

Library of Congress Cataloging-in-Publication Data


p. cm.
Includes index.
ISBN 978-1-93353-140-3

NSTA is committed to publishing material that promotes the best in inquiry-based science education. However, conditions of actual use may vary, and the safety procedures and practices described in this book are intended to serve only as a guide. Additional precautionary measures may be required. NSTA and the authors do not warrant or represent that the procedures and practices in this book meet any safety code or standard of federal, state, or local regulations. NSTA and the authors disclaim any liability for personal injury or damage to property arising out of or relating to the use of this book, including any of the recommendations, instructions, or materials contained therein.

Permissions
You may photocopy, print, or e-mail up to five copies of an NSTA book chapter for personal use only; this does not include display or promotional use. Elementary, middle, and high school teachers only may reproduce a single NSTA book chapter for classroom- or noncommercial, professional-development use only. For permission to photocopy or use material electronically from this NSTA Press book, please contact the Copyright Clearance Center (CCC) (www.copyright.com; 978-750-8400). Please access www.nsta.org/permission for further information about NSTA's rights and permissions policies.
Contents

Foreword .................................................................................................................. ix
  Elizabeth Stage

Introduction ............................................................................................................. xi
  Janet Coffey and Carole Stearns

Section 1
  Formative Assessment: Assessment for Learning .................. 1

  Chapter 1 ........................................................................................................... 3
  Improving Learning in Science With Formative Assessment
  Dylan Wiliam, Institute of Education, University of London

  Chapter 2 ......................................................................................................... 21
  On the role and impact of Formative Assessment on Science
  Inquiry Teaching and Learning
  Richard J. Shavelson, Yue Yin, Erin M. Furtak, Maria Araceli Ruiz-Primo,
  Carlos C. Ayala, Stanford Educational Assessment Laboratory, and Donald
  B. Young, Miki K. Tomita, Paul R. Brandon, and Francis M. Pottenger III,
  Curriculum Research and Development Group, University of Hawaii

  Chapter 3 ......................................................................................................... 37
  From Practice to Research and Back: Perspectives and Tools
  in Assessing for Learning
  Jim Minstrell, Ruth Anderson, Pamela Kraus, and James E. Minstrell,
  FACET Innovations, Seattle

Section 2
  Probing Students’ Understanding Through Classroom-Based
  Assessment ................................................................................................. 69
Chapter 4
Documenting Early Science Learning
Jacqueline Jones, New Jersey State Department of Education, and Rosalea Courtney, Educational Testing Service

Chapter 5
Using Science Notebooks as an Informal Assessment Tool
Alicia C. Alonzo, University of Iowa

Chapter 6
Assessing Middle School Students’ Content Knowledge and Reasoning Through Written Scientific Explanations
Katherine L. McNeill, Boston College, and Joseph S. Krajcik, University of Michigan

Chapter 7
Making Meaning: The Use of Science Notebooks as an Effective Assessment Tool
Olga Amaral and Michael Klentschy, San Diego State University—Imperial Valley Campus

Chapter 8
Assessment of Laboratory Investigations
Arthur Eisenkraft, University of Massachusetts, Boston, and Matthew Anthes-Washburn, Boston International High School

Chapter 9
Assessing Science Knowledge: Seeing More Through the Formative Assessment Lens
Kathy Long, Larry Malone, and Linda De Lucchi, Lawrence Hall of Science, University of California, Berkeley

Chapter 10
Exploring the Role of Technology-Based Simulations in Science Assessment: The Calipers Project
Edys S. Quellmalz, West Ed; Angela H. DeBarger, Geneva Haertel, and Patricia Schank, SRI International; Barbara C. Buckley, Janice Gobert, and Paul Horwitz, Concord Consortium; and Carlos C. Ayala, Sonoma State University

Chapter 11
Using Standards and Cognitive Research to Inform the Design and Use of Formative Assessment Probes
Page D. Keeley and Francis Q. Eberle, Maine Mathematics and Science Alliance
Section 3  
High-Stakes Assessment: Test Items and Formats.........227

Chapter 12 .................................................................231  
Assessment Linked to Science Learning Goals: Probing Student 
Thinking Through Assessment  
George E. DeBoer and Cari Hermann Abell, Project 2061 at AAAS; Arhonda 
Gogos, Sequoia Pharmaceuticals; An Michiels, Leuven, Belgium; Thomas Regan, 
American Institutes for Research, and Paula Wilson, Kaysville, Utah.

Chapter 13 .................................................................253  
Assessing Science Literacy Using Extended 
Constructed-Response Items  
Audrey B. Champagne, Vicky L. Kouba, University at Albany, State University of 
New York, and Linda Gentiluomo, Schenectady N.Y. School District

Chapter 14 .................................................................283  
Aligning Classroom-Based Assessment With High-Stakes Tests  
Marian Pasquale and Marian Grogan, EDC Center for Science Education

Chapter 15 .................................................................301  
Systems for State Science Assessment: Findings of the National 
Research Council’s Committee on Test Design for K–12 Science 
Achievement  
Meryl W. Bertenthal, Mark R. Wilson, Alexandra Beatty, and Thomas E. Keller, 
National Research Council

Chapter 16 .................................................................317  
From Reading to Science: Assessment That Supports and Describes 
Student Achievement  
Peter Afflerbach, University of Maryland

Section 4  
Professional Development: Helping Teachers Link Assessment, 
Teaching, and Learning..............................................337

Chapter 17 .................................................................341  
What Research Says About Science Assessment With 
English Language Learners  
Kathryn LeRoy, Duval County, Florida, Public Schools, and Okhee Lee, 
University of Miami
Chapter 18.................................................................................................................................357
Washington State’s Science Assessment System: One District’s Approach to Preparing Teachers and Students
Elaine Woo and Kathryn Show, Seattle Public Schools

Chapter 19.................................................................................................................................387
Linking Assessment to Student Achievement in a Professional Development Model
Janet L. Struble, Mark A. Templin, and Charlene M. Czerniak,
University of Toledo

Chapter 20.................................................................................................................................409
Using Assessment Design as a Model of Professional Development
Paul J. Kuerbis, Colorado College, and Linda B. Mooney,
Colorado Springs Public Schools

Chapter 21.................................................................................................................................427
Using Formative Assessment and Feedback to Improve Science Teacher Practice
Paul Hickman, science education consultant, Drew Isola, Allegan, Michigan, Public Schools, and Marc Reif, Ruamrudee International School, Bangkok

Chapter 22.................................................................................................................................447
Using Data to Move Schools From Resignation to Results: The Power of Collaborative Inquiry
Nancy Love, Research for Better Teaching

Volume Editors.........................................................................................................................465
Contributors .................................................................................................................................467
Index..............................................................................................................................................473
Foreword

Elizabeth Stage
Director, Lawrence Hall of Science
University of California, Berkeley

It is all too common to pick up a newspaper and see an article about student achievement (usually declining test scores) or district testing policies and the effects of No Child Left Behind on the allocation of instructional time. All around the country, annual testing for the purpose of accountability is dominating public conversations about education. This focus on accountability testing is just one of many assessment responsibilities teachers juggle daily, and probably the least important for supporting student learning. As the essays in this book attest, teachers also need to assess students to guide daily instructional decisions, to promote their further learning, and to assign grades. In a more perfect world, assessment for accountability and assessment for student learning would align, reinforcing one another. Unfortunately, more often than not, such synergy remains elusive.

In 2005, NSTA invited a distinguished group of researchers and teacher educators to share their current research and perspectives on assessment with an audience of teachers. As the conference demonstrated, a rich body of research on what works and what does not is available to inform teachers’ assessment practices. The conference also demonstrated the value of an open dialogue among researchers and teachers on practical applications of assessment research to practice. The goal of this book, with chapters by the conference presenters, is to share these research-based insights with a larger audience and to help teachers bring together different assessment priorities and purposes in ways that ultimately support student learning.

This book is also a call for greater teacher involvement in assessment discussions, particularly at the state and local levels. Just as we know from classroom-based research that teachers can gain great insight by listening carefully to their students, so too researchers and policy makers will be better informed by listening to teachers—to the questions they have, the
realities they face, and the dilemmas with which they struggle. Teachers should actively engage in conversations, participate in test design and item development, and help improve the assessment literacy of students and parents. Indeed, teachers’ voices are prominent in many of the research efforts described in this book; teachers co-authored many of the chapters. Insights from teachers will help generate strands of research that contribute to richer understandings of assessment practice and its ultimate influence on student learning. While no simple fixes exist for the seemingly divergent assessment purposes, by working together, teachers and researchers can design powerful assessment contexts that help all students reach deep levels of conceptual understanding and high levels of science achievement.
Introduction

Janet Coffey and Carole Stearns

In an era of accountability, talk of assessment often conjures up images of large-scale testing. Although it dominates attention, annual testing is only a small corner of what occurs in the classroom in the name of assessment. Assessment is the chapter test, the weekly quiz, the checking of nightly homework assignments. Assessment can be the observations made as students engage in an activity or the sense-making of student talk as they offer explanations. It is the teacher feedback offered on the lab report, provided to students as they complete an investigation or after they have completed a journal entry. As all of these things and more, assessment is a central dimension of teaching practice.

As the multiple images of assessment suggest, within any classroom, assessment takes on many forms, and must serve multiple purposes. These purposes include accountability and grading. Another important purpose that has received increasing attention is assessment that supports student learning, rather than solely documenting achievement. Different ways to talk about assessment have emerged. We can talk about its purpose, as we just did above. We can talk about the form assessment takes—the multiple-choice test, the portfolio, the alternative assessment, the written comments or oral feedback, or the piece of student work. Different uses of information gleaned from assessment have led us to talk about assessment of learning and assessment for learning, or, in assessment terminology, summative and formative assessment. All of these purposes, forms, and functions are important; all are at play in the classroom.

Over the past decade, the National Science Foundation (NSF) has funded numerous research efforts that seek to better understand assessment in science and math education at all levels; the various strategies and systems; and the variety of forms, roles, and contexts for assessment of and for student learning. NSF has also funded assessment-centered teacher professional development efforts and creation of models for assessment systems that seek synergy among different purposes. In 2005–2006, the National Science Teachers Association convened two full-day conferences to help
disseminate these NSF-funded research findings to practitioners. Many of the recipients of those grants shared their work at the conferences and have prepared chapters for this book in an effort to build connections between research and practice and to facilitate meaningful conversation.

Conversations between research and practice are not commonplace, yet greater exchange is essential. Practitioners help researchers better understand the terrain, including the practitioners’ underlying rationales for their everyday decision making. These insights from those “on the ground” can inform research and contribute to generative lines of questioning. Although starting points and perspectives may differ, ultimately the assessment research and practitioner communities are working toward the same goal: to better understand the relationships between assessment and learning in order to create classroom environments that support our students’ learning.

Researchers are afforded the luxury of stepping back; they can extract a part from the whole—the formative from the summative, for example. They can focus on particular strategies or activities, such as use of notebooks or assessment of lab reports. Teachers, on the other hand, need to make sense of assessment in all its complexity and juggle what may seem like competing priorities and purposes. There may even be times when the different roles teachers take on with respect to assessment appear to conflict: They are, at once, judge and juror, coach and referee. Teachers are asked to figure out ways to navigate these different roles and to align strategies to priorities. They are asked to implement assessment activities and strategies in such a way that a variety of purposes is served, and served well, while mitigating tensions that appear unavoidable.

Research does not hold all the answers. The research community still wrestles with very real and difficult issues that teachers face every day, such as equitable assessments, challenges associated with wide-scale professional development, and assessment designs that capture the complexity of disciplinary reasoning and understanding. As the education community makes progress on these fronts, new challenges and questions arise. No silver bullet exists, nor does a one-size-fits-all fix. However, research can offer insights into strategies and features that are particularly productive, and into frameworks that are particularly compelling.

The essays in this collection will introduce readers to some of the many voices in the national discourse on science assessment, a field currently at the crossroads of education and politics. The essays present authors’ deeply held values and perspectives about the roles of assessment and how assessment must not only provide accountability data but also support the learn-
Assessing science learning of students from different backgrounds. Readers will notice that many of the research studies are grounded in classroom practices and involve teachers as collaborators or in professional development settings. Practitioners’ expertise in understanding the complexity of classrooms is crucial to realizing the importance of assessment in deep science learning for all students.

You will not hear a message of consensus here. The research community does not speak in a unitary voice—beyond the claims that there exists a tight coupling between assessment and student learning and that events and interactions that occur in classrooms in the name of assessment do matter. This is not a “how-to” manual. You will not find polished strategies or assessments to try tomorrow in your classroom. Research cannot offer assistance in that form. Strategies, approaches, and frameworks will need modification and accommodation in order to be meaningfully integrated into your classroom and school. As you read, we encourage you to reflect on your own practice, consider your own priorities, and make sense of what you are learning in light of your own school community.

Organization of the Book
The chapters in this book are grouped into four sections: (1) formative assessment in the service of learning and teaching; (2) classroom-based strategies for assessing students’ science understanding; (3) high-stakes tests; and (4) assessment-focused professional development.

Each section begins with a brief introduction and overview of the included chapters. The section introductions also offer a set of framing questions intended to help readers identify important themes and construct take-home messages that are relevant to their own teaching environment and needs.

The first section, “Formative Assessment: Assessment for Learning,” introduces three perspectives on formative assessment: its role in improving student learning; research examining connections between a sequence of formative assessments and their impact on teaching and learning; and the importance of probing how students learn and their misconceptions. Many of the book’s central ideas are introduced in this section:

- Roles of assessment in teaching and learning,
- Characteristics of meaningful assessment items,
- Need for research to validate assessment practices,
- Significance of assessing both the knowledge and misconceptions of students,
• Value of assessing students’ ability to apply their knowledge, and
• Importance of assessment-focused professional development.

The opening chapter defines classroom-based formative assessment as an ongoing activity informing daily instructional decisions and accompanied by meaningful feedback to students. The author asserts that an essential precursor to raising student achievement in science is providing professional development that will help teachers improve their assessment practices, a topic addressed in many of the chapters and explored in great detail in Section 4.

A research study on correlations between use of embedded formative assessments, teacher practice, and student achievement is the subject of Chapter 2. The focus of the third chapter is the importance of knowing how students learn and the nature of their misconceptions. Readers will learn about tools the authors developed to gather and analyze this information.

The chapters in Section 2, “Probing Students’ Understanding Through Classroom-Based Assessment,” present specific classroom-based strategies for assessing students’ science knowledge and understanding. Several of these strategies are closely linked with students’ literacy and communication skills, primarily writing, but also drawing, reading, and oral communication. These chapters address the day-to-day issues that teachers confront, such as “How much do my students understand?” “What still confuses them?” “How can I encourage them to communicate more clearly?” and “What constitutes a good formative assessment?”

Several authors write about using familiar classroom artifacts such as students’ drawings and notebook entries for assessment purposes. There is a chapter on teaching students to construct reasoned scientific explanations based on their own observations and analysis of data. Secondary teachers may be particularly interested in the chapter on assessing laboratory work. One chapter reports a research study on the use of science notebooks to assess English language learners. (Chapters in later sections also address the needs of English language learners, one in the context of eliminating bias in test items [Chapter 12] and another in a large-scale study of correlations between the science achievement of non-native speakers and the amount of assessment-based professional development their teachers receive [Chapter 17].)

Many of the chapters in this section consider assessments based on familiar classroom routines and artifacts (e.g., science notebooks, lab reports, conversations with and among students) that, when observed through an assessment lens, reveal valuable information about what and how students
are learning. Other chapters in this section describe classroom-based assessment formats and items that were developed by researchers and subjected to field testing in multiple classroom settings. A team of developers describes a suite of formative assessment tasks designed to monitor student learning at several points during a multi-week unit of study. Another chapter introduces a technology-based assessment system developed to track students’ problem-solving skills as they interact with a computer simulation. This section concludes with a chapter offering teachers guidelines on constructing standards-based formative assessment probes.

Section 3, “High-Stakes Assessment: Test Items and Formats,” begins with an examination of the cognitive demands of several high-stakes test item formats. Authors focus on what students must know and be able to do to succeed on high-stakes tests and how teachers’ own classroom assessment can help students meet these challenges. The opening chapter takes readers through the process of designing and field testing items that are closely linked to specific standards-based learning goals. The next chapter analyzes constructed-response test items, a format commonly used in national and international tests, such as TIMSS and NAEP. The authors present sample items and detailed scoring guides to help teachers better understand how such items are scored. Another chapter provides teachers with guidelines for analyzing the content and format of high-stakes test items and creating closely aligned questions to use in their own classrooms.

Section 3 continues with a chapter summarizing the National Research Council’s (NRC) report on design principles for state-level science assessment systems. The authors discuss the goals of state-level assessment, calling attention to the distinct differences between these tests and the classroom-based assessments described in Section 2. The concluding chapter offers reflections by a literacy expert on high-stakes testing practices and test items in his field. He summarizes the lessons learned and offers some suggestions to science test developers.

In Section 4, “Professional Development: Helping Teachers Link Assessment, Teaching, and Learning,” authors describe several large-scale professional development initiatives that emphasize building assessment expertise. Programs in Seattle, Washington, Toledo, Ohio, Miami, Florida, and Colorado Springs, Colorado are discussed. While each had a different approach to professional development design, all included a research component investigating potential correlations between the teachers’ experiences and their student performance on high-stakes tests. Each study reports compelling data showing a positive correlation between teachers’ participation in the professional development efforts and student achievement on high-stakes science tests.
A chapter on a classroom observation research tool titled the Reform Teacher Observation Protocol (RToP) offers another approach to professional development. The authors discuss the use of this tool by secondary teachers to self-evaluate their classroom assessment practices. The final chapter describes strategies that school teams can use to analyze assessment data from multiple sources; including high-stakes tests, classroom-based assessments, and teacher observations, for the purposes of program evaluation and guiding instructional decisions.

* * *

This brief summary does little justice to the richness of the essays herein and to the multiple examples of meaningful science assessment practices they explore. The collection reflects work with socioeconomically and ethnically diverse populations to better understand the attributes of equitable assessment practices. While the authors may describe an assessment study conducted within a narrow context (science teachers will recognize the constraints required by a controlled experiment), the findings and recommendations are broadly applicable. For example, professional development programs in Seattle, Washington, offer many ideas equally relevant for schools and districts in other parts of the United States. Similarly the assessment potential of student notebooks extends far beyond classrooms in El Centro, California.

We hope that this book can be used to fuel the conversations about assessment sparked in the initial NSTA conference. From the informal interactions that occur among students and teachers to more formal exchanges, from item design to grading, and from classroom systems of reporting on progress to large-scale external state tests, fodder exists for deep and provocative discussion. In the essays that follow, readers have an opportunity to consider the issues closely and to reflect on the ways in which assessment can be better coordinated. We hope that, eventually, the entire system will become more synergistic in order to meet the many purposes of assessment while not neglecting or undermining any single one.

The editors are grateful to the researchers who contributed to this volume for their commitment to communicating their work to practitioners, the ultimate consumers of science assessment knowledge. We hope that readers will find many ideas that enrich their own understanding of the assessment landscape and help them better serve their students. We encourage teachers to actively engage in the national assessment conversation and to share insights they develop in their own classrooms.
Science education researchers, like science teachers, are committed to finding ways to help students learn science. Like teachers, we researchers start with an informed hunch about something that we think will improve teaching. Then we work with teachers and try out our hunch in real classrooms. If we get positive results, we share them with a wide range of educators. Sometimes we find out that our hunch does not work, and we try to figure out what went wrong so that we can improve it the next time. In other cases, we find that while the idea may have been good, the technique will not work in practice. In those cases, we continue our search for other ways to help improve students’ learning of science.

In reviewing the literature on assessment, Paul Black and Dylan Wiliam found strong evidence that embedding assessments in science curricula would lead to improved student learning and motivation (Black and Wiliam 1998; see also Wiliam, Chapter 1 in this book). Based on this finding, our team of teachers, curriculum and assessment developers, and science education researchers developed a series of formative assessments to embed in a middle school physical-science unit on sinking and floating. We wanted...
to see if this kind of assessment, which helps teachers to determine the status of students’ learning while a unit is still in progress, would improve sixth- and seventh-grade students’ knowledge and motivation to learn science. If it worked, we knew we might have a large-scale impact on teaching and learning.

In this chapter, we begin by describing what we mean by formative assessment and outline the potential and challenges of trying to implement and study this promising technique for scientific inquiry teaching. We then describe our study on formative assessment in middle schools, including some mistakes and wrong turns, and what we found when we tested our ideas experimentally. We conclude with future challenges in improving science education with formative assessment.

What Is Formative Assessment?

Formative assessment is a process by which teachers gather information about what students know and can do, interpret and compare this information with their goals for what they would like their students to know and be able to do, and take action to close the gap by giving students suggestions as to how to improve their performance. In this way, formative assessment is carried out for the purpose of improving teaching and learning while instruction is still in progress.

To clarify what we mean by formative assessment, consider the large-scale, high-stakes assessments that are carried out in all U.S. schools today. These types of assessments are summative in nature; that is, they provide a summary judgment about, for example, students’ learning over some period of time. The goal of summative assessment is to inform external audiences primarily for evaluation, certification, and accountability purposes. Since the federal No Child Left Behind legislation was passed in 2001, summative assessment has certainly received a great deal of publicity in the popular media and has, to a certain degree, swamped the important formative function of assessment.

By focusing on formative assessment, we hope to put assessment back into its rightful place as an integral part of the teaching-learning process. Formative assessment takes place on a continuous basis, is conducted by the teacher, and is intended to inform the teacher and students, rather than an external audience (Shavelson 2006). We view classroom formative assessment as a continuum ranging from informal formative assessment to
formal formative assessment. The position of a particular formative assessment technique on the continuum depends on the amount of planning involved, the formality of technique used, and the nature of the feedback given to students by the teacher. We focus on three important formative assessment techniques—(1) “on-the-fly,” (2) planned-for-interaction, and (3) embedded in the curriculum (Figure 2.1) and describe each in turn.

**Figure 2.1 Variation in Formative Assessment Practices**

<table>
<thead>
<tr>
<th>Informal Unplanned</th>
<th>Planned</th>
<th>Formal</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-the-Fly</td>
<td>Planned-for-Interaction</td>
<td>Embedded-in-the-Curriculum</td>
</tr>
</tbody>
</table>

**On-the-Fly Formative Assessment.** On-the-fly formative assessment occurs when “teachable moments” unexpectedly arise in the classroom. For example, teachers circulate between groups to listen in on conversations and make suggestions that give students new ideas to think about. A teacher might overhear a student in a small group investigating sinking and floating say that, as a consequence of an experiment just completed, “Density is a property of the plastic block. It doesn’t matter what the mass or volume is, the density stays the same for that kind of plastic.” The teacher recognizes that the student has a grasp of what density means for that block, and presents the student with other materials to see if she and her group-mates can generalize the density idea to a new situation. In this way, the teacher challenges the student to test her new idea by having her and her group measure the mass/volume relationships of a new material. Moreover, when satisfied that the students are onto something, the teacher calls for other students to hear what this group found out.

This vision of taking advantage of the “teachable moment” sounds a lot like good teaching, not necessarily assessment. This is exactly our point: Teaching and assessment are and should be considered as one and the same.
Rather than teachers planning assessment as a separate event during the class period, on-the-fly assessment is seamless with instruction and is based on the teacher capitalizing on opportunities as they arise to help students to move forward in reaching learning goals.

However, as we learned from our research, such on-the-fly formative assessment and action ("feedback") may be natural for some teachers but difficult for others. Identification of these moments is initially intuitive and then later based on cumulative wisdom of practice. Moreover, even if teachers can identify the moment, they may not have the confidence, techniques, or content knowledge to sufficiently challenge and respond to students.

**Planned-for-Interaction Formative Assessment.** In contrast, planned-for-interaction formative assessment is deliberate. Teachers plan for and craft ways to get information about the gap between what students know and need to know, rather than use questions just to “keep the show going” during an investigation or whole-class discussion. Consider, for example, teacher questioning—a ubiquitous classroom event. While developing a lesson plan, a teacher can prepare a set of “central questions” that get at the heart of the learning goals for that day’s lesson and that have the potential to elicit a wide range of student ideas. For example, these questions may be general (“Why do things sink and float?”) or more specific (“What is the relationship between mass and volume in floating objects?” “Can you give me an example of something really heavy that floats? Why do you think it floats?”). At the right moment during class, the teacher poses these questions to the class, and through a discussion the teacher learns what students know and allows different ideas to be presented and discussed. In this example, the teacher planned the assessment prompt in advance rather than waiting for unexpected opportunities to arise. Although not every student in class may respond to each question, the information gained from the students’ responses allows the teacher to act on the information collected by fine-tuning instruction or intervening with individual students.

**Embedded-in-the-Curriculum Formative Assessment.** Alternatively, teachers or curriculum developers may embed more formal assessments ahead of time in the ongoing curriculum to intentionally create “teachable moments.” These assessments are embedded after junctures or joints in a
unit where an important goal should have been reached before going on to the next lesson. Embedded assessments inform the teacher about what students currently know and what they still need to learn (i.e., “the gap”) so that teachers can provide timely feedback to students.

In their simplest forms, formal formative assessments are designed to provide information on important goals that students should have reached at critical joints in a unit before going onto the next lesson. In their advanced forms, formative assessments are based on a developmental progression of the ideas students have about a particular topic (such as why things sink and float). In contrast to the other two types of formative assessment, embedded assessments are more sophisticated because they are designed to collect critical information about student learning at the same time. The main difference between planned-for and embedded formative assessment is in the designer. Whereas planned-for assessment is usually done by the teacher as a part of the lesson-planning process, embedded assessments are usually designed by curriculum and assessment developers working with experienced teachers.

Embedded formative assessments are valuable teaching tools for at least four reasons. First, they are consistent with curriculum developers’ understanding of the curriculum and are therefore consistent with instructional goals. Second, assessment developers contribute technical expertise that increases the quality of the assessments. Third, the involvement of experienced teachers in developing embedded assessments means that they are practical and based on the wisdom of practice. And fourth, embedded assessments provide thoughtful, curriculum-aligned, and valid ways of determining what students know, rather than leave the burden of planning on the teacher.

Formal embedded assessments come “ready-to-use” as part of a preexisting curriculum, and instructional decisions made from them may improve students’ learning. Therefore, in our study, we sought to learn whether embedded formative assessments actually helped teachers close the learning gaps in their classrooms.

Potential and Challenges
Formative assessment is a potentially powerful teaching idea embodying knowledge and skills for creating and capitalizing on teachable moments. In the context of science education, formative assessment links teaching
and learning in the service of building students’ understanding of the natural world and of how the methods of science justify knowledge claims. In using formative assessments, we sought to move students from naive conceptions of the natural world to scientifically justifiable conceptions (“conceptual change”). To change their conceptions, students need to link what they find out through inquiry investigations to their current conceptions of the natural world and to change those conceptions when their evidence does not fit their “theory.” Formative assessment’s critical characteristic, then, lies in identifying learning gaps and providing immediate feedback to students that helps them close gaps.

This said, many teachers are in some ways skeptical about incorporating formative assessment substantively into their teaching practice, even when they know that it is important. Teachers have many questions about their role in formative assessment, and for good reason. For example, formative assessment creates a conflict with the teacher’s traditional grade-giving role in summative assessment. How can the teacher on the one hand ask students to lay bare their understanding of a concept and at the same time have the responsibility for giving the student a grade? In other cases, teachers may have only experienced summative assessment when they were students themselves, or in their teacher education programs. Consequently, they may not have personal experience with the ways that formative assessment can improve the quality of teaching and learning. Other questions arise as well. Should teachers really change their beliefs about their role as assessors? Why should teachers change their practices to accommodate a yet unproven teaching technique? Will our emphasis on formative assessment eventually fade away as have other reform techniques?

Clearly, teachers’ skepticism is appropriate; part of the science education researcher’s role is to test out new (or not so new) techniques to see if they stand up to scientific scrutiny. To this end, our team designed and conducted a study that put formative embedded assessment to the test.

**Embedding Formative Assessment in a Science Curriculum**

Our study of formative embedded assessment addressed two central research purposes: first, to learn how to build and embed formative assessments in science curricula and, second, to examine the impact of formative assessments on students’ learning, motivation, and conceptual change.
Building and Embedding Formative Assessments in Science Curricula

As noted above, we sought to move students from naive conceptions of the natural world to scientifically justifiable ones. To this end, we wanted students to link what they were finding out through investigations to their conceptions about the natural world. The intent was for students to change those conceptions when their evidence didn’t fit their “theory.”

We embedded formative assessments in the Foundational Approaches in Science Teaching (FAST) curriculum unit on the properties of matter—more specifically, buoyancy (Pottenger and Young 1992). As a first step, we identified the goals for the unit. The main goal was for students to develop, through a series of inquiry investigations, a relative density-based explanation for sinking and floating (or, as we came to call it during the study, “Why things sink and float” or “WTSF”). We then worked from the goals backward to the beginning of the unit, identifying key junctures between lessons (“investigations”) where important goals needed to be met. We then inserted assessments to provide information about student performance.

Despite our well-conceived plans, in the end, a seemingly straightforward process of developing formative assessments was anything but straightforward. We made some wrong turns and learned from our mistakes.

Pilot Study: From Embedded Formative Assessments to Reflective Lessons

Our basic idea was to develop and embed formative assessments where the “rubber hit the road”—that is, at critical curricular joints where students’ conceptual understanding was expected to develop from a simple level to a more sophisticated one. In this way, teachers would know whether students were advancing in their knowledge as the curriculum progressed. We expected that assessments embedded at the critical joints would provide timely information to (a) help teachers and students locate the levels of students’ understanding, (b) determine whether students had reached the desired level, (c) diagnose what students still needed to improve, and (d) help students move to the next level.

At each critical joint, we created a set of assessments designed to tap different kinds of knowledge that students should construct in learning about sinking and floating. There were facts (e.g., density is mass per unit volume—declarative knowledge) and procedures (e.g., using a balance scale to measure the mass of an object—procedural knowledge). But most impor-
tant, and often implicit in curricula, was the use of this declarative and procedural knowledge in inquiry science to build a model or mini-theory of why things sink and float (e.g., a model of relative densities—schematic knowledge). Consequently, we embedded assessments of these types of knowledge at four natural joints in a 10-week unit on buoyancy. The assessments served to focus teaching on different aspects of learning about mass, volume, density, and relative density. Feedback on performance focused on problematic areas revealed by the assessments.

In order to embed assessments that were based on research and that could identify in a valid and reliable way what students know, we created four extensive assessment “suites” (combinations of individual assessments—graphing, short answer, POE [predict-observe-explain], and PO [predict and observe]). These assessments covered the declarative, procedural, and schematic knowledge underlying buoyancy. Each suite included multiple-choice (with space for students to justify their selections) and short-answer questions that tapped all three types of knowledge. We also included a substantial combination of concept maps (structure of declarative knowledge), performance assessments (procedural and schematic knowledge), predict-observe-explain assessments based on lab demonstrations (schematic knowledge), and/or “passports” verifying hands-on procedural skills (e.g., measuring an object’s mass).

Three brave teachers volunteered to try out this extensive battery of embedded assessments in a pilot study. After the completion of the pilot study, the teachers warned us that the original formative assessments were too time-consuming and the amount of information obtained from them was overwhelming. Our lead pilot-study teacher, who was also a member of our assessment team, gently pointed out the problems that pilot-study teachers faced using our assessment suites. She suggested that perhaps there could be only a few assessments that directly led to a single, coherent goal, such as knowing why things sink and float. She pointed out that FAST provided ample opportunity for teachers to observe and provide feedback to students on their declarative and procedural knowledge. She urged us to focus on schematic knowledge and on students’ developing an accurate mental model of why things sink and float in the assessment suite.

Moreover, Lucks (2003) viewed and analyzed videotapes of the pilot study teachers using the assessment suites. She found that our teachers were treating the “embedded assessments” more as external tests that were some-
thing apart from the curriculum—in other words, as *summative* assessment—rather than using the formative assessments as a way to find out what the students were learning. Thus, the teachers treated the new assessments like any other test that they were required to give to the students during the year, rather than as opportunities to increase their students’ learning.

Based on the thoughtful feedback we received from the teachers and the researcher, we revised our initial embedded assessments, greatly reducing their numbers and focusing in on the overarching goal of explaining “why things sink and float.” Afterward, when talking with teachers, we no longer spoke of embedded assessments, which we thought would trigger their stereotypes about assessments. Instead, we started calling them “Reflective Lessons” to emphasize their function as a component of the teaching and learning process.

**The New Generation of Formative Embedded Assessments: The Reflective Lessons**

A second look at the FAST unit and the information collected during the pilot study led us to a developmental progression of student ideas, which then became the basis for redesigning the original embedded assessment suites into Reflective Lessons (Figure 2.2, p. 30). This progression was aligned to the unit and based on different conceptions students have as they develop an understanding of sinking and floating. These conceptions develop from naive (e.g., “things with holes in them will sink”) to scientifically justifiable conceptions (e.g., “sinking and floating depend on the relative densities of the object and the medium supporting the object”).

Although Figure 2.2 may appear quite complicated, the ideas behind it are straightforward and consistent with students’ different ideas about sinking and floating. Before instruction, students have all different kinds of ideas about sinking and floating, such as that heavy things sink, flat things float, things with air in them float. We would place these ideas at Level 1 or “Naive Conceptions.” As students progress through the unit, they complete investigations that apply either mass or volume to sinking and floating; that is, a single uni-dimensional factor (Level 2), holding all else constant. Next, students simultaneously apply mass and volume, or multiple uni-dimensional factors, to explain sinking and floating (Level 3). Afterward, students integrate mass and volume into density, a single bi-dimensional factor, in their explanations (Level 4). Finally, students consider
the object’s density and the liquid’s density, or multiple bi-dimensional factors (Level 5), in their explanations (Yin 2005).

The final Reflective Lesson suites are shown at their critical junctures in Figure 2.3. Two types of Reflective Lessons were embedded in the unit. Each of the type one Reflective Lessons included a sequence of the following activities: (a) graphing and interpreting evidence and drawing conclusions about WTSF (“Why things sink or float”), (b) applying knowledge learned to predict and explain what would happen in a new situation (Predict, Ob-
serve, Explain), (c) writing a brief explanation about why things sink and float, and (d) predicting and observing a surprise phenomenon to introduce the next set of lessons. The second type of Reflective Lesson was concept mapping, which encouraged students to make connections between the concepts they learned.

The Reflective Lessons were designed to enable teachers to (a) elicit students’ conceptions, (b) encourage communication of ideas, (c) encourage argumentation (comparing, contrasting, and discussing students’ conceptions), and (d) reflect with students about their conceptions. In this way, teachers could guide students along a developmental trajectory that they had in hand from naive conceptions of sinking and floating to more scientifically justifiable ones (Figure 2.2).
The Experimental Study

To test whether the final Reflective Lessons could help students improve learning, motivation, and conceptual change, we conducted a small experiment. We randomly assigned 12 teachers to teach either the regular inquiry curriculum (control group—6 teachers) or the curriculum with the Reflective Lessons included (experimental group—6 teachers). Teachers in the experimental group attended a training workshop with the researchers, curriculum developers, and one of the pilot teachers. During the training, teachers participated in the Reflective Lessons as students, talked about the process of the lesson, and then practiced teaching the Reflective Lessons themselves with lab school students. Teachers in the control group also attended a training workshop that oriented them to the study and invited them to share their assessment practices, among other things.

In the study, we gave pretests and posttests to the students in both groups. We examined the effect of the Reflective Lessons by comparing improvement made by the two groups, regarding students’ motivation, achievement, and conceptions of sinking and floating (Figure 2.4) (Yin 2005).

Since the Reflective Lessons integrated formative assessment ideas, curriculum goals, and teachers’ input, we expected that students in the experimental group would benefit from the Reflective Lessons and show higher learning gains than the control group. To our surprise, our findings did not support this conjecture. We found no statistically significant differences between average performance in the control and experimental groups. That
is, students in the experimental group and control group did not differ, on average, on motivation, learning, or conceptual change. This finding persisted even after we accounted for differences among students’ achievement and motivation before the study began.

Despite the fact that the study did not come out as expected, we learned a lot about how teachers actually used the Reflective Lessons in their classrooms. In each group, teachers varied substantially in producing differences in students’ motivation, learning, and conceptual change. In viewing classroom videos we found that although the Reflective Lessons (embedded assessments) were implemented by teachers in the experimental group, not all the teachers used them effectively to give students feedback or modify teaching and learning (Ruiz-Primo and Furtak 2006, 2007). That is, among the teachers in the experimental group, those teachers whose students had higher learning gains relied more on the other two types of assessment techniques—on-the-fly and planned-for-interaction assessment—rather than on the Reflective Lessons.

To give an idea of the differences among teachers, let us consider two teachers in the experimental group, Gail and Ken. Gail took an active role in using the Reflective Lessons with her students. She would build knowledge with students by challenging their ideas, asking them for empirical evidence to justify their ideas, and making clear how a model of sinking and floating was emerging. The Reflective Lessons created teachable moments for her, which she then took advantage of with informal assessment techniques. Ken, in contrast, relied on the Reflective Lessons themselves to help the students learn and looked at the activities as discovery learning; that is, he depended on the students to develop their own understandings with limited teacher intervention (Furtak 2006). He reasoned that it was not his role to act on the students’ ideas about sinking and floating and to guide the students or tell them the answers; rather it was up to students to discover for themselves why things sink and float.

In Figure 2.5, page 34, we see the developmental trajectory for a typical student from Gail’s class and another from Ken’s. While Gail’s student progressed along the trajectory, Ken’s student held to her original explanation. The achievement test scores for the two students reflected the differences

---

1These names are pseudonyms. We use male and female names for writing ease (e.g., to avoid he/she, his/her). We did not find gender differences in teaching effects in our study.
Concluding Comments

As we know, when any new reform idea comes along, there is a lot of hype. Moreover, teachers are expected to pick up the new “tools” and implement the ideas perfectly on the first try, after they have been trained (briefly!) to do so. Even though we worked intensely with our experimental teachers to learn how to use Reflective Lessons and provided follow-up during the experiment, the kinds of knowledge, belief, and practice changes we wanted to bring about—conceptual changes—needed much more time. Those in learning (Gail’s student: pretest 15 and posttest 36; Ken’s student: 23 and 23, respectively) (Yin 2005).
teachers who already believed in and had already incorporated some of the techniques in their practice that we sought to build in the experimental group performed largely as we had hoped. However, those teachers whose beliefs were somewhat different took even longer to acquire the habits of mind and teaching techniques required to use Reflective Lessons (formative assessment) effectively.

We continue to believe that formative assessment practices hold promise for improving science inquiry teaching, and for improving students’ motivation, learning, and conceptual change. However, if we are to put formative assessment to the test fairly, we need time to work with teachers on their formative assessment knowledge, beliefs, and practices. Once a reasonable level of expertise has been reached, that is the time to try the experiment again (and again and again). If successful, we may have something that would help improve science education; if not, we know not to pursue this aspect of reform further. Perhaps not surprisingly, we are currently engaged in a replication (hopefully with appropriate improvements) of the experiment. Stay tuned!

References
Index

Page numbers in boldface type refer to figures or tables.

A
AAAS (American Association for the Advancement of Science), 108, 117, 194, 195, 227, 228, 231–232, 233, 235, 258, 267, 413, 414
Abell, C. H., 227, 231
Accountability
assessment and, 319, 410
NCLB and, 303
standards-based, 232
for student collaborative learning, 12
for teacher change, 17
Accountability assessments, 168
Active Chemistry, 148, 149
Adequate yearly progress (AYP), 303, 322
Advanced Placement tests, 6
Afflerbach, P., 229, 317
Alonzo, A., 70
Amaral, O., 70, 117
American Association for the Advancement of Science (AAAS), 108, 117, 194, 195, 227, 228, 231–232, 233, 235, 258, 267, 413, 414
American Federation of Teachers, 233
America’s Lab Report, Investigations in High School Science, 150
Anderson, R., 2, 37
Anthes-Washburn, M., 71, 145
Assessing Science Knowledge (ASK) Project, 71–72, 169–190
activities of, 176–178
assessment triangle as theoretical framework of, 171–176, 172
cognition, 172, 173
interpretation, 175, 175
observation, 173, 174
background of, 169–171
benchmark assessments of, 183–184
challenges of incorporating new assessment paradigms into classroom practices, 185–188
classroom culture, 187–188
coverage, 186
reflection, 187
time, 185–186
trust, 186–187
embedded assessments of, 178–183, 190
quick write, 178–179
response sheets, 181–183, 182
science notebook entries, 179–181, 180
field test centers for, 171
funding for, 169
goals of, 171
I-Checks, 170–171, 183–184
organization of, 171
purpose of, 169
valuing progress rather than achievement in, 184–185
Assessment, defined, 308–309
Assessment design as model of professional development, 409–425. See also Science Teacher Enhancement Project-unifying the Pikes Peak region
Assessment for learning, 6–13. See also
Formative assessment
cost-benefit analysis of, 5–6
effect on student achievement, 5–6, 44–45
integrated assessment, 18
to keep learning on track, 13–14
perspectives and tools in, 37–56
strategies for, 7–13
activating students as learning resources for one another, 12–13
activating students as owners of their learning, 11–12
clarifying and sharing learning intentions and success criteria, 10–11
engineering effective classroom discussions, questions, activities, and tasks, 7, 7–8
providing feedback, 8–10
teacher learning communities for implementation of, 14–18
Assessment for learning cycle, 39–43, 45
acting with purpose, 39–40, 42
planning based on findings, 42
targeting needs, 42
developing skills and classroom culture for, 42–43
gathering information, 39, 41
anticipating student responses, 41
choosing and implementing appropriate strategy, 41
determining learning goal, 41
implementation of, 50–52
interpreting information, 39–40, 41–42
determining needs to move learning forward, 42
identifying problems and strengths in student thinking, 41
repeating until students achieve learning goal, 40
Web-delivered tools for support of, 52–55, 58–66 (See also Diagnoser Tools)
Assessment linked to content standards, 231–251
background of, 232–235
development of, 235–250
aligning assessment items to standards, 238–240
clarifying content standards, 235–238, 239
pilot testing: using student data to improve items, 240–250, 241
Project 2061, 227, 231–251
stakeholders’ needs fulfilled by, 234–235
Assessment probes, defined, 206–207. See also Formative assessment probes
Assessment triangle, 162, 162–163
use in Assessing Science Knowledge Project, 171–176, 172
Atlas of Science Literacy, 194, 238, 258, 267, 413
Audiences for assessment, 324–326, 325
Authentic assessment, 329
Ayala, C. C., 21, 71, 84, 191
AYP (adequate yearly progress), 303, 322
B
Ball, D. L., 5
Beatty, A., 228, 301
Bellina, Jr., J. J., 436
Benchmarks
creating explanation assessment tasks based on, 108, 108–109
development in STEP-uP project, 413–414
Berkeley Evaluation and Assessment Research Center, 177
Bertenthal, M. W., 228, 301
Biologica, 194
Black, P., 14, 21, 45, 46, 169
Bloom, B., 328
Brandon, P. R., 21
Buckley, B., 71, 191
C
California Standards Test–Science Subtest, 139–140
Calipers simulation-based science assessments, 71, 194–201
alignment with national standards, 195, 200
development of, 194–199
for ecosystems, 198–199, 199, 200
for forces and motion, 195–198, 196
principled assessment design approach to, 195
goals of, 194
pilot testing of, 200
promise of, 201
technical quality of, 200–201
Carr, E., 393
Champagne, A. B., 228, 253
Chinn, C. A., 159
Choice, in process of teacher change, 17
Classroom Assessment and the National Science Education Standards, 396
Classroom culture
Assessing Science Knowledge (ASK) Project and, 187–188
Copyright © 2008 NSTA. All rights reserved. For more information, go to www.nsta.org/permissions.
assessment of learning cycle and, 42–43
challenges in transformation of, 188
“Classroom Observation Protocol,” 397
Classroom technology, 52–53
simulations, 191–201
Classroom video records, 431–432
Classroom-based assessment, 69–72, 168, 233
aligning to high-stakes tests, 292–299, 294–298
for English language learners, 348–349
Assessing Science Knowledge (ASK) Project, 71–72, 169–190
discussions, questions, activities, and tasks to elicit evidence of student learning, 7, 7–8
documenting early science learning, 69–70, 73–81
formative assessment probes, 72, 203–225
to identify teachable moments, 320
importance of high-quality assessments, 318–321
laboratory investigations, 71, 145–164
lessons from other disciplines, 320–321
rubrics for, 129–137
science notebooks, 70, 83–97, 117–140
student self-assessment, 332–333
summative vs. formative, 22, 168
technology-based simulations: Calipers project, 71, 191–201
traditional paradigm of, 169
written scientific explanations, 70, 101–113
zone of proximal development and, 319, 319–320
Class-size reduction and student achievement, 4–5
Clement, J., 48
Clymer, J. B., 10
Coffey, J., xi
Cognition
assessment of higher-order thinking skills, 283, 304
in assessment triangle, 162, 162, 172, 172
cognitive processes used by scientists, 159–160
metacognitive approaches to instruction, 118, 127
Model of Student Cognitive Processes, 126
Cognitive and affective learning outcomes, 330–331
Cognitive demands of constructed-response items, 255, 256
on PISA, 262, 262–263
on TIMSS, 267–269, 268
of high-stakes state tests, 283–284
Collaborative inquiry among teachers, 447–462
building foundation for, 461–462
connecting data to results, 450–451, 451, 452
establishing times for, 462
guiding questions for, 447–448
impact on student achievement, 448–450, 449
to improve students’ graphing skills, 456–459
for instructional improvement, 460–461
school culture for, 461
sources of data for, 462
use of Data Pyramid, 453, 453–456
Collaborative learning, 12–13
“two stars and a wish” format for, 13
Colorado Science Model Content Standards and Benchmarks, 413
Concept maps, 28
Conceptual storylines, in STEP-uP project, 412–414, 414–418
Conceptual strand maps, 238, 239
Concord Consortium, 195
Connected Chemistry project, 193
Constructed-response items, 253–271
analysis of, 257
to assess science content vs. practices, 256–257
cognitive demands of, 255, 256
determining correct responses to, 257
extended vs. short, 253
on high-stakes tests, 287, 287
on international assessments, 254–269
PISA, 258–263, 259, 260, 262
INDEX

placing in context, 257–258, 271
scoring of, 270
Cost-benefit analyses, 3, 4–6
of assessment for learning, 5–6
of class-size reduction, 4–5
Courtney, R., 69–70, 73
CRESST (National Center for Research
on Evaluation, Standards, and
Student Testing), 362, 370
Criteria for learning, 10–11
Critical Friends protocols, 420, 423
CSAP (Colorado Student Assessment
Program), 285–286, 286, 289,
290–291, 338, 409, 419, 424
CTS. See Curriculum Topic Study–
developed assessment probes
Curriculum developers/researchers,
benefits of assessment linked to
content standards for, 234–235
Curriculum Topic Study (CTS)–
developed assessment probes, 204,
205
analysis of, 218–219
as basis for further inquiry into student
ideas, 220
deconstruction of, 219–220
development of, 211–214, 212, 213
field testing of, 209
publication of, 214
relation to learning goals, 209–210
scaffold for design of, 211, 212
teacher’s reflection on, 222–225
teaching informed by responses to,
220–222, 221
vs. traditional assessment items, 207,
207–211, 208
two-tiered format of, 214
to uncover students ideas, 214–215,
215–218
Curriculum-embedded assessment. See
Embedded assessment
Czerniak, C. M., 338, 387, 399

De
Declarative knowledge
cognitive demands of, 255, 256
on PISA, 262
on TIMSS, 268, 269
embedded formative assessment of,
27–28
in National Science Education
Standards and Benchmarks, 258
DeNisi, A., 8
“Density” labs, 148–150, 149
Development of high-quality assessments,
18
Developmental approach to assessment,
312–313
Developmental “storylines,” 125
Diagnoser Tools, 2, 53–55, 58–66
Developmental Lessons, 61–62
effective teacher use of, 53–55
Elicitation Questions, 54, 59, 60
Facet Cluster, 59–60, 61
Learning Goals, 58–59
Prescriptive Activities, 65–66, 67
Question Sets, 55, 62–63, 63, 64
Teacher Report, 63–64, 65
Diranna, K., 447
Disabilities, NCLB and assessment of
students with, 304
diSessa, A., 48
Documenting early science learning,
69–70, 73–81, 74
collecting evidence that shows
understanding of groups of
children’s drawings, 78, 79
record of class discussion, 78, 79
collecting forms of evidence over a
period of time for, 76, 77
collecting variety of forms of evidence
for, 74–75
drawing, 74, 75
drawing and dictation, 75, 75
photographs, 75, 76
five-stage process of, 79–81
principles for, 74
Dweck, C. S., 11

E
Eberle, F. Q., 72, 203
Ecologist, 161
Education reform, value for money in, 3,
4–6
Effectiveness assessments, 168
INDEX

477

Eisenkraft, A., 71, 145
Elementary and Secondary Education Act (ESEA), 302
Elementary Reading Attitude Survey, 330
Elementary Science Study (ESS), 83
ELL students. See English language learner students
Embedded assessment, 21–22, 23, 24–35, 169, 292. See also Assessment for learning cycle
to assess declarative, procedural, and schematic knowledge, 27–28
in Assessing Science Knowledge (ASK) Project, 71–72, 178–183, 190
building and embedding of, 27
creation of assessment suites, 27–28
at critical curricular joints, 27–28
design of Reflective Lessons suites, 29–32, 30, 31
experimental study of Reflective Lessons, 32, 32–34, 34
outcome of study of, 34–35
pilot study of, 27–29
in STEP-uP project, 410
in TAPESTRIES project, 393, 394, 395
English language learner (ELL) students, 341–351
aligning classroom assessment with high-stakes assessments for, 348–349
assessment accommodations for, 304, 348, 350
assessment in their home languages, 341, 350
designing science and literacy assessment instruments for, 346–347
assessment results, 347, 348
reasoning task, 347
science test, 346
writing prompt, 346–347, 352–355
how to assess science and literacy achievement of, 349–350
inquiry-based science for, 342–345, 344, 345
low science achievement of, 348
NCLB and, 304, 341, 348, 350
ESEA (Elementary and Secondary Education Act), 302
ESS (Elementary Science Study), 83
Extended constructed-response items, 253–271
analysis of, 257
to assess science content vs. practices, 256–257
cognitive demands of, 255, 256
determining correct responses to, 257
on high-stakes tests, 287, 287
on international assessments, 254–269
PISA, 258–263, 259, 260, 262
placing in context, 257–258, 271
scoring of, 270
vs. short constructed-response items, 253
use in classroom, 287

F

FACET Innovations, 53, 66
Facets of student thinking, 48
Falconer, K., 436
FAST (Foundational Approaches in Science Teaching), 27, 28, 29
F-CAT (Florida Comprehensive Assessment Test), 338
Feedback to improve teacher practice, 427–433
Feedback to students, 8–10, 167
in assessment for learning cycle, 45
from other students, 13
on science notebooks, 125–126, 128–129
on written scientific explanations, 106–107
Fellows, N., 287
“Fishbowl” techniques, in STEP-uP project, 420, 423
5-E Learning Model, use in TAPESTRIES project, 387, 393–395
Classroom Tools for, 394, 402–406
Flexibility, in process of teacher change, 16–17
Florida Comprehensive Assessment Test (F-CAT), 338
Assessing Science Knowledge (ASK) Project, 71–72, 169–190
assessment for learning cycle, 39–43, 45

ASSESSING SCIENCE LEARNING

477
INDEX

classroom-based, 69–72, 168, 233
connecting practice to research, 43–52
developing Web-delivered tools to support assessment for learning cycle, 52–55, 58–66
(See also Diagnoser Tools)
extending practice to virtual community of colleagues, 48–50
implementing strategy vs. adopting practice, 50–52
moving from “misconceptions” to “facets of student thinking,” 48
moving from teacher curiosity to funded research, 46–47
personal style vs. sharable practice, 47–48
reasons for lack of, 46
continuum of techniques for, 23, 23
cost-benefit analysis of, 5–6
definition of, 22, 44
development in STEP-uP project, 421–423
effect on student achievement, 5–6, 44–45
embedded-in-the-curriculum, 21–22, 23, 24–35, 169 (See also Embedded assessment)
formal and informal, 6, 22–23, 25
to improve teacher practice, 427–433
integrated assessment, 18
to keep learning on track, 13–14
moving beyond “did they get it,” 38–39
on-the-fly, 23, 23–24
perspectives and tools in, 37–56
planned-for-interaction, 23, 24
potential and challenges of, 25–26
science notebooks as tool for, 83–97, 117–140
strategies for, 7–13
activating students as learning resources for one another, 12–13
activating students as owners of their learning, 11–12
clarifying and sharing learning intentions and success criteria, 10–11
engineering effective classroom discussions, questions,
activities, and tasks, 7, 7–8
providing feedback, 8–10
vs. summative assessment, 22, 168
teacher learning communities for implementation of, 14–18
teachers’ skepticism about use of, 26
uses of, 168–169
Formative assessment probes, 72, 203–225
Curriculum Topic Study–developed probes, 204, 205
analysis of, 218–219
as basis for further inquiry into student ideas, 220
deconstruction of, 219–220
development of, 211–214, 212, 213
field testing of, 209
publication of, 214
scaffold for design of, 211, 212
teacher’s reflection on, 222–225
teaching informed by responses to, 220–222, 221
vs. traditional assessment items, 207, 207–211, 208
two-tiered format of, 214
to uncover students ideas, 214–215, 215–218
definition of, 206–207
vs. tasks, 206–207
upfront part of backward design, 206
FOSS (Full Option Science System), 71, 84, 171, 176, 183, 187, 390, 393–394, 409, 410, 414, 417–418
Foundational Approaches in Science Teaching (FAST), 27, 28, 29
Frederiksen, J. R., 10
Full Option Science System (FOSS), 71, 84, 171, 176, 183, 187, 390, 393–394, 409, 410, 414, 417–418
Fulwiler, B. R., 371
Funded research, 46–47
Furtak, E. M., 21
G
Gentiluomo, L., 228, 253
Glynn, S., 118, 126
Gobert, J., 71, 191
Gogos, A., 227, 231
Grades, 26, 168
related to rubrics for evaluating lab reports, 155, 156–157
Gradualism, in process of teacher change, 16
Graphic assessments on high-stakes tests, 287–288, 288
use in classroom, 289
Graphing skills, collaborative inquiry for improvement of, 456–459
Grogan, M., 228, 283
Guided inquiry, 160–161

Haertel, G., 71, 191
Handelsman, J., 429
Haney-Lampe professional development model, 387, 388
Harlen, W., 127
Harrison, C., 14
Hickman, P., 338, 427
Hiebert, J., 427
Higher-order thinking skills, 283, 304
High-stakes tests, 227–229
aligning classroom assessment to,
292–299, 294–298
for English language learners,
348–349
extended constructed-response items on,
253–271
extrapolations from reading to science,
317–335
influence of, 321–324
linked to content standards, 231–251
No Child Left Behind Act and mandated tests, 232–233, 301,
302–304, 322
state science assessments, 301–315
traditional test items on, 322–323,
322–323
types of assessment items on, 283,
284–292
constructed-response questions, 287,
287
graphic assessments, 287–289, 288
multiple-choice questions, 284–286,
285, 286
performance assessments, 289–292,
290–291
Washington Assessment of Student Learning, 338, 357–376
Hill, H. C., 5
Ho, P., 118
Horizon Research, 397, 398
Horwitz, P., 71, 191
Hubble, E. P., 145

Hunt, E., 48, 53

Imbalances in assessment, 317–318,
324–335
assessment done to or for students vs.
assessment done with and by students, 332–333
assessment of knowledge, skills, and strategies vs. how students use
this knowledge, 328–330
cognitive and affective learning outcomes and characteristics,
330–331
demands for teacher/school accountability vs. professional development
to develop expertise in assessment, 334
formative and summative assessments,
327–328
meeting needs of different audiences and purposes of assessment,
324–326, 325
process and product assessments,
326–327
Imperial Valley, California, Mathematics–Science Partnership, 139
Inquiry-based science
creating explanation assessment tasks for, 109, 112
developing students’ competence in, 118
for English language learners, 342–345,
344, 345
essential features of activities for, 147
guided vs. open inquiry, 160–161
impact of formative assessment on,
21–35
laboratory investigations for, 71,
145–164
scaffolding student initiative and responsibility in, 344–345, 345
scaffolding writing for, 91–96, 92–96,
119–121, 120
in TAPESTRIES project, 394
Insights kits, 409, 410
Instructional planning principles,
117–118
Integrated assessment, 18
International assessments, 253–254
constructed-response items on,
PISA, 258–263, 259, 260, 262
development of, 254
importance to teachers, 254
Interpretation
  in assessment for learning cycle, 39–40, 41–42
  in assessment triangle, 162, 162, 172, 175
Inverness Research Associates, 365
Iowa Test of Basic Skills (ITBS), 10
Isola, D., 338, 427, 435
ITBS (Iowa Test of Basic Skills), 10
J
Jenkins, A., 5
Jepsen, C., 4
Jones, J., 69–70, 73
Justified multiple-choice questions, 285–286, 286
K
Katz, A., 129
Keeley, P. D., 72, 203, 413
Keeping Score for All, 304
Keller, T. E., 228, 301
Kentucky Core Content Test, 287, 287, 293, 294
Klentschyi, M., 70, 117, 119, 121, 128
Kluger, A. N., 9
Knowing What Students Know: The Science and Design of Educational Assessment, 171
Kouba, V. L., 228, 253
Krajcik, J. S., 70, 101
Kraus, P., 2, 37
Kuerbis, P. J., 338, 409
KWL chart, 393–394, 395
L
Laboratory investigations, 71, 145–164
  aspects that promote student learning, 146
  assessing student performance in, 152–157
  grading, 155, 156–157
  proficiency criteria for rubrics, 155–157, 157–159
  in relation to goals of task, 154–155
  rubrics, 152, 153–154, 164
  student rubrics for self-evaluation, 155
  assessment triangle for, 162, 162–163
  components of assessment of, 145
  to create student scientists, 159–162
  definition of, 150
  “density” labs, 148–150, 149
  design principles for, 151–152
  disparities in access to, 152
  essential features of inquiry activities, 147
  example of inquiry-based lab activity, 147–148
  goals of, 146, 150–151
  evidence that students have accomplished, 146
  improvement of, 146–152, 163–164
  isolation of typical labs from flow of science teaching, 151
  logical placement in curriculum, 152
  NRC report on, 150–151
Learning goals. See also Standards in assessment for learning cycle, 40, 41
  assessment linked to, 231–251
  Bloom’s taxonomy of educational objectives, 328
  clarification of, 235–238
  in Diagnoser Tools, 58–59
  formative assessment probes and, 209–210
  laboratory investigations and, 154–155
  necessity and sufficiency criteria for, 238
Learning performances, 109–110, 110, 308
Learning progressions, 306–308, 307
Learning-for-learning’s sake, 328
Lee, C., 14
Lee, O., 337, 341
LeRoy, K., 337, 341
Lesson plans, evaluation of, 398–399
Levacic, R., 5
Li, M., 84, 119, 121, 140
Long, K., 71, 167
Love, N., 339, 447, 452
Lucks, M. A., 28
M
Maclsaac, D., 436
Making Sense of Secondary Science: Research Into Children’s Ideas, 413
INDEX

Malhotra, B. A., 159
Malone, L., 71, 167
Marshall, B., 14
Maryland Student Assessment, 323
Marzano, R., 125, 130, 136
Massachusetts Comprehensive Assessment System (MCAS), 284–285, 285, 288, 288
Mass–volume ratios, 148–149, 149
MCAS (Massachusetts Comprehensive Assessment System), 284–285, 285, 288
McNeill, K. L., 70, 101
McTighe, J., 130, 136, 292, 419
Metacognitive approaches to instruction, 118, 127
Michiels, A., 227, 231
Minstrell, J., 2, 37
Minstrell, J. E., 2, 37
Model of Student Cognitive Processes, 126
Model-It, 193
Molina De La Torre, E., 121, 128
Mooney, L. B., 338, 409
Motivation to Read Profile, 330
Motivations for learning, 330–331
Multiple-choice questions on high-stakes tests, 284–286, 285, 286
justified, 285–286, 286
Mundry, S., 447
Muth, D., 118, 126

N

NAEP (National Assessment of Educational Progress), 228, 253–255
National Academy of Sciences, 150
National Assessment of Educational Progress (NAEP), 228, 253–255, 346
cognitive demands of, 255, 256
definitions of principles, practices, and performances in, 254–255
determining students’ motivation to learn, 331
Reading Framework for, 318, 328
Science Framework for, 254, 255
National Center for Research on Evaluation, Standards, and Student Testing (CRESST), 362, 370
National Mathematics Advisory Panel, 233, 251
National Oceanic and Atmospheric Administration (NOAA), 376
National Science Foundation (NSF), xi–xii, 46, 148, 169, 204, 301, 331, 357–358, 387, 397, 410, 448
National Science Teachers Association (NSTA), xi, 211, 214, 317
“Nation’s Report Card,” 331
NCLB. See No Child Left Behind Act
Nelson-Denny Reading Test, 323
NetLogo, 193
No Child Left Behind Act (NCLB), 3, 22, 192, 232–233, 322, 357
adequate yearly progress formulas of, 303
requirement for multiple measures of student achievement, 303
state science assessment and, 301, 302–304, 309
students with disabilities or limited English proficiency and, 304, 341, 348, 350
NOAA (National Oceanic and Atmospheric Administration), 376
Notebooks. See Science notebooks
NSF (National Science Foundation), xi–xii, 46, 148, 169, 204, 301, 331, 357–358, 387, 397, 410, 448
NSTA (National Science Teachers Association), xi, 211, 214, 317
Observation
INDEX

in assessment triangle, 162, 162, 172, 173
of teachers by peers, 17–18
Ohio Graduation Test, 449
Ohio Proficiency Tests, 388
Olsen, J., 435
Olson, J. K., 129
On-the-fly formative assessment, 23, 23–24
Open inquiry, 160–161
O’Rourke, A., 420
O’Rourke, W., 420

P
Parents, benefits of assessment linked to content standards for, 235
Pasquale, M., 228, 283
Pathfinder Science, 159
Peer observation of teachers, 17–18
Performance assessments, 109–110, 110, 308
on high-stakes tests, 289–290, 290–291
use in classroom, 292
Pickering, D., 125, 130, 136
PISA. See Programme for International Student Assessment
Planned-for-interaction formative assessment, 23, 24
Pollock, J., 125
Pottenger III, F. M., 21
Probes, defined, 206–207. See also Formative assessment probes
Problematic thinking of students, 38–39, 41, 43–44, 203
moving from “misconceptions” to “facets of student thinking,” 48
Procedural knowledge, 126
cognitive demands of, 255, 256
on PISA, 262
on TIMSS, 268
embedded formative assessment of, 27–29
Process assessments, 326
Product assessments, 326–327
Professional development, 4, 14, 52, 204, 337–339
assessment design as model of, 409–425

Q
Quellmalz, E. S., 71, 191

R
Reading assessment, 318–335
INDEX

imbalances in, 317–318, 324–334
traditional methods of, 322–323, 322–323
zone of proximal development and, 319, 319–320
Reading Self-Concept Scale, 330
Reasoning interviews, for English language learners, 347
“Reflective assessment,” 10–11
Reflective Lessons
design of, 29–32, 30, 31
experimental study of, 32, 32–34, 34
Reformed Teaching Observation Protocol (RTOP), 338, 429–433
applications of, 435–436
benefits of, 431
categories of classroom observations in, 430
definition of, 430
development of, 429
purpose of, 429
scoring for, 431
Training Guide for, 430, 437–446
video records of classroom practices, 431–432
Regan, T., 227, 231
Reif, M., 338, 427, 435
Rivkin, S. G., 4
Role of assessment, 168–169
Rowan, B., 5
RTOP. See Reformed Teaching Observation Protocol
Rubrics, 129–137
design of, 130
in STEP-up project, 420, 421, 422
for evaluating lab reports, 152, 153–154, 164
grading, 155, 156–157
proficiency criteria, 155–157, 157–159
relation to goal of task, 154–155
student rubrics for self-evaluation, 155
for evaluating teachers’ lesson plans, 398–398
for evaluating writing of English language learners, 346, 352–355
for poster project, 129, 129–130
for science notebook entries, 122–124, 124, 136–137
for scientific explanation tasks, 103, 112–113, 115–116
for use across science concepts, 130, 131
Ruiz-Primo, M. A., 21, 84, 119, 121, 140
S
Scaffolding writing in inquiry process, 91–96, 92–96, 119–121. See also Science notebooks
current research on, 140
Schank, P., 71, 191
Schematic knowledge
cognitive demands of, 255, 256
on PISA, 262
on TIMSS, 268
embedded formative assessment of, 28
Scholastic kits, 393, 394
Science and Technology for Children (STC), 393, 394, 409, 410, 414, 415–416
Science Curriculum Topic Study: Bridging the Gap Between Standards and Practice, 204, 413
Science for All Americans, 231, 413
Science Inquiry Framework, 343, 344
Science Inquiry Matrix, 345, 345
Science K–10 Grade Level Expectations: A New Level of Specificity (Washington State), 357, 358
Science literacy, 117–118
extended constructed-response items for assessment of, 253–271
Science Matters: Achieving Scientific Literacy, 413
Science notebooks, 70, 83–97, 117–140
in Assessing Science Knowledge (ASK) Project, 179–181, 180
assessment template for, 122, 123
benefits of use as instructional strategy, 118
challenges teachers face in assessing student knowledge and understanding from, 126–129, 138
application of results, 128
interpreting students’ writing and deciphering their thinking, 128
lack of content knowledge, 127–128
making effective use of assessments, 128

ASSESSING SCIENCE LEARNING 483
skills for effective feedback, 128–129
time constraints, 127, 138
using equitable assessment process, 128
current research on use of, 139–140
decision points for use of, 85–96
content and structure of notebook entries, 87–91, 88, 90
purpose of having students write in notebooks, 85–87, 86
scaffolding writing in inquiry process, 91–96, 92–96, 119–121, 120
summary of, 97
effect on student achievement, 119, 140
examples of student entries in, 132–135, 133–135
as formative assessment tool for content and literacy, 85, 119, 121–124, 123, 124
initiating use of, 122
language arts and, 86, 86–87, 118–119
in preparation for Washington Assessment of Student Learning, 368–371
professional development to enhance teachers’ assessment of, 137–139
research bias in study of, 83–84
rubrics for assessment of, 136–137
scoring guide for, 122–124, 124
sense-making writing in, 89–91, 90
teacher feedback on, 125–126, 128–129
typical components of, 84, 84
use in STEP-uP project, 422–423
variety of formats of, 84
Science Teacher Enhancement Project—unifying the Pikes Peak region (STEP-uP), 409–425
aligning with Colorado Science Standards, 413, 420–421
assessment leadership team of, 411
background of, 410–411
conceptual storylines developed by, 412–414, 414–418
assessment storylines parallel to, 422–423
development of formative assessments in, 421–423
field testing materials produced by, 423
guiding principles of assessment work in, 416–419
impact on student achievement, 424
implementation of, 423–424
initial planning for, 411–412
professional development plan of, 411–423
leadership course: phase one, 412–414
leadership course: phase three, 421–423
leadership course: phase two, 416–421
purposes of, 410–411
rubric design in, 420, 421, 422
teacher participants in, 424–425
use of embedded assessments, 410
User’s Guide for, 423
Science Teaching Action Research Project, 132
Scientific explanation. See Written scientific explanations
Scoring rubrics. See Rubrics
Self-assessment by students, 332–333
rubrics for evaluating lab reports, 155
7E instructional model, 150, 151
Shannon, C. E., 161
Shavelson, R., 2, 21, 84, 119, 121, 140
Show, K., 338, 357
Simpson, D., 48
Simulations, 71, 191–201
Calipers simulation-based assessments, 194–201
promise of simulation-based assessments, 201
value and uses of, 193–194
Slavin, R., 12
Songer, N., 118
SRI International, 177, 195
Standards. See also Learning goals
alignment of Calipers simulation-based assessment with, 195, 200
assessment items aligned to, 231–251
in assessment of scientific thinking, 130–131
clarification of, 235–238
conceptual strand maps for, 238, 239
connections among ideas in, 237–238
creating explanation assessment tasks based on, 108, 108–109, 110
integration in STEP-uP project, 413, 420–421
key ideas of, 235–236
National Science Education Standards,
41, 58, 79, 101, 112, 192, 194,
209, 231–232, 235, 257, 258,
358, 393, 410, 413
organizing around big ideas, 305
rubrics to assess science notebook
entries based on, 136–137
state science assessments and, 304–308
learning performances, 308
learning progressions, 306–308, 307
Washington State science standards
and assessment, 358–361, 359,
360
State science assessment, 301–315. See
also High-stakes tests
aligning classroom-based assessment
with, 283–299
developmental approach to, 312–313
of English language learner students,
348–349
NCLB and, 301, 302–304, 309
professional development and teacher
competencies for, 313–315, 314
standards-based, 304–308
learning performances, 308
learning progressions, 306–308, 307
of students with disabilities or limited
English proficiency, 304
system of, 309–312
characteristics of high-quality
system, 310, 311
coherence of, 310
framework for, 310, 312
variations in, 309
Washington Assessment of Student
Learning, 338, 357–376
STC (Science and Technology for
Children), 393, 394, 409, 410, 414,
415–416
Stearns, C., xi
STEP-uP. See Science Teacher
 Enhancement Project-unifying the
Pikes Peak region
Stiles, K. E., 447
Stimpson, V., 48
Strategic knowledge, cognitive demands
of, 255, 256
on PISA, 262
on TIMSS, 268, 269
Struble, J. L., 338, 387
Student achievement
assessment for learning and, 5–6,
44–45, 192
benefits of improvements in, 3
class-size reduction and, 4–5
collaborative inquiry and, 448–450,
449
of English language learners, 347–349,
348
judged by single test scores, 324
NCLB and, 302–304
in science, 3
science notebooks and, 119, 140
STEP-uP project and, 424
student diversity and, 318
TAPESTRIES project and, 398–400
teacher subject knowledge and, 5
Students
assessing scientific thinking of,
130–132
challenges teachers face in assessing
knowledge and understanding
of, 126–129
developing scientific literacy of,
117–118
as learning resources for one another,
12–13
motivations for learning, 330–331
as owners of their own learning, 11–12
principles for maximizing opportunity
to learn, 117–118
problematic thinking of, 38–39, 41,
43–44, 203
moving from “misconceptions” to
“facets of student thinking,” 48
providing feedback to, 8–10
as scientists, 159–162
self-assessment by, 332–333
written scientific explanations by, 70,
101–113
Summative assessment, 22, 26, 29, 168,
169, 170–171, 203–204, 210–211,
233, 292, 327. See also Tests
in Assessing Science Knowledge
(ASK) Project, 183–184
Support for teacher change, 17
Systems for State Science Assessment, 302
TAPESTRIES. See Toledo Area Partnership in Education: Support Teachers as Resources to Improve Elementary Science

Teachable moments, 23–24, 25, 320

Teacher learning communities (TLCs), 14–18

peer observation in, 17–18

principles for establishing and maintaining, 15–18

accountability, 17

choice, 17

flexibility, 16–17

gradualism, 16

support, 17–18

training leaders of, 17

Teachers

adoption of reform teaching techniques by, 428–429

benefits of assessment linked to content standards for, 234

challenges in assessing student knowledge and understanding, 126–129

collaborative inquiry among, 447–462

effect of subject knowledge on student achievement, 5

effective use of Diagnoser Tools by, 53–55

evaluating lesson plans of, 398–399

focusing on quality of teaching by, 427–428

instructional planning principles for, 117–118

professional development of (See Professional development)

providing feedback that moves learning forward, 8–10

skepticism about use of formative assessment, 26

video records of classroom practices of, 431–432

Technology in the classroom, 52–53

simulations, 191–201

Templin, M. A., 338, 387

Test developers and administrators, benefits of assessment linked to content standards for, 234

Testing English-Language Learners in U.S. Schools, 304

Tests. See also Summative assessment vs. formative assessments, 6, 29, 168, 169, 192, 210, 292

high-stakes, 227–229

aligning classroom-based assessment with, 292–299

extended constructed-response items on, 253–271

linked to content standards, 231–251

No Child Left Behind Act, 232–233, 301, 302–304

from reading to science, 317–335

systems for state science assessment, 301–315

types of tasks on, 283, 284–292

The Data Coach's Guide to Improving Learning for All Students: Unleashing the Power of Collaborative Inquiry, 447, 452

ThinkerTools, 10, 194

Thinking Works, 388, 393

Time constraints, 127, 138, 185–186

TIMSS. See Trends in International Mathematics and Science Study

TLCs. See Teacher learning communities

Toledo Area Partnership in Education: Support Teachers as Resources to Improve Elementary Science (TAPESTRIES), 387–400

application phase of, 387, 396–398

“Brainstorming Wheel,” 397

monthly professional development sessions, 396–398, 407

role of Support Teachers, 387–388, 389–390, 396

background of, 388

definition of assessment in, 393

effectiveness with regard to student achievement, 398–400

lesson plans, 398–399

professional development model, 399

embedded assessments in, 393, 394, 395

follow-up phase of, 388

obstacles encountered in, 399

planning phase of, 387, 388–392

project staff retreat, 388–389

school administrators and principals, 392

Support Teachers, 389–390, 391, 392
INDEX

purpose of, 388
schools and universities participating in, 388
training phase of, 387, 393–396
content-based sessions, 393–396
informational sessions, 393
Summer Institutes, 387, 393
use of 5-E Learning Model, 387, 393–395
Classroom Tools for, 394, 402–406
Tomita, M. K., 21
Trends in International Mathematics and Science Study (TIMSS), 4, 228, 253–254, 257, 346, 428, 432
adapting test items for classroom use, 296–298, 296–299
cognitive demands of, 267–269, 268
reading and writing knowledge and skills required for, 264–267, 266
science knowledge and practices assessed by, 267
Trust, 186–187
“Two stars and a wish” format, 13

U
Uncovering Student Ideas in Science, 211
Understanding by Design, 151
Using Data Project, 448–451. See also Collaborative inquiry among teachers

V
Validity of test items, 240
Valle Imperial Project in Science, 121
van Zee, E., 48
Video records of teaching practices, 431–432
Vignoles, A., 5
Virtual performance assessments, on high-stakes tests, 289–290, 290–291
Vygotsky, L., 319

W
Washington Assessment of Student Learning (WASL), 338, 357–376
background of, 357–358
fifth graders’ achievement on science test of, 361–363, 362
professional development to support teachers in preparing students for, 363–366, 364
Expository Writing and Science Notebooks classes and Supplementary Writing Curriculum, 368–371, 371
Initial Use classes and Supplementary Curriculum Guides, 366–368, 368
Prep classes, 372–375, 373, 374
science content courses, 375–376
sample Scenario of, 378–385
state science standards and, 358–361
Essential Academic Learning Requirements, 358, 360
Grade Level Expectations, 359–361, 360
science symbol, 358, 359
Systems, Inquiry, and Application Scenarios, 361
Weaver, W., 161
WestEd, 195
White, B. Y., 10
White, M. A., 10
Wiggins, G., 419
William, D., 1, 10, 14, 21, 45, 46, 169
Wilson, M. R., 228, 301
Wilson, P., 227, 231
Woo, E., 338, 357
Writing in Science: How to Scaffold Instruction to Support Learning, 371
Written scientific explanations, 70, 101–113
assessment of, 103–106
element of strong explanation, 104, 104–105
element of weaker explanation, 105, 105–106
sample task for, 103, 104
scoring rubrics for, 103, 112–113, 115–116
benefits for student learning, 102
components of, 102–103
claim, 102–103
evidence, 103
reasoning, 103
creating explanation tasks for, 108–112
<table>
<thead>
<tr>
<th>Step 2</th>
<th>Unpacking scientific inquiry practice, 109</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 3</td>
<td>Creating learning performances, 109–110, 110</td>
</tr>
<tr>
<td>Step 4</td>
<td>Writing assessment task, 110–111, 111</td>
</tr>
<tr>
<td>Step 5</td>
<td>Reviewing assessment task, 111–112</td>
</tr>
<tr>
<td>Step 6</td>
<td>Developing specific rubrics, 112, 116</td>
</tr>
</tbody>
</table>

- Importance of, 101–102
- Instructional framework for, 102–103
- Providing feedback on, 106–107

**Y**
- Yin, Y., 21
- Young, D. B., 21

**Z**
- Zone of proximal development, 319, 319–320