documenting the cyclical and extended process of professional growth and the building of relevant practice experiences (McMillan 2007). This occurs through continuous monitoring of the process towards student-oriented goals, which, in the case of teacher education (the context of this study), occurs through providing formative feedback to help student teachers gain insights into performance that are valuable for their professional growth and individual learning needs. This confirms the tendency to consider summative assessment as a final product-based activity and formative assessment as an intermittent in-course activity which is process based. As such summative assessment often tends to measure easily defined aspects of knowledge and/or learning and relates to the ability to evaluate or measure a specified product.

Formative assessment is not so easy to perform or quantify because it tends to focus on a process: coming to know rather than knowing per se (Stiggins and Chappuis 2005). Summative assessment appears less problematic because it assumes the product can be 'held still and measured' to determine the extent of knowing at a given point in time. It is not difficult to see then that talking about assessment is very much caught up with that which is being assessed. If that which is being assessed moves and changes (i.e. is a process as opposed to something that is assumed to be static and/or singular and discrete), then attempting to measure it is likely to be difficult. Such measures, by their very nature, will therefore be less precise and more variable in comparison to something that is less dynamic.

We argue that formative implies 'coming to understand' because it is about how that understanding comes to be *formed*. So the notion of 'coming to' is really about the process through which 'something' is formed. If, as in the case of the study reported in this chapter, the 'something' to be learnt is science teaching through a teacher education programme, then that which is formed could well be described as student teachers' knowledge of practice. However, as so much of the literature illustrates, teacher knowledge is not a static product and so cannot be easily assessed or measured in a summative fashion—at least not if it is to represent the true complexity of that knowledge. Rather, our approach to assessing student teachers' developing knowledge of practice is through a careful examination of the process of forming.

In this study the forming of knowledge of practice was facilitated though formative interactions (Black 2009). The study examines formative assessment in learning to teach science through a structured programme of teacher education based, in part, on the use of Content Representations (CoRes; see Appendix 1 and Loughran et al. 2004, 2006) as a heuristic tool for student teachers. The use of CoRes offers possibilities for exploring formative assessment because they are based on questions that help participants to develop answers without undue concern about arriving at a specified end point (Nilsson and Loughran 2011). Student teachers' responses to working with CoRes offer insights into how, through formative interactions in a teacher education programme, student teachers' learning about science teaching might be made more visible for teacher educators and student teachers alike.

Formative Assessment: Improving Student Teachers' Learning About Science Teaching

Formative assessment is recognised as one of the most powerful ways of enhancing student motivation and achievement in both classroom and large-scale assessments (Black and Wiliam 1998a, b; McMillan 2003). However, compared to the school sector, formative assessment has received relatively little attention in teacher education. Sadler (1989) argued that 'Formative assessment is concerned with how judgment about the quality of students' responses (performances, pieces, or works) can be used to shape and improve students' competence by short circuiting the randomness and inefficiency of trial-and-error learning' (p. 120). Clearly the same applies to teacher education. More recently, Black (2009) stressed that formative assessment:

> might be a rare example of a happy marriage between research evidence, learning theory, innovative practice, and national policy. However, there is the dark side, of misinterpretations, of tensions between formative and summative practices, and of ambiguity about the place of formative practice within pedagogy as a whole. (p. 3)

As formative assessment is often perceived as a simple, rather uncritical way of gathering information from students and using it to improve learning, formative assessment strategies could easily be interpreted as being a set of tips and tricks for teachers without clarity about the place of assessment in a comprehensive theory of teaching and learning (Black 2009). In providing feedback to their student teachers, teacher educators can encourage self-assessment by asking questions that help them judge their own work. Therefore, developing ways of working with student teachers using a formative assessment framework is important as it can help to establish ways of better understanding, and valuing, formative assessment more generally.

Tillema (2009) asserted that assessment of learning is an important vehicle for supporting student teachers' competence in teaching because it allows them to control their own learning by helping them identify strengths and weaknesses in a continuous, nonthreatening way. Sadler (1998) was of the view that peer assessment is an important complement to self-assessment because students may accept from one another criticisms of their work which they would not take seriously if made by their teacher.

Stiggins and Chappuis (2005) noted that when students are involved in collecting evidence of their achievement and setting goals for future learning, they also develop insights into themselves as learners. They further highlighted four conditions that must be satisfied for assessment for learning:

1. 'Assessment development must always be driven by a clearly articulated purpose' (p. 14), that is, students need to know what is to be learnt.

2. 'Assessments must arise from and accurately reflect clearly specified and appropriate achievement expectations' (p. 15), that is, teachers are clear about the intended learning and teach to it intentionally.
3. 'Assessment methods used must be capable of accurately reflecting the intended targets and are used as teaching tools along the way to proficiency' (p. 15), that is, teachers have a variety of assessment alternatives from which to select.
4. 'Communication systems must deliver assessment results into the hands of their intended users in a timely, understandable, and helpful manner' (p. 17), that is, an effective communication system provides regular diagnostic information to the teacher and frequent descriptive feedback to the learner.

Considering the fact that it has been well recognised that teachers' knowledge of practice is largely tacit (Korthagen and Kessels 1999; Polanyi 1966), it is not surprising that in order to recognise and articulate such knowledge, scaffolding and support is essential. Doing so matters for experienced teachers but is perhaps even more important for beginning teachers if they are to set reasonable goals for, and expectations about, the knowledge that they aim to develop as learners of teaching.

Formative Interactions

Research on formative assessment (Black 2009; Black and Wiliam 1998a, b, 2009; Black et al. 2002; Sadler 1989) of student teachers' learning to teach, and in particular, formative assessment of their knowledge of the how, why and what of the content to be taught, is limited. Black and Wiliam (1998a) demonstrated that informed use of formative assessment can lead to improvements in student learning. However, they also noted that such practice was rarely found. More recently, they emphasised the importance of interactive dialogues between teachers and students and between students themselves as an aspect of the social construction of knowledge (Black and Wiliam 2009). Black (2009) further noted that the core activity of assessment for learning is the involvement of learners in formative interaction with their teachers and with one another, 'Only through such activity can they become actively engaged in their own learning, and so acquire the confidence and skill needed to become effective learners' (p. 3).

A formative interaction is one 'in which an interactive situation influences cognition, i.e., it is an interaction between external stimulus and feedback, and internal production by the individual learner' (Black and Wiliam 2009, p. 11). Dialogic interactions between learners and teachers have also been highlighted by Pryor and Crossouard (2008). They suggested that in such dialogues the questions teachers ask are different to those more commonly seen in classrooms because they tend to not know the answer. Pryor and Crossouard characterised these as 'helping questions' rather than 'testing questions'. As helping questions they are seen to offer feedback which is exploratory or provocative, prompting further engagement as opposed to correcting mistakes.

In the context of teacher education, it could be asserted that assessment activities that build on formative interactions between student teachers and teacher educators offer new ways of better understanding (science) teaching and learning and how student teachers' knowledge of teaching science is being formed. However, enabling the kinds of classroom interactions that support formative assessment of student teachers' learning to teach science is a challenge that requires teacher educators to have knowledge of student teachers' concerns for teaching specific science content, as well as being able to respond appropriately to these concerns. A helpful way of thinking about this is through Wiliam and Thompson's (2007) suggestions for effective formative assessment which they conceptualised as consisting of five key strategies:

1. Clarifying learning intentions and sharing criteria for success
2. Engineering effective classroom discussions, questions and learning tasks that elicit evidence of learning
3. Providing feedback that moves learners forward
4. Activating students as the owners of their own learning
5. Activating students as instructional resources for one another

There is an obvious synergy between the conditions described above by Stiggins and Chappuis (2005) and Wiliam and Thompson's (2007) key strategies. Formative assessment in the context of teacher education clearly requires evidence about student teachers' learning being used to adjust instruction to better meet student teachers' needs, that is, teaching should be adaptive and responsive to student teachers' learning needs. In this chapter we offer an expansive example of a formative interaction in a Primary Science Teacher Education Programme in Sweden. Through the formative interactions outlined and explained in this chapter, that which is being formed (i.e. learning about teaching science) is based on a framework that was developed in order to make formative assessment a real and active touchstone for the programme.

Formative Assessment Framework Used in This Study

In assessment for learning, the assessment purpose is to provide teachers and students with information they need in order to make decisions that will bring about more learning. When they are involved in collecting evidence of their achievement and setting goals for future learning, students develop insight into themselves as learners (Stiggins and Chappuis 2005). Chappuis (2005) noted that student goal setting, self-reflection and self-assessment help them to understand the expectations for the task and the steps necessary to meet their learning goals. In teacher education, such ideas should resonate with the purpose of learning for student teachers in the complex business of learning about teaching.

In the project reported in this chapter, a framework was developed in order to help a group of student teachers (n = 24) learn how to make explicit the how, why and what of their science teaching and to stimulate self-reflection and self-assessment on that which they aimed to develop (i.e. form) as beginning professionals.

In the beginning of the semester, the student teachers were introduced to the purpose and learning intentions of the course using CoRes as a tool to help them set goals and reflect on and self-assess important aspects of their science teaching and, hence, their developing knowledge of practice. During this phase of the project, all student teacher participants were asked to establish their own learning goals for the semester and together with their teacher educator (the first author) formulate the expectations of their achievement. Next, the student teachers (in pairs) chose a specific science topic (chemistry or physics) to teach at a Science Learning Centre (SLC) at the university and completed an initial CoRe to guide their thinking for that specific lesson—the CoRe was therefore a tool to help prepare them for teaching.

In this initial CoRe the student teachers were asked to think about that which they considered to be the 'Big Ideas' associated with teaching the given topic based on such things as their knowledge of the content, their knowledge of students' understandings, the teaching procedures (and particular reasons for using these) and their specific ways of ascertaining students' understanding or confusion around these ideas (i.e. responding to the prompts in the left-hand column of the CoRe; see Appendix 1).

During the seminars in the SLC (n = 12 lessons), a teacher educator participated by giving immediate feedback to the student teacher pairs on their teaching as well as video recording the lessons for later shared observation and discussion. Following each seminar the student teachers in groups of six participated in a stimulated-recall session using the video recordings (in all, four seminars), where teaching sequences from each lesson were observed and reflected on in order to provide formative feedback to each other.

During these seminars, the CoRes were used as tools to stimulate student teachers' self-reflection and self-assessment about their teaching activities. Hence, the aim of the stimulated-recall seminars became the basis for formative interaction between the student teacher participants and their teacher educators. An important pedagogical emphasis for these interactions was to create cognitive conflict or cognitive challenge rather than to simply supply answers or tell student teachers 'what to do better'. As such, the approach to formative interaction was to purposefully enhance the social construction of knowledge and invite participants to reflect (and then act) on their own learning about science teaching.

Although involvement in these activities was compulsory, assessment was exclusively for formative not summative purposes (i.e. student teachers were not graded or tested on the nature of quality of their involvement and interactions). Importantly, the formative interactions were conducted in such a way as to reinforce the centrality of a safe learning environment in which respect for individual student teachers' needs and concerns was paramount. As such, the formative interactions were conducted in ways aimed at diminishing any sense of competition or threat between the student teacher peers and/or the teacher educators (i.e. attempts were made to address negative affective situations and to focus on issues and concerns, not on individuals). After the stimulated-recall seminar, the student teachers were asked to revise their CoRe and to formulate critical aspects of their teaching that they felt needed to be addressed before the next teaching of that specific lesson.

In the next phase of the project, the student teachers taught the same science lesson once again (during their 6 week school practicum) using the revised CoRe and the formative feedback from the seminar to reflect on and improve their teaching practice. In the last phase of the project, the student teachers in groups of six (same groups as before) again participated in a group seminar where they shared (through formative interactions) their teaching and learning experiences from their school practicum based on their reflections from their initial learning intentions, the specific lesson taught in the SLC and their revision of the CoRe.

A brief overview of the phases of the project is outlined below:

Phase 1: Seminar clarifying the learning intentions and introducing a CoRe
Phase 2: Completing initial CoRe and teaching a science lesson in the SLC
Phase 3: Stimulated-recall seminar using video-recorded lessons to stimulate student teachers' self-assessment and offer formative feedback through the use of formative interactions
Phase 4: Revising the CoRe for the specific science lesson and teaching science during the 6-week school practicum
Phase 5: Seminar, formative interactions based on teaching experiences and initial learning intentions and their final learning outcomes

By providing student teachers with the appropriate tools for assessment (i.e. CoRes, opportunities for self-assessment, formative interactions), their approach to the teaching of their chosen science topic and the reasons for that approach were able to be analysed in ways that demonstrated the extent to which participants' learning about science teaching appeared to develop over time. Hence, the CoRe functioned as a tool to problematise the content as well as the pedagogy (connected to their teaching in the SLC and school practicum) in a way that provoked their thinking about that which they considered to be important in the teaching of their chosen science topic and why. Furthermore, the quality of the formative assessment was evident in the student teachers' reflections on their developing knowledge of science teaching and the possibilities they envisaged for its future development (i.e. they identified that which they needed to know and to think further about what that might mean for teaching new topics as encouraged through the use of Big Ideas and the prompts in the CoRe). The specific formative assessment framework described above was developed to stimulate formative interactions and self- and peer assessment throughout the course of a semester.

Research Design and Collection of Data

The student teachers who participated in this study were in a three-and-a-half-year pre-service primary teacher education programme, one year of which was devoted to mathematics and science. During their second and third semesters, they had various basic courses in mathematics, physics, chemistry and biology to prepare them for teaching in primary schools. At the time of the study, participants were in their third term, which contained fundamental courses of science, a science teaching method course involving planning and conducting a lesson with a group of students in the SLC at the university and 6 weeks of practicum in a primary school. Hence, in terms of their total programme, they had finished some of their science courses, but their science teaching experiences were somewhat limited.

The formative interactions between student teachers and the teacher educators during Phases 3 and 5 of this project could be seen as seminars designed around facilitating meaningful interaction between the pairs of student teachers (who taught the lesson being observed on video), their peer student teachers in their particular seminar group and their teacher educators. Therefore, the data collection consisted of an audio recording taken from the formative interactions in a particular stimulated-recall seminar (in this case, Phase 3 of the project). Within the collection of data, the student teachers' CoRes were also considered as evidence of their changed (or not) thinking of their teaching and learning activities as they moved between their practice as seen in the video, their thinking about their teaching and their students' learning and the goals and expectations they had set through the construction of their CoRes.

To capture the essence of a formative interaction, an extended transcript of one stimulated-recall session is offered (in the next section) from the many conducted in the programme. As is immediately obvious through the transcript, formative interactions can become very weighty tomes. Hence, the one offered was selected because, with accompanying analysis (in this case from Phase 3 of the framework), it was of a manageable size in terms of the constraints of words for a chapter in a book.

The transcript itself has been marginally adjusted in terms of language so that the written form offered flows smoothly and is readable in ways that maintain fidelity with the original verbal forms. In order to highlight particular aspects of the formative interaction, the accompanying analysis (based on coding the data using Wiliam and Thompson's (2007) five key strategies) illustrates how, by thinking about the intent inherent in formative interactions, that which is being formed (these student teachers' developing knowledge of practice) is able to be brought into clear focus. As a consequence, the process of formative assessment through the specific approach of formative interaction is able to not only be illustrated but also coded and analysed through one extensive worked example.

The dual column layout of the analysis that follows is specifically designed to allow the reader to maintain an uninterrupted flow of the event (formative interaction) while at the same time being able to decide how to engage with sections of the interaction in concert with the authors' analysis. As such, it is designed to give the reader choice in ways of engaging with the text and the accompanying analysis. For some, there will be a desire to read the transcript uninterrupted by the authors' analysis; for others the ability to move between formative interaction, coding and analysis will be preferable.

In the first instance, analysis is based on coding for the five key strategies proposed by Wiliam and Thompson (2007). Each of these key strategies is highlighted in italics once to show the link between the coding and the deeper level of analysis. As the framework of the formative interaction unfolds through the second column, analytic insights sit side by side with the transcript in line with the utterances that bring to the surface ‘that which is being formed’. Analysis of the coded data could clearly occur in a variety of ways, two of which include the teacher educator’s intentions and the student teachers’ learning. Initially, following coding, both of these were applied. However, for this chapter the lens of student teacher’s learning has been highlighted as, in addition to the length of the transcript, two layers of analysis created a very dense text and tended to distract the reader. As a result of narrowing the analytic focus, as is clear in the text, multiple features of the framework are evident but not all attract comment.

Overall, the transcript offers one complete formative interaction with insights into coding and analysis in order to offer a concrete example of the important features of an interaction that is formative. In so doing, formative interaction as a process is outlined in such a way as to draw attention to that which is being formed—student teachers’ developing knowledge of science teaching and learning—thus making the process much more explicit and clear, both for the student teacher participants and their teacher educator(s).

Following the worked example of a formative interaction, a discussion of the salient features of the analysis is offered through which fundamental issues of formative interactions are considered in light of the literature about and expectations of the nature and potential of formative assessment.

Analysis and Findings

In the video-recorded teaching session which is analysed below, two student teachers’ (Alan and Jane) video-recorded teaching of a lesson with a group of students aged 9–10 years is being observed. The lesson lasted for 2 h and contained different experiments about ‘air’ in which the students were involved in making predictions, observing and explaining different phenomena. In the CoRe Alan and Jane developed before this lesson, their Big Ideas (see Appendix 1 for an outline of a CoRe) were ‘air is something’, ‘air takes space’ and ‘there is a difference between cold and warm air’. In the video-recorded lesson, they conducted experiments, had students develop mind maps and managed whole group discussions to engage their students and stimulate them in their learning.

In the stimulated-recall session based on observing the video recording, Alan and Jane reflected on the lesson together with their teacher educator and another two pairs of student teachers. The formative interaction that follows begins with Alan and Jane offering a brief introduction to the (about to be observed) video recording of their teaching experience (i.e. they present the aim of the lesson, the group of students and the different activities). Alan begins as follows:

> We had a very big group of 9–10 year olds. They were very interested in what we did but I actually thought that the session was quite messy. We had a number of different experiments for them on air. There was a boy in the class [pointing out on the screen one particular student in the class] who had some difficulties and he could not settle down … yes he was quite disruptive.

Original formative interaction text	Analytic points of interest
Alan: We started with a mind map as planned and the students had quite good ideas about what air is. They started to talk about molecules. There was a boy who said that there is no air in water but then a girl said that there is oxygen in water because fish breathe Mary: I would have been very nervous if the students had started to ask all these questions because I would not know how to answer Mandi: It was quite cool that the students started to talk about the oxygen in the air Ida: I think you made a good mind map but you could have built more on their ideas … I mean you could have started to discuss about the different things that they bring up T. Ed: How come you didn't build on that? Mandi: I think that it is about our own self-confidence. I wasn't sure how to approach it then but now I see that I could have done it in another way	This formative interaction opens with the 'stimulated recall group' initially being drawn to the ways in which the teaching is structured and how students' responses can create issues for them in working out how/whether to 'stick' to the lesson plan (clarifying learning intentions, sharing criteria for success). Their learning intentions were to uncover their students' prior knowledge through the use of a mind map as their planned teaching procedure. As the transcript illustrates, there is variation among the participants in their ability to work with students' ideas which in turn influences their self-confidence and therefore their views of success. The opening to this formative interaction (through watching the teaching on video) illustrates how these student teachers readily opened up about their concerns with student discussion and questioning. What is being formed in this interaction is the recognition of the value of drawing on students' prior knowledge
T. Ed: What do you think about the introduction? Mandi: I think that they were good at discussing with each other Mary: I was impressed that you got such a good response from the students and that they seemed to be so spontaneous Mandi: Yes and I have noticed that students are mostly very constructive in the way they think. I think that as teachers we are too often quite regimented and want to keep things the way we have planned for the lesson I think that we … well at least I, need to be more open minded and really build on the students' ideas. If we don't know the right answer we don't want to just guess and so we go all quiet, but the students don't seem to care about that. They are much freer in the way they think Mandi: Yes, and I liked that you used open-ended questions. That is a good way to simulate students' reasoning. I was also thinking of what you said when the student mentioned that air is oxygen … you could have said that yes, there is oxygen in air but there are several gases in air and oxygen is only one of them. I think that shows that no matter what you say, it never really covers everything That's something I think we learn through this	This is an example of how the teacher educator, through directing the interaction towards assessing the activity, engineers effective classroom discussions and questions that elicit evidence of learning in order to help them judge their own work. The student teachers are stimulated to identify strengths and weaknesses in the teaching situation and share their experiences by giving feedback to each other. For example, Mandi provides feedback that moves her colleagues forward. She highlights the importance of the students coming to see that air consists of several gases (oxygen is one of them)Through this discussion she comes to see that teaching is not a linear process in which all students' ideas are covered. Her knowledge about science teaching is being shaped and formed through recognition of the need to be open to students' ideas as opposed to simply sticking to the script of a lesson plan

(continued)

(continued)

Original formative interaction text	Analytic points of interest
T. Ed [Looking at the video]: Here you have that student who tells you that air is wind and that there are atoms … what do you think about that?	The teacher educator pushes the interaction towards a more content-specific interactive situation by using an external stimulus (i.e., the boy's comment on the video). Alan's response goes to the core of why teaching science is not linear; students' previous knowledge is different and capturing all students' ideas and building on these is a challenge. This was something that he did not pay attention to before and so the formative interaction created an opportunity for him to form this knowledge about science teaching and how it can be influenced by students' existing knowledge and beliefs. His understanding is formed from his perspective as a teacher as well as that of his students as learners of science
Alan: Yes I think that was hard when he mentioned atoms and I was not sure if I was going to build further on that. I mean, just because he mentions atoms we can't start to talk about only that and just leave the other things we are talking about. That was something that I did not think of before and that just came up spontaneously. I did not want to go too deeply into just one idea. I wanted to include all of the students but now I really see that they have so many different ideas so it is almost impossible to involve them all. Well, that is really a challenge. I can see that now	
Mandi: Yes that is true. I also heard him mention atoms and so even in trying to get out all of the students' ideas we still lose some. I mean, if you started to talk about atoms just because this boy mentioned them then you would lose some others. But it could also be that as you actually put his word "atoms" up in the mind map and then only said there are atoms with no further explanations maybe for other students it was boring. I mean, I think you could have mentioned something like "everything consists of atoms"	
Cathy: I think you made a very good mind map but I think it could've been even more constructive if you had used the word in the mind map to discuss with the students and not only write down what they said. Like that you might have missed out on some good opportunities to create good discussion with the students. You should have had more explanations like that, you know, if they mentioned oxygen you could have had built on that further and talked about why oxygen is important and where we find it. You could also have started to talk about the relationship between air and oxygen. It's hard to know what the students might say. But you know much more than they do and that is always a thing that makes us feel sure about what we say and do—we have more content than them	Cathy's feedback about the mindmap created discussion with the students that was important and gave Alan an idea of how to meet students' knowledge needs and include students' ideas in his teaching. The knowledge that is being formed through this interaction is the need to move beyond an activity as a static learning event in order to draw out the various aspects of the science concept, which is important in challenging science as the delivery of information

(continued)

(continued)

Original formative interaction text	Analytic points of interest
T. Ed: So why didn't you build on these students' ideas? Is it because you lacked the subject matter knowledge or something? What was the problem?	The teacher educator uses this to create questions to help the student teachers reconsider their content knowledge
Jane: I think that we are uncertain about what to do, do we know? Will there be enough time to work with this idea properly? If I start with that … where will it all end up? I think that we don't have enough teaching experience yet to make these decisions about what direction the lesson will go. And sometimes you also lose the students if you go on too fast and that is also a difficult decision … should I stop here or should I go further? We are still uncertain in our roles as teachers and at least I feel that I am afraid and I do not dare to take that last step and move beyond my content knowledge. I don't want to say the wrong things, and this is so clear here that we're trying to avoid going into deep discussions. We should actually have built on that boy's ideas in a better way	Jane highlights again that teaching is not linear and how this influences her confidence and knowledge of content and the way that both are so strongly connected. What is being formed here is an insight into her being a teacher and a co-learner and that successful science teaching is not about knowing all the answers to science questions and transferring that to the students. What is also formed through this element of the interaction is recognition of a process in which teachers must support students in being responsible for their own learning
Alan: We need the deep content knowledge and we also need to have a good overview of the whole which makes running a discussion a very delicate balance between our surface and our deep knowledge. I mean, in order to understand a phenomenon in a deeper way I absolutely need the small pieces. But it is a little bit like the chicken and the egg, what do we need first … and this is something that we draw on in our own teaching of science to our students. Should we be stuck in the concept of atoms or do we want the students to have a bigger understanding of the role of the atom in the whole world. But it is funny, we have all this discussion but look at them, they actually start to talk about nanophysics. You learn a lot through doing this	Through this aspect of the interaction, Alan questions the nature of his content knowledge ("as a delicate balance between surface and deep knowledge") and the value of being confident to open up a level of uncertainty about specific content knowledge. What is being formed here is an understanding that, just as science teaching is not linear, neither is science content simply a matter of facts
Jane: Yes but I still think that I need to consider how to explain the concept of atoms. Could I say that an atom is just a thing that builds up the world or should I go deeper into it? But I don't think about that directly when I get the question	

(continued)

(continued)

Original formative interaction text	Analytic points of interest
Mandi: That is probably because as teachers we don't think of atoms only as pieces that build up matter. We think of an atom as having protons, neutrons and electrons and then we get drawn into all of that. If we simplify it so much that we just say they are small pieces that build up other things we are probably not scientifically correct. We know it in a more advanced way … I mean we know that atoms create ions and all that Mary: Yes but that is a skill that we as teachers have and then how to transform it for the students is something we need to build up over time. We need teaching experience to be able to learn about that. We have a very deep knowledge but in the same way it is so important for us to be able to transform the knowledge so that we stimulate the students' interests. I don't yet have that knowledge; I need to transform the content in a way that keeps the students interested and also makes them learn T. Ed: Yes isn't that what PCK is about? You will learn that through your teaching experience. Because even though you know how to explain what to do and say in ten different ways there might be another one that forces you to be flexible and use new strategies. What do you think? Mandi: That is why it is so constructive to ask the students about different things because then you get their explanations and you get also a good view of what they think and then you can build further on their ideas. Like what you had in your video when that 8 year old explained that there is more air under the paper. Then we can listen to their explanations and use that Alan: There was one boy who was very clever with his explanations and I really used him to build my own thinking and my own explanations	Mandi provides feedback noting the difference between the nature of what the teacher knows and what the teacher teaches. She draws attention to the tension between knowing the "right facts" (i.e., that atoms have protons, neutrons and electrons, etc.) and making concepts simpler (but not scientifically correct) for students. What is being formed here is an extension on the previous recognition of the difference between knowing science and content, and knowing how to teach that content for enhanced student understanding. Mary pushes this further, noting that developing teacher knowledge is challenging. The dialogue has helped her to understand the importance of "transforming the content in a way that keeps the students interested and also makes them learn." Even though this learning intention was communicated by her teacher educators in the beginning of the project it did not become a real part of her learning intention until she had experienced a situation that better connected it to her own teaching and learning journey. The teacher educator extends the interaction to involve issues such as the importance of taking risk, helping students be active participants and building on students' ideas

(continued)

(continued)

Original formative interaction text	Analytic points of interest
Jane: I was impressed that they really understood all the instructions Cathy: Sometimes I think that we underestimate the students just because they are young. I mean here [gesturing to the television] we see that they come up with brilliant ideas. When we really look closely at these teaching situations it makes you reflect on why we always try to facilitate thinking for them, but it is not that difficult for them. We see that they know much more than we think. But I think that it would not have helped if you as teacher educators [looking to the two teacher educators] had told us that young students are good at science. No I am sure that we needed to experience that by ourselves Mary: I don't know why we always think that the students do not know science and that we need to make it easy for them Alan: Perhaps it is because we did not know all these things when we were 10 years old. I mean, I don't know about nanophysics and all that. But now I understand that I need to know so much and also know that students are not always interested in the environment and greenhouse effect, they want to know about nanophysics and all that new science that's coming up	Jane illustrates a concern she appears to have about aspects of students' capabilities in understanding instructions. Cathy highlights that students know much more than their teachers may give them credit for. She notes that knowing about students' capabilities, prior knowledge and interests is something that she needs to experience for herself in order for it to really impact her understanding of teaching. What is being formed here is an understanding of students as valuable resources for her learning about science teaching
Mary: How come you asked them to do a hypothesis by themselves? Why didn't you get them to work in pairs? Jane: We didn't want them to listen to each other but instead really reflect on the experiments by themselves, that is why they did the experiments together Mandi: Sometimes I think that our supervising teachers get too involved with what the students are doing. They seem to be afraid that we will make mistakes, but on the other hand it is through the mistakes that we actually learn. I think it is important both for us and for the students to understand that it does not matter if we say that we don't know everything or if we make a mistake. Both the students and we can learn from that	Mary provides feedback by asking a helping question that is exploratory and prompts further engagement by Jane in relation to what she needs to reconsider about her way of making students think. This aspect of the interaction illustrates well how students can act as *instructional resources for one another*. Mandi comes to understand the importance of being an active learner. What is being formed is recognition of the value of learning from mistakes; this is a crucial aspect of learning about teaching science as being interactive and instructive rather than rigid and formulaic

(continued)

(continued)

Original formative interaction text	Analytic points of interest
T. Ed [Looking at the video]: This is a wonderful example, when you Alan struggle with how to explain what happened and then a boy says that, "Can't you just say that the air is in the way?" What did you think here? Mandi: I feel that your lesson is quite structured and that the students don't get much freedom to work by themselves Alan: Yes we tried to use a lot of mixed methods but you are right that perhaps we used a lot of traditional "lecturing". First we let them try and then we lectured and then they did the next experiment and discussed that and then we handled the explanations Mandi: Yes I would never have done that by myself but it is interesting to see that you actually seem to catch the students even though you, in some sense, decided what they were going to do. We often think that "lecturing" to young students is bad but here we see that it actually works … well you mix it up with experiments, but still you really are in charge of the lesson Alan: Yes but they did the experiments by themselves so we just tried to tie it up through our lecturing. That is actually something that I would not have planned to do several months ago. I mean, lecturing to primary kids is supposed to be "bad", isn't it? But we tried it out here in this way that we did it and they seemed to be with us Jane: Yes but even though the students had quite some space to do the experiments by themselves I still feel that we had control over the lesson Mary: Yes we can see that [through this episode] and I am happy that I got the chance to see you do this because I am not sure that I would have planned my lesson in this way but we see that it worked very well. I would be afraid that the lesson would be too messy Alan: We could have failed too if we had talked too much. Then we would have risked losing them	The teacher educator pushes for deeper reflection about how the student teachers take up (or not) opportunities to build on students' ideas. Mandi offers feedback by providing a provocative comment that prompts further engagement by Alan and Jane to clarify the purpose of the lesson. Alan reconsiders their way of structuring the lesson, explaining the underpinning reasons for why they structured the lesson in such a way. Mandi's comment seems to be important in moving his learning forward. What is being formed here is an understanding that just doing a lot of activities is not a sufficient purpose for learning; the Big Ideas and the reasons (teacher's intent) for why students need to learn these ideas become apparent—both major aspects of their CoRes that perhaps become more tangible and meaningful through this type of interaction
T. Ed: I think that there is a difference between having a lecture like a monologue and having a lesson where you involve the students in the discussions. It becomes more a sort of an interactive lecture. Was that your idea?	The teacher educator pushes further about teaching methods to prompt deeper reflection about how the student teachers interact with students in a teaching situation

(continued)

(continued)

Original formative interaction text	Analytic points of interest
Jane: I think that it is important to help the students understand that air is not only something that we work on here and now in this lesson, it is a concept that they need to know about in their everyday life. That's actually something I have thought about before but now when we have done it I see that their everyday life is not always the same type of everyday life that we as adults have	Jane's insight into students' everyday lives illustrates an assessment of the thinking associated with that which may be a taken-for-granted assumption. Jane is forming an understanding that challenges the tacit assumption that students interpret the world around them the same way as their teachers
Mary: Yes that is why I think that working with air is so good because you can talk about so many different things connected to science. You can start to talk about atoms, you can talk about space and why there is no air in space and then you can start to talk about a vacuum in the space, you can easily get into pressure and build further on that to finally end up with the weather. Yes that is really cool to see. I wish that we would have chosen air instead of electricity	Mary uses the feedback to build her ideas around what and how to teach about air. What is being formed is a recognition that difficult concepts can be 'opened up for exploration' in meaningful ways, and that such an approach is not limited to any one concept or situation
Jane: Yes it is a great theme to work with, it allows you to go in several different directions depending on where the students might lead you	
Mandi: It was great to see the others' lessons. It gives you a lot of ideas	
T. Ed: Yes and in what way did you use the CoRe?	The teacher educator asks a question about the CoRe in order to push Alan to clarify the purpose of the lesson by considering the prompts in the CoRe as a way of becoming more consciously aware of his teaching. What is being formed here is the value of linking the 'what' and the 'why' of teaching. The interaction indicates awareness among the student teachers that they are active learners, they are moving their learning forward in forming their knowledge of what they assess as needing to know better (i.e., their learning needs)
Alan: I think that it is important that the students are motivated to learn and we as teachers must create that motivation. I think that working with the CoRe like we did here made me reflect on why I chose the different activities but also why different things are important for students to learn. I learnt a lot from thinking carefully about these aspects of the CoRe for my teaching	
Mandi: Yes because this way of looking at our activities really makes us reflect on what we need to do better but also what was good. But it is really important that we just don't forget this now but instead use it in the future	

(continued)

(continued)

Original formative interaction text	Analytic points of interest
Mary: Yes, I think that the CoRe was like a spotlight that helped you focus on questions that you had not really thought of before. Like why the content is important to learn and what exactly, through the Big Ideas, the students should learn. I feel that the CoRe is a way to explore our thinking and create new ideas and to give an indication of where to go with the direction instead of just giving out the knowledge	Mary suggests that the CoRe was an important tool for her in forming her knowledge of teaching and learning science and as such, she shares a criterion for success. Her thoughtful closing of the dialogue indicates that she has formed an understanding of how to approach content through Big Ideas in order to promote students' learning. The interaction suggests that her professional learning has been catalysed and enhanced. The outcome of the interaction has helped her to form new ways of exploring the development of her knowledge of science teaching

The formative interaction outlined above was designed to serve two distinct purposes. The first was to illustrate that Black and Wiliam's (2009) notion of a formative interaction—an activity carried out between teachers and their students and between students themselves (or in this case, teacher educators and student teachers)—can be made concrete in the research literature through using a real-world example. In so doing, the formative interaction outlined above opened up to scrutiny that which Black (2009) asserted about such an interaction, that is, that the core activity of assessment for learning requires the involvement of learners in formative interaction with their teachers and with one another if they are to be engaged in their own learning and, as a consequence, acquire the confidence and skill to become effective learners.

Tillema (2009) noted that multi-perspective assessment in mentored learning and in tutorial relations may well be undervalued in teacher education. The extended formative interaction above indicates that receiving feedback from multiple perspectives (peers and teacher educators) can indeed foster the learning about teaching process for beginning teachers and help them start to form a deeper understanding about what it means to develop their knowledge of practice. If that is to be taken seriously in a teacher education context where the development of beginning teachers is crucial to the overall quality of school teaching and learning, formative interactions can now be seen as central to any assessment of learning about knowledge of practice. It could be argued, then, that formative interactions should be an explicit component of assessment in teacher education.

The second purpose of the explication of the formative interaction described above was to focus serious attention on aspects of the science education research literature that have been emphasised as important in forming student teachers' knowledge of science teaching. For example, the interactions illustrated how student teachers came to understand the importance of their content knowledge, but more importantly they illustrated the value in going beyond what the student teachers knew and in examining what content knowledge they needed to explore further in order to genuinely build on students' ideas and questions. As such, the interaction indicates an awareness of teaching as anything but linear and formulaic. The formative interaction made clear that they recognised that content knowledge alone was not enough to teach science. They recognised the teaching role as requiring knowledge and skills in transforming science content into meaningful and educative learning experiences for their students (and themselves).

Conclusion

In the context of higher education, Yorke (2003) asserted that there is a need to move towards a theory of formative assessment in order to consider some of the implications for pedagogic practice. Yorke (2003) referred to Vygotsky's (1978) 'zone of proximal development' which, broadly stated, is the region between a student's existing problem-solving ability and the ability to solve more complex problems given guidance and support from a more skilled person. Pryor and Crossouard (2008) also offered a conceptualisation of formative assessment that takes into account sociocultural learning theories in which teachers and learners seek to respond to student work by making judgments about that which comprises good learning. They further asserted that, at its best, formative assessment could be seen as a kind of scaffolding whereby the teacher plays a crucial role in enabling learners to do with help that which they would not have been able to do alone. Black and Wiliam's (2009) position about formative assessment, theory and practice is made clear when they conclude from their analysis of the relation between formative assessment and other broader theories of pedagogy that 'it is clear that the complexity of the situations in which formative feedback is exchanged is such that it could only be understood in terms of several theoretical perspectives required to explore the different types of issues involved' (p. 28).

Sadler's (1989) definition of formative feedback (i.e. that students are able to monitor continuously the quality of what is being produced during the act of production itself and that they have a repertoire of alternative strategies from which to draw at any given point) is important because it has been well noted in the formative assessment literature that the nature of self-assessment and feedback is a major concern. Student teachers need to be able to judge the quality of what they are producing and be able to regulate what they are doing during the doing of it: 'In order to learn from practical experience it is reasonable to suggest that experiences must be reflected and reasoned upon' (Nilsson 2008, p. 4).

The message inherent in the results of the research reported in this chapter is that there are positive learning outcomes through the use of formative interaction because it offers a way of developing and assessing science student teachers' professional learning in ways that they personally engage with and value. Wiliam and Thompson (2007) suggested that effective formative assessment consisted of five key strategies: clarifying learning intentions and sharing criteria for success; engineering effective classroom discussions, questions and learning tasks that elicit evidence of learning; providing feedback that moves learners forward; activating students as the owners of their own learning; and activating students as instructional resources for one another. The extended example of a formative interaction offered in this chapter highlights the importance of acknowledging these strategies in developing and applying valid tools for formatively assessing the development of student teachers' knowledge of science teaching and learning over time.

Tillema (2009) suggested that assessment is an important vehicle for supporting student teachers' learning about teaching because it allows them to control their own learning by helping them identify strengths and weaknesses in a continuous, nonthreatening way. The findings of this study indicate that the use of a holistic tool such as CoRe can act as a trigger to encourage student teachers on their own and through purposeful interactions with others to identify strengths and weaknesses and further begin to develop knowledge of science teaching and learning by using it as a catalyst for personal assessment of their own developing skills, knowledge and ability. Therefore, the use of formative interactions can be the basis for powerful personal assessment with relation to active engagement in one's own learning.

Black's (2009) notion that the core activity of assessment for learning is the involvement of learners in formative interaction—with their teachers and with one another—was the centre piece of Phases 3 and 5 of this project. During the seminars in these two phases, as the data above illustrate, the teacher educators discussed and analysed events, questions and activities to provide responses and advice for further action and learning. Formative interaction provided student teachers with important information about their identified learning needs in order to modify their own teaching approaches. Through the formative interaction, issues for enhancing student teachers' learning about science teaching and learning were unpacked in ways that demonstrated and offered insights into the extent of their developing knowledge of practice.

In the formative dialogues in this project, student teachers were expected to contribute to whole class and group dialogue based on their experience and knowledge in order to collaboratively develop their knowledge of science teaching and learning. The formative interaction offered illustrated how student teachers were provided with opportunities to interact with, and receive feedback about, their learning. In the stimulated-recall sessions, the teacher educator asked helpful *why* and *how* questions designed to encourage the student teachers to explain their ideas more fully and to encourage their peers to be actively involved in the learning. The formative interaction was a mechanism to support and encourage that learning in a meaningful way.

Korthagen (1993) noted that teacher educators need to stimulate student teachers during their teacher education programmes to internalise dispositions and skills in order to study their teaching and to become better at teaching over time. This project illustrates that student teachers can be encouraged to form understandings of the complexity of science teaching through formative interactions. As the extended example of the formative interaction highlights, these student teachers came to see science knowledge as complex and as a consequence, the importance of reflection on how to approach content through different Big Ideas in order to promote students' learning began to emerge in meaningful ways for the participants. They also formed a deeper understanding of the classroom as they began to see how the use of different teaching approaches and alternative ways of explaining phenomena influenced students' learning in different ways.

As has been stated a number of times, this chapter was designed to explore one way of formatively assessing student teachers' learning about teaching science based on the use of formative interactions. It has made clear how, through formative interaction, student teachers' professional learning can be catalysed and enhanced. By connecting the research on formative assessment with the development of student teachers' learning to teach science, this chapter has brought into sharp focus how the explication of formative interactions can inform approaches to science teacher education.

Appendix 1: Full Core Template

Content:			
Age of the children:	Big Idea A	Big Idea B	Big Idea C
What do you intend students to learn about this idea?			
Why it is important for students to know this?			
What else do you know about this idea (that you do not intend students to know yet)?			
What difficulties/limitations are connected with teaching this idea?			
What is your knowledge about students' thinking which influences your teaching of this idea?			
What other factors influence your teaching of this idea?			
What teaching procedures will you use and what are the particular reasons for using these to engage with this idea?			
What specific ways do you have of ascertaining students' understanding or confusion?			

References

Black, P. (2009). Looking again at formative assessment. *Learning and Teaching, 30*, 3–5.

Black, P. J., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–71.

Black, P. J., & Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. London: King's College London School of Education.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box*. London: King's College London School of Education.
Chappuis, J. (2005). Helping students understand assessment. *Educational Leadership, 63*(3), 39–43.
Korthagen, F. A. J. (1993). Two modes of reflection. *Teacher and Teacher Education, 9*(3), 317–326.
Korthagen, F. A. J., & Kessels, J. (1999). Linking theory and practice: Changing the pedagogy of teacher education. *Educational Researcher, 28*(4), 4–17.
Loughran, J. J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching, 41*, 370–391.
Loughran, J. J., Berry, A., & Mulhall, P. (2006). *Understanding and developing science teachers' pedagogical content knowledge*. Rotterdam: Sense.
McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice, 22*(4), 34–43.
McMillan, J. (2007). *Formative classroom assessment: Theory into practice*. New York: Teachers College.
Nilsson, P. (2008). *Learning to teach and teaching to learn—Primary science student teachers' complex journey from learners to teachers.* Dissertation, Linköping Studies in Science and Technology Education No 19.
Nilsson, P., & Loughran, J. (2011). Understanding and assessing primary science student teachers' pedagogical content knowledge. *Journal of Science Teacher Education*. doi:10.1007/s10972-011-9239-y.
Polanyi, M. (1966). *The tacit dimension*. Garden City: Doubleday.
Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education, 34*(1), 1–20.
Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144.
Sadler, R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education, 5*(1), 77–84.
Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory Into Practice, 44*(1), 11–18.
Taras, M. (2008). Summative and formative assessment: Perceptions and realities. *Active Learning in Higher Education, 9*(2), 172–192.
Tillema, H. H. (2009). Assessment for learning to teach: Appraisal of practice teaching lessons by mentors, supervisors, and student teachers. *Journal of Teacher Education, 60*(2), 155–167.
Vygotsky, L. S. (1978). Mind in society: *The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah: Lawrence Erlbaum Associates.
Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45*, 477–501.

Chapter 18
Changing Assessment Practices in Science Classrooms

Chris Harrison

> Perhaps no area of education policy is as contentious—or as consistently newsworthy—as assessment… (Mansell et al. 2009, p. 4)

Throughout all sectors of education, assessment that is more interactive, that is teacher or student led and that encourages teacher differentiation and the nurturing of student skills is being endorsed (Boud and Falchikov 2004; Brooks and Tough 2006; James and Mansell 2009; Harrison and Howard 2009). In recent years, assessment in the UK and elsewhere has also seen many changes, including a shift in the focus of attention (Black and Wiliam 1998a; James and Mansell 2009), away from the technicalities of test construction towards approaches that focus on student learning. Part of this has arisen from criticism of high-stakes testing (Harlen and Deakin-Crick 2004; Brooks and Tough 2006), while the influence of several research programmes on formative assessment indicates alternative perceptions of pedagogy and learning (Bell and Cowie 2001; Black et al. 2002, 2003; Hutchinson and Hayward 2005).

There is an extensive body of evidence, which describes and explains the effectiveness of assessment for learning (AfL) as a pedagogical tool beginning with the review by Black and Wiliam in 1998 and through various projects mainly carried out in the school sector over the last decade (Black et al. 2002, 2003; Harrison and Howard 2009; James et al. 2007). Through an assessment for learning approach, short cycles of assessment, feedback and changes to teaching take place and directly affect students' learning and progress. This AfL approach is also geared to train the learner to be more self-regulating (Harrison 2011; Sadler 1989) through regular experience of self- and peer-assessment opportunities and within classrooms that foster a dialogic approach (Harrison and Howard 2009).

The last two decades have also seen many changes in the ways that schools are managed, inspected and financed in the UK, with school improvement and effectiveness measures becoming key factors in the daily life of schools. Such pressures influence the

C. Harrison (✉)
King's College, London
e-mail: christine.harrison@kcl.ac.uk

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_18,

way that teachers decide to teach, how they manage the curriculum and the ways in which they interact with their students. The introduction of the National Curriculum in England and Wales in 1989 and its assessment procedures affected teaching strategies in science classrooms (Fairbrother et al. 1995; Hacker and Rowe 1997; Russell et al. 1995). All three studies reported a reduction in the range of teaching strategies employed by teachers, with a movement away from pupil-centred teaching towards a more didactic approach. In Hudson and Smith's survey, over 70 % of the sample admitted that they could not teach in the way they would select because of recent changes in the curricular and assessment demands of the National Curriculum. The teachers felt they were being constrained by the content overload of the curriculum and the introduction of new assessment procedures that had been externally imposed.

While there have been large- and small-scale moves to implement AfL practice in schools in England and elsewhere, the reports from school inspectors (OfSTED 2007, 2010), government agencies (DCSF 2007) and researchers (Carless 2005; Smith and Gorrard 2005) indicate that the implementation is sporadic and underdeveloped. Instead, the accountability demands of summative assessment have driven many teachers to 'teach to the test' (Mansell 2007; Popham 2001). This chapter sets out to explore teachers' formative and summative assessment practices in science classrooms and outline some of the complexities that teachers face when wanting to change their assessment practices. Historically, for most examples of curriculum change, assessment has been essentially an afterthought and the role and significance of assessment in curriculum development has been undervalued and underinvested (Black et al. 2006). At the same time, there is evidence that classroom assessment becomes less formative and more summative in response to high-stakes testing (Pollard et al. 2000), and so formative assessment is crowded out or downplayed by teachers because of the dominance of summative styles of assessment (Carless 2006). Harlen (2006) suggests that formative and summative assessments are potentially complementary, while Broadfoot and Black (2004) indicate that ways of working that link the two purposes together need to be actioned to help teachers deal with the current assessment dilemma many find themselves in. Other researchers emphasise that formative and summative assessments should be dealt with separately (Pellegrino et al. 2001; Simpson 1990), since their purposes are so different and therefore cause difficulties and misinterpretations if dealt with simultaneously. This chapter attempts to begin the important role of providing the detail of what happens when teachers begin to make changes in their assessment practices and attempts to unveil some of the tensions and synergies that prevail when they make these changes within their classrooms and within their school contexts.

The KREST Project

The King's Researching Expertise in Science Teaching (KREST) project was a collaborative action research project with science teachers designed to support them in strengthening their classroom assessment practices. This work is based on a 3-year project led by King's College London in collaboration with the Weizmann Institute, Israel, and the Institute of Education. Six domains of science education were selected, with each country designing and investigating three domains with groups of science teachers in England and Israel. The focus for this approach is evidence-based professional development (PD), where the process of collection, analysis and reflection on evidence arising from classrooms provides the basis and motivation for teachers to transform their practice (Harrison et al. 2008). This science education project provided a new approach to professional development for science teachers. It draws on an extensive research base (Bell and Gilbert 1996; Hoban 2002; Loucks-Horsley et al. 2003; Shulman 1987), which puts teacher learning at the centre of the PD agenda and has been put into practice by researchers and science teachers.

KREST started by exploring the idea of teachers and researchers, comparing their views on what good teaching might look like in a number of areas of science education—so-called accomplished or expert science teaching. This process of sharing guild knowledge began to encourage science teachers to be more reflective about their own practice and that of others, so that they came to develop a collaborative understanding of what good teaching is and how they might try and improve their current practice. The essence of this approach was providing teachers with an avenue for discussion, and so meetings were arranged at approximately 6-week intervals in which teachers provided evidence of their developing practice and in which they supported and challenged the practice of colleagues.

Through the PD sessions and documenting evidence from their own classrooms to bring to these meetings, teachers began to formulate questions and descriptions of teaching and learning that prompted professional dialogue, reflection and critique, which in turn led to teacher learning. Using this approach, the teachers were able to see and discuss concrete examples of their own classrooms and those of colleagues on a similar quest, from which they could consider alternative approaches that might develop their own practice. The following transcript is taken from PD meeting 2 where teachers were discussing what they had tried in their own classrooms from the action plans they drew up in PD meeting 1.

Chloe	So I tried to do the 'wait time' strategy. I remember hearing about it on my PGCE but it was last session when I really thought why don't I do that. It makes sense. I told them (the class) what I was going to do and how it was going to help them think more and talk more instead of waiting for me to tell them the answer.
Derwi	Did it work? I know mine would just sit there or talk about something else.
Chloe	No. (Laughter) Well it did and it didn't. More of them were willing to answer and they did say more but I still ended up putting the ideas together for them and it was still the main two or three (students) answering.
Sally	Yes. That was what happened with my Year 10s but it got better the more I tried to do it.
Researcher	Can you say some more about that Sally?

Sally	Not sure if it was me or them or both but it just got better. Felt better. More started to join in and I had to do less. Things came up such as mixups between what is happening in osmosis and I managed to get them to sort it out rather than me correct it in their books later.
Derwi	Didn't it take a lot of time though?
Sally	Not really 'cos it got sorted out rather than having to keep coming back to it.
Derwi	I did say I was going to try it but I just don't feel it can work at our place. It's not like Sally and Chloe's schools and they (the students) just wouldn't … probably couldn't do it. I know my colleagues would think I would be barmy to even try.
Sally	I did do it with my best class.
Researcher	What do others think? Can 'wait time' be done with all classes?

The professional dialogue opens up an avenue for teachers to reflect on and justify their actions in their own classrooms. The teachers could weigh up the risks involved and so begin to balance their classroom aspirations alongside their own or perceived institutional expectations. Through this process, they were able to witness the reality of ideas in practice and then make decisions to adopt and adapt specific ideas and then trial and evaluate these in their own classrooms. By promoting reflection on lesson outcomes within group discussions, teachers made public what is usually hidden. These revelations of what certain decisions were made, and why, helped the teachers strengthen both their understanding of the nature of effective classroom assessment and their resolve to make such practices work in their own institutions. The science teachers constructed pedagogic knowledge by clarifying their own understanding and so reinforced their pedagogical identity as they interacted with others. They also identified common goals and talked through problems as they arose, which helped build their professional community. The PD sessions gave teachers both the language with which to reflect on their own work and a clear framework to assist them in their reflections.

Teachers documented their ideas in a portfolio both as part of the activities in the PD meeting and back in their schools when reflecting on previous meetings or planning for future ones. The PD session discussions and the portfolio entries provided a record of how a teacher practised, developed and self-evaluated his/her competence in a specific area of science teaching and demonstrated how a teacher practised the skills acquired from the PD programme within the classroom context (Joyce and Showers 2002), sometimes with one class but more often with several different classes.

The following transcript is taken from interview 2 with Derwi, where he is using his portfolio entry from PD meeting 1 to explain his developing practice to the interviewer.

Derwi	I said here, "Hearing Chloe and Sarah have success with 'wait time', I did feel I may have missed an opportunity to move AfL forward. With my classes. I had ditched the 'wait time' idea for mini whiteboards and that wasn't working as the boys saw these as an excuse to write and draw silly things instead of helping them have a go at an answer. Hearing Sally say

> she'd only tried it with her best class made me feel better." So I did try it with my top set Y8 and it was really good. I used 'talking partners' like Aisha did and they really took to it. So that was a real move forward for me because I had been dubious in the first meeting and getting back to school convinced me it couldn't work but now I have got it working with this class. Not sure I can do it with others but I can do it with them.

Through the portfolio and PD meeting discussions, Derwi was able to challenge his ideas and come back to consider possibilities for change in his classroom. He had gone away after the first meeting seemingly confident and eager to try things, but in the reality of his school, these intentions faded and various concerns constrained these changes. Hearing how other teachers had tried ideas, and particularly how they had been careful about which classes to try these ideas with, gave Derwi the confidence to try things out with a specific class. Without these discussions and the support of his peers, Derwi would probably have never attempted to make the changes that he did with classroom assessment practices.

Compiling a portfolio was valuable because it enabled each teacher to show his/her reflections on a particular lesson or activity, including how it might be taught differently on a future occasion. Reflection in this way enables teachers to assess their own ability to teach science, and it prompts them to reflect critically on their progress by identifying evidence of each element of that progress as their practice evolves. Such endeavours seem risky at the outset, but the collegiality, support and guidance from peers play a major part in driving this approach forward. The construction of portfolios, alongside the dialogue in the PD sessions, helped teachers improve their science teaching and to be more explicit about the progress they had made and the effects this had on their learners and classrooms.

Sadler's (1989) use of guild knowledge is pertinent here. He argues that teachers use a sense of what it means to be very good at something and translate such knowledge, 'guild knowledge', into everything they assess. The final grade awarded mirrors how closely they approximate what it means to be good at that particular skill, competency or understanding. Where formative assessment is involved, the teacher encourages the student, through peer marking and discussion, to enter that guild, empowering the learner to take a role in the assessment process. Such changes dramatically alter how classrooms function as assessment becomes an embedded practice, used by both teacher and students, to monitor progress and set learning targets. This evolution of assessment-driven learning is a far cry from the type of science classroom that relies predominantly on end-of-topic or end-of-year tests as the main form of assessment.

Pedagogic Decisions

Teaching is a highly personal activity where teachers bring together and make sense of notions of curriculum, pedagogy and assessment. While teachers do have some autonomy in the way that they choose to work in classrooms, they are increasingly required to demonstrate accountable outcomes (Brooks and Tough 2006), and this can lead to teachers limiting their range of pedagogical practices (Harlen and Deakin-Crick 2004). Clearly, there are implications both for teacher training and for professional development of teachers if schools are to strengthen their classroom assessment practices by including more formative approaches within their pedagogy.

To understand classroom practice and some of the reasoning behind how and why teachers make decisions about what they decide to do in the classroom, we need to consider who a teacher is and how teachers interact with their social setting. In order to do this, we need a theoretically driven model of teaching in context (Enyedy et al. 2006). There already exists a substantial tome of literature that explores how the beliefs of teachers affect the decisions they make in practice for both experienced (Nespor 1987) and inexperienced teachers (Pajares 1992). Much of this focuses on how teachers use their previous experience of classrooms to make sense of new situations and dilemmas as these arise. So, for both novice and experienced teachers, beliefs about lesson planning, assessment and evaluation influence the actions and decisions made in the classroom scenario (Enyedy et al. 2006). If context and experience strongly influence practice, then this suggests that it may be difficult to bring about change in practice as the 'status quo' of teachers' existence confines the interpretation of any new pedagogic ideas within the realms of previous ideas. This suggests that radical change in practice may be difficult to achieve, which has massive implications for professional development (PD) programmes.

Throughout their teaching career, teachers take part in professional development. This generally takes the form of a course or programme that the teacher participates in but can also result from working with colleagues within their own school or from personal endeavour by individuals in their own classrooms. In recent times, many PD programmes have adopted an approach to teacher change that conceptualises professional development from a personal growth perspective (Clarke and Hollingsworth 2002), where teachers come to make sense of what they do and how they might do things differently. At the heart of this approach is professional learning and teacher autonomy. Such an approach differs markedly from that offered historically, where teacher change has been approached through offering workshop opportunities where teachers could acquire or master predetermined skills and knowledge (Clarke and Hollingsworth 2002). The latter approach results in a technocratic skills-based approach to professional development (Kennedy 2005) and has received much criticism in the literature (Fullan and Stiegelbauer 1991; Howey and Joyce 1978; Wood and Thompson 1980) and was perceived by Guskey (1986) as a deficit model of 'fixing teachers'. With a personal growth approach comes a change in agency with the focus and drive coming from the teacher, and the purpose here is not to change teachers but for teachers to be actively involved in changing themselves. This is a much more personal and proactive approach to professional development than previously envisaged and involves professional reflection and action (Schön 1983).

Clarke and Hollingsworth (2002) draw attention to the 'idiosyncratic and individual nature of teacher growth' (p. 965) and the importance for both researchers and professional development trainers to understand more about teacher change. Teachers are regularly bombarded with ideas intended to improve their classroom practice but find that many of these ideas fail to come to fruition or get displaced by competing priorities (Harrison 2005). Effective PD needs to be designed to provide opportunities for professional development that are centred on classroom practice (Joyce and Showers 1988) and allow time and support for teacher reflection and learning. Through these processes new practices can be evolved (Priestly and Syme 2005), moulded and honed from existing classroom practice (Hoban 2002). Teachers need to familiarise themselves with new ideas and also understand the implications for themselves as teachers and for their learners gradually in the classroom before they accept or reject them. This involves them reshaping their own beliefs about what science teaching and science learning is, especially if the new practices suggest they 'go against the grain' (Cochran-Smith and Lytle, 1999).

Bell and Gilbert's work (1996) on the Learning in Science Project (LISP) suggests that there are three facets to teacher development that need to be considered to promote teacher learning and changes in practice. These are personal, social and professional development. Teachers use their beliefs about curriculum, pedagogy and assessment combined with their current knowledge of their students to decide both whether new ideas are worthwhile and how they might mould their current practice to take in these changes. This is neither an easy nor a simple move to make and is likely to result in different outcomes for different teachers. This is because teachers have complex beliefs about learning as well as a range of expectations and aspirations for their learners and for themselves and sometimes teaching dilemmas arise from the intersection of these beliefs and their identity. An example of this arose in one of our early formative assessment projects (Black et al. 2003), when one of the teachers was eager to improve classroom talk but was reticent to allow his students to work in groups, as he believed they would not focus on the task in hand without him leading the discussion. So while the teacher wanted to bring change in the classroom, his perception of his role in the teaching-learning process initially inhibited him from making that move. For several months, the teacher tried to improve classroom talk by working on his questions and working on 'wait time' (Rowe 1974), and it was only when he recognised that these actions in themselves were insufficient to build the type of classroom talk he wanted, that he was able to reconsider and try to integrate more group work in his lessons.

So, if teacher development needs to be construed from a professional growth perspective, then it needs to be planned and designed with teacher autonomy in mind, and its effectiveness needs to be considered from a teacher-learning viewpoint. My belief is that effective PD needs to provide an opportunity for teacher dialogue and reflection so that learning about how new practices can evolve or be moulded from existing classroom practice. In other words, professional development needs a formative approach that allows each teacher to self-regulate their own development. At the same time, such changes require collaborative endeavour to both provide support and offer possibilities of what can be achievable in the reality of the classroom.

Research evidence on teacher professional transformational change concurs that deep-rooted changes are difficult and generally take considerable time and effort to achieve (Fullan 2003). Senge and Scharmer (2001) argue that creating a system that facilitates such change requires action on three levels:

1. Establishing a shared statement of purpose and a shared set of guiding principles
2. Developing infrastructures that support community building
3. Undertaking collaborative projects that focus on key change issues that create concrete projects for further deepening common purpose and improving infrastructures (p. 242)

These ideas helped form the framework and design for the PD programme that we undertook within the KREST project.

This project provided the impetus for teachers to investigate a number of interrelated issues, namely, how AfL might be strengthened in their classroom, the role that formative assessment takes alongside how they make judgements of students' attainment, why and how these decisions foster or limit learning experiences in their classrooms and how they deal with the pressures they encounter when sharing assessment information with parents, students, teacher colleagues, senior leadership teams and inspectors. It has enabled the project teachers to consider what the 'ideal' assessment situation might be for them and to see how far they might reconcile the achievement of this ideal with the policy constraints and practical realities that they face on a day-to-day basis in their schools. The problem lies in unravelling the complexities of this reconciliation to produce a workable system of teacher assessment.

The teachers found it challenging to put the ideas developed during the PD sessions into practice when they returned to their classrooms. It required a full understanding about why they were bringing in new practices. This involved examining and perhaps changing their views about what constituted effective science teaching. Teachers' identity is an important factor in terms of how they negotiate their role within their school community, and this has a direct impact on their practice (Enyedy et al. 2006). They also needed to justify to their students and other colleagues their reasons for changing practice.

Teachers initially found the evidence-based approach to PD difficult due to time constraints, problems of acceptability of new approaches with colleagues in school and reticence to work new ideas into their existing practice, so that different overall practice emerged. However, when they looked back on their experiences, many of the teachers recognised that the concerns they had had were not as critical as first envisaged:

> The idea of having to bring evidence was scary but, in reality, it's been the thing that has helped me see what I am doing and not doing to help my students learn. (Chloe, Interview 3)
>
> I wanted to change how I did assessment in my classroom but the pressures from our senior leadership's approach to assessment seemed unsurmountable at first. Little by little, I found ways of squeezing in more AfL, while satisfying the examination gods. (Tracey, Interview 3)
>
> Creating a portfolio and justifying to yourself and others that how you were teaching was helping (students learn) seemed impossible but now I see it as an inevitable and necessary step in helping me understand what really counts in terms of classroom assessment. (Aisha, Interview 3)

Conclusion

The KREST project demonstrated that asking teachers to recognise and collect evidence from their own classrooms reduced dependency on an external expert coach and established a more autonomous approach to professional development. It also helped to foster a teacher-learning community that provided support for the teachers in their learning both within the timescale of the project and beyond. Being able to share difficulties and achievements with their peers provided the impetus for teachers to take risks when they returned to their own classrooms. This was crucial to their professional growth. The researchers noted that, on the whole, the teachers felt positive and satisfied with the programmes and as such:

- Enhanced their acquaintance with particular domains of science teaching and learning that fitted with an AfL approach
- Improved their pedagogical content knowledge
- Improved their practical teaching knowledge
- Heightened their sensitivity to students' understanding and progress
- Empowered teachers as professional learners

The main findings within this domain of KREST were that teachers could strengthen their formative assessment practices at the same time as they carried out periodic summative assessments. The reason why this was possible, despite strong pressures from within their schools to focus on summative assessment, was the essential part in aiding professional learning played by the sociocultural practices that were engaged in through the PD programmes, which strengthened their resolve to include a more formative approach in their classrooms. In order to do this, the teachers needed to recognise good practice within a domain, make sense of its complexities and understand the effects and synergies of various aspects of practice as they came to find their own ways of establishing such practice within their own institutions.

This way of working took considerably more time than had been envisaged at the start of the project as teachers needed to be involved in planning, actioning and evidencing practice as well as analysing and reflecting on their teaching and that of others. Many researchers have commented on the slow pace of teacher change (Fullan 2005; Hargreaves 2005), and the KREST project documents some of the reasons why this is inevitable, as teacher identity is challenged by new ideas from professional development programmes invading the classroom domain. What was clear was that even within the contentious area of changes in classroom assessment practices, the resolve and professional bonds that formed within the teacher community steadied the situation sufficiently to prevent teachers rejecting new ideas as untenable and provided breathing space for them to suspend disbelief of new ideas so that there was sufficient time and opportunity for new practice to develop at a reasonable pace. This approach to PD meant that the teachers did not become risk averse and instead were energised to take action as they took control of the change within their own classrooms.

Much of the literature on teacher development focuses on the training methods used and the design of programmes with effectiveness outcomes sometimes measured by student achievement or, more often, by teacher confidence. Making changes in classroom practice is a highly complex series of events, and when teachers take part in professional development, much of the change process that they go through and the variability in implementation of the professional development goals of the programme remain undocumented. By looking in some detail at the professional development of science teachers on an Assessment for Learning Project, we explored why and how a focus on change in assessment practices requires professional development programmes that support professional dialogue and encourage teacher autonomy in order for teachers to develop as reflective practitioners.

Much of the early literature on professional development artificially narrows and simplifies the path towards professional growth, often resulting in a stepwise 'catch-all' approach to professional development, which engages only some aspects of teacher beliefs and identity. The KREST project highlighted why there is a need to focus on teacher change from a professional growth perspective because changes in assessment practice may impinge on teacher beliefs about pedagogy and curriculum as well as assessment. This has implications for both training new teachers and the professional development of teachers.

References

Black, P., McCormick, R., James, M., & Pedder, D. (2006). Learning how to learn and assessment for learning: A theoretical inquiry. *Research Papers in Education, 21*(2), 119–132.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice, 5*(1), 7–73.

Black, P. J., & Wiliam, D. (1998b). *Inside the black box*. London: NFER.

Black, P. J., & Wiliam, D. (1998c). Developing a theory of formative assessment. In J. Gardener (Ed.), *Assessment and learning* (pp. 81–100). London: Sage.

Black, P., Harrison, C., Lee, C., Marshall, B., & Dylan, W. (2002). *Working inside the black box: Assessment for learning in the classroom*. London: King's College London Department of Education and Professional Studies.

Black, P., Harrison, C., Lee, C., Marshall, B., & Dylan, W. (2003). *Assessment for learning: Putting it into practice*. London: Open University.

Bell, B., & Cowie, B. (2001). *Formative assessment and science education*. Dordrecht: Kluwer.

Bell, B., & Gilbert, J. (1996). *Teacher development: A model from science education*. London: Falmer.

Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles, Policy and Practice, 11*(1), 7–26.

Boud, D., & Falchikov, N. (2004). *Beyond formative and summative assessment: Developing a new agenda for assessment for lifelong learning.* Second Biannual Joint Northumbria/EARLI SIG Assessment Conference, University of Bergen, Norway.

Brooks, R., & Tough, S. (2006). *Assessment and testing: Making space for teaching and learning*. London: IPPR.

Carless, D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education, 12*(1), 39–54.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*(2), 219–233.

Clarke, D. J., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*, 947–967.

Cochran-Smith, M., & Lytle, S. (1999). Teacher learning in communities. In A. Iran-Nejad & C. D. Pearson (Eds.), *Review of research in education*. Washington, DC: American Educational Research Association.

DCSF. (2007). *Assessment for Learning (AfL) 8 schools project*. http://nationalstrategies.standards.dcsf.gov.uk/node/97897. Accessed 29 Aug 2010

Enyedy, N., Goldberg, J., & Welsh, K. (2006). Complex dilemmas of identity and practice. *Science Education, 90*(1), 68–93.

Fairbrother, R. W., Dillon, J., & Gill, P. (1995). Assessment at Key Stage 3: Teachers' attitudes and practices. *British Journal of Curriculum and Assessment, 5*(3), 25–31 and 46.

Fullan, M. (2005). *Leadership and sustainability: System thinkers in action*. Thousand Oaks: Corwin Press.

Fullan, M., & Stiegelbauer, S. (1991). *The new meaning of educational change* (2nd ed.). New York: Teachers College.

Guskey, T. R. (1986). Staff development and the process of teacher change. *Educational Researcher, 15*(5), 5–12.

Hacker, R. J., & Rowe, M. J. (1997). The impact of National Curriculum development on teaching and learning behaviours. *International Journal of Science Education, 19*(9), 997–1004.

Hargreaves, A. (2005). Educational changes take ages: Life, career and generational factors in teachers' emotional responses to educational change. *Teaching and Teacher Education, 21*, 967–983.

Harlen, W., & Deakin-Crick, R. (2004). *Testing, learning and motivation*. Cambridge: Assessment Reform Group.

Harlen, W. (Ed.). (2006). *The role of teachers in the assessment of learning*. London: ARG.

Harrison, C. (2005). Teachers developing assessment for learning: Mapping teacher change. *International Journal of Teacher Development, 9*(2), 255–263.

Harrison, C. (2011). Classroom assessment. In J. Dillon & M. Maguire (Eds.), *Becoming a teacher: Issues in secondary teaching* (4th ed., pp. 222–235). London: Open University.

Harrison, C., & Howard, S. (2009). *Inside the primary black box: Assessment for learning in primary and early years classrooms*. London: GLAssessment.

Harrison, C., Hofstein, A., Eylon, B. S., & Simon, S. (2008). Evidence-based professional development of science teachers in two countries. *International Journal of Science Education, 30*(5), 577–591.

Hoban, G. F. (2002). *Teacher learning for educational change*. Buckingham: Open University.

Howey, K. R., & Joyce, B. R. (1978). A data base for future directions in in-service education. *Theory Into Practice, 27*, 206–211.

Hutchinson, C., & Hayward, L. (2005). The journey so far: Assessment for learning in Scotland. *Curriculum Journal, 16*(2), 225–248.

James, M., & Mansell, W. (2009). *Assessment in schools—fit for purpose?* London: ESRC/TLRP.

James, M., Black, P., Carmichael, P., Drummond, M., Fox, A., MacBeath, J., & McCormick, M. (2007). *Improving learning how to learn: Classrooms, schools and net-works*. London: Routledge.

Joyce, B., & Showers, B. (1988). *Student achievement through staff development*. White Plains: Longman.

Joyce, B., & Showers, B. (2002). *Student achievement through staff development*. White Plains: Longman.

Kennedy, A. (2005). Models of continuing professional development: A framework for analysis. *Professional Development in Education, 31*(2), 235–250.

Mansell, W. (2007). *Education by numbers: The tyranny of testing*. London: Metheun.

Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies, 19*, 317–328.

Ofsted. (2007). *Annual report of Her Majesty's Chief Inspector 2006–2007*. London: Her Majesty's Stationery Office.

Ofsted. (2010). *Standards and Quality 2009/10. The Annual Report of Her Majesty's Chief Inspector of Schools*. London: Her Majesty's Stationery Office.
Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307–333.
Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
Pollard, A., Triggs, P., Broadfoot, P., McNess, E., & Osborn, M. (2000). *What pupils say: Changing policy and practice in primary education*. London: Continuum.
Popham, W. J. (2001). Teaching to the test. *Educational Leadership, 58*(6), 16–20.
Priestly, M., & Syme, H. (2005). Formative assessment for all: A whole school approach to pedagogic change. *The Curriculum Journal, 16*(2), 475–492.
Rowe, M. B. (1974). Wait time and rewards as instructional variables, their influence on language, logic and fate control. *Journal of Research in Science Teaching, 11*, 81–94.
Russell, T. A., Qualter, A., & McGuigan, L. (1995). Reflections on the implementation of National Curriculum science policy for the 5-14 age range: Findings and interpretations from a national evaluation study in England. *International Journal of Science Education., 17*(4), 487–492.
Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144.
Schön, D. (1983). *The reflective practitioner*. London: Temple Smith.
Senge, P., & Scharmer, O. (2001). Community action research: Learning as a community of practitioners. In P. Reason & H. Bradbury (Eds.), *Handbook of action research: Participative inquiry and practice* (pp. 238–249). London: Sage.
Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22.
Simpson, M. (1990). Why criterion-referenced assessment is unlikely to improve learning. *Curriculum Journal, 1*(2), 171–183.
Smith, E., & Gorard, S. (2005). "They don't give us our marks": The role of formative feedback in student progress. *Assessment in Education: Principles, Policy & Practice, 12*(1), 21–38.
Wood, F. H., & Thompson, S. R. (1980). Guidelines for better staff development. *Educational Leadership, 37*(5), 374–378.

Chapter 19
Assessment: Where to Next?

Deborah Corrigan, Cathy Buntting, Richard Gunstone, and Alister Jones

Introduction

Early in Chap. 1 we noted that '[i]n this volume, a range of authors explore assessment philosophies and practices and possibilities from different socio-cultural contexts and across educational levels, from early childhood through to tertiary level'. We then presented a frame for the contributions and gave a very brief outline of the central foci of each. At the end of Chap. 1, in the final paragraph, we observed:

> The concluding messages expressed in each of the chapters in this volume also provide a basis for consideration of where the gaps might be in thinking about assessment in science education and research. In the final chapter of this volume we therefore offer our analysis of what these gaps are, and suggest possible fruitful areas for further investigation in order to enhance assessment's role in relation to science education policy, curriculum and pedagogy.

And so we turn now to these gaps, to those things that preceding chapters have indicated are 'still to be done' in terms of the valuing of assessment in science education on a number of fronts, but particularly in terms of pedagogy, curriculum and policy. We also are clear that the substantial lack of student voice in the preceding chapters is, by its omission from these chapters, another indicator of a gap. This gap is of such significance it is difficult to overstate its importance.

D. Corrigan (✉) • R. Gunstone
Centre for Science, Mathematics and Technology Education,
Faculty of Education, Monash University, Building 6, Wellington Road,
Clayton, VIC 3800, Australia
e-mail: debbie.corrigan@monash.edu; richard.gunstone@monash.edu

C. Buntting • A. Jones
Faculty of Education, University of Waikato, Gate 1 Knighton Road,
Hamilton 3240, New Zealand
e-mail: BUNTTING@waikato.ac.nz; a.jones@waikato.ac.nz

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6_19,

We choose to make the following points of 'what still needs to be done' in a succinct manner. Our intent here is only to introduce and outline and not in any way to attempt some sort of developed research framework or agenda. We have chosen this path because we recognise that there will be many ways of addressing any of the points that follow.

The outline that follows is in two sections—'Policy and Curriculum and the Value of Assessment' and 'The Voice of the Student in the Assessment Agenda: An Issue of Pedagogy'. We intend that each be read with recognition that it is crucial that science educators in all contexts (classrooms, systems, research) take greater responsibility in attending to the agenda of valuing assessment and recognising its fundamental impact on all other aspects of science education.

Policy and Curriculum and the Value of Assessment

There is clearly an increasing focus on accountability in education, at all of the individual, school, system and country/jurisdiction levels. This has raised the profile of and importance for policy makers of international comparative achievement tests such as TIMMS and PISA (see, e.g. Fensham et al., this volume). The demand from policy makers for similar indicators of achievement has seen an increase in many countries of national testing in a similar vein to TIMSS and PISA. It remains unclear how such indicators provide an accountability in line with the educational goals of any given school, system, jurisdiction or country. It is commonly the case that the goals of an education system are broader than student learning of specific subject (e.g. science) content. It is appropriate, indeed essential, that more work be undertaken to develop indicators of achievement that reflect such goals. This is needed both to give more valid possibilities for considering the achievements of a system and to make cross system comparisons more valid.

Millar (this volume) begins to pose questions such as 'What will we accept as evidence of the achievement or non-achievement of any given learning objective?' as he suggests that 'assessment becomes the operational definition of the objective'. Such a perspective may help policy makers, system bureaucrats and curriculum developers think about the fundamental importance of closer alignment of assessment practices and the learning objectives that are valued as worthwhile in science education. Too often and in too many contexts, curriculum development and assessment development are quite separated, and current approaches essentially mean that well-established assessment practices dictate which learning objectives that are valued. This is most often the case when the primary purposes of assessment are for certification and hence high stakes (and the demands for public accountability) rather than supporting learning (where the focus is on achievements of the individual). Paradoxically, it is in such certification/high-stakes contexts where separation of curriculum development and assessment development is often the most extreme, even though logic would suggest these are the contexts where it is most important for the two to be integrated. The science education research community needs to do much more to develop reliable and valid assessment approaches of a wide range of types, which are of high quality and credibility, and allow assessment of the full range of valid goals for science education. Only then can comparisons be made with validity and confidence.

There has been much commentary in this volume about the need for and importance of a contextualised approach to curriculum. The assessment of such contextualised approaches has been problematic for many of the reasons cited above and elsewhere in this volume; the development of assessment instruments as suggested above must be a priority for the future.

In the last several decades, science has evolved and shifted in its practices, for example, it has become more interdisciplinary and collaborative and less reductionist. We argue that it is clear science education has not evolved in a similar way. Of particular relevance to this volume is that assessment of science learning has rarely evolved to incorporate the interdisciplinarity and collaborative nature of modern science. Indeed, the historic view of science is the antithesis of interdisciplinary, collaborative, less reductionist approaches, but the assessment of this historical view has persisted in science education. This lingering and less relevant view of science suggests science education has lost touch with the dynamic nature of science. Many of the curriculum examples provided in this volume suggest a vibrancy in curriculum development that unfortunately is not translated into the assessment practices of such curricula.

The Voice of the Student in the Assessment Agenda: An Issue of Pedagogy

When we talk about pedagogy, for many this can mean teaching. However, based on Loughran's definition (Corrigan et al., this volume), pedagogy is the relationship between learning and teaching. The implication of this more broadly defined view of pedagogy is that the voice of the learner is situated centrally in any discussion about pedagogy.

As pointed out by Fensham and Rennie (this volume), 'each student is individually entitled to the assessment of their learning' and such assessment should be as authentic as possible using a variety of modes. What is not clear from current assessment practices, both broadly and specifically in science education, is how such an entitlement is realised, particularly given that students rarely have any say about the form this entitlement should take. Fensham and Rennie have proposed that students should be maintaining a portfolio of their achievements. Such a notion may gain more acceptance if we personalise this argument and ask, *How would I like my life's work be presented and re-presented (or presented in different ways)?*

In this century, most education professionals in most contexts, be they academics or teachers, administrators or system bureaucrats, are asked to maintain some form of portfolio of work in the form of a performance management plan. To do this one sets goals and targets, details achievements and outputs and makes judgements about how successful the pursuit of the goals and targets has been. Then one engages with a supervisor or mentor who makes some form of summative judgement about these things and suggests possibilities to encourage further development. None of this plan or portfolio is independent of who each individual is as the individual's interests and motivations provide some direction for how they shape their endeavours, despite the requirements any employment may place upon them. All have some choice in what is included or not included in the portfolio because of its relevance to each individual.

Documenting one's life work in a portfolio is a dynamic and ongoing process and clearly not just something done at the end of a period of activity. It requires reflection on personal achievements and reflection on personal change—how one is now compared to a month, a year or a decade before. It makes statements about how one's knowledge, skills and abilities have developed, both holistically and specifically.

It is revealing to take these arguments about portfolios and personal development and contemplate them in the context of classrooms. At all levels of formal education, students' current stage of achievement is determined in formative ways within the teaching and learning situation and often determined in a summative way through the use of assignments, tests and examinations at senior levels of school. However, there is rarely any opportunity for students to determine for themselves the ways in which they can present and represent their own achievements. In other words, the ways in which education professionals now commonly find their own learning and development being fostered by the processes of personal documentation of and reflection on their own learning are very rarely applied to the learning of the students those professionals are responsible for.

Some possibilities for such self-determined approaches exist in formative assessment processes that focus on providing feedback so that the learner can better understand the relationship between her or his current performance and the desired performance. However, even here the parameters of what is expected are set by the teacher. Students have little chance to develop any sense of autonomy over what to present and represent in terms of their learning achievements as fundamental decisions are made for them. If, as Fensham and Rennie suggest, students are individually entitled to the assessment of their learning, it is not clear how this occurs in systems where students have little voice in determining what they have learnt and how they can represent their individual learning. The dynamic and ongoing nature of documenting one's life work as an education professional does not seem to be a part of the assessment process for documenting a student's learning journey in school.

What is also apparent from the example above is that a portfolio representing learning achievements (of an education professional or a student) is geared towards a profile of learning rather than a single and specific score that is derived from a variety of tasks and modes of assessment. For example, an area that is often under-represented in the consideration of assessment in science learning is performance assessment and the role it can play in providing teachers with observable evidence of student understanding of particular ideas. Such performance assessment in science is not limited to 'doing practical/experimental activities' but includes the representation of science ideas through role play, the creation of images, the development of an oral argument or debate and so on. With the use of modern technology, such observable performances can be recorded to provide evidence of important forms of student learning in science.

In developing an individual student's profile, the classroom teacher is the person most able to undertake such assessment in the most authentic way. In other words, we must trust in the professional knowledge of teachers in making judgements about what their students have achieved and what potential these students have for future achievements. Helpful here is Abell and Siegel's (2011) model of science assessment literacy, which details the range of assessment knowledge, types and skills that teachers need to create an assessment-centred learning environment. At the heart of this model is the teacher's view of learning, which itself is influenced by a core set of values and principles about science learning and assessment that guide decision making: 'These values and principles interact with four categories of science teacher knowledge of assessment—assessment purposes, what to assess, assessment strategies, and assess interpretation and resulting actions—which also interact with each other in practice' (p. 211).

From this model, there is a clear indication that there are specific assessment practices associated with science, since the assessment practices are dependent on a teacher's science knowledge and pedagogical content knowledge. Of course, more general aspects of teaching and learning such as pedagogical knowledge and knowledge of learners are also important, but it is the values, principles and beliefs a teacher holds about science, science learning and assessment that should take precedence. This points to the need for professional development and other teacher education programmes to provide insights into:

- Views of science learning where the learner actively constructs her or his understanding and how assessment needs to provide evidence of individual understanding and learning
- The inclusion of the student's voice in choosing how her or his work is assessed
- How assessment information is interpreted and the subsequent process of making decisions about teaching and assessment
- Consideration of whether assessment accurately reflects the desired learning outcomes and whether it is equitable and authentic

Some Future Possibilities

It probably comes as no surprise to the reader that this volume has raised more questions that it has answered. What are the priorities in assessment—accountability or improved student learning? Are these two priorities mutually exclusive? What might more authentic assessment of curriculum innovations such as contextualised science and scientific literacy look like? How can students of science be given more control over what they learn and the evidence they provide that they have learnt it?

How can we assess science practices holistically and authentically? What is the role and function of new technologies in assessing students' learning in science? How can disconnects between curriculum and assessment policy be addressed at the level of the classroom when teachers continue to face pressure to 'teach to the exam'? How does the interplay between real and virtual in-school and out-of-school experiences impact on such assessment? Associated with each of these questions is also how to foster competence and capability in teachers for emergent assessment practices.

There is therefore still much to be done by the science education community if we are to demonstrate the value of assessment as it plays out in the pedagogy, curriculum and policy arenas of school science. What is presented in this final chapter is just a beginning in the discussion, as not only is the nature of knowledge itself changing, but so is the way we acquire it, process it, store it, retrieve it and use it. To determine what science knowledge of worth is—and how to assess it—remains a significant challenge.

Reference

Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching* (pp. 205–221). Dordrecht: Springer.

Author Biographies

Mike Askew is the foundation chair professor of primary education at Monash University (Australia). Prior to that, he was chair and professor of mathematics education at King's College London. He has directed many research projects including the 'Effective Teachers of Numeracy in Primary Schools', 'Raising Attainment in Numeracy' and 'Mental Calculations: Interpretations and Implementation'. He was deputy director of the 5-year Leverhulme Numeracy Research Programme, examining teaching, learning and progression in number from age 5 to age 11. The findings from such research have influenced education policy in England and abroad. His most recent book is *Transforming Primary Mathematics* (2012, Routledge).

Beverley Bell is an associate professor in the Faculty of Education at Waikato University (New Zealand). She has been a teacher of science in secondary schools and a national curriculum developer and has extensive experience of supervision of master's and doctoral research. Her research interests include science education; teaching, learning and assessment in general; and teacher development/education/learning. Her most recent book is *Theorising Teaching in Secondary Classrooms: Understanding Our Practice from a Sociocultural Perspective*.

Paul Black is emeritus professor of science education at King's College London (UK). He worked as a physicist in Birmingham University for 20 years before moving to a chair in science education in London. He has made many contributions to curriculum development for science in the Nuffield Curriculum Projects at primary and secondary levels, as leader for the national science surveys of the Assessment of Performance Unit in the 1980s, and to research into learning and assessment. He was chair of the UK government's Task Group on Assessment and Testing in 1988, which formulated advice on the new national assessment system. He has served on three assessment advisory groups of the US National Research Council as visiting professor at Stanford University. His work on formative assessment with Dylan Wiliam and colleagues at King's has had widespread impact.

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6,

Cathy Buntting holds a research position within the Wilf Malcolm Institute of Educational Research at the University of Waikato, Hamilton (New Zealand). She has a master's degree in biochemistry and a Ph.D. in science education and was responsible for managing the initial development of the New Zealand Biotechnology Learning Hub. Her research interests are in science and technology education, including biotechnology education. Her current research focus is on how digital technologies can enhance student learning in these subject areas.

Audrey Champagne is professor emerita at the University at Albany, State University of New York (USA). The focus of her research has been on students' understanding of science principles, especially non-canonical principles that persist even after extensive exposure to formal science instruction. She has served on framework committees for the US National Assessment of Educational Progress and the IEA TIMSS. These activities have provided opportunity to investigate the alignment of assessment tasks with national standards and goals for science education and to consider how the structure of assessment tasks influences the responses they provoke in students.

Rebecca Cooper is a science educator in the Faculty of Education, Monash University (Australia). She works with pre-service and in-service science teachers, and her research interests include considering how science teachers and science teacher educators develop pedagogical knowledge throughout their career, improving the quality of science teaching to increase student engagement and working with teachers on promoting values in their science teaching in an effort to better understand the development of scientific literacy with students. Prior to working at Monash University, Rebecca taught physics, science and mathematics in secondary schools for many years.

Deborah Corrigan is an associate professor in science education and the deputy dean of the Faculty of Education at Monash University. After working as a chemistry and biology teacher, she worked at Monash University in chemistry and science education, particularly in teacher preparation. Her research interests include industry and technology links with chemistry curricula and curriculum design and science education policy. However, her main research interests remain improving the quality of chemistry and science education so that it is relevant to students and improving the professional practice of teachers and other industry professionals.

Bronwen Cowie is an associate professor and director of the Wilf Malcolm Institute of Educational Research, Faculty of Education, and the former director of the Centre for Science and Technology Education Research at the University of Waikato (New Zealand). She has been a teacher of science in secondary schools. She has directed a number of large national research and development projects and has extensive experience in classroom-based research. Her research interests include the nature of assessment for learning interactions in science and technology classrooms, student voice, culturally responsive pedagogy, curriculum implementation and the role of ICTs in learning science.

Peter Fensham is emeritus professor of science education, Monash University, and an adjunct professor in the School of Mathematics, Science and Technology Education, Queensland University of Technology (Australia). His interests centre on curriculum reform and assessment in science education and on issues of equity in education.

Angela Fitzgerald is a lecturer in the Faculty of Education at Monash University (Australia). She joined the Monash team in mid-2010, having previously worked as a secondary school science teacher and completed doctoral research in the area of primary science education. In particular, Angela's research interests focus around working with pre-service and in-service teachers in a variety of ways to improve the quality of science learning and teaching in primary schools.

Marilyn Fleer holds the foundation chair of early childhood education at Monash University (Australia). She is currently the president of the International Society for Cultural Activity Research (ISCAR). Her research interests focus on early years learning and development, with special attention on pedagogy, culture, science and technology.

Elvira Folmer is a senior curriculum researcher in the research department of SLO, the Netherlands Institute for Curriculum Development. Her research interests include curriculum monitoring and evaluation. She is involved in the evaluation of the renewal of mathematics and science education in senior secondary education and is also project leader of the European Science Education Curriculum Research (SECURE), a comparative study of mathematics and science curricula in ten European countries.

Richard Gunstone is emeritus professor of science and technology education, Monash University (Australia). His current major research projects involve the ways emotion is involved in the scientific/technological learning of very young children and the exploration/rethinking of the school education of prospective scientists. He is editor in chief of the forthcoming *Encyclopedia of Science Education* (Springer).

Chris Harrison is a senior lecturer in science education at King's College London (UK). She taught science in schools for 13 years and ran the science teacher training course at King's for many years. Her research interests range across science education, assessment and teacher learning but with a particular focus on assessment for learning.

Alister Jones is a research professor and deputy vice chancellor at the University of Waikato, Hamilton (New Zealand). Prior to this, he was dean of education, and he is a past director of both the Wilf Malcolm Institute of Educational Research and the Centre for Science and Technology Education Research at the university. His current roles include visiting professorships at King's College London and the Hong Kong Institute of Education and membership of the APEC working group on science and math teacher education. He is also director of the New Zealand Science Learning Hub.

Wilmad Kuiper is a professor at the Freudenthal Institute for Science and Mathematics Education at Utrecht University (Netherlands). He is also head of the research department of SLO, the Netherlands Institute for Curriculum Development. SLO is an independent, non-profit organisation, bridging the contexts of policy, research and practice. His research interests include curriculum evaluation, particularly regarding mathematics and science education, and the consistency between (science) curriculum renewal and assessment.

Desmond Mene Lee-Hang is a senior lecturer in the Faculty of Science at the National University of Samoa (Samoa). He completed his doctorate in education from the University of Waikato in 2011. He has been a science teacher since 1996. He has also been involved for a number of years in science curriculum development and curriculum reviews and assessments in Samoa. His research interests include science education, culturally appropriate formative assessments, Samoan language issues in science education, and the influence of culture on science learning, teaching and assessment.

John Loughran is the foundation chair in curriculum and pedagogy and dean of the Faculty of Education, Monash University (Australia). John was a science teacher for 10 years before moving into teacher education. His research has spanned both science education and the related fields of professional knowledge, reflective practice and teacher research. John was the co-founding editor of *Studying Teacher Education* and is an executive editor for *Teachers and Teaching: Theory and Practice*. His most recent books include *Understanding and Developing Science Teachers' Pedagogical Content Knowledge* (Loughran et al. 2012, 2nd Edition), Sense Publishers; *What Expert Teachers Do* (2010), Allen & Unwin/Routledge; and *Developing a Pedagogy of Teacher Education: Understanding Teaching and Learning About Teaching* (2006), Routledge.

Hongming Ma is currently a lecturer in the Faculty of Education at Monash University (Australia). She used to work as a secondary vocational school teacher in China. Her research interests are related to cross-cultural understanding of issues in science education. Specifically, the issues include the nature of science and its role in school science curriculum at different levels, affective learning in science and relevant teacher pedagogical knowledge, and cultural influence on people's understanding of the nature of science and on affective dimensions of science teaching and learning.

Robin Millar is Salter's professor of science education at the University of York (UK). He moved to York in 1982, after working for 8 years as a secondary school science teacher. His main research interests are in teaching and learning science (especially physics) at secondary school level and curriculum design and development with a particular focus on the development of scientific literacy. He has directed several large national research and curriculum development projects and was an author of the influential *Beyond 2000* report. He was president (1999–2003) of the European Science Education Research Association and a member of the PISA Science Expert Group for the 2006 and 2015 studies.

Pernilla Nilsson is an associate professor in science education at Halmstad University (Sweden). She is also chair of the board of the Swedish Association of Research in Science Education. During the last 10 years, she has been involved with organising pre-service and in-service science teacher education programmes, which focus on scientific literacy, practical work and stimulated reflection. In 2002, she also initiated the work with building up a science learning centre at the university, which has since then been the context for her research as well as teaching practices for pre-service and in-service teacher education. Her research interests concern pre-service and in-service teachers' professional learning, such as the development of pedagogical content knowledge (PCK) in primary science and factors that influence that development. Her research also concerns scientific literacy, early childhood science, self-study research and different forms of assessment such as formative interaction and self-assessment.

Wout Ottevanger is a senior curriculum researcher in the research department of SLO, the Netherlands Institute for Curriculum Development, with specific focus on (science and mathematics) curriculum evaluation. Until recently, he also was a member of the staff at the VU University Amsterdam, with special attention towards the development of education in an international context with partners in countries such as Tanzania, Ghana and Yemen.

Gloria Quiñones is a lecturer in the Faculty of Education at Monash University (Australia). She has researched family and teaching practices, children's play, science and technology in early childhood and primary education settings, in both Mexico and Australia. Her doctoral research focused on the role of emotions in children's everyday life across family and early childhood settings.

Léonie J. Rennie is a professor of science and technology in the Science and Mathematics Centre at Curtin University in Perth (Australia) and adjunct research professor in science communication at the University of Western Australia. Her research interests focus on learning and assessment in science and technology education in integrated and out-of-school contexts and the promotion of scientific literacy. She is the author of over 200 book chapters and journal papers and co-editor/co-author of several books, including *Knowledge that Counts in a Global Community: Exploring the Contribution of Integrated Curriculum* (2012) and *Integrating Science, Technology, Engineering, and Mathematics: Issues, Reflections, and Ways Forward* (2012).

Jinwoong Song is a professor in the Department of Physics Education at Seoul National University (Republic of Korea) and has previously been the director of BK21 Science Education for the Next Society (SENS) group and the president of East-Asian Association for Science Education (EASE). His research interest focuses on students' learning of and teachers' practice of secondary science (especially physics) in its social, historical and philosophical contexts. He is also interested in informal science education, history of science education and East Asian science education. Beyond science education, he has been actively engaged with international collaboration and exchanges of (pre-service and in-service) teachers and teacher education programmes.

Index

A
AAAS. *See* American Association for the Advancement of Science (AAAS)
ACARA. *See* Australian Curriculum, Assessment and Reporting Authority (ACARA)
Achievement, 4, 8, 12, 18, 25–27, 29, 33–42, 46–48, 55, 56, 71, 74–90, 142, 185, 191, 210, 211, 214, 217, 218, 220, 223, 249, 282, 285–304, 319, 320, 327–330, 354–356, 360–363. *See also* Performance
Affective, 15–17, 26, 27, 60, 86, 156, 184, 198–201, 241–244, 297, 299, 302, 330
Alternative conceptions, 14, 252, 262, 309, 311, 317
American Association for the Advancement of Science (AAAS), 120, 135, 142, 144
Argumentation, 74, 80, 301
Askew, M., 6, 7, 169–180, 217
Assessment
  assessment profile, 23
  authentic assessment, 71, 74, 75, 78, 87, 90, 91, 216, 363
  classroom assessment, 33–51, 249–263, 348, 350–352, 354, 355
  contextualised assessment, 20–22, 25
  diagnostic assessment, 311, 314
  modes of assessment, 71, 74–75, 362
  peer assessment, 214, 327, 331, 347
  potentive assessment, 239, 242–244
  school-wide assessment, 43, 103
Assessment as learning, 2, 103
Assessment for learning, 2, 34, 47–49, 103, 232, 239, 325, 327–329, 341, 343, 347, 356. *See also* Formative assessment
Assessment literacy, 220, 221, 227, 363
Assessment of learning, 2, 5, 14, 24, 103, 114, 325, 327, 341
Astronomy, 309, 311, 317
Attitudes, 3, 7, 16, 17, 36–39, 60, 89, 122–127, 129–131, 133, 146, 156, 157, 170, 185, 189–191, 193–198, 200, 201, 209, 216, 240–244, 286, 293, 301, 303
Australia, 6, 8, 13, 17, 21, 23–27, 29, 36, 153, 156, 161, 219, 234–237, 268, 282, 309
Australian Academy of Science, 308, 309, 323
Australian Curriculum, Assessment and Reporting Authority (ACARA), 17, 69, 70, 80, 169

B
Backwash, 103
Beliefs, 3, 4, 6, 80, 84, 115, 153–168, 176, 180, 183, 187, 223, 250, 272, 304, 335, 352, 353, 356, 363
Bell, B., 8, 103, 213, 249, 251–253, 267283, 314, 347, 349, 353
Biology, 47, 73, 104–105, 107–108, 110–114, 131, 140, 141, 146, 157–160, 185–188, 193, 200, 218, 273, 331
Black, P., 5, 7, 12, 34, 47, 48, 55, 71, 73–75, 90, 91, 101, 103, 207–227, 233, 249, 267, 285, 304, 308, 314, 326–328, 341–343, 347, 348, 353
Buntting, C., 1–9, 12, 33–51, 359–364

C
Case study, 6, 8, 91, 307–323
Cassrooom assessment. *See* Assessment

D. Corrigan et al. (eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, DOI 10.1007/978-94-007-6668-6,

Champagne, A., 6, 119–, 226
Chemistry, 35, 36, 47, 72–73, 75, 80, 86, 104–105, 107, 109, 111–114, 127, 131, 140, 157, 166, 185–188, 193, 200, 330, 331
China, 7, 183–202, 226
Chinese Taipei, 36, 287, 288, 290–294
Classroom, 1, 23, 33–51, 67, 71, 101, 119, 153, 170, 187, 207–227, 249–263, 267–283 285, 307, 327, 347–356, 360
Cognition, 7, 133, 149, 240, 242–244, 328
Cognitive, 15, 17, 21, 23, 24, 27, 36, 50, 129, 132, 133, 146–149, 155, 156, 158, 162, 163, 165, 193, 200, 201, 210, 234, 238, 238, 240–242, 244, 297, 299, 303, 330
Competence, 15, 17, 19–27, 77, 220, 302, 325, 327, 350, 364
Competency, 4, 25, 166, 167, 200, 351. *See also* Key competencies
Conceptual understanding, 57, 78, 92–93, 107, 109, 131, 239, 254, 307, 315, 317, 319–321, 323
Confidence, 22, 28, 62, 64, 90, 91, 179, 214, 223, 224, 232, 240, 280, 307, 323, 328, 336, 341, 351, 356, 361
Constructivist, 311, 322
Content knowledge, 7, 15, 24, 37, 48, 61, 71, 73, 80, 198, 200, 232, 234, 235, 245, 279, 307, 335, 336, 341, 342. *See also* Subject matter knowledge
Context-based, 5, 18, 20, 21, 47, 49, 70, 75–79, 8992–96, 104, 107
Contextualised assessment. *See* Assessment
Cooper, R., 6, 7, 153–168, 219
Correlation, 28, 61, 63, 75, 81, 302
Corrigan, D., 1, 6, 7, 153, 155, 219, 359, 361, 366
Cowie, B., 7, 34, 36, 42, 45, 48–50, 103, 249–263, 270, 271, 280, 308, 314, 347
Creativity, 6, 7, 49, 64, 111, 155, 164, 169–180, 188, 216, 250
Cultural-historical, 232, 235–239
Culturally-appropriate, 261, 267, 271, 272, 274, 281–283
Culture, 6, 8, 28, 29, 123, 180, 201, 213, 220, 223, 257, 268, 269, 271, 272, 282, 282, 286, 287, 289–295, 299–303
Curriculum, 1, 12, 33, 57, 69–98, 101–116, 119, 154, 169, 183, 208, 234, 250, 267, 299, 309, 348, 361
Curriculum development, 2, 4, 5, 58, 60–65, 102, 103, 105, 348, 360, 361
Curriculum integration, 86. *See also* Integrated curriculum
Curriculum learning, 14

D
Denmark, 25, 26
Descriptive, 61, 173, 326, 328
Design, 5, 9, 18, 19, 22, 46, 47, 49, 60, 61, 63, 65–67, 90, 91, 131, 133–137, 140–142, 148, 149, 164, 184, 185, 189–193, 196, 210, 225, 226, 233, 252, 289, 290, 331–333, 354, 356
Diagnostic assessment. *See* Assessment
Didaktik, 25
Differentiation, 325, 347
Disciplinary, 28, 78, 79, 86, 128, 134, 137, 140, 189, 191, 192, 194–196, 198, 200
Diversity, 7, 8, 74, 86, 97, 138, 199, 213, 218, 249–263

E
Economic, 12, 24, 25, 72, 83, 129, 272
Elementary, 66, 179, 297. *See also* Primary
Empirical, 59, 61, 63, 81, 104, 107, 121, 128, 134, 135, 154, 155, 177, 180, 238, 253
Engagement, 8, 35, 37–39, 45, 46, 74, 78, 81, 85, 210, 217, 223–225, 242, 245, 249, 256, 260, 285–304, 328, 338, 339, 343
England, 9, 64, 65, 78, 80, 175, 176, 208, 217, 219, 222, 225, 226, 348, 349
English National Curriculum, 57–59
Evaluate, 47, 61, 97, 132, 161, 190, 192, 200, 214, 274, 275, 308, 309, 319–322, 326, 350
Evaluation, 88, 96, 105–107, 111, 114, 115, 121, 154, 161, 173, 185, 186, 200, 212, 226, 275–280, 303, 319, 352
Everyday views, 308
Evidence-based practice, 58
Examinations, 5, 46, 56, 72, 102, 153, 183–202, 217, 267, 298, 317, 326, 354, 362. *See also* Exit examinations
Exemplar, 47, 50, 67, 173
Exemplary, 16, 18, 21, 70, 85, 114, 130
Exit examinations, 46, 103, 104, 106, 113–116. *See also* Examinations

F
Feedback, 2, 8, 34, 41, 48, 55, 81–83, 165, 201, 209, 211–215, 221, 233, 253–255, 258, 260–262, 270–276, 278–282, 285, 308, 317, 322, 323, 326–331, 334, 335, 337–343, 347, 362
Feed-forward, 270–276, 278–282

Fensham, P., 5–7, 11–29, 34, 35, 37, 38, 40, 41, 69–98, 103, 105, 183, 219, 221, 226, 249, 302, 303, 322, 360–362
Finland, 29, 38, 39, 169, 287
Fit-for-purpose, 64
Fitzgerald, A., 8, 76, 225, 307–323
Fleer, M., 7, 221, 231–245, 252, 257
Folmer, E., 6, 91, 101–116
Formative assessment, 2, 7–9, 43, 50, 55, 70, 72, 75, 76, 91, 154, 165–168, 185, 193, 199, 207–227, 232, 233, 238, 253–255, 267–283, 285, 314, 315, 317, 320, 323, 325–344, 347, 348, 351, 353–355, 362. *See also* Assessment for learning
Formative interactions, 213, 269, 280, 326–344
Funds of knowledge, 256–262

G

GCSE. *See* General certificate of secondary education (GCSE)
Gender, 16, 22, 26–28, 208
General certificate of secondary education (GCSE), 61, 63, 64, 217, 223, 225
Germany, 21, 25, 26, 104, 120
Gunstone, R., 1–9, 76, 155, 225, 307–323, 359–364

H

Harrison, C., 9, 102, 212–214, 227, 304, 347–356
Harvard, 57
High school, 61, 66, 125, 126, 128, 144, 185, 186, 282, 296, 297. *See also* Secondary
High stakes, 2–4, 6, 7, 33, 46, 55, 57, 73, 75–76, 78, 90, 103, 104, 116, 183, 198–200, 202, 207, 209, 211, 215–222, 224, 227, 267, 280, 285, 347, 348, 360
Hong Kong, 26, 27, 38, 39, 287–294, 302, 303

I

ICT. *See* Information and communication technology (ICT)
IEA. *See* International Association for the Evaluation of Educational Achievement (IEA)
Indicators, 18, 49, 57, 184, 299–302, 359, 360
Indigenous knowledge, 258
Information and communication technology (ICT), 37, 303
Infrastructure, 220, 300, 301, 354
Inquiry, 15, 35, 36, 38, 69, 70, 72–74, 103, 127, 129–133, 141, 143, 146, 147, 149, 157, 175, 186, 189–191, 195, 197, 198, 201, 235, 296, 309, 323
In-service teacher education. *See* Teacher education
Instruction, 41, 65, 71, 89, 129, 134, 193, 200, 208–211, 226, 227, 238, 249, 273, 275, 292, 294, 329, 338
Integrated curriculum, 86, 87. *See also* Curriculum integration
Interactions, 3, 19, 33, 64, 71, 111, 120, 164, 195, 209, 231, 249, 269, 285, 310, 326. *See also* Formative interactions
Interest in science, 15–17, 27, 28, 38, 70, 188, 261, 298, 302
International Association for the Evaluation of Educational Achievement (IEA), 11, 12, 14, 24, 29, 34
International studies. *See* Programme for International Student Assessment (PISA); Trends in International Mathematics and Science Study (TIMSS)
Intrinsic, 20, 78, 161, 200, 299
Investigations, 9, 43, 47, 63, 70, 73–76, 78, 94, 131, 136, 148, 153, 154, 158, 164, 165, 217, 224, 225, 235, 242, 259, 262, 263, 289, 290, 317–320, 359. *See also* Practical work
Israel, 9, 25, 349

J

Japan, 169, 287–294, 300, 302, 303
Jones, A., 1–9, 12, 33–51, 226, 359–364
Judgement (judgment), 2, 7, 47, 56, 57, 85, 90, 96–97, 116, 172, 173, 177, 179, 180, 207, 214–216, 219, 220, 222, 224, 227, 232, 241, 243, 245, 278, 301, 327, 342, 354, 361–363

K

Key competencies, 4, 34, 41, 44, 45, 50
King's College London, 178, 349
Knowledge, 1, 13, 34, 59, 69, 107, 119, 155, 170, 183, 209, 232, 249, 267, 285, 307, 325, 349, 362
Knowledge of worth, 3–5, 69, 183, 364
Knowledge society, 265
Korea, 8, 29, 36, 39, 169, 285–299, 302, 303
Kuiper, W., 6, 91, 101–116, 226

L

Learning, 1, 11–29, 34, 55, 69, 102, 121, 153, 169, 184, 207, 231–245, 249, 267, 285–304, 307, 325–344, 347, 360
Learning outcomes, 4, 5, 25, 45, 48, 49, 56, 58–60, 65, 66, 71, 76, 86, 91, 170, 175, 250, 267, 283, 303, 331, 342, 363
Learning progression, 134–135, 141. *See also* Progress in learning
Lee Hang, D., 8, 103, 213, 267–283
Literacy, 141, 300, 302, 309, 316
Logic, 163, 360
Longitudinal, 105–107, 178, 233
Loughran, J., 1, 8, 103, 213, 323, 325–344, 361

M

Ma, H., 7, 183–202, 211, 226
Mathematics, 4, 11, 33, 69, 104, 119, 156, 169–180, 189, 212, 233, 286, 333
Measurement, 11, 16, 63, 64, 86, 105, 119, 123, 128–130, 134, 138, 143, 144, 159, 235
Mentor, 275, 362
Millar, R., 5, 16, 55–67, 91, 102, 104, 116, 210, 226, 360
Moderation, 90, 91, 154, 217, 219, 223–227, 293
Monash University, 156
Māori, 39, 42, 251, 256, 259–261, 268, 272
Motivation, 4, 42, 46, 47, 192, 197, 198, 215, 224, 250, 289, 293, 294, 303, 327, 340, 349, 362
Multi choice (multiple choice), 17, 20, 60, 66, 107, 192, 218

N

National Assessment of Educational Progress (NAEP), 6, 119–149
National Research Council (NRC), 61, 80, 120, 134, 135, 137, 141–144
National science standards, 120
National testing, 25, 26, 218, 360
Nature of science, 3, 4, 21, 23, 35, 37, 39, 56, 59, 70, 71, 74, 79, 80, 85, 125, 128–131, 133, 134, 140, 143, 146, 147, 148, 155, 258, 361
The Netherlands, 6, 21, 77, 80, 103–105, 114
New Zealand, 5, 12, 33–41, 43–47, 49, 50, 226, 250–252, 256, 258–260, 267, 268, 271, 272, 277, 280, 282
Nilsson, P., 8, 103, 213, 323, 325–344
No Child Left Behind (NCLB), 120
Norway, 26
NRC. *See* National Research Council (NRC)
Numeracy, 4, 34, 41, 44, 45, 47, 49, 50, 178

O

Observation, 45, 48, 61–64, 81, 85, 101, 107, 108, 114, 128, 132, 154, 156, 161, 173, 187, 190, 202, 221, 231, 233–236, 243, 254, 259, 260, 289, 290, 310, 314–316, 318, 330
Organisation for Economic Co-operation and Development (OECD), 11, 12, 14–15, 18, 21, 24, 29, 34, 37–40, 45, 77, 82, 287, 289, 292–294
Ottevanger, W., 6, 91, 101–116

P

Pacific, 8, 267–269
Paper and pencil. *See* Tests
Pasifika, 39, 42, 268, 272
PD. *See* Professional development (PD)
Pedagogical content knowledge (PCK), 3, 4, 7–9, 37, 48, 49, 263, 317, 337, 355, 363
Pedagogical knowledge, 8, 9, 363
Pedagogy, 1–5, 7–9, 34, 35, 43, 60, 71, 180, 184, 188, 207–227, 231–245, 250, 251, 256, 259, 262, 263, 283, 327, 331, 342, 347, 351–353, 356, 359–364
Peer assessment. *See* Assessment
Performance, 11, 12, 18, 22–24, 26, 29, 36, 39, 47, 49, 55, 57, 75, 76, 82, 85, 90–93, 98, 106, 107, 112–114, 119, 120, 123, 124, 128, 132, 133, 135, 136, 140, 143, 144, 148, 154, 171, 177–179, 186, 190, 199, 216, 220, 238, 244, 245, 287, 294, 302, 317, 326, 327, 361–363. *See also* Achievement
Performance tasks, 91, 142, 170, 178–179
Physical sciences, 28, 35, 131, 139, 140, 146, 148
Physics, 11, 19, 20, 35, 36, 46, 57, 66, 77–80, 86, 104–105, 107, 109–110, 112–114, 127, 140, 141, 157, 165, 185–188, 193, 200, 201, 330, 331
PISA. *See* Programme for International Student Assessment (PISA)
Policy, 1, 3–6, 9, 11, 13, 28, 29, 33, 34, 41, 44–46, 49, 50, 67, 93, 103, 103, 106, 114, 119–123, 134, 141, 142, 144, 145, 153–154, 157, 200, 202, 208, 217, 267, 286, 304, 325, 327, 347, 354, 359–361, 364

Portfolio, 74, 75, 82, 185, 190, 218, 219, 223, 225, 234, 350, 351, 354, 361, 362
poststructuralist, 234
Potentive assessment. *See* Assessment
Practical work, 59, 73, 303. *See also* Investigations
Practice. *See* Teacher practice
Pre-service teacher education. *See* Teacher education
Primary, 5, 7, 8, 23, 24, 33–36, 41, 44–48, 50, 73, 78–81, 120, 128, 145, 178, 179, 185, 212, 216, 232, 233, 299, 307–323, 329, 331, 332, 339, 360. *See also* Elementary
Principles, 6, 21, 38, 40, 47, 48, 59, 70–71, 74, 77, 79, 82, 84, 91, 102, 106, 111, 112, 124, 125, 127, 130–133, 135–138, 140, 143, 144, 145, 148, 154, 167, 186, 190, 207, 209, 211–217, 222, 224, 227, 235, 244, 245, 259, 295, 354, 363
Professional development (PD), 8, 9, 23, 35–38, 41, 45, 47, 50, 90, 154, 177, 219, 221, 227, 267, 274, 280, 349–356, 363
Professional learning, 35, 111, 341, 342, 344, 352, 355
Programme for International Student Assessment (PISA), 4, 5, 8, 11–29, 33, 34, 36–43, 45, 49, 50, 58, 77, 82, 119, 226, 286–289, 292–294, 299, 302–304, 360
Progress in learning, 2, 134–135, 141. *See also* Learning progression
Project work, 18
Public understanding of science, 37, 299, 300
Purpose, 2–4, 7, 12, 25, 34, 47, 49, 50, 55–56, 64, 71, 73–76, 79, 86, 88, 90, 91, 96, 97, 101, 103, 121, 128, 132, 154, 158, 179, 184–186, 189, 192, 195, 197–199, 207, 209, 214, 215, 220–224, 226, 250, 254, 261–262, 268, 275, 280, 283, 291, 295–296, 308, 319, 328–330, 339–341, 348, 352, 354, 360, 363

Q

Quality assurance, 55, 90, 116
Quality science teaching, 28, 157
Queensland, 6, 78, 91, 92, 94, 153–154, 157, 160, 161, 165, 167, 219
Questionnaires, 14, 16, 17, 26, 106, 187, 280
Quiñones, G., 7, 221, 231–245, 257

R

Reform, 4, 7, 24, 58, 73, 77, 102–106, 109, 115, 177, 183–202, 285, 304
Relevance of Science Education (ROSE), 16, 17, 77
Reliability, 11, 14, 18, 65, 81, 97, 115, 116, 163, 215–218, 223, 233
Rennie, L., 5–7, 18, 21, 23, 29, 40, 69–98, 103, 105, 219, 361, 362
Research, 1, 13, 34, 55, 70, 106, 120, 153, 169, 190, 213, 232, 250, 267, 293, 309, 327, 347, 361
Role play, 87–89, 315–317, 321, 363
ROSE. *See* Relevance of Science Education (ROSE)

S

Samoa, 8, 267–283, 272, 274–277, 280, 281
Sampling, 26, 27, 63, 215
School, 1, 12, 33–51, 55, 69, 103, 120, 153, 170, 183, 207, 231, 249, 267, 285, 307–323, 327, 347, 362
Science as process, 158, 163
Science beyond the classroom, 165
Science curriculum. *See* Curriculum
Science for all, 61, 77, 250, 251
Science knowledge, 6, 13–17, 20, 21, 27, 74, 76, 77, 82, 119, 129, 142, 144, 239, 249, 258, 259, 281, 343, 363, 366. *See also* Scientific knowledge
Science literacy, 123–145
Science teacher, 6, 8, 9, 18, 20, 21, 25, 57, 64, 91, 128, 153–168, 183–202, 212, 225, 251, 262, 269, 270, 273, 274, 281, 282, 296, 323, 329, 344, 348–350, 356, 363
Science teaching, 1, 9, 14, 15, 20, 21, 24, 27, 28, 70, 77, 80, 81, 90, 157, 161, 169–180, 250, 289, 307, 308, 323, 326–336, 338, 341, 343, 348–351, 353–355
Science, technology and society (STS), 132, 188, 189, 194, 196
Science, technology, engineering and mathematics (STEM), 124
Scientific inquiry. *See* Inquiry
Scientific knowledge, 15, 35, 37, 38, 61, 85, 124, 234, 295, 297, 298, 303. *See also* Science knowledge
Scientific literacy, 13, 15, 16, 21, 22, 24, 28, 35–37, 39, 41, 50, 61, 62, 77, 128, 129, 146, 281, 299, 300, 363

Secondary, 4–6, 8, 11, 18, 23, 33, 34, 36, 37, 41, 44, 46–48, 50, 61, 75, 79–82, 93, 94, 103–106, 113, 114, 153, 154, 157, 158, 160, 161, 167, 173, 183–187, 202, 217, 222, 233, 242, 267–269, 276, 281, 294, 298, 325. *See also* High school
Singapore, 169, 287
Skills, 7, 24, 35, 70, 103, 119, 169, 185, 208, 249, 281, 303, 319, 328, 347, 362
Social justice, 169, 250, 283
Sociocultural, 3, 6, 172, 298, 342, 355
Socioeconomic status (SES), 23, 29
Socio-scientific issues, 5, 18, 22
Song, J., 8, 213, 285–304
State, 3, 13, 65, 69, 119, 153, 169, 184, 207, 235, 253, 267, 344
Statistical analysis, 18, 114, 302
Student teacher, 180, 275, 278–280, 326–335, 339–346. *See also* Teacher education
Subject matter knowledge, 143, 336. *See also* Content knowledge
Summative assessment. *See* Assessment
Sweden, 8, 176, 219, 329
System-level, 3, 7, 208
Systems, 3, 14, 35, 55, 69, 101, 120, 154, 170, 183, 207–227, 241, 280, 285, 326, 353, 360

T

Taiwan, 25, 287, 289, 302, 303
Teacher education
  in-service, 275
  pre-service, 8, 174, 178, 179, 269, 275, 276, 279–280, 331 (*see also* Student teacher)
Teacher practice, 354
Teaching, 1, 14, 33, 55, 70, 103, 122, 154, 169–180, 184, 207, 232, 249, 268, 289, 307, 326, 347, 363
Technology, 12–16, 27, 28, 35, 38, 63, 77–80, 82, 87, 90, 123, 124, 127, 132, 141, 156, 185, 188, 189, 194, 196, 299–301, 363
Testing, 11–20, 22, 24–28, 41–43, 46, 47, 56, 69, 73, 75, 82, 85, 86, 89, 121, 123, 132, 172–174, 176, 177, 179, 200, 210, 218–220, 226, 239, 325, 328, 347, 348, 360
Tests
  paper and pencil, 17–19, 21, 24, 73–75, 116, 186, 190, 198, 200
  written, 11, 85–86, 103, 217
TIMSS. *See* Trends in International Mathematics and Science Study (TIMSS)
Transform, 142, 285, 337, 349
Transformative education, 170
Trends in International Mathematics and Science Study (TIMSS), 4, 5, 8, 11, 12, 14–20, 23–29, 33–38, 41–43, 49, 77, 119, 128, 226, 286–291, 299, 302–304, 360
Twenty-First Century, 20, 28, 37
Twenty-First Century Science, 5, 58, 60, 61, 63, 64, 78

U

United States, 36, 286
University of Waikato, 250, 256
US National Science Standards, 58

V

Validity, 17, 18, 75, 82, 84, 116, 128, 140, 156, 166, 215–218, 220, 222, 223, 225–227, 233, 271, 361
Values, 3, 22, 40, 56, 73, 113, 122, 154, 170, 184, 208, 250, 269, 294, 334, 362
Variables, 16, 23, 42, 63, 130, 133, 163, 191, 318, 326
Victoria, 6, 153–155, 157, 158, 160, 166, 167, 219
Vygotsky, L., 172, 212, 231, 238, 240, 242, 342

W

Wait time, 349, 350, 353
Wales, 18, 78, 80, 217, 219, 220, 348
Written tests. *See* Tests