

CHAPTER 15

Systems for State Science Assessment: Findings of the National Research Council's Committee on Test Design for K–12 Science Achievement

Meryl W. Bertenthal
Indiana University

Mark R. Wilson
University of California, Berkeley

Alexandra Beatty
National Research Council

Thomas E. Keller
National Academy of Sciences¹

Shortly after the passage of the No Child Left Behind Act of 2000 (NCLB), the National Science Foundation asked the National Research Council (NRC) to form a committee to provide advice and guidance to states and their test development contractors on the design and implementation of quality science assessments to meet the requirements of the law. This chapter summarizes some of the findings and

¹Meryl Bertenthal served as the study director and Mark Wilson was the chair of the committee that authored the report on which this chapter is based. Alexandra Beatty served as a program officer to the committee. Thomas Keller served as the chair of the state science supervisors' panel that advised the committee.

CHAPTER

15

SECTION 3: HIGH-STAKES ASSESSMENT

recommendations of the NRC Committee on Test Design for K–12 Science Achievement and is based on the committee’s final, book-length report, *Systems for State Science Assessment* (NRC 2006).² While the committee’s report is targeted at state policy and education decision-makers, many of the findings are of interest to teachers and others involved in the improvement of science education and assessment in their states. In this chapter we try to connect a few of the committee’s findings and recommendations to issues of interest to teachers.

There are inherent difficulties in summarizing a whole book in a single chapter. One such difficulty is providing the kind of specific detail and exemplification that the reader might be looking for. Readers who seek additional examples or who are interested in learning more about the suggestions and approaches mentioned here are encouraged to consult the committee’s full report, which is available from the National Academies Press.³

Many of the points made in the NRC report and in this chapter may apply to educational assessment in other areas. In particular, the principles that framed the committee’s thinking about science assessment could guide assessment in other subject areas as well. However, there are aspects of science as a discipline—for example, the abstract nature of many of the concepts that students are expected to learn and the emphasis on scientific inquiry and investigation in many state standards—that present challenges specific to science assessment. Thus, in designing high-quality science assessment, states will need to focus on both the general precepts of sound educational measurement and the features that are unique to science assessment.

No Child Left Behind and Science Assessment

NCLB is an extension of the 1994 reauthorization of the Elementary and Secondary Education Act (ESEA). (ESEA moves beyond that law because it affects all public schools, districts, and students and because it includes science in its requirements for standards and assessments.)

NCLB rests on four pillars: (1) a set of challenging content and achievement standards for student performance; (2) assessments to measure the

²We are grateful to the committee members and to the advisory groups of science teachers, science education supervisors, and state assessment directors that assisted the committee for their contributions to the report. We acknowledge and thank them for their intellectual contribution to this chapter.

³The report can be viewed online at <http://books.nap.edu/catalog/11312.html>

achievement of all students relative to the state standards; (3) flexibility for states and districts to implement instruction and curriculum so that all students can achieve the standards; and (4) accountability measures that move schools toward greater effectiveness in promoting achievement. In this chapter, as in the committee's report, the focus is on the first two pillars: standards and assessment.

Under NCLB, states are required to raise the reading, mathematics, and science achievement level of all students and to narrow the achievement gap in these subjects among students of different backgrounds. To attain these goals, states are required to have a set of challenging standards, expand their standardized testing programs, analyze and report assessment results in specific ways for accountability purposes, and ensure that all students have the opportunity both to be taught by qualified teachers and to learn in well-performing schools.

NCLB puts in place strong accountability measures for schools and districts and imposes sanctions on schools when their students do not make adequate yearly progress (AYP). So that states work diligently toward these goals, there is a rigorous timetable for state compliance with NCLB regulations. Currently, science is not part of AYP formulas, but NCLB requires that results from science assessments be reported publicly along with other information on student achievement.⁴ This means that science assessment under NCLB will still carry high stakes for schools and school districts. Individuals and schools will get their own results, and they, as well as other stakeholders, will be able to compare their results with those of other schools. It was the consensus of the NRC committee that teachers will tailor their instruction to the assessments and, therefore, the quality of the assessments will be of paramount importance.

Under NCLB, states must have challenging academic content and achievement standards for science in place by 2005–2006. States must begin measuring student attainment of their standards in 2007–2008 through assessments that are fully aligned with the standards. The assessments must meet accepted professional standards for technical quality for the purposes

⁴As this chapter was being completed, several bills were being introduced in Congress to require the inclusion of science assessment results in AYP formulas. The committee that authored the report on which this paper is based suggested that such a step should not be undertaken without careful consideration of both the positive and negative implications of including or excluding science from AYP formulas.

CHAPTER

15

SECTION 3: HIGH-STAKES ASSESSMENT

for which they will be used. The committee found that meeting the technical quality and alignment provisions of the law could present serious challenges for states. For example, the committee found that some states may list so many standards that tests cannot measure them all. Further, most state assessments fail to measure adequately the cognitive complexity or “depth of knowledge” described in state standards.

NCLB also specifies that states’ assessment systems must incorporate multiple up-to-date measures of student achievement, including measures that assess higher-order thinking skills and understanding of challenging content. Science assessments are to be administered annually to all students, including those with disabilities and those who are not fluent in English, at least once in each of three grade bands, 3–5, 6–9, and 10–12. States are required to make reasonable accommodations for students with disabilities and limited English proficiency to allow them to participate in the assessments, and they must have in place alternate assessments for students who cannot participate in the regular assessment even with accommodations. [For more information about issues related to assessing students with disabilities or English language learners see, for example, *Keeping Score for All* (NRC 2004); *Testing English-Language Learners in U.S. Schools* (NRC 2000); and pages 136–141 of the committee’s report (NRC 2006).]

Under NCLB, states’ assessment systems may include many different types of assessment strategies and be comprised of a uniform set of assessments statewide or a combination of state and local assessments. Regardless of the form that a state’s assessment system takes, the results must be reported publicly and be expressed in terms of its academic achievement standards. The results must be reported in the aggregate for the full group of test takers, be disaggregated for specified population groups, and provide information that is descriptive, interpretive, and diagnostic at the individual level.

Standards

Content standards are considered fundamental to science education and science assessment because they define what is to be taught and what kind of performance is expected. For standards to serve this function well, they must communicate clearly to everyone concerned—teachers, assessment developers, students, parents, and policy makers—what students are expected to know and be able to do. The NRC committee found that standards will communicate best if they are clear, detailed, and complete; reasonable in scope;

rigorous and scientifically correct; and built around a conceptual framework that reflects sound models of student learning (NRC 2006, Chapter 4).

The committee also found that many state content standards and, consequently, the curricula aligned to them contain too many disconnected topics. They found that there is not enough attention paid to how a student's understanding of a topic can be supported and enhanced from grade to grade or across subject disciplines. As a result, topics often appear to receive repeated coverage over multiple years with few opportunities for students to see connections between a topic under study and other related topics, thus giving students an incomplete foundation for further knowledge development.

In general, good state standards should be organized and elaborated in ways that clearly specify what students need to know and be able to do and how their knowledge and skills will develop over time with competent instruction. Science standards should be structured around central principles of science that represent the foundation for the concepts, theories, principles, and explanatory schemas for phenomena in a discipline. These central principles, which are sometimes referred to as the “big ideas” or the enduring understandings (see Wiggins and McTighe 1998) of science, are the foundation for the principles, theories, and explanatory schemas within a discipline. While there is no universally accepted list of “big ideas,” and states may have to decide on their individual foci, possible examples might be evolution, Newton's laws, or kinetic molecular theory.

Organizing standards around big ideas is a fundamental shift from the organizational structure that many states use in which standards are grouped under subject areas or discrete topics. The committee found that a potential positive outcome of the reorganization of state standards from discrete topics to big ideas is the opportunity for a shift in emphasis from breadth of coverage to depth of coverage around a relatively small set of foundational principles and concepts. These foundational principles and concepts become the focused target of instruction and can be progressively refined, elaborated, and extended over time (NRC 2006, p. 3).

The committee identified two ideas that emerge from the science education research literature—learning progressions and learning performances—that would be useful for states to adapt in organizing, elaborating, and critiquing their standards in order to make them better able to guide curriculum, instruction, and assessment. These two ideas would also be useful

CHAPTER

15

SECTION 3: HIGH-STAKES ASSESSMENT

for teachers, curriculum designers, and assessment developers because they provide the specificity and organizational framework that is missing in most standards documents but that is needed for coherence among curriculum, instruction, and assessment.

Learning Progressions

Learning progressions, which can be developed for lessons, units of study, yearlong courses, or an entire K–12 experience, describe in words and examples what it means to develop greater understanding of an idea or set of ideas. They are anchored on one end by students' prior knowledge and on the other by the expectations for learning over time with instruction. Learning progressions propose the intermediate understandings or benchmarks between anchor points that contribute to building a more sophisticated understanding and serve as targets for curriculum, instruction, and assessment.

The committee emphasizes that learning progressions are not developmentally inevitable; more than one path leads to competence. The pathways that individual students or groups of students follow depend on many things, including the knowledge and experience that they bring to the task, the quality and content of the instruction that supports their learning, and the nature of the specific tasks that are part of the experience. Nonetheless, some paths are followed more often than others. Organizing standards to reflect the more typical ways in which deepening understanding develops provides a structure around which states, curriculum developers, and teachers can organize learning. Also, this organization can provide clues about the types of assessment tasks that can be used to shed light on students' achievement at different points along the progression.

Ideally, learning progressions should be based on research about how competence develops in a particular domain. However, for many aspects of science learning, the research literature is incomplete and research findings may have to be supplemented with the judgments of expert teachers and others with knowledge of how students learn science and what is known more generally about cognition and science learning.

What might a learning progression look like? Figure 15.1 shows a learning progression for atomic molecular theory that was developed as an example for the NRC committee (Smith et al. 2006).

The learning progression is organized around two sets of big ideas. The first set (M1, M2) consists of big ideas about (a) the properties of matter and material kinds, (b) their constancies and changes across transformations, and (c) the role of measurement, modeling, and argument in knowing. This first set of big ideas can be introduced in the earliest grades and elaborated throughout schooling at the macroscopic level. For example, students move from describing the properties of material kinds to learning about density as the ratio of weight over volume, and from conservation of material kind and weight during melting to conservation of mass across all phase changes. In middle school and high school, the atomic-molecular theory can be introduced as a second set of big ideas. This second set of ideas (AM1, AM2, AM3, AM4) builds on the first because understanding the atomic-molecular theory depends to a great extent on the macroscopic big ideas studied earlier (e.g., the understanding that matter has weight and occupies space) and, at the same time, it provides deeper explana-

Figure 15.1 Learning Progression for Atomic Molecular Theory

Six Big Ideas of Atomic Molecular Theory That Form Two Major Clusters

M1. Macroscopic Properties: We can learn about the objects and materials that constitute the world through measurement, classification, and description according to their properties.

M2. Macroscopic Conservation: Matter can be transformed, but not created or destroyed, through physical and chemical processes.

AM1. Atomic-molecular theory: All matter that we encounter on Earth is made of less than 100 kinds of atoms, which are commonly bonded together in molecules and networks.

AM2. Atomic-molecular explanation of materials: The properties of materials are determined by the nature, arrangement, and motion of the atoms and molecules of which they are made.

AM3. Atomic-molecular explanation of transformations: Changes in matter involve both changes and underlying continuities in atoms and molecules.

AM4. Distinguishing data from atomic-molecular explanations: The properties of and changes in atoms and molecules have to be distinguished from the macroscopic properties and phenomena they account for.

CHAPTER

15

SECTION 3: HIGH-STAKES ASSESSMENT

tory accounts of macroscopic properties and phenomena. They also enable the further elaboration of macroscopic understandings. (p. 12)

Learning Performances

The committee found that most state science standards describe science content knowledge—and the general cognitive demands relative to that knowledge—without providing operational definitions of what it means to know or understand. Therefore, standards must be elaborated before they can serve as a basis for instruction or assessment. If they are not elaborated clearly, teachers and assessment developers can infer from almost any standard an array of meanings that draw on many different combinations of content knowledge and cognitive demand.

Learning performances are a way to reformulate a scientific content standard in terms of the scientific skills and practices that *use* that content, such as being able to define terms, describe phenomena, use models to explain patterns in data, construct scientific explanations, or test hypotheses (Perkins 1998; Reiser et al. 2003). By defining the most important skills and practices for which the knowledge is used, it is possible to connect the conceptual statements in the content standards with assessment performances in which students can demonstrate their understanding.

Articulated learning performances act as guides for designing assessments and can serve as the link between classroom and large-scale assessments by defining what students who meet the standard would be capable of doing with their knowledge. Assessment tasks could then be developed specifically to elicit evidence that students can use their knowledge as described in the standard. Learning performances can be targeted broadly by state and district tests and at a more fine-grained level in classroom assessments.

Assessment

Assessment is a systematic process for gathering information about student achievement. It provides important information for many different purposes that are important to the education system, including guiding instructional decision-making in the classroom, holding schools accountable for student achievement, and monitoring and evaluating educational programs. It is also the way that teachers, school administrators, and state policy-makers exemplify their goals for student learning. While assessment can do all of these things, it must be designed specifically

to serve the particular purpose or purposes for which the results will be used. For example, an assessment that is designed to provide information about student achievement that is diagnostic, descriptive, and interpretive would need to test students' understanding of a few key concepts deeply and thoroughly. On the other hand, an assessment that is used to inform policy-makers about the effectiveness of the overall education system would need to cover broadly all of the topics deemed important by decision-makers. Neither of these strategies could provide results that are valid for the purposes of the other (NRC 2006, Chapter 2).

A System of Assessment

The committee concluded that a single assessment strategy would not, by itself, meet the requirements of NCLB as it could not provide results that would be valid and reliable for all of the purposes identified under the law. The committee therefore recommended that states develop a system of science assessment that collectively would meet the various purposes of NCLB and provide education decision makers with assessment-based information that is appropriate for each specific purpose for which it will be used. The system that each state develops in response to NCLB necessarily will vary according to the state's goals and priorities for science education and its uses for assessment information. For example, a state might choose to develop a single test in which students take a common core assessment that provides individual results, along with an assessment with a matrix-sampling design that provides information about the achievement of groups of students across a broad content domain. Or a state might choose to combine standardized classroom assessments that provide diagnostic, descriptive, and interpretive information with an external assessment that shows the progress that all students are making toward achieving state standards or for program evaluation purposes. Or a state may decide to give up a statewide test and instead use one of many different models in which results from local, district, or state assessments are combined, aggregated, and reported for specific purposes.⁵ When multiple assessment strategies are used, they should be designed from the beginning to function as part of a coherent system of assessment.

⁵Examples of these are summarized in the committee's report (pp. 31–37) and are illustrated in detail in papers commissioned by the committee. These papers are available online at: www7.nationalacademies.org/bota/Test_Design_K-12_Science.html

CHAPTER

15 ● SECTION 3: HIGH-STAKES ASSESSMENT

A coherent system of standards-based science assessment is horizontally coherent: curriculum, instruction, and assessment are all aligned with the standards; target the same goals for learning; and work together to support a student's developing science knowledge, understandings, and skills. It is vertically coherent: All levels of the education system—classroom, school, school district, and state—are based on a shared vision of the goals for science education, of the purposes and uses of assessment, and of what constitutes competent performance. The system is also developmentally coherent: It takes into account how students' science understanding develops over time and the scientific content knowledge, abilities, and understanding that are needed for learning to progress at each stage of the process (Wilson 2005).

If a collection of tests is not coherent, the information that is produced can yield conflicting or incomplete information and send confusing messages about student achievement that are difficult to untangle. If discrepancies in achievement are evident, it is hard to determine whether the tests in question are measuring different aspects of student achievement—and are useful as different indicators of student learning—or whether the discrepancy is an artifact of assessment procedures that were not designed to work coherently together. Gaps in the information provided by the assessment system can lead to inaccurate assumptions about the quality of student learning or the effectiveness of schools and teachers and can bring about interventions that may not be necessary. In Figure 15.2 we list some of the characteristics of a high-quality, coherent assessment system. But what would a coherent system look like? Table 15.1, page 312, is a framework for an assessment system adapted from the one developed by the state of Maine in 2003 (NRC 2003). We include it here to illustrate how multiple assessments can be incorporated coherently into a single assessment system.⁶

⁶The system has since been modified and so the description included here does not describe the current (2007) Maine assessment system.

Figure 15.2 Characteristics of a High-Quality Science Assessment System

The following are characteristics of an assessment system that could provide valid and reliable information to the multiple levels of the education system and could support the ongoing development of students' science understanding:

1. incorporates assessments that are closely aligned to the standards that guide the system and is structured so that all elements are coherent with the goals, curriculum materials, and instructional strategies of the science education system of which it is a part;
2. includes a range of measurement approaches and multiple measures of achievement that provide a variety of evidence to support educational decision-making at different levels of the system;
3. contains measures that assess student progress over time rather than relying solely on one-time, large-scale testing opportunities;
4. is useful in the sense that the assessment results are made accessible and are reported in a timely manner to those who need them;
5. fits into a larger education system that provides the necessary resources for the development, operation, and continued improvement of both the assessment system and the education system when assessment results indicate improvement is necessary;
6. provides systematic, ongoing professional development for teachers and others on current science assessment practices, the uses and limitations of assessment results, and processes for developing and using sound assessments; and
7. is systemically valid—that is, it promotes in the education system desired curricular and instructional changes that result in increased learning and not just improvement in test scores.

Source: National Research Council (NRC). 2006. *Assessment in support of instruction and learning: Bridging the gap between large-scale and classroom assessment*. Washington, DC: National Academy Press, p. 28.

CHAPTER

15

SECTION 3: HIGH-STAKES ASSESSMENT

Table 15.1 Framework for an Assessment System

	Primary purpose of the assessment	Who selects or develops the assessment?	Who scores the assessment?
Classroom Assessment	Informing teaching and learning	Individual teacher	Individual teacher
School or District Assessment	Informing and monitoring	Groups of teachers and administrators	Groups of teachers (and others)
State Assessment	Monitoring and evaluating programs to ensure accountability	Groups of administrators, administrators, and/or policy-makers	Scorers outside the district
Assessment System	Informing teaching, monitoring and evaluating, certification	District assessment leadership	Both internal and external

The assessment system in Table 15.1 was designed to meet three principal goals: (1) provide high-quality information about student performance to inform teaching and learning, (2) monitor schools and administrative units and hold them accountable for their success at making sure students meet the state standards, and (3) certify that students have met the state standards.

A Developmental Approach to Assessment

In keeping with the committee's conclusion that science education and assessment should be based on a foundation of how a student's understanding of science develops over time with competent instruction, the committee called for a developmental approach to science assessment (see Masters and Forster 1996, 1997). A developmental approach means gathering evidence of the development of students' understanding over time, as opposed

to only at a specific point in time. This approach recognizes that science learning is not simply a process of acquiring more knowledge and skills, but rather a process of progressing toward greater levels of competence and understanding as new knowledge is linked to existing knowledge and as new understandings build on and replace earlier conceptions.

A hallmark of a developmental approach is the use of multiple assessments that collectively provide information about students' learning over time. Students' performances on these multiple assessments are compared against a pre-constructed progress map or learning progression that describes what students are expected to learn and how that learning is expected to unfold as the student progresses through the instructional material. Records are maintained about performance on each assessment and are combined and viewed as a whole to estimate students' level of achievement in a particular area of learning. The estimate of a student's current location on the progress map or learning progression serves as a guide to the kinds of learning experiences that would help them develop the knowledge, understanding, and skills necessary to progress.

Professional Development and Assessment

Regardless of the information it is designed to collect, assessment, on its own, cannot improve student learning—it is the way in which assessment and the results are used that can accomplish that goal. For an assessment process to function as it should, everyone who uses assessment results needs to be assessment literate. That is, they need to have a clear understanding of the purposes and limitations of assessment, goals and purposes for assessments, the ways in which different assessments function, and how to interpret and use assessment results appropriately. Those who need to develop their assessment literacy includes everyone from elected officials at the highest levels to school board members to parents to students to teachers. The teachers need to develop their knowledge of assessment the most because they are in the best position to use the results to improve student learning.

Although teachers may not be involved in designing their state's tests, they should understand the kinds of data that are produced by these tests and how the results can and will be used. Further they should have sufficient understanding of the technical properties of the assessments to put the results in context and to link them to other pieces of information they have about their students.

CHAPTER

15

SECTION 3: HIGH-STAKES ASSESSMENT

Assessment activities require that teachers have, in addition to a deep understanding of the content domains they are teaching, knowledge about how to develop tasks that are valid and useful in the classroom and how to use the results of these assessments in instructionally supportive ways. They must also understand the uses and limitations of external assessment, such as large-scale statewide tests, and be cognizant of the ways in which such assessment affects their teaching. Figure 15.3 is a list of some of the things that assessment-literate teachers should know and be able to do.

Figure 15.3 Assessment Competencies for Teachers

Teachers should be able to

1. choose assessment methods appropriate for instructional decisions,
2. develop assessment methods appropriate for instructional decisions,
3. administer, score, and interpret the results of both externally produced and teacher-produced assessment methods,
4. use assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement,
5. develop valid pupil grading procedures that use pupil assessments,
6. communicate assessment results to students, parents, other lay audiences, and other educators, and
7. recognize unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

Source: American Federation of Teachers, National Council on Measurement in Education, and National Education Association. Standards for teacher competence in educational assessment of students. Available at www.lib.muohio.edu/edpsych/stevens_stand.pdf

The committee found that inservice teachers need to have the opportunity, and take the responsibility, to develop their assessment literacy. Taking classes, participating in professional development opportunities, and

keeping up with issues surrounding assessment are some of the ways that teachers can develop their own assessment literacy. By organizing and participating in opportunities to discuss student work that results from assessment, teachers can develop a common understanding of what competence in a domain looks like and how to support student progress toward achieving it.

Conclusions

Designing high-quality science assessment is an important but difficult goal to achieve. Science assessment must target the knowledge, skills, and habits of mind that are necessary for science literacy, and it must reflect current scientific knowledge and understanding in ways that are scientifically accurate and consistent with the ways in which scientists understand the world. It must assess understanding of science as a content domain and an understanding of science as a specific way of thinking. It must also provide evidence that students can apply their knowledge appropriately and that they are building on their existing knowledge and skills in ways that will lead to more complete understanding of the key principles and big ideas of science. Adding to the challenge, competence in science is multifaceted and does not follow a singular path. Science assessment must address these complexities while also meeting professional technical standards for reliability, validity, and fairness for the purposes for which the results will be used.

The committee concluded that the goal of developing high-quality science assessment will only be achieved through the combined efforts of scientists, science educators, developmental and cognitive psychologists, experts on learning, law and policy makers, and educational measurement specialists working collaboratively rather than separately.

References

- American Association for the Advancement of Science (AAAS). 1993. *Benchmarks for science literacy*. New York: Oxford University Press.
- American Education Research Association (AERA). 2003. *Research points: Standards and tests, keeping them aligned*, L. B. Resnick, ed. Spring (1,1). Washington, DC: AERA.
- Masters, G., and M. Forster. (1996, 1997) *Developmental assessment: Assessment resource kit*. Hawthorne, Australia: Australian Council on Educational Research Press.
- National Research Council (NRC). 1996. *National science education standards*. National Committee on Science Education Standards and Assessment. Center for Science,

CHAPTER

15

SECTION 3: HIGH-STAKES ASSESSMENT

- Mathematics, and Engineering Education. Washington, DC: National Academy Press.
- National Research Council (NRC). 2000. *Testing English-language learners: Report and workshop summary*. Board on Testing and Assessment, Committee on Educational Excellence and Equity. Washington, DC: National Academy Press.
- National Research Council (NRC). 2003. *Assessment in support of instruction and learning: Bridging the gap between large-scale and classroom assessment*. Board on Testing and Assessment, Committee on Science Education K–12, and the Mathematical Sciences Education Board. Washington DC: National Academy Press.
- National Research Council (NRC). 2004. *Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessments*. Committee on Participation of English Language Learners and Students with Disabilities in NAEP and Other Large-Scale Assessments. Washington, DC: National Academies Press.
- National Research Council (NRC). 2006. *Systems for state science assessment*. Committee on Test Design for K–12 Science Achievement. M. R. Wilson and M. W. Bertenthal, eds. Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.
- Perkins, D. 1998. What is understanding? In M. S. Wiske (Ed.), *Teaching for understanding: Linking research with practice*. San Francisco: Jossey-Bass.
- Reiser, B. J., J. Krajcik, E. Moje, and R. Marx. 2003. Design strategies for developing science instructional materials. Paper presented at the National Association for Research in Science Teaching meeting, Philadelphia (March).
- Smith, C., M. Wiser, A. Anderson, and J. Krajcik. 2006. (Focus Article of combined double issue of journal): Implications of research on children’s learning for standards and assessment: A proposed learning progression for matter and atomic-molecular theory. *Measurement* 14 (1&2): 1–98.
- Wiggins, G. P., and J. McTighe. 1998. *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Wilson, M. 2005. *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum.