SCIENCE EDUCATORS' ESSAY COLLECTION

Everyday Assessment in the Science Classroom

Edited by J Myron Atkin and Janet E. Coffey

NSTApress®
NATIONAL SCIENCE TEACHERS ASSOCIATION
Arlington, Virginia

NSTA press
NATIONAL SCIENCE TEACHERS ASSOCIATION

Claire Reinburg, Director
J. Andrew Cocke, Associate Editor
Judy Cusick, Associate Editor
Betty Smith, Associate Editor

Art and Design Linda Olliver, Director
NSTA Web Tim Weber, Webmaster
Periodicals Publishing Shelley Carey, Director
Printing and Production Catherine Lorrain-Hale, Director
Nguyet Tran, Assistant Production Manager
Jack Parker, Desktop Publishing Specialist
Publications Operations Erin Miller, Manager
***sci*LINKS** Tyson Brown, Manager
David Anderson, Web and Development Coordinator

National Science Teachers Association
Gerald F. Wheeler, Executive Director
David Beacom, Publisher

Science Educators' Essay Collection
Everyday Assessment in the Science Classroom
NSTA Stock Number: PB172X

05 04 03 4 3 2

Library of Congress Cataloging-in-Publication Data
Everyday assessment in the science classroom / J. Myron Atkin and Janet E. Coffey, editors.
p. cm.— (Science educators' essay collection)
Includes bibliographical references and index.
ISBN 0-87355-217-2
1. Science—Study and teaching (Elementary) —Evaluation. 2. Science—Study and teaching (Secondary)—Evaluation. 3. Science—Ability testing. I. Atkin, J. Myron. II. Coffey, Janet E. III. National Science Teachers Association. IV. Series.
LB1585.E97 2003
507'.1—dc21 2003000907

Contents

Acknowledgments

We wish to acknowledge our debt to Rodger Bybee and Harold Pratt. A National Science Teachers Association (NSTA) "yearbook" was Rodger's vision, and he served as the editor of the first volume, *Learning Science and the Science of Learning*. Rodger and Harold helped to identify classroom assessment as a focus for a subsequent volume. Their discussions were the impetus for this collection. Many people within the NSTA organization supported the launch of the annual series and this volume—particularly Gerry Wheeler, executive director of NSTA, and David Beacom, NSTA publisher. We also thank Carolyn Randolph, current NSTA president, for her support.

The authors of the separate chapters have been extraordinarily responsive to suggestions from us and from the reviewers and have made their revisions in a timely manner. Their dedication to improving science education and their desire to engage in a dialogue with and for teachers are inspiring.

We also thank the following individuals for their reviews of this volume:

Ann Fieldhouse, Edmund Burke School, Washington, D.C.
Patricia Rourke, Potomac School, McLean, Virginia
Rick Stiggins, Assessment Training Institute, Portland, Oregon

The reviewers provided thorough and thoughtful feedback. The volume is better for their efforts.

Special acknowledgments are due for Claire Reinburg and Judy Cusick at NSTA. Claire initiated early communication and helped with logistics from the outset. Judy Cusick has provided support in innumerable ways, not least as careful production coordinator for the entire volume. Claire and Judy have guided us smoothly through the process, coordinated reviews, and served as invaluable resources. We thank them both for their time and effort.

Although the reviewers provided constructive feedback and made many suggestions, they did not see the final draft before release. Responsibility for the final text rests entirely with the chapter authors and the editors.

About the Editors

J Myron (Mike) Atkin is a professor of education and human biology at Stanford University. He directs the National Science Foundation–supported Classroom Assessment Project to Improve Teaching and Learning (CAPITAL) at Stanford. He formerly chaired the National Research Council committee that prepared an addendum to the National Science Education Standards on assessment in the science classroom. In addition to having taught at the elementary and secondary school levels, he has served as dean of education at both the University of Illinois and at Stanford. For much of his career, he has emphasized the central role of teachers in designing high-quality science education programs and now focuses on strengthening that role in any comprehensive system of assessment and accountability.

Janet E. Coffey, a former middle school science teacher, currently works with teachers on the Classroom Assessment Project to Improve Teaching and Learning (CAPITAL) at Stanford University. CAPITAL is a National Science Foundation–funded effort that seeks to better understand teachers' assessment practices as they strive to improve their practices. She worked at the National Research Council as a staff member on the development of the National Science Education Standards. She earned her Ph.D. in science education from Stanford University.

Introduction

J Myron Atkin and Janet E. Coffey

The assessment that occurs each day in the science classroom is often overlooked amidst calls for accountability in education and renewed debates about external testing. We believe that daily assessment should be moved to the foreground in these broad discussions and receive greater attention in policy circles. Research points to the positive influence that improved, ongoing classroom assessment can have on learning. Documents that offer visions for science education—such as the *National Science Education Standards* (NRC 1996) and materials associated with Project 2061 (AAAS 1993)—strongly echo that sentiment.

Too often, assessment is used synonymously with measurement (in the form, for example, of tests and quizzes). The misconception that they are the same minimizes the complexity and range of purposes of assessment. As teachers are aware, assessment in any classroom is rich and complicated. It includes tests and quizzes, of course, but also students' other work, students' utterances made while conducting lab investigations, and class discussions in which students share their explanations for what they have observed. It raises issues of quality and of what counts as evidence for learning. It happens in reflection, in exchanges that occur countless times a day among teachers and students, and in feedback on work and performance. When we reduce assessment to a specific event or "thing"—the test, the lab practical, the grade—it is easy to overlook the interactive assessment that occurs each day in the classroom.

Assessment operates to improve student learning, not solely measure it, when it is used to move the student from his or her current understanding to where the student would like to be (or where the teacher would like the student to be). To cross that gap, the teacher and student must *use* feedback from assessments. Quality and use of information become crucial in that process. Sometimes the way to bridge that gap is clear, with obvious starting and destination points. Sometimes, however, it is rendered more complex when the destination point is not so clear, as in inquiry-based science investigations. In inquiry learning, students ask their own questions and face multiple paths to answering those questions. Students must continually reflect on what they are doing and ask themselves, Where am I now? Where do I need to go? How do I get there? What is my evidence?

Everyday assessment is local and contextual. It depends more on the skills, knowledge, attentiveness, and priorities of teachers and students than on any particular set

of protocols or strategies. Opportunities for meaningful assessment occur numerous times each day—within interactions, in conversations, by way of observations, and even as part of traditional assessment. A test review or a discussion about criteria of good lab reports each provides opportunities for assessment-related conversations. When the K–12 teacher attends closely to questions and responses and observes students as they engage in inquiry, he or she will gather important assessment information on a daily basis.

The challenge for teachers and students is to maintain a classroom assessment culture that supports learning for *all* students. In this culture, teaching and assessment are so closely entwined that the two become difficult to differentiate; students engage actively and productively in assessment; clear, meaningful criteria exist; and teachers provide high-quality, regular feedback. In classrooms with supportive assessment cultures, the focus is on learning rather than on grades, on progress rather than on fixed achievement, on next steps rather than on past accomplishments. Achievement and accomplishments serve an assessment purpose, as can grades; however, they alone do not assessment make.

In any discussion of assessment, issues associated with teaching, learning, and curriculum quickly arise, as do questions of equity, fairness, and what counts as knowing and understanding. These interconnections, after all, constitute one reason that the topic of assessment is so important. In this volume, the authors address the interconnections and provide guidance and illuminate challenges—at the same time that they maintain a sense of the bigger picture. Although our primary audience is classroom teachers, we also hope that this book will be informative and useful to those who work with teachers, in either professional development or teacher education capacities, and to school administrators, program designers, policy makers, and parents.

In Chapter 1, Paul Black frames the assessment that occurs every day in the classroom within the broader realm of assessment and its multiplicity of purposes. For everyday assessment to contribute to improved learning, he argues, action must be taken based on evidence. This action could come in any number of forms, not the least of which include altering teaching, modifying curriculum, or providing students with useful feedback. Even the most informative data do little good to students and teachers if the data do not feed into learning activities. Black provides an overview of some theoretical assumptions embedded in a view of assessment that supports learning, and he highlights issues related to assessment in science classrooms. He also addresses the tensions and synergistic opportunities that exist when trying to manage the many purposes assessment must serve.

In Chapter 2, Anne Davies discusses the critical role teachers play in ongoing assessment. Teachers not only identify and articulate learning goals, they also find samples of work to meet those goals, consider the type of work that would serve as evidence of attainment, and assist students in developing an understanding of the goals. Her discussion highlights the relationships among assessment, curriculum,

and teaching. As we see in many of these chapters, the three topics are, at times, difficult to distinguish from one another.

Cary Sneider, author of Chapter 3, examines what can be gained through careful consideration of student work, which includes what students do, say, and produce. Through reflecting on his own career as a teacher, which began as an Upward Bound tutor, his work as a curriculum developer, and now his present position in an education department at a science museum, Sneider makes a case for tuning into what students do and say to gain insights into the nature of their ideas and understandings. He reminds us that to teach is not necessarily to learn. Assessment serves as a critical link between the two processes. In this essay, Sneider encourages teachers to move beyond the use of written responses and final products and to see the value of listening to, conversing with, and observing students as they engage in activities as well.

The focus of Chapter 4 is on assessing inquiry science. Richard Duschl offers a metaphor of "listening to inquiry" to guide teachers as they support inquiry in their classrooms. The organic nature of inquiry excludes more traditional tests as a means of helping students move forward in pursuit of investigation. With inquiry, there are few clear destinations or explicit goals toward which to steer. Duschl looks at the design of learning environments that engage students in thinking about the structure and communication of scientific ways of thinking. Deliberations about what counts as data, evidence, and explanation would give inquiry a voice, to which student and teacher alike can listen.

Chapter 5 explores issues related to questioning in the classroom. Specifically, Jim Minstrell and Emily van Zee highlight the importance of developing and asking questions that help students consider scientific phenomena and elicit thoughtful and in-depth responses that reveal insight into student understandings. The authors discuss ways in which these insights can be used by teachers to plan additional activities, modify existing ones, and inform future questions. A message that emerges in this chapter is the important role that subject matter plays when teachers are listening to student responses and following those responses with further questions.

Much of the classroom assessment literature focuses on the roles of teachers. Chapter 6 shifts the focus to students. Janet Coffey addresses the integral role students can play in everyday classroom assessment. Student participation in and with assessment activities can help clarify and establish standards for quality work and help students to identify the bigger picture of what they are learning. Through lessons learned from a middle school program where students had opportunities and expectations to participate actively in assessment, Coffey identifies ways to support students as they become self-directed with respect to assessment.

Mark Wilson and Kathleen Scalise take on issues related to grading in Chapter 7. Grades quickly can become the centerpiece of any discussion of assessment at the classroom level. Wilson and Scalise discuss the meanings that underlie grades, the information they convey, and what they often represent. Grades, they argue, often reflect a teacher's perception of a student's effort rather than what the student has

learned. Any time a large amount of information is reduced to a single letter grade, much of the useful information behind the grade gets lost. The authors offer a framework for assessment tools that can generate useful assessment data for classroom purposes and for reporting purposes. Assessment tools such as the ones they share can yield useful and high-quality assessment data for teachers, students, parents, and other interested parties. An element of their model includes teacher "moderation meetings," where teachers discuss student work, scores, and interpretations. These deliberations can provide powerful professional development opportunities.

In Chapter 8, Mistilina Sato discusses assessment-related professional development for teachers. She advocates for a change in the teacher's image from that of monitor and maker of judgments to that of manager of learning opportunities. For lasting change, she argues, teacher professional development must go beyond learning new strategies and skills to take into consideration the teacher as a person. Sato points out that teachers enter the classroom as individuals with beliefs, values, backgrounds, assumptions, and past experiences that shape who they are in the classroom and the actions they take. Reform efforts that overlook these personal dimensions of teachers will find minimal success. She shares lessons from a National Science Foundation–funded assessment effort currently underway with Bay Area middle school science teachers.

Lorrie Shepard explores the assessment landscape beyond the classroom in Chapter 9. Specifically, she discusses the ways in which large-scale assessment could be redesigned to heighten contributions to student learning. Even within the realm of large-scale, external testing, a myriad of purposes clamor for attention. Due to constraints such as time and cost, these assessments often take the form of traditional tests. Shepard points out that all tests are not the same. The intended purposes of a test shape the content, criteria for evaluating, and technical requirements; all needs cannot be met through a single test. Shepard calls for external tests to embody important learning goals, such as those set forth in the National Research Council's *National Science Education Standards* (1996) or AAAS's *Benchmarks for Science Literacy* (1993). Examples of actual efforts underway in districts and states help show possibilities for lessening tensions.

James Rutherford provides a historical overview of educational assessment and reform in science education in Chapter 10. The shifting focus of school, district, and national reform efforts has made sustained attention to any one initiative difficult and frustrating at best. Assessment is no exception. We are quick to react to "crises"—real or imagined—rather than take proactive steps, with a long-term view and guidance from solid research literature, toward higher quality science instruction. Rutherford proposes that teachers, parents, and others within the educational community assess assessments by asking critical questions about the assessments as well as the information they provide. He concludes his chapter by offering questions for all of us to consider. In doing so, he sets a frame for use of this book as a tool for professional development. Generating discussion among teachers about some of the

ideas raised and addressed in any or all of these chapters would be a valuable outcome of the book.

As Rutherford indicates, this collection may raise more questions than it answers. We, too, hope that this volume contributes to the practical and professional development needs of teachers. We hope it illuminates the importance of attending to everyday assessment, raises issues worthy of reflection and consideration, and offers some practical suggestions. Thinking and acting more deliberately with regard to the ongoing assessment that occurs each and every day in the classroom can go a long way to making our classrooms more conducive to learning for all students.

References

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

National Research Council (NRC). 1996. *National science education standards*. Washington, D.C.: National Academy Press.

The Importance of Everyday Assessment

Paul Black

Paul Black is Emeritus Professor of Science Education at King's College London. He retired in 1995, having spent much of his career in curriculum development and research in science education. In 1987–1988 he was chair of the government's Task Group on Assessment and Testing, which set out the basis for the United Kingdom's national testing. More recently, he has served on advisory groups of the National Science Foundation and is a visiting professor at Stanford University. His recent research with colleagues at King's has focused on teachers' classroom assessments. This work has had a significant impact on school and national policies in the UK.

The Context: Conflicts and Synergies of Purpose

If assessment is understood in a broad sense—that is, to signify all those processes and products that provide evidence about what is happening—it is immediately evident that in education, assessments are all-pervasive. They influence, even rule, the context within which teachers work, but are also an essential part of the everyday minutiae of teachers' work with their students. This broad interpretation leads to a need to impose some structure on any discussion, so I start here with a discussion of purposes of assessment.

Three main purposes can be distinguished. Assessment can serve *accountability*, *certification*, or *learning*.

- For *accountability*, the evidence has to be broad in scope and designed to highlight needs that might be met through policy actions. This purpose can be served by testing samples of the student population, as well as by collecting a range of other data so that interpretation might be served by exploring interrelationships. Such a broad program might be called an evaluation, to distinguish it from assessment, which is seen as only one of its components. Such surveys as the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study (TIMSS), and the Program for International Student Assessment (PISA) are examples. However, an *accountability* exercise can be designed for a different purpose—to drive improvement by linking the results to public exposure and other rewards or punishments. It is usual, albeit unnecessary, to test every student.
- For *certification*, the audience is the individual students and those who care about them, together with potential employers and those controlling admission to the further stages of education. For both *accountability* and *certification,* the evidence is often limited to results of formal written tests drawn up and marked by agencies external to the school. However, it is possible, and indeed normal in some countries, to use evidence provided from within schools to help meet these purposes. The collection of evidence need not then be limited to formal, timed tests; the broader term assessment is then appropriate.

- For *learning,* the dominant term can be assessment, but evaluation and diagnosis sometimes creep in. The purpose is clear enough in principle, while action based on the evidence can range from minute-by-minute feedback to adjustment of the lesson plans for next year.

Over all of this spectrum, the concepts of *reliability* and *validity* are central. To claim reliability, one has to be sure that if the student were to take a parallel form of the same assessment on another occasion, the result would be the same. This claim is rarely supported by comprehensive evidence. For certification, and internal school decisions about tracking and grading, weak reliability is serious because the effects on students are hard to reverse. In assessment for learning, reliability is harder to achieve, but it is less serious an issue provided that the teacher's approach is flexible so that the effects of wrong decisions can be discerned and corrected within a short time.

Validity is a more serious issue. The key to the concept is that the inferences that are made on the basis of an assessment result are defensible (Messick 1989). In the case of a formal written test, any inference that goes beyond saying that the student did well on this test on this date requires justification. But inferences do go far beyond such limitation, for it is often assumed that the student could do well in any test in any part of the subject on any occasion in the future, understands the nature and structure of the subject, is competent in exercising the discipline of the subject, and is well-equipped to benefit from more advanced study.

One source of the limitation on validity is that formal tests have to be short and inexpensive to administer and mark. However, many aspects of understanding and competence can only be displayed over extended periods of time in unconstrained conditions. It is often claimed that a written test might serve as a surrogate for such activity, but those claims usually lack empirical support (Black 1990; Baxter and Shavelson 1994). For this reason, significant efforts have been invested in developing a broader range of methods of assessment—for example, new types of evidence for performance (e.g., notebooks written by students about a science investigation) and new procedures to improve and attest to the evidence that teachers can gather from their more extensive interactions with their students. However, what then needs to be made clear is whether these innovations are designed to serve *accountability* and *certification*, or to serve *learning*, or to serve all three purposes.

Demands of accountability are often seen to impose such high-stakes pressures on teachers that assessment for learning is not seriously considered. When the public believes that wholly external written tests are the only trustworthy evidence for students' achievement, and then teachers believe that the pressures to succeed can only be met by ad hoc and inadequate methods of learning, dominance of the accountability purpose seems inevitable. Yet critical scrutiny of the claims for the superiority of formal tests is not supported by careful scrutiny of their reliability and validity. It has been demonstrated that when any such test is newly introduced, performances will rise steadily for a few years and then level out. If the test is then changed, perfor-

mance will drop sharply and then rise again to level off in a further few years (Linn 2000). Thus, the high-stakes performances that ensue are, at least in part, artifacts of the pressures of the particular test being used rather than valid educational measures. Ironically, the belief of most teachers that they have to "teach to the test" rather than teach for sound understanding is also unjustified, for there is evidence that the latter strategy actually leads to better performances even on the tests to which the former, narrow teaching approach is directly aligned.

The problem for policy makers is to stand back from current assumptions and radically rethink their approach to reconciling the different purposes of assessment. A key feature of any such appraisal must be to strengthen teachers' own skills in assessment so that the public can have confidence in the capacity of teachers to serve all three purposes in a valid and rigorous way. The problem for teachers and schools is to improve practices, to be clear about the purposes that those practices are designed to serve, and to resolve any conflicts as best they can within today's constraints.

How Do People Learn?

Three common assumptions about learning, which have their origins in behaviorist psychology (Collins 2002), are that (1) a complex skill can be taught by breaking it up and teaching and testing the pieces separately; (2) an idea that is common to action in many contexts can be taught most economically by presenting it in abstract isolation so that it can then be deployed in many situations; and (3) it is best to just learn about new things first and not try for understanding—that will come later. A test composed of many short, "atomized," out-of-context questions and the practice of "teaching to the test" are both consistent with those assumptions.

Contemporary understanding of the ways that children learn looks at the process quite differently (Wood 1998). A first important lesson is illustrated by the following quotation:

> *… even comprehension of simple texts requires a process of inferring and thinking about what the text means. Children who are drilled in number facts, algorithms, decoding skills or vocabulary lists without developing a basic conceptual model or seeing the meaning of what they are doing have a very difficult time retaining information (because all the bits are disconnected) and are unable to apply what they have memorized (because it makes no sense).* (Shepard 1992, 303)

Current "constructivist" theories focus attention on the models that a learner employs when responding to new information or to new problems. Even for the restricted task of trying to memorize something, one can do better if one already has some scheme built on relevant understanding and tries to link the new knowledge with existing patterns. It appears that memory is rather like a filing cabinet—that is, the storage is only useful insofar as the filing system makes sense so that one knows where to look.

More generally, learning always involves analyzing and transforming any new information. Piaget stressed that such transformation depends on the mind's capacity to learn from experience—within any one context, we learn by actions, by self-directed problem-solving aimed at trying to control the world. Abstract thought evolves from such concrete action. It follows that

> *… teaching that teaches children only how to manipulate abstract procedures (e.g., learning how to solve equations) without first establishing the deep connections between such procedures and the activities involved in the solution of practical concrete problems (which the procedures serve to represent at a more abstract level) is bound to fail.* (Wood 1998, 9)

Here, context is important. An individual's general capacity for abstract thought may be exhibited in, say, family relationships but be quite absent in, say, physics concepts. It is also evident that transformations of incoming ideas can only be achieved in light of what the learner already knows and understands, so the reception of new knowledge depends on existing knowledge and understanding. It follows that

> *… learning is enhanced when teachers pay attention to the knowledge and beliefs that learners bring to a learning task, use this knowledge as a starting point for instruction, and monitor students' changing conceptions as instruction proceeds.* (Bransford, Brown, and Cocking 1999, 11)

Research in the learning of science has shown that many learners resist changes in their everyday and naive views of how the natural world works, despite being able to play back the "correct" science explanations in formal tests. So teaching must start by exploring existing ideas and encouraging expression and defense of them in argument, for unless learners make their thinking explicit to others, and so to themselves, they cannot become aware of the need for conceptual modification. The next step is to find ways to challenge ideas, usually through examples and experiences that are new to pupils and that expose the limitations of their ideas. It follows that assessment for learning must be directed at the outset to reveal important aspects of understanding and then be developed, within contexts that challenge pupils' ideas, to explore responses to those challenges.

Such classroom activities can be a basis for learning development at a more strategic level. Research studies have shown that those who progress better in learning turn out to have better self-awareness and better strategies for self-regulation than their slower learning peers (see, e.g., Brown and Ferrara 1985). Thus self-assessment becomes an important focus of assessment for learning. Peer-assessment also deserves priority, for it is by engaging in critical discussion of their work with their peers that learners are most likely to come to be objective about the strengths and weaknesses of their work. The main message is that students need to understand

what it means to learn. They need to monitor how they go about planning and revising, to reflect on their learning, and to learn to determine for themselves if they understand. Such skills enhance metacognition, which is the essential strategic competence for learning.

When the teacher starts from where the learners are, helps them to take responsibility for their learning, and develops peer- and self-assessment to promote metacognition, the teacher becomes a supporter rather than a director of learning. This idea was taken further by Vygotsky (1962), who emphasized that because learning proceeds by an interaction between the teacher and the learner, the terms and conventions of the discourse are socially determined, and its effectiveness depends on the extent to which these terms and conventions are shared. His influence can be seen in the following statement:

> *Participation in social practice is a fundamental form of learning. Learning involves becoming attuned to the constraints and resources, the limits and possibilities, that are involved in the practices of the community. Learning is promoted by the social norms that value the search for understanding.*
> (Bransford, Brown, and Cocking 1999, xii)

Wood, Bruner, and Ross (1976) developed this approach by introducing the metaphor of "scaffolding"—the teacher provides the scaffold for the building, but the building itself can only be constructed by the learner. In this supportive role, the teacher has to discern the potential of the learner to advance in understanding, so that new challenges are neither too trivial nor too demanding. Vygotsky called the gap between what learners can do on their own and what they can do with the help of others the "zone of proximal development." One function of assessment is to help to identify this zone accurately and to explore progress within it.

All of this shows how important it is for the teacher to develop a classroom discourse through which all students learn to internalize and use the language and the norms of argument used by scientists to explain phenomena and to solve problems (Bransford, Brown, and Cocking 1999, 171–75). Thus, the way students talk about science, both in informal and formal terms, is important formative assessment material for teachers (Lemke 1990).

Because a learner's response will be sensitive to the language and social context of any communication, it follows that assessments, whether formative or summative, have to be very carefully framed, both in their language and context of presentation, if they are to avoid bias (i.e., unfair effects on those from particular gender, social, ethnic, or linguistic groups). The importance of context is also a far-reaching issue. For example, tests that ask questions about mathematics that might be used in society in general might be failed by a student who can use the same mathematics in a familiar school context, and vice versa (Boaler 2000).

The discussion in this section has been limited to the cognitive aspect of links between assessment and student response. Other important elements will be explored in the next section.

Assessment for Learning

In the broadest sense of the word, assessment is something that we do all the time. We encounter a new situation, make a judgment about the meaning of what is happening, and decide what to do next. The evidence of our encounters continually shapes and reshapes our actions. Our actions may be more effective if we are flexible—that is, if we are prepared to modify our intentions in the light of events. They might also be more effective if we probe the situation carefully in order to ensure that we understand what is going on before jumping to conclusions.

All of this applies in particular to life inside the classroom. The teacher has some understanding of the state of the students' learning, and must decide what to do next. This understanding is bound to be imperfect, but it can be refined by setting up activities through which the students will provide more evidence. So the cycle is to evoke or explore, to interpret the feedback, and then to modify the teaching actions.

The key to formative assessment lies in this flexibility—the capacity to change what was planned in order to meet the needs exposed by the evidence. The prospects are improved by finding ways to so elicit evidence that key features of the learning are illuminated; this can be called *assessment for learning*. However, there is little point in doing this if the evidence is not used to fashion what happens next; only when such refashioning occurs does the assessment become *formative assessment* (Black and Wiliam 1998). It is necessary to stress this feature; some teachers believe that they are engaged in formative assessment when, even though they are listening to their students, they then proceed with a lesson plan despite what they have heard.

The concept of assessment is a very general one—in the classroom it is happening all the time. When the looks on the faces of the students, or their written work, or their oral answers to a question are appraised, the teacher is assessing. A quiz or written test is also an occasion for assessment, but it is only one among many possibilities. As outlined above, the quality of the assessment feedback will depend on the quality of the interventions that evoke that feedback. It is here that the theories of learning become relevant. A question that asks about a technical term—for example, "What is the unit of current?"—serves a very different purpose from one that probes understanding—for example, "Does the current get used up as it goes through the light bulb?" The former question looks for recall, and there is little that can be done with the response apart from noting whether or not it is correct.

The latter question probes rather deeply, for it has a basis in research evidence that the notion of "current used up" is a common misconception. One way for a teacher to respond is to listen to an answer and then to tell the class the right answer; such action, however, is not responsive to the evidence. A second way is to explore opinions among a class to stimulate a discussion about the concept of current, which

could lead to a test with a simple circuit with ammeters on either side of the bulb. This second way is formative, for it explores understanding, involves students actively in the learning process, and follows the learning principle that effective learning starts from where learners are and helps them to see the need to change. Furthermore, insofar as discussion is evoked, the learning is in the context of a discourse in a learning community rather than being a one-way transmission. Thus, there is an intimate connection between good formative assessment and the implementation in the classroom of sound principles of learning.

Similar arguments can be applied to other learning activities, notably the marking of written work, the use of peer- and self-assessment, the possibilities for the formative use of written tests, and so on. As teachers change to make formative assessment a constant feature of their work, they will inevitably be changing their roles as teachers. They have to be more interactive with their students, and they have to give them more responsibility for learning. This leads to a change in role, from directing students to empowering them (Black et al. 2002).

Assessment and the Student

Change in the role of teachers must lead, in the formative classroom, to changes in the roles of students. One type of change will be cognitive. As questions become more searching, and as the classroom routine is altered to depend more on the active involvement of the learner, students will find that they have to think more and take responsibility for doing more of the work themselves. Because formative work requires the elicitation of students' ideas, they will also have to be more willing to expose these ideas and to submit them to discussion and challenge by their peers as well as by their teachers. This calls for a change in the expectations of students, and such change may disconcert many, who are likely to resist. Thus it becomes important to build a supportive environment. Students must learn to listen to one another, to respect one another's opinions, and to understand that learning works through exploration and challenge, not by rewarding those who are right and labeling those who are wrong.

However, learning is not just a cognitive exercise—it involves the whole person. The need to motivate pupils is evident, but it is often assumed that motivation should consist of extrinsic rewards, such as merits, grades, gold stars, and prizes. Ample evidence challenges this assumption. If a learning exercise is seen as a competition, then everyone is aware that there will be losers as well as winners; those who have a track record as losers will see little point in trying. Thus, the problem is to motivate everyone, even though some are bound to achieve less than others. In tackling this problem, teachers need to realize that the type of feedback they give is very important. Many research studies support this assertion, as the following citations attest:

- Pupils told that feedback "will help you to learn" learn more than those told that "how you do tells us how smart you are and what grades you'll get"; the difference is greatest for low attainers (Newman and Schwager 1995).

- Those given marks as feedback are likely to see the marks as a way to compare themselves with others (ego-involvement); those given only comments see such feedback as helping them to improve (task-involvement). The latter group outperforms the former (Butler 1987).
- In a competitive system, low attainers attribute their performance to lack of "ability" and high attainers attribute their performance to effort. In a task-oriented system, all attribute their performance to effort, and learning is improved, particularly among low attainers (Craven, Marsh, and Debus 1991).
- A comprehensive review of research studies of feedback showed that feedback improved performance in 60 percent of the studies. In the cases where it was not helpful, the feedback turned out to be merely a judgment or grading with no indication of how to improve (Kluger and DeNisi 1996).

In general, feedback in the form of rewards or grades enhances ego rather than task involvement. It can focus pupils' attention on their "ability" rather than on the importance of effort, thereby damaging the self-esteem of low attainers and leading to problems of "learned helplessness" (Dweck 1986). Feedback that focuses on what needs to be done can encourage all students to believe that they can improve. Such feedback enhances learning, both directly through the effort that can ensue and indirectly by supporting the motivation to invest such effort.

Assessment across Subjects

Everyday assessment is not an abstract idea; it is a concrete activity that the science teacher conducts in and through the stuff of science education. While there are generic principles applicable to any learning, practical implementation is bound to be different in the teaching of science and the teaching of, say, history.

The formulation of insightful oral or written questions and the subsequent development of dialogue through which students become involved in their own learning are essential components of formative assessment. For example, a useful question about heat transfer can be based on a picture of three imaginary children arguing about the melting of a snowman. The scenario is that the sun is shining, and there is a breeze blowing, and child A suggests that they wrap a black coat around their snowman to stop the sun from melting it. Child B objects that the black coat will warm up the snowman, and child C says that it all depends on the wind. The class can be asked to say what they think about the arguments of these three children. The question is conceptually rich in that it can be used to open up discussion of radiation, conduction, and convection. But it has two other features. One is that it has the potential to elicit a well-known misconception—namely, that a coat actively warms you up rather than reducing the outward flow of heat. The second feature of the question is that it is likely to interest the children because it uses a context, and a practical need for decision, with which they might identify. Such knowledge, about the way children might think and might be interested, is called *pedagogical content*

knowledge. Science teachers can only plan and conduct an exercise of this type on the basis of understanding of the concepts involved and of the relevant pedagogical content knowledge.

In this snowman example, there is a set of "correct" ideas and a "right answer," albeit a complex one. More generally, there is in any intellectual discipline a spectrum of possible issues, with those issues having well-defined answers lying at one end and those with a variety of good answers lying at the other. Emphasis on inquiry works at the latter end. It calls for a pedagogy that promotes investigations in which students have to exercise initiative and follow a variety of paths and for which the criteria of achievement are concerned with the quality of strategies and arguments rather than with attainment of a single, well-defined outcome. The type of formative "scaffolding" needed here is more akin to that of the teacher of English who is helping students to develop their individual writing styles than to that of the same teacher concerned with basic rules of punctuation.

The current trend of reform in science education is to shift the balance of learning toward the open-ended. This calls for emphasis on inquiry skills, but there is a further and broader aim, which is to engage students in discussion of the effects of the advances of science on society, issues that are, or ought to be, of evident importance to them. Such issues are almost always areas of controversy. A discussion of, say, genetic variation might start with clarification of the concepts, but will change style when the discussion moves on to whether to develop genetically modified crops and the sale of genetically modified foods. What matters here is the quality of the arguments that are deployed, and the teacher has to guide students, in the midst of controversy, both to argue and to listen carefully (Osborne 2000). Teachers in the social sciences might be more experienced in such work than their science colleagues.

Teachers have to be flexible in varying their classroom styles to promote different types of learning. Indeed, many science teachers who are accustomed to dealing with "fixed-answer" topics find it very hard to cope with discussions for which the aim is to help learners be critical about the quality of the arguments, rather than about the correctness of the outcome. The shift from "delivery" modes of learning to interactive modes is an essential step in developing this flexibility. As hinted above, science teachers struggling with the new challenges presented by the standards reforms might find help through study of the classroom practices of colleagues who teach other subjects.

Summary of the Issues

Two main themes emerge from this chapter. The first is that tensions and synergies exist between different purposes of assessment, raising problems for both policy makers and classroom teachers. It is obvious that insofar as there is tension, priority must be given to assessment for learning—the measurement of the outcome of schooling can hardly be more important than the means to secure that outcome.

If this is accepted, then all concerned must find ways to achieve a new equilibrium between the formative and summative purposes, which involves finding ways in which summative systems can serve rather than damage good learning. One obstacle to such achievement has been revealed by many surveys of teachers' classroom practices, for these have shown that assessment practices are one of the weakest aspects of many teachers' work. In part this may well be due to the dominance of summative tests, which provide many teachers with models for assessment—models that are unhelpful because they are not designed as aids to everyday learning.

So there is a need to reconstruct, to build up, everyday assessment so that it serves its first purpose—to elicit and to serve the needs of each learner. A secondary reason why such a development is of central importance is that only when teachers are more skilled and more confident in their own assessment practices might it be possible for them to command the public confidence that is needed so that they can play a strong part in any reconstruction of assessment and testing as a whole.

The second theme of this chapter is the necessity for reform of the classroom as a learning environment. Four principles have been set out for the design of learning environments (Bransford, Brown, and Cocking 1999, xvi–xvii): They should be *learner-centered*, *knowledge-centered*, *community-centered,* and strong in *assessment for learning*. To create a *learner-centered* environment, the teacher must start from where the learner is and work by interaction and formative feedback to promote learning. Teachers must also attend to the learners' understanding of how they can learn and the learners' confidence that they can all learn.

For a school environment to be *knowledge-centered*, the focus has to be on a coherent approach to developing important knowledge and skills, an approach in which all students can grasp what the purposes and values of science are and can feel that they are actually taking part, albeit in a modest way, in sharing in the practices of scientists. So active participation of all is clearly essential.

To be *community-centered*, the environment must again be one in which active participation by everyone in the serious business of learning about science is a priority. Through lively discourse in the classroom, the understanding of all students is advanced and refined and their power to participate in scientific argument is developed. The dominance of competition through testing, which creates winners and losers, and of labeling pupils as "bright" or "dull," ought to be replaced by a shared belief that all students can make progress—and can help one another to do so, as well.

Formative assessment is an essential ingredient for each of these aspects of a learning environment.

References

Baxter, G. P., and R. J. Shavelson. 1994. Science performance assessments: Benchmarks and surrogates. *International Journal of Educational Research* 21: 279–98.

Black, P. J. 1990. APU science: The past and the future. *School Science Review* 72 (258): 13–28.

Black, P., and D. Wiliam. 1998. Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan* 80 (2): 139–48.

Black, P., C. Harrison, C. Lee, B. Marshall, and D. Wiliam. 2002. *Working inside the black box: Assessment for learning in the classroom.* London, UK: King's College London, Department of Education and Professional Studies.

Boaler, J. 2000. Introduction: Intricacies of knowledge, practice and theory. In *Multiple perspectives on mathematics teaching and learning*, ed. J. Boaler, 1–17. Westport, NJ: Ablex.

Bransford, J. A., A. Brown, and R. Cocking. 1999. *How people learn: Brain, mind, experience and school.* Washington, DC: National Academy Press.

Brown, A. L., and R. A. Ferrara.1985. Diagnosing zones of proximal development. In *Culture, communication and cognition: Vygotskian perspectives,* ed. J. V. Wersch, 273–305. Cambridge: Cambridge University Press.

Butler, R. 1987. Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest and performance. *Journal of Educational Psychology* 79 (4): 474–82.

Collins, A. 2002. How students learn and how teachers teach. In *Learning science and the science of learning*, ed. R. W. Bybee, 3–11. Arlington, VA: National Science Teachers Association.

Craven, R. G., H. W. Marsh, and R. L. Debus. 1991. Effects of internally focused feedback on enhancement of academic self-concept. *Journal of Educational Psychology* 83 (1): 17–27.

Dweck, C. S. 1986. Motivational processes affecting learning. *American Psychologist (Special Issue: Psychological science and education)* 41 (10): 1040–1048.

Kluger, A. N., and A. DeNisi. 1996. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 119 (2): 254–84.

Lemke, J. L. 1990. *Talking science: Language, learning and values.* Norwood, NJ: Ablex.

Linn, R. L. 2000. Assessments and accountability. *Educational Researcher* 29 (2): 4–14.

Messick, S. 1989. Validity. In *Educational Measurement* (3rd ed.), ed. R. L. Linn, 13–103. London: Macmillan.

Newman, R. S., and M. T. Schwager. 1995. Students' help seeking during problem solving: Effects of grade, goal, and prior achievement. *American Educational Research Journal* 32 (2): 352–76.

Osborne, J. 2000. Science for citizenship. In *Good practice in science teaching: What research has to say,* eds. M. Monk and J. Osborne, 225–40. Philadelphia, PA: Open University Press.

Shepard, L. A. 1992. Commentary: What policy makers who mandate tests should know about the new psychology of intellectual ability and learning. In *Changing assessments: Alternative views of aptitude, achievement and instruction*, eds. B. R. Gifford and M. C. O'Connor, 301–28. Boston and Dordrecht: Kluwer.

Vygotsky, L. S.1962. *Thought and language.* New York: Wiley.

Wood, D. 1998. *How children think and learn: The social contexts of cognitive development* (2nd ed.). Oxford: Blackwell.

Wood, D., J. S. Bruner, and G. Ross. 1976. The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry* 17: 89–100.