Using National Standards to Improve K–8 Science Curriculum Materials

Jo Ellen Roseman
Mary Koppal
American Association for the Advancement of Science

The Elementary School Journal
Volume 109, Number 2

0013-5984/2008/10902-0002$10.00

Abstract

With increased emphasis on state standards and assessments in science, one might question whether there is still a role for national science standards and benchmarks. This article makes a case for the importance of national guidelines and identifies key functions performed by the National Research Council's *National Science Education Standards* and the American Association for the Advancement of Science's *Benchmarks for Science Literacy*. We describe how national guidelines provide a vision of what science education could be and its value to all students; consider the implications of aligning curriculum, instruction, and assessment to these guidelines; discuss the quality of current state standards; and consider what the national reform efforts have to offer states and school districts seeking to improve teaching and learning in K–8 science.

Based on what is known about national and international measures of academic achievement, the majority of U.S. students are destined to graduate from high school without even a basic understanding of core concepts and skills in science. Nor is it likely that most students will meet their states' standards for science achievement as measured by statewide assessments mandated under the federal No Child Left Behind (NCLB) legislation.

Although NCLB requires that the tests be aligned to each state's content standards, there are serious concerns about the quality of the science standards used to select curriculum materials and develop assessment in many states. Furthermore, the lack of uniformity in content and quality across 50 sets of state science standards has resulted in curriculum frameworks and textbooks that are unfocused and ineffective in supporting student learning (Schmidt, McKnight, & Raizen, 1997). Given these concerns, what contribution can the national standards make? Do the recommendations in *Benchmarks for Science Literacy* (American Association for the Advancement of Science [AAAS], 1993) and in the *National Science Education Standards* (National Research Council [NRC], 1996) still matter in the era of NCLB?

In this article we argue that national standards and benchmarks not only matter but, indeed, are essential to the long-term improvement of U.S. science education and to the students it serves. By performing several critical functions—providing a vision of what science education could be and building consensus for that vision, modeling what high-quality standards should look like, and guiding the development of a new generation of curriculum materials and assessments that can help all students achieve those standards—the national reform efforts have much to offer states, districts, and schools seeking to improve their science education. We briefly describe those critical functions, consider the condition of current state standards, and then focus on the unique role that national standards can play in shaping a new generation of science curriculum materials designed to give all K–8 students a solid foundation for future science learning.

Critical Functions of National Standards

Science content standards developed at any level should provide a guiding vision for K–12 science education built on a broad consensus of the scientific and education communities and the public at large. Standards should set goals for learning for all students at specific grade levels while at the same time recognizing that the current education system may not provide adequate support for all students to reach those goals. The learning goals articulated in standards should also provide educators with a framework for designing curriculum, instruction, and assessment that are aligned to those standards and to each other. In the following sections, we consider how the national standards documents fulfill these functions.

A Common Core

One of the earliest visions of what every high school graduate should know and be able to do in science, mathematics, and technology was laid out by Project 2061 of the American Association for the Advancement of Science (AAAS) in its publication *Science for All Americans* (AAAS, 1989). Representing the work of scientists, mathematicians, engineers, and technologists and reviewed by more than a thousand educators and others, *Science for All Americans* identifies the knowledge and skills that constitute adult literacy in four domains: biological and health sciences, mathematics, physical and information sciences and technology, and social and behavioral sciences. It describes the nature of the scientific endeavor and defines science literacy broadly, emphasizing the connections among ideas in the natural and social sciences, mathematics, and technology, and identifies habits of mind, important episodes in the history of science, and common themes that cut across disciplines and can serve as tools for thinking about and in them. *Science for All Americans* also includes chapters on effective learning and teaching, reforming education, and next steps toward reform.

To help educators develop a K–12 curriculum that would enable all students to reach the literacy goals proposed in *Science for All Americans*, Project 2061 collaborated with education researchers, teachers, scientists, mathematicians, and engineers to publish *Benchmarks for Science Literacy* (1993), a statement of expectations of what students should know at the ends of grades 2, 5, 8, and 12. Following the same topical organization as *Science for All Americans*, *Benchmarks* concentrates on the common core of learning that contributes to the science literacy of all students while acknowledging that most students have interests and abilities that go beyond that common core, and some have learning difficulties that must be considered.

In 1996, the National Research Council (NRC) published its vision of science literacy in the *National Science Education Standards* (*NSES*) (NRC, 1996), drawing extensively on the learning goals in *Benchmarks*. Both *Benchmarks* and *NSES* are the result of large-scale collaborative efforts involving experts in science, technology, and education as well as policy makers; business and community leaders; and parents, families, and members of the public. Both documents were reviewed in draft form by thousands of individuals and organizations and reflect their input.

There is considerable overlap between the content recommendations in *NSES* and in *Benchmarks;* in the introduction to *NSES*, the authors describe their view: "The National Research Council of the National Academy of Sciences . . . believes that use of *Benchmarks for Science Literacy* by state framework committees, school and school-district curriculum committees, and developers of instructional and assessment materials complies fully with the spirit of the content standards" (NRC, 1996, p. 15).

The two documents share a commitment to promoting a common core of ideas and understandings that all students should know and to reducing the sheer number of topics included in the curriculum. With a few exceptions, both documents take similar approaches to the placement of ideas and skills within grade ranges and levels of difficulty and detail. Where differences exist between the two sets of national guidelines, ongoing research into student learning and Project 2061's own work in clarifying the key ideas in each benchmark, specifying expectations for students, and developing theoretical K–12 learning trajectories are helping to resolve them.

Together, *Science for All Americans*, *Benchmarks*, and *NSES* have helped to shape the national debate on science education over the past 20 years. The national standards documents have also been a major resource for developers of state and local science standards, providing both a rationale for science literacy and a conceptual framework upon which the states could model their own standards. To help in the implementation of the recommendations in *Benchmarks* and *Science for All Americans*, AAAS's Project 2061 has published subsequent print and electronic tools that shed light on how to interpret and apply its learning goals to curriculum, instruction, and assessment (see www.project2061.org for more information). We refer to some of these tools later in the article. Likewise, the NRC has also published additional documents to aid in the use of its standards (see http://www7.nationalacademies.org/bose/ for more information).

A Guide for Curriculum Development

To guide the development of curriculum materials and assessments, the goal statements in the national standards documents have been conceptualized, organized, and expressed to make their intent clear. The statements reflect a significant and coherent set of knowledge and skills that, taken together, convey some important ideas along with selected supporting details. The goal statements are written at a level of specificity that is discrete enough to inform users without overly fragmenting knowledge. They present subject matter that is accurate and appropriate for students' ages and developmental levels.

Significant. Because there is more valuable and interesting knowledge than students can reasonably be expected to learn in the K–12 years, having a process and criteria for choosing learning goals is essential. For example, the science literacy goals presented in *Science for All Americans*, from which the grade-level learning goals in *Benchmarks* were derived, were selected because of their considerable intrinsic and cultural value and also because they were likely to be useful to individuals in their career and educational pursuits and to enhance their personal and public lives.

Coherent. The importance of coherence to learning has been well documented; part of what characterizes expertise is having considerable interconnected knowledge and knowing when it can be applied (Chi, Feltovich, & Glaser, 1981). Knowledge of relations among ideas and of the underlying conceptual framework that makes those relationships meaningful is what, according to Bruner (1960/1995), enables learners to integrate new ideas into what they already know (p. 335). When standards present a coherent "story" for the topics to be learned, they help make those necessary connections explicit: from idea to idea, from grade to grade, and from particulars to deeper structures. Newmann, Smith, Allensworth, and Bryk (2001) argue for curriculum coherence as a key contributor to instructional programs that are themselves coherent and, as a result, promote higher student achievement.

For the developers of *Benchmarks* and *NSES*, the challenge of maintaining coherence while also drastically reducing the information for students to learn was daunting. Although *Science for All Americans* lays out a coherent story of the knowledge, skills, and habits of mind that constitute adult science literacy, in *Benchmarks* (and *NSES*) these topics are subdivided into discrete ideas distributed into sequential K–12 grade bands. To help educators consider the connections among K–12 learning goals within and across grades, Project 2061 has published the *Atlas of Science Literacy* (AAAS, 2001, 2007), a two-volume collection of conceptual strand maps. Each *Atlas* map graphically displays the relationships among key ideas, derived from *Science for All Americans* and *Benchmarks*, that contribute to an understanding of an important topic such as the structure of matter or basic functions of the human body (see http://www.project2061.org/publications/atlas/default.htm for sample maps).

Specific. Learning goals need to be broad enough to provide context and coherence and yet precise enough to make clear what is to be learned. Of course, different users of standards require different levels of specificity. For example, policy makers and others with broad system-wide responsibilities are likely to require far less specificity, whereas developers of curriculum materials or assessments would need to clarify even further what each goal statement or a related set of goal statements expects students to know (DeBoer, 2008).

To balance the need for context and specificity, the learning goals in *Benchmarks* often contain significant detail, but the details all relate to a much smaller set of basic ideas. For example, consider the following goal for students in grades 6–8: "Atoms and molecules are perpetually in motion. Increased temperature means greater average energy of motion, so most substances expand when heated. In solids, the atoms are closely locked in position and can only vibrate. In liquids, the atoms or molecules have higher energy of motion, are more loosely connected, and can slide past one another; some molecules may get enough energy to escape into a gas. In gases, the atoms or molecules have still more energy of motion and are free of one another except during occasional collisions" (AAAS, 1993, p. 78). By combining statements about the expansion of substances and the arrangement and motion of their component atoms or molecules into a single learning goal, it is clear that students should understand something about the *relationship* between observable phenomena and the molecular explanation for them.

Age appropriate. The grade placement of learning goals in both *Benchmarks* and *NSES* has been informed by the insights of researchers and teachers. Much of the available research on student learning is summarized in chapter 15 of *Benchmarks* and is reflected in the learning progressions depicted in the conceptual strand maps published in Project 2061's *Atlas of Science Literacy* (AAAS, 2001, 2007).

Accurate. Both *Benchmarks* and *NSES* are the products of close collaboration between the scientific and science education communities. Although scientists can speak to the correctness of the ideas and skills specified in a goal, it is also necessary to take into account the grade placement of the learning goal and its cognitive demands on students. The statements in *Benchmarks* express learning goals (particularly those for the earlier grades) in language that is "intended to signal the nature and sophistication of understandings to be sought" (AAAS, 1993, p. xiii). Therefore, *Benchmarks* presents the idea that atoms "stick together" as a "correct enough" concept for middle school students to learn in contrast to the more technically accurate, but far more difficult to understand, idea of chemical bonds. Similarly, both *Benchmarks* and *NSES* present a relatively small number of technical terms to avoid confusing scientific accuracy with the use of technical language for its own sake, especially in learning goals aimed at students in the lower grades.

Next Steps for Standards

Although *Benchmarks* and *NSES* reflect the knowledge base and prevailing views that existed in the field at the time of their publication, neither document purports to be the final word on K–12 science learning goals. Since their release more than a decade ago, *Benchmarks* and *NSES* have been studied and used widely, and, not unexpectedly, users have identified areas where they would like more guidance. In addition, new findings on science teaching and learning have provided a better sense of what it takes for students to learn. At the same time, the number of hours in the school day available for science has dropped (Center on Education Policy, 2006). As a result, there is concern that, although *Benchmarks* and *NSES* have greatly reduced the quantity of information that students are to learn, they have not gone far enough, nor have they identified those learning goals that are of highest priority (NRC, 2007).

As debate on NCLB continues, some are calling for states to adhere to new voluntary national science education standards, but no mechanism for developing such standards has been proposed (Olson & Hoff, 2006). One option would be to use this renewed interest in standards as an occasion to reconcile differences between *Benchmarks* and *NSES* such as was done in the 2009 NAEP Science Framework. States would have a single set of recommendations on which to rely. In the meantime, local and state educators and curriculum developers and publishers must decide what to include in (or exclude from) their curriculum.

To help in prioritizing the ideas in the national standards documents, maps in *Atlas of Science Literacy* (AAAS, 2001, 2007) display connections among ideas and topics and can be used to determine which are most important and central to understanding others. Those ideas and topics that connect to the greatest number and variety of other ideas are likely to be worth retaining. Those that connect to fewer other ideas might be deferred. Although this kind of theoretical approach to making curricular decisions will have to do in the short term, empirical studies are under way to develop interventions to test how students make progress in learning different topics. These findings about learning progressions are likely to shed light on which ideas are most central, the grade-by-grade sequencing of those ideas, the kinds of instructional activities that are most effective and the time students need to learn them, and whether learning those ideas helps students to learn related ideas more efficiently (NRC, 2007). As states and school districts go about the work of developing their own guidelines for K–12 science, they will want to take advantage of the principles and conceptual framework that underpin both *Benchmarks* and *NSES*, as well as the insights that current research can provide.

Problems with State Standards

Although most states claim to have drawn on *Benchmarks* and *NSES* in developing their own standards, many have interpreted the national recommendations for science learning quite freely. The American Federation of Teachers (AFT), for example, reports that, although state science standards have improved over time, many students are being evaluated with high-stakes state tests that are aligned to fundamentally weak learning goals. Another common problem, according to the AFT study, is a lack of specificity that results in guidelines that are too vague to be useful. Consider the following example: "Students should be able to use basic science concepts to help understand various kinds of scientific information" (American Federation of Teachers, 2003, p. 7). Researchers in another study conducted by Achieve, an organization created by the nation's governors and business leaders, found that many states have not taken the steps necessary to ensure that their standards focus on ideas that are significant and age appropriate: "states have added concepts and skills to try to 'cover' everything, without making the tough choices about what is most important for students to learn. In other cases, standards are repeated grade after grade with no signal of the progression of knowledge and skills that should mark students' academic growth" (Achieve, Inc., 2002, p. 6). Other studies of state standards have pointed to a lack of coherence and problems with accuracy (Gross et al., 2005).

A brief review by Project 2061 of how states treat significant ideas on an important topic—the particle model of matter—illustrates the problem. Project 2061's study of state science standards revealed that only 19 states have clearly identified the ideas that students should learn about the particle model of matter. Those ideas include the following concepts that are recommended in *Benchmarks* as essential for all middle school students to learn:

> All matter is made up of atoms, which are far too small to see directly through a microscope. The atoms of any element are alike but are different from atoms of other elements. Atoms may stick together in well-defined molecules or may be packed together in large arrays. Different arrangements of atoms into groups compose all substances.
>
> Atoms and molecules are perpetually in motion. Increased temperature means greater average energy of motion, so most substances expand when heated. In solids, the atoms are closely locked in position and can only vibrate. In liquids, the atoms or molecules have higher energy, are more loosely connected, and can slide past one another; some molecules may get enough energy to escape into a gas. In gases, the atoms or molecules have still more energy and are free of one another except during occasional collisions. (AAAS, 1993, p. 78)

Thirty-one states have no goal at all for these ideas or only vague statements such as these describing what fifth graders in Arkansas should be able to do, but not the knowledge needed to accomplish the tasks: "Identify the relationship of atoms to all matter . . . . Explain how heat influences the states of matter of a substance: solid, liquid, gas, and plasma . . . . Demonstrate the effect of changes in the physical properties of matter . . . . Model the motion and position of molecules in solids, liquids, and gases in terms of kinetic energy" (Arkansas Department of Education, 2005). Some states have placed goals for this topic at inappropriate grade levels, such as California's expectation that third-grade students should be able to understand that "each element is made of one kind of atom and the elements are organized in the periodic table by their chemical properties" or Louisiana's requirement that fifth-grade students be able to "describe the structure of atoms and the electrical charges of protons, neutrons, and electrons" (Herrmann Abell, 2006).

Placing learning goals at inappropriate grade levels creates significant problems for science teaching and learning. Research documents persistent learning difficulties that students of all ages have with ideas about the particulate nature of matter (Kesidou & Roseman, 2002). Holding students accountable for these ideas in the elementary grades—far earlier than research suggests that they are able to understand them—could foster rote memorization without real understanding. But there is also a danger in waiting until high school to teach ideas about the particle model of matter. Because students typically study biology first and then move on to chemistry, they would not be adequately prepared to learn the many biology ideas—about the molecular basis of heredity or matter transformation in photosynthesis and cellular respiration—that depend on their understanding of the particulate model of matter. The key point is that the placement of learning goals requires careful thinking through of relevant conceptual, developmental, and practical concerns and ultimately needs to be tested empirically.

The variability of state science standards also presents considerable challenges for curriculum developers and publishers. Differences in the ideas that states choose to include in their standards and in the grades to which standards are assigned make it difficult to produce materials that properly scaffold students' conceptual development and, at the same time, align with 50 different sets of standards. For developers of materials that attempt to integrate science ideas and practices over several grades, the challenges can be particularly acute. We discuss the implications of states' decisions about standards in more detail later.

It will not be easy for states to solve these problems. Most states have limited time and resources to spend on the standards-development process. Few states are able to replicate the breadth and depth of scientific and educational expertise or the investment in building support for standards among all stakeholders that characterized the creation of *Benchmarks* and *NSES*. What is more, state standards documents are products of a political process that defines their purpose, content, authorship, ratification, and implementation. One need only consider recent public debates over the treatment of topics related to evolution and natural selection in the standards for Kansas, Ohio, Georgia, and Pennsylvania, among others, to see that state content standards are shaped by a wide range of forces other than the needs of students, teachers, and schools. In short, as the National Science Board (2004) has reported, state science standards "vary greatly in detail, degree of focus, specificity, clarity, and level of rigor" (pp. 1–19). As a result, they are not equally useful for guiding curriculum, instruction, or assessment.

Problems with Textbooks

Based on data derived from the Third International Mathematics and Science Study (TIMSS), Schmidt et al. (2001) conclude that content standards have their most direct effect on what is taught through the influence they exert on textbooks. Clearly, then, the poor quality of and wide variation among state standards have serious implications for what happens in the science classroom. When each state includes in its standards only a few additional topics not found in the standards of other states, in the aggregate these extra topics add considerably to the total amount of content that publishers believe they must include in their textbooks in order for them to be marketable in as many states as possible. Because textbooks must meet the specifications of each state, differences in learning goals from state to state make the job of developing curriculum materials aligned to standards much more difficult. In a study prepared by the Association of American Publishers, the industry justifies its "typical 750-page textbook" this way: "It is a common misperception that textbook publishers determine the content of the instructional materials they publish. Thus it is sometimes asked why publishers just can't produce a textbook that covers less content and therefore is lighter weight. The reality is, textbook content is dictated by the state or local school system that purchases the books, not the publishers" (Association of American Publishers, 2003).

Studies of textbook quality have shown that when textbooks become encyclopedic in size and style, their effectiveness as tools for teaching and learning is undermined. In their analysis of science education systems in 50 countries, the TIMSS researchers concluded that U.S. science curricula and textbooks lack focus and emphasize breadth over depth of coverage, thus contributing to U.S. students' poor performance in science (Schmidt et al., 1997).

Project 2061's evaluations of middle and high school science textbooks confirmed the TIMSS findings and went beyond them by looking closely at the extent to which the textbooks would help students achieve the learning goals in *Benchmarks* and *NSES*. By examining how closely textbooks' content aligned with the goals and how effective textbooks' instructional strategies were likely to be, the Project 2061 evaluations provided a fairly comprehensive picture of the strengths and weaknesses of the materials available. Applying a set of criteria for judging alignment and instructional quality (Roseman, Kesidou, & Stern, 1997), the Project 2061 studies evaluated a range of materials, including the most widely used commercially published textbooks as well as more innovative materials developed with funding from the NSF. All textbooks except one were published from 1991 to 2000. The results from the evaluation identify problems that have implications for elementary as well as middle school materials and point to shortcomings where attention from curriculum researchers, developers, and publishers can have the most benefit for teachers and students (AAAS, 2002, 2005; Kesidou & Roseman, 2002; Stern & Roseman, 2004).

Content Alignment

The Project 2061 evaluation of middle school science textbooks examined each textbook to judge the extent and nature of its treatment of key ideas for important topics in physical, life, and earth sciences that are specified as learning goals for all students in AAAS's *Benchmarks* and the NRC's *NSES*. For each topic—the kinetic molecular theory, flow of matter and energy in ecosystems, and processes that shape the earth—expert reviewers looked for instances in the student and teacher editions where the key ideas were treated in text, activities, assessments, or in notes or wraparound material intended for the teacher. Key ideas for the physical science topic, for example, included a set of basic assumptions of the kinetic molecular theory and their use in explaining thermal expansion and changes of state:

> *Idea a: All matter is made up of particles called atoms and molecules (as opposed to being continuous or just **including** particles).*
>
> This idea has two aspects: (1) that matter is particulate (rather than continuous), and (2) that the particles (atoms or molecules) *are* the matter (rather than the commonly held incorrect idea that particles [atoms or molecules] *are contained* in matter). In principle, the curriculum material could teach this idea in terms of particles, without making the connection to atoms and molecules. In such cases, the content analysis part of an evaluation report should note that the material addresses this idea in terms of particles rather than in terms of atoms and molecules. If the material introduces the ideas in terms of particles and only later makes the connection to atoms and mol-

ecules, the coherence segment of the content analysis part should examine how well the material links atoms and molecules to particles. The treatment of subatomic particles goes beyond this idea.

Idea b: These particles are extremely small—far too small to see directly through a microscope.

The idea is that atoms and molecules are far too small to see through a *light* microscope. Images of atoms obtained with scanning tunneling microscopes and the actual size of atoms and molecules go beyond this idea.

Idea c: Atoms and molecules are perpetually in motion.

The idea is that atoms and molecules of *all* matter are *perpetually* in motion. A complete content match to this idea would make clear that molecules of solids, liquids, *and* gases are in motion.

Idea d: Increased temperature means greater molecular motion, so most substances expand when heated.

This idea has three components: (1) the relationship between temperature and molecular motion; (2) thermal expansion of solids, liquids, and gases; and (3) the connection between changes in the motion of molecules (with increased temperature) and thermal expansion. The link between increased temperature and average energy of motion of molecules (a more sophisticated idea) does not serve as a basis for this analysis. The curriculum material should neither be held accountable for presenting the link between increased temperature and average energy of motion, nor penalized for including it.

Idea e: There are differences in the arrangement and motion of atoms and molecules in solids, liquids, and gases. In solids, particles (1) are packed closely, (2) are (often) arranged regularly, (3) vibrate in all directions, (4) attract and "stick to" one another. In liquids, particles (1) are packed closely, (2) are not arranged regularly, (3) can slide past one another, (4) attract and are connected loosely to one another. In gases, particles (1) are far apart, (2) are arranged randomly, (3) spread evenly through the spaces they occupy, (4) move in all directions, (5) are free of one another, except during collisions.

This idea focuses on the differences in proximity, arrangement, motion, and interaction of atoms and molecules of solids, liquids, and gases. The nature of interactions between and within molecules (for example, types of bonds) goes beyond this idea.

Idea f: Changes of state—melting, freezing, evaporating, condensing—can be explained in terms of changes in the arrangement, interaction, and motion of atoms and molecules.

This idea focuses on the *molecular* explanation of changes of state. Descriptions of changes of state only in terms of heat transfer are not sufficient for alignment. Descriptions of phase diagrams go beyond this idea. (AAAS, 2002).

Although nearly all of the key ideas for each of the three topics were presented in the textbooks, the books did not focus on them. The sheer volume of information and the inclusion of a great many unrelated ideas, technical terminology for its own sake, and pointless detail usually obscured the most important ideas.

Overall, the textbooks lacked a meaningful narrative to weave the key ideas into a coherent story—such as by showing how they can be used to explain a variety of observable phenomena—and rarely tried to extend previously encountered ideas to new contexts or to connect different applications of the same idea to each other. The sequence in which ideas were presented often appeared to be arbitrary rather than based on the logic of the discipline or on what is known about how students learn the ideas. For example, in some textbooks, "explanations of changes of state in the earlier units rely on the ideas that particles are in motion and increased temperature means increased molecular motion, although these ideas are explicitly introduced only in later units on heat and energy" (Kesidou & Roseman, 2002, p. 528).

Instructional Quality

Project 2061 contends that for students to learn the important science concepts and skills identified in *Benchmarks* and *NSES*, curriculum materials must do much more than merely present the ideas and expect teachers to apply a generic set of instructional strategies to convey those ideas to students. As described above, the middle school science textbooks evaluated in the Project 2061 study fell short in their presentation of the key ideas used as the basis for the evaluation. The instructional strategies the textbooks used were judged to be equally unfocused and inadequate to support effective teaching and learning. Using a set of evaluative criteria based on the available research on the difficulties many students have learning the key ideas, the Project 2061 study applied the criteria to each instance in which the ideas were presented in the textbooks.

The criteria take into account what is known about successful science teaching and learning, particularly when the goal is to help a diverse range of students achieve at least basic science literacy. Accordingly, the criteria attend to aspects of instruction that recognize and deal with individual differences among students and enable teachers to identify, address, and build on the ideas and skills their students already have. Specifically, the Project 2061 evaluation study used the criteria to consider "whether and how materials make their purposes explicit and meaningful to students, take account of student preexisting knowledge (both conceptual and cultural), provide a variety of phenomena and representations to make abstract ideas plausible and intelligible to students, guide student interpretation and reasoning, and help teachers create a learning-based classroom environment where all can succeed" (Kesidou & Roseman, 2002, p. 525).

Project 2061's textbook evaluation also examined the quality of the assessment resources and activities provided as part of each textbook program. Reviewers looked carefully at the match between the knowledge targeted in a benchmark or standard and the knowledge tested by an assessment question or task. They considered whether the assessments would measure what students actually know about the key ideas rather than their ability to memorize definitions and formulas and whether the assessments would provide teachers with useful feedback for modifying their instruction.

Findings

Of the nine multiyear middle school science programs evaluated in the Project 2061 study, not one was judged to be effective in helping students learn key ideas identified as important to science literacy in *Benchmarks* and *NSES*. Among the most troubling shortcomings found in the materials was the failure to help students relate scientific concepts they study to a range of appropriate phenomena or to use key science ideas to explain real-world phenomena. For example, research has shown that through hands-on activities, demonstrations, discussions, and other strategies, students can be helped to view scientific concepts as both plausible and useful explanations of real-world events and observations (Anderson & Smith, 1987; Champagne, Gunstone, & Klopfer, 1985; Strike & Posner, 1985). And yet Project 2061's evaluation revealed that, "Although all [textbook] programs claimed to be activity-based and indeed included several hands-on activities in each chapter, activities only occasionally targeted key ideas and were often add-on features that were neither well integrated with the rest of the text nor explicitly linked to the key ideas" (Kesidou & Roseman, 2002, pp. 533–534).

The evaluation also revealed that textbooks rarely took account of students' beliefs that can interfere with learning, did little to scaffold students' efforts to make meaning of phenomena and representations, and offered little to encourage students to develop and use their science ideas. The textbooks also failed to provide teachers with support that would enable them to interpret students' work or to help

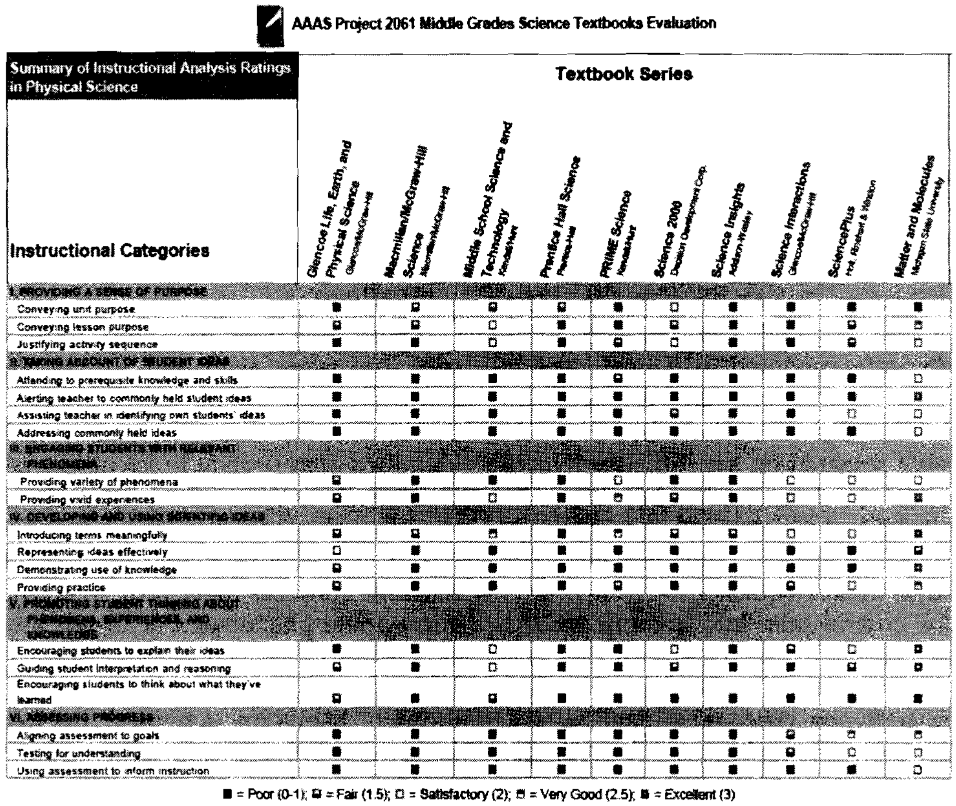

FIG. 1.—Summary of AAAS Project 2061's ratings of instructional quality in physical science textbooks used in the middle grades.

students overcome common misconceptions or fill in gaps in their prerequisite knowledge. Although many of the textbooks contained handsomely rendered drawings, diagrams, and other illustrative material, a large proportion of these were inaccurate, mislabeled, confusing, or peripheral to the central ideas that were the focus of the lesson (Kesidou & Roseman, 2002; Kurth & Roseman, 2001; Roseman & Caldwell, 2001). Assessments included in or with the textbooks—that is, assessments embedded throughout a unit, end-of-chapter quizzes, or end-of-unit tests—were poorly aligned to the content in benchmarks and standards or were focused largely on relatively trivial aspects of the content. The evaluation found that even when assessments were aligned to the content, many tasks were "incomprehensible and . . . not likely to reveal students' difficulties" (Stern & Ahlgren, 2002, p. 905). As a result, according to the study, many of the middle school materials did not "assist teachers in interpreting students' responses or in using these responses to change the instruction. The end-of-instruction assessments might be useful for teachers to grade their students but not to monitor what students actually know about core ideas as a feedback to instruction" (Stern & Ahlgren, 2002, p. 905).

For the physical science topic—the kinetic molecular theory—Figure 1 displays the scores for instructional quality across all nine of the middle school materials evaluated in the Project 2061 study. Only Matter and Molecules, a stand-alone unit developed in the 1980s as part of a research project investigating how to improve student learning, was rated as effective in helping students understand the set of ideas that are central to the topic.

Implications for Elementary Science

In addition to its evaluation of middle-grades curriculum materials (which encompassed materials intended for use in grades 6 through 9), Project 2061 also examined high school biology textbooks in a subsequent study. However, there has been no systematic evaluation of the textbooks, kits, and other materials commonly used for elementary school science. Nevertheless, given that a few commercial publishers control nearly 70% of the market for K–4 science textbooks (Weiss, Banilower, McMahon, & Smith, 2001) and that most of these also publish the middle and high school textbooks that were rated poorly in the Project 2061 evaluations, it is likely that materials used in the lower grades have the same kinds of deficiencies as materials for the upper grades. As for elementary materials developed with NSF funding, most were already under way before the publication of *Benchmarks* and *NSES*, and, as a result, were not designed with those goals in mind.

The lack of good science and mathematics textbooks has especially serious implications for K–5 students. For example, research has shown that the less content-specific training that teachers receive, the more likely they are to rely on textbooks as their primary source of instructional guidance in the classroom (Schmidt et al., 1997). In general, elementary teachers receive far less preparation in science and mathematics than middle or high school teachers. Furthermore, many early childhood curriculum programs take an integrated approach, which may lead teachers to emphasize those disciplines in which they feel most comfortable, often at the expense of mathematics and science (Copley & Padrón, 1999).

To begin to fill the need for high-quality elementary science materials, Project 2061 has been identifying, developing, and testing a set of components that can be used to create materials that are well aligned with the learning goals in *Benchmarks* and *NSES* and meet Project 2061's criteria for supporting effective instruction. Some of these components are being designed especially for use in elementary curricula and focus on topics such as light, the solar system, and processes that shape the earth. The components include (1) summaries of existing research on how students learn key ideas in the targeted benchmark, (2) phenomena and representations that can make key ideas plausible and intelligible to students, and (3) questions that can help elicit students' ideas and scaffold and assess their understanding of phenomena and representations (Kesidou, Kurth, Willard, Caldwell, & Wilson, 2004). Useful for researchers, curriculum developers, and teachers, these well-aligned and carefully screened components can be incorporated into classroom activities and into curriculum materials that are being used to investigate student learning.

A New Generation of Science Curriculum Materials

Findings from Project 2061's textbook evaluations, from the TIMSS analysis, and from other studies, although pointing to serious shortcomings in materials, have also stimulated interest in developing the knowledge base, principles, and processes that can lead to the design of new and more effective science curriculum materials. Collaborations between researchers and curriculum developers have begun to yield both theoretical and practical results, and the methods and criteria Project 2061 used to analyze curriculum materials and their alignment to standards are being discussed in the field and adapted for use in curriculum development (Heller, 2001; Krajcik, McNeill, & Reiser, 2008).

The Role of Research

One of the most important outcomes of Project 2061's textbook evaluations has been the realization that curriculum design must take place within the context of research on students' learning. Calling for "new partnerships between researchers and curriculum developers," Project 2061's study of middle school materials points out that there are "many topics in the science curriculum that have hardly been studied" (Kesidou & Roseman, 2002, p. 541). One example of how development teams have tried to attend to these research needs by incorporating them into their curriculum design strategies can be seen in the IQWST (Investigating and Questioning Our World through Science and Technology) project led by researchers at the University of Michigan, Northwestern University, and the Weizmann Institute of Science in Israel (Center for Highly Interactive Classrooms, Curricula and Computing in Education, 2007; Shwartz et al., 2008, in this issue). Although IQWST is aimed at middle school students, the design principles and processes that have guided its development can easily be adapted for use by developers of elementary materials. Funded through the NSF's instructional materials development program, the IQWST curriculum focuses on national content standards for life, earth, and physical science at the middle school level. Project 2061's involvement with the IQWST development team has been extensive.

While taking advantage of the existing research on student learning to inform their curriculum design, the developers of IQWST also conducted their own research with students and teachers throughout the design process. Project 2061 analyzed the IQWST materials in draft form and provided extensive feedback and suggestions for revisions. By systematically testing their assumptions about goals, pedagogy, and classroom implementation within the context in which the materials would actually be used, the IQWST design team was able to think through all aspects of the material more carefully and take advantage of feedback on the material from a variety of sources:

> We collected a variety of data sources that allowed us to evaluate the curriculum materials in terms of their learning goals alignment and pedagogical supports, and the enactments of the curriculum in terms of classroom focus on student learning, engagement in scientific practices, and participation in project-based pedagogy. The data sources included student pre and posttests, student artifacts, field notes, selected classroom videos, teacher feedback, Project 2061 review, and content expert feedback. We used these data sources to identify and triangulate design issues to be addressed in subsequent curriculum revisions. (Krajcik et al., 2008)

Because of this research effort, each unit in IQWST is being designed for students at a particular grade based on an analysis of the data and on an iterative process of pilot testing in the classroom, revising based on feedback, further testing, and so on. The developers have been willing to pay attention to what they were learning and to make the kinds of mid-course corrections that were needed. Similarly, the IQWST units are being conceptualized and tested to ensure that the needed prerequisites within a grade band are in place. As a result, when the final units are completed, they will fit together in a carefully thought through sequence and will not lend themselves to being "mixed and matched" for use in any grade or in any sequence.

A Focus on Learning Goals

Developers of the next generation of science curriculum materials are also putting benchmarks and standards—and the instructional strategies that can help students meet them—at the center of their efforts. The IQWST development team uses the term "learning-goals driven design" to describe its design approach (Krajcik et al., 2008), but the description could apply equally well to the process used by the Interactions in Physical Science (formerly known as Constructing Ideas in Physical Science or CIPS) development team, another university-based curriculum project funded by NSF (Goldberg, Bendall, Heller, & Poel, 2005). For both teams, their first step required careful consideration and selection of the content learning goals, which then informed all subsequent phases of the curriculum design and evaluation process. For the IQWST team, this involved "unpacking" the content standards to consider "what content is important, as well as what aspects are suitable for middle school students by examining common student difficulties, prior understandings needed to build the target understanding, and aspects of prior conceptions that may pose challenges" (Krajcik et al., 2008).

Because both development teams aim for greater depth of understanding rather than mere coverage, their selection of goals had strategic implications. The Interactions team initially focused on a set of 63 benchmarks to be addressed in a 1-year course. However, after clarifying the goals and considering the time and effort it would take for students to achieve a deep understanding of all of the targeted ideas as defined by the Project 2061 instructional criteria, the developers realized that "if we included too many related benchmarks from other chapters, we would risk having teachers rush through the curriculum activities and leave inadequate time for students to make sense of the content benchmarks" (Heller, 2001). In the end, the Interactions developers pared down their original list of 45 targeted ideas to just 17, concluding that there was simply not enough instructional time in 1 year for more.

In addition to being explicit about content standards that materials address, the developers of Interactions and IQWST have also made a commitment to applying the Project 2061 content and instructional criteria to the task of curriculum design. Both have used the criteria to guide their initial conception and design of the components of their materials and as an analytical tool to evaluate the alignment and likely effectiveness of those components at various points along the way. For example, once the developers of Interactions had selected the learning goals for their material, they then applied the Project 2061 criteria to the activities they planned to use. At first, they found it difficult to "give up our favorite activities or approaches to teaching particular content ideas," even when those activities did not meet Project 2061's content and instructional criteria (Heller, 2001). Nevertheless, they were able to use the criteria to evaluate their planned activities, place them along a learning trajectory, and create several learning cycles to target a small set of related benchmarks.

In its role as consultant to both development teams, Project 2061 used its content and instructional criteria to provide expert analysis of the Interactions and IQWST materials in draft form and extensive feedback and suggestions for revisions. For example, in reviewing an IQWST unit focused on chemical reactions, Project 2061 pointed out that "given the importance of the explanatory power of atoms, it is important to develop this idea through several instances, rather than the limited number described below. While this aspect of the learning goal was not something 'taken on' by the unit developers, we think they should consider doing so" (Roseman, 2003).

Following up on Project 2061's recommendations, which drew on both an analysis of the discipline and the learning research, IQWST developers revised their unit: "Project 2061 argued that we should introduce the particle nature of matter earlier and use it across the multiple different contexts .... In debriefing interviews, pilot teachers agreed that the particle nature of matter and the concepts of molecules could be brought in earlier, and might increase students' understanding of substances and chemical reactions .... We revised the unit to include three learning sets with a focus on substance and property, chemical reactions, and conservation of mass, with the particle nature of matter integrated throughout each learning set" (Krajcik et al., 2008).

Two of the Project 2061 instructional criteria reflect the cognitive apprenticeship learning model and are used to look for evidence that a material helps students make use of scientific ideas by explicitly modeling that use and providing principles and examples of how to judge it. The criteria also are used to consider whether a material provides students with guided practice at first, followed by practice in which guidance is gradually decreased. Project 2061's analysis helped the IQWST developers see where they needed to articulate more specifically, for example, what students' explanations of chemical reactions should include: "Project 2061's review of our learning performances specifically articulated a concern that the learning performances had not clearly defined what is meant by 'explain.' . . . [While] in our discussions, we had unpacked scientific explanation . . . into three components: claim, evidence, and reasoning . . . . We realized this unpacking was not provided as an explicit structure in either the student or teacher materials . . . . [Our data revealed] that students had difficulty with evidence and reasoning aspects of constructing scientific explanations" (Krajcik et al., 2008). The Project 2061 analysis also revealed problems with the material's assessments and helped the developers identify the source of the problems—the need to specify more precisely the ideas to be learned. According to the developers, by identifying and addressing these problems, they were able to create "consistency across the unit including the assessment measures, which are often a neglected portion of the design process" (Krajcik et al., 2008).

Because of the developers' careful focus on specific learning goals and their use of Project 2061's content and instructional criteria, each unit in Interactions and IQWST has been precisely designed for students at a particular grade based on the available research, attention to instruction that supports the relevant learning goals, and an interactive process of pilot testing in the classroom, revising based on feedback, further testing, and so on. For example, the 1-year Interactions curriculum was initially tested in grades 7, 8, and 9, but the developers eventually settled on students in grades 7 and 8 as the most appropriate target population. Through their own research and through Project 2061's analyses of their materials, both development teams have taken on at least some of the analytical and empirical studies that the National Research Council has identified as necessary for evaluating curricular effectiveness (NRC, 2004).

Coherence

Curriculum developers must also take account of the growing body of research that demonstrates the importance of coherence to the development of students' science understanding. Coherence has many aspects—narrative, cognitive, pedagogical, as well as disciplinary—all of which have a role in creating more effective learning experiences (Roseman, Stern, & Koppal, 2008). Curriculum developers are investigating ways to organize and present science concepts coherently, design more coherent curriculum and curriculum materials, and make use of instructional strategies that foster students' ability to connect and apply their knowledge of science concepts and skills. As a first step, many are using *Science for All Americans*, *Benchmarks*, and *NSES* to identify and clarify the significant science ideas as the foundation for lessons, units, materials, or an entire K–12 curriculum focused on a coherent set of important science ideas.

To help developers specify and characterize connections among these ideas and essential prerequisites, the conceptual strand maps published in the *Atlas of Sci-*

ence Literacy (AAAS, 2001, 2007) provide a model for nearly 100 topics that are central to science literacy. Based on the *Atlas* maps, developers can then create even more detailed maps that make up the learning sequences for their materials. For developers of elementary level materials, the *Atlas* strand maps serve as a visual reminder of the larger K–12 context for learning in the early grades: the ideas and skills acquired in the first few years must set the stage for more sophisticated ideas to be encountered in later grades.

Policy Implications

It is important to acknowledge that even with the best possible standards in place, neither standards nor textbooks that are aligned to them are sufficient on their own to accomplish the kinds of changes that are needed in science teaching and learning at any level of schooling. Social and economic inequalities continue to have a significant effect on the educational outcomes and career aspirations of far too many U.S. children. And yet much can be accomplished if time and resources are invested wisely to improve the science curriculum materials used in classrooms. Focusing on the curricular issues raised in this article, we offer a few modest recommendations about how we think federal and state education policies can make a difference.

Draw on National Benchmarks and Standards

Despite recent federal action to encourage voluntary national science standards, most standards-setting activity in the near term is likely to take place at the state level. And although neither *Benchmarks* nor *NSES* claims to be a perfect set of goals, there is far greater respect for the national standards documents and for their conceptualization and placement of learning goals than for the standards of any state. This respect, of course, does not take the place of well-designed, empirical tests to determine more precisely what, how, and when students are best able to learn specific science ideas. But given the pace of research and the current resources devoted to it, educators, curriculum developers, publishers, and policy makers are unlikely to have the benefit of empirical support for the range of decisions they must make about the goals, strategies, and outcomes of science education. At the very least, the recommendations in the national documents, which have been carefully vetted and received wide acceptance, can help to overcome the incoherence that results from 50 sets of state standards of highly variable quality. Consider, for example, the map in Figure 2 that illustrates how states' decisions about the grade placement of two sets of ideas might, in theory, affect their ability to adopt the IQWST integrated middle school curriculum, which is being developed to be compatible with the national documents. Figure 2 shows that if states interpret their standards rigidly, only five of them would be able to adopt this curriculum, which addresses the topics of the particulate model of matter and light and matter interactions in the sixth grade. Rather than revising their standards independently yet again, states might do well to join together to adapt the national benchmarks and standards as regional guidelines, thereby exerting more influence on the content of textbooks and other kinds of curriculum materials available for adoption. At the very least, states should allow districts and schools more flexibility in their textbook adoption policies, making it possible for them to adopt curriculum materials that are aligned to national benchmarks and standards as well as to their own state standards. Federal and private funding agencies can do much to support a more consistent goals-based approach by encouraging the use of national standards (or regional standards based on them) through their science education program guidelines and solicitations.

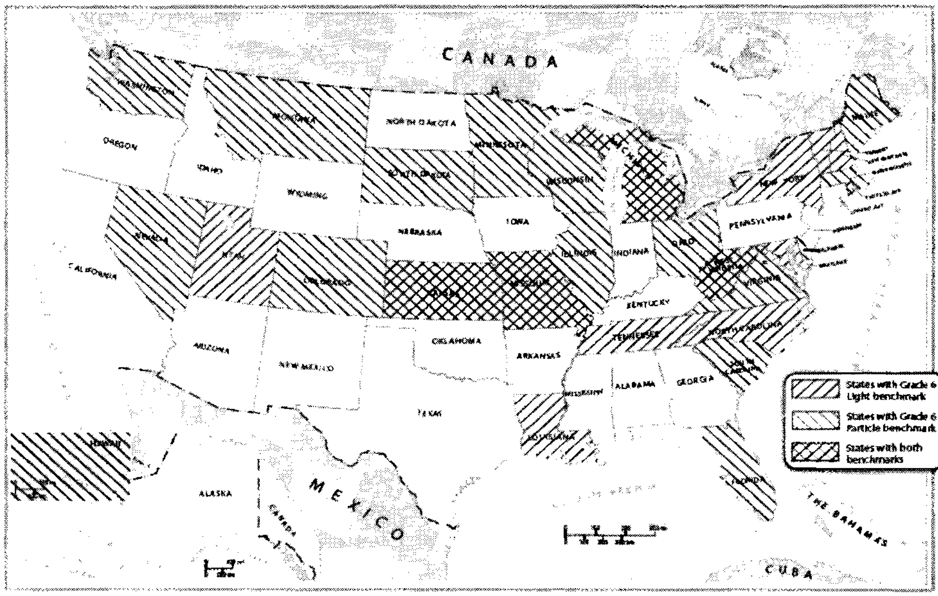

FIG. 2.—Grade levels at which states place standards for the particle model of matter and for interactions of light and matter.

Increase the National Investment in Curriculum Research and Development

Little progress can be made without better and more comparable data on what works, why, and with whom. Enormous complexities are involved in conducting the research needed and interpreting the results, but it is essential that researchers meet this challenge and develop high-quality data that can inform the many decisions in which curriculum plays a role.

In its review of hundreds of studies of mathematics curricula, the National Research Council found that no single method or study design could adequately evaluate effectiveness. Instead, the council advocated the use of "multiple methods of evaluation, each of which should be a scientifically valid study" (NRC, 2004, p. 5). Given the time and resources needed to conduct large-scale empirical studies of student learning, perhaps a two-step approach is required. Analytical studies using methods such as those developed by Project 2061 for its textbook evaluations could be applied first to predict which materials were most likely to be effective. A series of next steps would involve small- and large-scale empirical studies of those promising materials.

Whatever approach is taken, it is important that all involved be realistic about the time and resources needed to produce high-quality materials and to evaluate their effectiveness. It will require significant and more consistent long-term funding from federal government agencies such as the NSF and the Department of Education, from private foundations and professional organizations, and from state agencies. Research on the effectiveness of curriculum materials should be viewed as an essential step at all stages of the curriculum development process rather than just at the end, and budgets and timelines for development projects should reflect that.

The science curriculum is just one part of the larger science education system, and the curriculum problems described in this article require systemic solutions in which all elements of the education system are focused on the learning goals educators want students to achieve. In addition to textbooks, assessments must also align with the goals and must fit into a coherent, purposefully designed K–12 curriculum. And, finally, teachers who have the preparation and support to help their students achieve those goals and communities that understand the goals and are committed to reform are also needed.

References

Achieve, Inc. (2002). *Staying on course: Standards-based reform in America's schools: Progress and prospects* [On-line]. Available: http://www.achieve.org/files/5YearReportfinal.pdf, p. 6.

American Association for the Advancement of Science. (1989). *Science for all Americans.* New York: Oxford University Press.

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

American Association for the Advancement of Science. (2001). *Atlas of science literacy.* Washington, DC: Author.

American Association for the Advancement of Science. (2002). *Middle grades science textbooks: A benchmarks-based evaluation* [On-line]. Available: http://www.project2061.org/publications/textbook/mgsci/report/index.htm

American Association for the Advancement of Science. (2005). *High school biology textbooks: A benchmarks-based evaluation* [On-line]. Available: http://www.project2061.org/publications/textbook/hsbio/report/crit-used.htm

American Association for the Advancement of Science. (2007). *Atlas of science literacy* (Vol. 2). Washington, DC: Author.

American Federation of Teachers. (2003, June). *Setting strong standards* (Item No. 39-0175). Washington, DC: Author.

Anderson, C. W., & Smith, E. (1987). Teaching science. In V. Richardson-Koehler (Ed.), *The educator's handbook: A research perspective* (pp. 84–111). New York: Longman.

Arkansas Department of Education. (2005). Science curriculum framework: Physical science: Matter: Properties and changes 5–8 [On-line]. Available: http://arkedu.state.ar.us/curriculum/pdf/science_framework_k-8_011006.pdf

Association of American Publishers, Advisory Committee for Textbook Specifications. (2003, June). *Report to the Advisory Committee for Textbook Specifications: Analysis of the overweight backpack phenomena* [On-line]. Available: http://www.publishers.org/SchoolDiv/issues/issues_06_SizeWeight/LGOverweightBackPhenom.pdf

Bruner, J. S. (1995). Retrospective: On learning mathematics. *Mathematics Teacher,* **8**(4), 330–335. (Reprinted from *Mathematics Teacher,* **53,** 610–619, December 1960).

Center for Highly Interactive Classrooms, Curricula and Computing in Education. (2007). *Investigating and questioning our world through science and technology (IQWST).* Manuscript in preparation. Ann Arbor: University of Michigan.

Center on Education Policy. (2006, November). *Ten big effects of the No Child Left Behind Act on public schools.* Washington, DC: Author.

Champagne, A., Gunstone, R., & Klopfer, L. (1985). Instructional consequences of students' knowledge about physical phenomena. In L. West & A. L. Pines (Eds.), *Cognitive structure and conceptual change* (pp. 61–90). Orlando, FL: Academic Press.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Sciences,* **5,** 121–152.

Copley, J. V., & Padrón, Y. (1999). Preparing teachers of young learners: Professional development of early childhood teachers in mathematics and Science. In *Dialogue on early childhood science, mathematics, and technology education* (pp. 117–129). Washington, DC: American Association for the Advancement of Science.

DeBoer, G. E. (2008). Assessment linked to science learning goals: Probing student thinking through assessment. In J. Coffey, R. Douglas, & C. Stearns (Eds.), *Assessing learning: Perspectives from research and practice.* Arlington, VA: National Science Teachers Association Press.

Goldberg, F., Bendall, S., Heller, P., & Poel, R. (2005). *Interactions in physical science.* Armonk, NY: It's About Time.

Gross, P., Goodenough, U., Haack, S., Lerner, L., Schwartz, M., & Schwartz, R. (2005). *The state of state science standards.* Washington, DC: Thomas B. Fordham Institute.

Heller, P. (2001, February). *Lessons learned in the CIPS (Constructing Ideas in Physical Science) curriculum project.* Paper presented at the AAAS Conference on Developing Textbooks That Promote Science Literacy, Washington, DC [On-line]. Available: http://www.project2061.org/events/meetings/textbook/literacy/heller.htm

Herrmann Abell, C. (2006). [How science standards in 19 states treat the topic of the particle nature of matter]. Unpublished raw data.

Kesidou, S., Kurth, L., Willard, S., Caldwell, A., & Wilson, P. (2004, April). *Creating components for the next generation of curriculum materials at the intersection of research and practice.* Symposium presented at the annual meeting of the National Association for Research in Science Teaching, Vancouver.

Kesidou, S., & Roseman, J. E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. *Journal of Research in Science Teaching,* **39**(6), 522–549.

Krajcik, J., McNeill, K., & Reiser, B. (2008). Learning-goals-driven design model: Developing curriculum materials that align with national standards and incorporate project-based pedagogy. *Science Education,* **92,** 1–32.

Kurth, L. A., & Roseman, J. E. (2001, March). *Findings from the high school biology curriculum study: Molecular basis of heredity.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, St. Louis.

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

National Research Council. (2004). *On evaluating curricular effectiveness: Judging the quality of K–12 mathematics evaluations.* Washington, DC: National Academies Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K–8.* Washington, DC: National Academies Press.

National Science Board. (2004). *Science and engineering indicators 2004* (Vol. **1,** NSB 04-1; Vol. **2,** NSB 04-1A). Arlington, VA: National Science Foundation.

Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis,* **23**(4), 297–321.

Olson, L., & Hoff, D. (2006, December 13). Framing the debate. *Education Week,* p. 22.

Roseman, J. E. (2003, June 11). [Chemistry unit content analysis findings]. Unpublished data.

Roseman, J. E., & Caldwell, A. (2001, March). *Findings from the high-school biology curriculum study: matter and energy transformations.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, St. Louis.

Roseman, J. E., Kesidou, S., & Stern, L. (1997, November). *Identifying curriculum materials for science literacy: A Project 2061 evaluation tool.* Paper presented at a colloquium of the National Research Council on Using the National Science Education Standards to Guide the Evaluation, Selection, and Adaptation of Instructional Materials, Washington, DC.

Roseman, J. E., Stern, L., & Koppal, M. (2008). *A method for analyzing the coherence of biology textbooks: A study of its application to the topic of matter and energy in four textbooks.* Manuscript submitted for publication.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning.* San Francisco: Jossey-Bass.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). *A splintered vision: An investigation of the U.S. science and mathematics education.* Dordrecht: Kluwer.

Shwartz, Y., Weizman, A., Fortus, D., Krajcik, J., & Reiser, B. (2008). The IQWST experience: Using coherence as a design principle for a middle school science curriculum. *Elementary School Journal,* **109,** 000–000.

Stern, L., & Ahlgren, A. (2002). Analysis of students' assessments in middle school curriculum materials: Aiming precisely at benchmarks and standards. *Journal of Research in Science Teaching,* **39**(9), 889–910.

Stern, L., & Roseman, J. E. (2004). Can middle-school science textbooks help students learn important ideas? Findings from Project 2061's curriculum evaluation study: Life science. *Journal of Research in Science Teaching,* **41**(6), 538–568.

Strike, K., & Posner, G. S. (1985). A conceptual change view of learning and understanding. In L. West & A. L. Pines (Eds.), *Cognitive structure and conceptual change* (pp. 211–231). Orlando, FL: Academic Press.

Weiss, I. R., Banilower, E. R., McMahon, K. C., & Smith, P. S. (2001, December). *Report of the 2000 national survey of science and mathematics education.* Chapel Hill, NC: Horizon Research.