NAEP 2009 Science Framework Development:

Issues and Recommendations

NAEP Science Issues Paper Panel

Audrey Champagne, Panel Chair
University at Albany, State University of New York

Kathleen Bergin
National Science Foundation

Rodger Bybee
Biological Sciences Curriculum Study

Richard Duschl
Rutgers University

James Gallagher
Michigan State University

November 22, 2004

Developed for the
National Assessment Governing Board

Table of Contents

NAEP 2009 SCIENCE FRAMEWORK DEVELOPMENT: ISSUES and RECOMMENDATIONS

INTRODUCTION

In September 2004, the National Assessment Governing Board (NAGB) convened a Panel to develop a Science Issues Paper to serve as a springboard for discussion of issues likely to be engaged in the design of a new Science Framework for the National Assessment of Educational Progress (NAEP). The new Science Framework will guide the development of the NAEP science assessment to be administered beginning in 2009. The panel's charge was to identify issues emerging from the nation's requirements for a science literate citizenry and perspectives on science literacy in documents developed by the U.S. and international science education communities. Issues identified by the panel and recommendations for their resolution are contained in this paper.

The Science Issues Paper will frame the deliberations of the Framework Steering and Planning Committees, convened by WestEd and the Council of Chief State School Officers (CCSSO) under contract to the National Assessment Governing Board. These committees are charged with drafting the 2009 Science Framework for approval by NAGB. The new Science Framework will describe in broad terms the science understandings and abilities to be assessed, guide the development of test specifications, and identify issues to be considered for special studies.

The essential challenge facing the Framework Committees is the breadth of science content, understandings, and abilities[1] that might be included in the Science Framework. This paper identifies factors—contextual and theoretical—to be considered in the choice of content for the Science Framework, examines perplexing issues that the selection process raises, and makes recommendations for approaches to their resolution.

THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

The National Assessment of Educational Progress is authorized by Congress and funded by the federal government; it is the only nationally representative and continuing assessment of what America's students know and can do. For more than 30 years, NAEP has been charged with collecting and reporting information on student achievement in mathematics, reading, science, U.S. history, writing, and other subjects. NAEP assessments were conducted on an annual basis until 1981, when they became biennial. Originally, assessments were given to students at ages 9,

[1] In the science education community, abilities, skills, and competencies are terms used to describe what students can do. It is unclear if the terms all mean the same thing or have distinct definitions. Definition and consistency in their use will contribute to understanding the Science Framework.

> 13, and 17, but beginning in 1983, NAEP has tested students at grades 4, 8, and 12.
>
> NAEP reports provide descriptive information about student performance in various subjects, including basic and higher order skills, and comparisons of performance by race/ethnicity, gender, type of community, and geographic region. They also show relationships between achievement and certain background variables, such as time spent on homework or educational level of parents. Results are based on samples of students and are reported for the nation, states, and for districts that volunteer for the Trial Urban District Assessment (TUDA). (NAGB, 2004).

Since the publication of the NAEP Science Framework in 1991, science education has been in a state of ferment. NAEP Science Assessments administered in 1996, 2000 and 2005 are based on this Science Framework completed just after the publication of *Science for All Americans* (SFAA) (American Association for the Advancement of Science [AAAS], 1989) and concurrent with the development and review of the *National Science Education Standards* (National Research Council [NRC], 1996). The vision of science education contained in *SFAA* and *NSES* have been further elaborated by their parent organizations. The vision and elaborations have guided states in the development of state standards, tests, and curricula. Data from international tests—trends in International Mathematics and Science (TIMSS) and the Programme for International Student Assessment (PISA)—have provided the nation and the states the opportunity to compare the science performance of U.S. students with the performance of students worldwide.

Advances in cognitive science have developed strategies to identify types of knowledge and cognitive processes[2] underlying students' performance on science test items and tasks. These strategies in consort with developments in measurement theory can be applied to the assessment of abilities that heretofore were difficult to measure, the ability to inquire, for instance (Mislevy, 1993; Mislevy, 2003).

Science is a priority under the No Child Left Behind Act (NCLB). Under the provisions of the Act, states receiving Title 1 funding must develop academic content standards in science by 2005-06 and implement aligned assessments based on those standards by 2007-2008. The science assessments must be administered at least once in each of three grade spans: 3-5, 6-9, and 10-12. Unlike the NCLB requirement that states receiving Title I funding must participate in the biennial NAEP assessments in reading and mathematics at grades 4 and 8, states may choose voluntarily to participate in the Science NAEP.

---

[2] Assessments provide data from which inferences are made about the contents and processes of the mind. The science education and psychology communities use different language to describe what is assumed to be stored in the human mind and how what is stored is processed. Maintaining the distinction between what is measured and inferences about mind that are inferred poses a challenge to the Framework Committees as does establishing a common language to distinguish the observed from the inferred.

NCLB, the visions of science education contained in national standards, the performance of U.S. students on international assessments, advances in cognitive science and measurement, and the demands of the public, higher education, the private sector, and the military must be considered in making decisions about the content and form of the Science Framework. Inevitably issues will arise during the decision-making process. The resolution of these issues involves weighing costs and benefits. Conflicting perspectives will necessitate compromise.

An overarching issue the Framework Committees must recognize is the danger of designing a Framework that is too far ahead of the current thinking of stakeholders in science education, science educators, the general public, higher education, the private sector, and the military. Furthermore, decisions about the content contained in the Framework must be tempered by the existing state of science education, resources that will be available for its improvement, and the quality of the psychometric tools available to the assessment. Committee members must be mindful that the Framework be neutral regarding matters of curriculum and instruction.

THE SCIENCE ASSESSMENT FRAMEWORK

The NAEP Science Framework will describe in broad strokes the understandings and abilities that will be measured on the science assessments to be administered beginning in 2009 and in future assessments. The Framework will guide the development of test specifications which define in greater detail the understandings and abilities to be measured by the test. (See Figure 1.) Typically test specifications detail tests' contents in terms of the

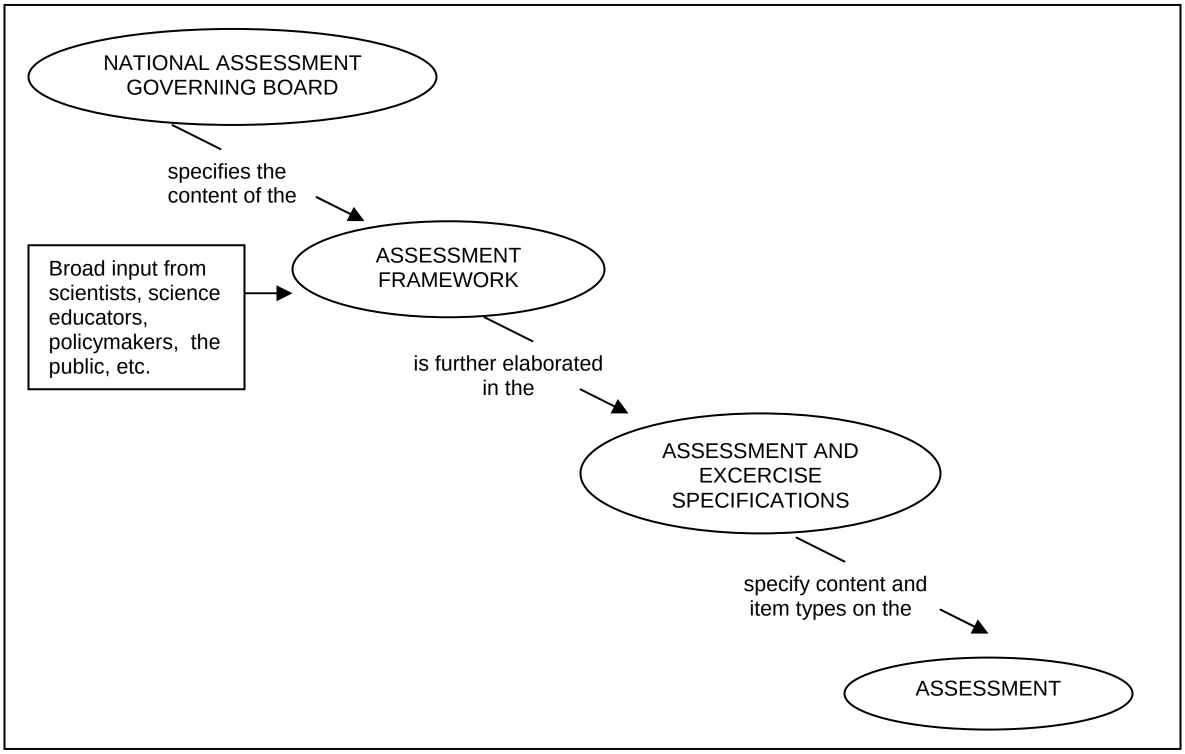
1) proportion of items of different types selected—response, constructed response, performance ("hands-on" and simulations),
2) proportion of items measuring each topic (for instance life or physical science) identified in the Framework, and
3) proportion of items measuring each ability (for instance, higher-order thinking or inquiry) identified in the Framework.

The Framework represents the gross anatomy of the content that will be measured by the assessment instrument or instruments[3]. The Specifications contain the fine anatomy of the content.

Science presents unique challenges to the design of large-scale assessments. While these challenges must be acknowledged in the design of the Framework, so too must measurement theory and practice be challenged to test those features of science that are difficult to measure in large-scale assessments.

[3] The 1996/2000/2005 Science Assessments were designed using a single framework for the grades 4, 8, and 12 assessments. A possibility for the 2009-2019 Assessments is to use three framework designs, one for the grade 4, one for grade 8, and one for the grade 12 assessment.

Figure 1. NAEP Assessment Design Sequence



Central to the development of the Framework is the decision about the content, understandings and abilities, to be included in the Framework and ultimately be measured by the test. The National Assessment Governing Board has the authority to determine the content. The Board's decision will be made with the advice of the Framework Steering Committee and the Framework Planning Committee with extensive review and comment by the general public and other stakeholders. The Framework Planning Committee will draft the Science Framework for NAGB. The Steering Committee will provide policy guidance to the Planning Committee.

The Framework contents will be influenced by the expectations for student achievement contained in national standards, the performance of U.S. students on international assessments, advances in cognitive science and measurement, and the influence of NCLB on science education in the states.

Framework Content
U.S. Standards and International Assessment Frameworks

The *National Science Education Standards* (NRC, 1996), *Science for All Americans* (American Association for the Advancement of Science, 1989) and documents elaborating on them contain expectations for what U.S. students should learn as a result of their K-12 science experiences. The Frameworks for the TIMSS and PISA assessments describe international expectations for students' understandings and abilities. Taken as a whole, these documents illustrate the considerable breadth of content for education in the natural sciences. This presents the Framework Committees with the task of considering the implications of the breadth and consequently the depth of science content for the Science Framework.

> ***Issue 1:*** *How broad should the content contained in the Science Framework be and what are the implications of the choice of breadth and depth?*
>
> ***Recommendation:*** *The choice of content breadth and depth must be made in consideration of its implications for national science education policy and practice and informed by national standards, international assessment, cognitive science, and measurement theory.*

While considerable similarities exist in the expectations contained in the U.S. standards and international test frameworks, there are differences in emphasis and differences in expectations for students at different points in the education continuum. (Neidorf, Binkley, & Stephens, in press). The Framework Committees must consider the implications of these differences for the Science Framework.

Central to the decision of the content contained in the Science Framework are the expectations for students' understandings, and abilities contained in the *National Science Education Standards,* (NRC, 1996), *Science For All Americans* (AAAS, 1989), and *Benchmarks for Science Literacy (*AAAS, 1993).

National Science Education Standards (*NSES*) and Project 2061—The National Research Council *NSES* and Project 2061 documents expand substantially the content base of expected science attainments of students. This broader view has begun to have an impact on student achievement in some schools, but its impact has been limited (Hollweg & Hill, 2003). However, this broader view deserves attention in the Science Framework.

The *NSES* were developed under the auspices of the National Research Council with the support of the U.S. Department of Education and others. The process of development of the standards took place over a period of six years and involved scientists, science educators from universities, and classroom teachers. A significant feature of the process is that the standards underwent extensive and thorough reviews required by the National Research Council.

The *NSES* include eight dimensions of science content, four of which are quite familiar to most teachers and to members of the general citizenry. The familiar dimensions of science include Inquiry[4] Physical Science, Biological Science, and Earth and Space

[4] The reader of this paper and the members of the Framework committees are cautioned to be aware of the many meanings of inquiry in the context of science education and the measurement of students' science achievement. Each time the word is encountered the reader should ascertain the intended meaning.

Science. The dimensions that are less familiar to many teachers and the public in general are Unifying Concepts and Processes, Science and Technology[5], Science in Social and Personal Perspectives, and History and Nature of Science.

Each of these dimensions is described in some detail in the published statement of the *National Science Education Standards* (NRC, 1996, pp. 103 – 207). The standards for grades 9 – 12 provide the most elaborated descriptions of what each entails. (Table A-1 in Appendix A provides a telegraphic synopsis of the content standards contained in *NSES*.)

Project 2061 was initiated by F. J. Rutherford, then Chief Education Officer of the American Association for the Advancement of Science (AAAS). Project 2061 has produced several documents that have received wide distribution and have influenced the science education community. The two books that have had the greatest impact on curriculum and instruction are *Science for All Americans* (AAAS, 1989) and *Benchmarks for Science Literacy* (AAAS, 1993). *Science for All Americans* grew out of a series of discussions by scientists about what would be important for America's youth to have learned by the end of high school. Their approach was to take a fresh look at what would be appropriate as a basis for long-term transformation of science teaching to bring it in line with needs of adults deep in the twenty-first century. The motive behind this work was to set a vision for science education that would be beneficial for youth of today and tomorrow who will be living when Halley's comet returns in 2061. (Hence the project name.) *Benchmarks for Science Literacy* took a major step further and described appropriate levels of attainment by the end of grades 2, 5, 8, and 12. In these documents, a framework for organizing the curriculum is described in twelve topical chapters: Nature of Science, Nature of Mathematics, Nature of Technology, The Physical Setting, The Living Environment, The Human Organism, Human Society, The Designed World, The Mathematical World, Historical Perspectives, Common Themes, and Habits of Mind. (A synopsis of these content dimensions is shown in Table A-2 in Appendix A.)

It is clear that there are differences between the National Research Council (NRC, 1996 & 2000) and Project 2061 statements concerning the content that is appropriate for America's youth to understand and be able to do. Also, there are important points of agreement. By combining and examining the two as shown in Table 1, comparing and contrasting the content recommendations contained in NSES and Project 2061 becomes easier.

[5] Technology has many meanings in the context of assessment and science education. Technology is a tool for the administration, scoring and reporting of large scale assessments. Technology refers to laboratory tools used in science to make measurements and record data. Technology refers to expectations for understandings and abilities students are expected to gain from their K-12 science experiences especially those related to the work of engineers and the process of engineering design. While in science education technology is often characterized as the application of science to human problems, the engineering community takes a broader and different view of its work.

Table 1. Comparison of Science Content Recommendations between *National Science Education Standards* and *Science for all Americans,* Project 2061

National Science Education Standards	***Science for All Americans,* Project 2061**
Unifying Concepts and Processes	Common Themes Habits of Mind
Inquiry	Inquiry (as a component of Nature of Science) Habits of Mind
Physical Science Earth and Space Science	The Physical Setting (encompassing both physical science and earth and space science)
Life Science	The Living Environment The Human Organism Human Society
Science and Technology	Nature of Technology The Designed World
Science in Personal and Social Perspectives	The Human Organism Human Society The Designed World Habits of Mind
History and Nature of Science	Historical Perspectives Nature of Science Habits of Mind
	Nature of Mathematics The Mathematical World

Examining this table, it becomes clear that the authors of these two reform agendas for science in America's schools have retained some important aspects of the content all students are expected to learn and have strongly recommended addition of some new components.

The familiar topics of Physical Science, Life Science, and Earth and Space Science have been retained, and although they have undergone some modifications, these still retain their recognizable character. The major change lies in the intention that students are expected to understand science principles and be able to apply their science knowledge.

In addition, both reform documents recommend students understand and develop appropriate skills related to inquiry, the nature of science, and key principles from the historical development of key scientific ideas. Inquiry, the nature of science, and the history of science ideas have had limited emphasis in science classes and on science assessments. Furthermore, these tend to remain marginalized in textbooks and university science teacher education programs, where the conceptual content of science dominates. Therefore, at the present time, most teachers lack both the knowledge base and the instructional resources to incorporate this content into their instruction. Both reform documents place renewed emphasis on technology to show the place of technology in our world and to enable more application of science knowledge in daily life.

Project 2061 authors used five major criteria in determining what should be included as science content in their recommendations. These were: utility, social responsibility, intrinsic value of the knowledge, philosophical value, and childhood enrichment

(AAAS, 1989, pp. xix – xx). Their emphasis on utility and social responsibility provided an important foundation for application of science knowledge by learners. In *NSES*, the emphasis on science and technology and science in personal and social perspectives represents a strong stance for moving the science curriculum in schools to greater emphasis on applications of science. Table 1 shows how these two content dimensions included in *NSES* are matched by equivalent features in Project 2061. The language is different, but the intent is the same—a strong recommendation is made that educators should foster understandings of technology as a means of deepening the utilization of science knowledge in comprehending, and acting effectively on, important societal affairs, including health and environment.

Science for All Americans also emphasizes the need "to teach less content in order to teach it better" (AAAS, 1989, p. xviii) and *NSES* echoes this recommendation. Which of the science understandings and abilities contained in the *NSES* and *Science for All Americans* shall be contained in the NAEP Science Framework? How shall the differences among the standards be identified systematically?

> ***Issue 2:*** *On what basis should decisions be made by the Framework developers about inclusion of content from the National Science Education Standards and Science for All Americans?*
>
> ***Recommendation:*** *As the members of the Framework Steering and Planning Committees consider their task, a central responsibility will be to comprehend the emphasis given to the different content of science as it is defined in the National Science Education Standards and in Project 2061 documents. As a beginning point, members of the Framework Committees will need to become knowledgeable about the recommendations contained in NSES and Project 2061 Benchmarks for the elementary, middle, and high schools, and learn how the new, broader scope of science content knowledge, its applications, and related inquiry skills are placed in a developmental framework. The depth and complexity of this task should not be underestimated. The criteria used by Rutherford and Ahlgren (utility, social responsibility, intrinsic value of the knowledge, philosophical value and childhood enrichment) have a role in decisions by the Framework Steering and Planning Committees, as it did for their predecessors.*
>
> ***Issue 3:*** *An additional important issue in development of the Framework will be the inclusion of dimensions of science literacy that appear in the National Science Education Standards and Project 2061 documents. Examples of these dimensions are nature of science, habits of mind, and science in personal and social perspectives.*
>
> ***Recommendation:*** *Dimensions will need to be defined carefully and criteria used to make decisions about their inclusion as part of the Framework explicit.*

> ***Issue 4:*** *What influence should expectations for students' understandings and abilities expressed in the National Science Education Standards and Project 2061 documents have on the NAEP Science Framework?*
>
> ***Recommendation:*** *Framework Steering and Planning Committees members should give careful attention to the recommendations contained in NSES and Project 2061 as they conduct their work.*

Influence of international assessments, TIMSS and PISA—The American public, policy makers, private sector executives, and school personnel have high levels of interest in international comparative assessments. For science, this interest centers on Trends in Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA). This interest likely will intensify with release of TIMSS and PISA 2003 results in December 2004 and with science being the primary domain for PISA 2006.

To provide an understanding of issues associated with the identification and selection of scientific understandings and abilities for the NAEP Framework, it is necessary to summarize briefly the orientation and emphasis for TIMSS and PISA.

The *Trends in International Mathematics and Science Study* (TIMSS) is conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). In 1995, TIMSS assessed the mathematics and science achievement of a half-million students in over 40 countries at three grade levels, which correspond roughly to the U.S. 4^{th} and 8^{th} grades and the last year of secondary school. TIMSS 1995 included several components: a curriculum analysis of 50 countries; a videotape study of 8^{th} grade mathematics classrooms in Germany, Japan, and the United States; a case study analysis of Germany, Japan, and the United States; and surveys of teachers and students to explore the context in which learning and teaching take place. The curriculum analysis related the mediocre performance of U.S. students to U.S. mathematics and science curricula, which it described as "a mile wide and an inch deep. (Schmidt, McKnight & Raizen, 1997)"

TIMSS was repeated in 1999 for grade 8 when 38 countries participated. TIMSS was conducted again in 2003 when 25 countries participated at grade 4 and 48 countries participated at grade 8. One distinguishing aspect of TIMSS is its use of school science and mathematics curricula as a foundation for the assessment. The background discussion uses a model of the intended, taught, and learned curriculum. For example, there is an analysis of frameworks and standards (the intended curricula), a video analysis of teaching and review of textbooks (the taught curricula), and the TIMSS assessment (the learned curricula).

The *Programme for International Student Assessment* (PISA) of the Organisation for Economic Co-operation and Development (OECD) is designed to assess how well 15-year-old students apply and use what they have learned both inside and outside of school.

PISA is the capstone of a large OECD education indicators program. PISA's outcome measures serve as a cross-national comparison on indicators of the quality and potential of entrants to a nation's workforce. PISA surveys mathematics, reading, and scientific literacy every three years, with one domain as a primary focus in each cycle. In 2000, PISA assessed over 250,000 students in 32 countries (including 28 OECD member countries). Literacy was the primary domain. In 2003 mathematics was the primary domain, and in 2006 science will be the primary domain. PISA also administers student and principal background questionnaires to explore the social and economic context of the learning environment and students' attitudes toward learning.

The unique feature of PISA centers on its orientation toward the general literacy of the students. It is not primarily an assessment based on school science programs. To clarify issues and make recommendations about scientific understandings and abilities for the NAEP Framework, it is necessary to compare those components of the TIMSS and PISA Frameworks and assessments. The TIMSS 2003 Framework and PISA 2006 Scientific Literacy Framework best serve this purpose in reference to a new NAEP Framework. Table 2 presents the content domains for both TIMSS and PISA.

Table 2. Content Domains for TIMSS 2003 and PISA 2006

TIMSS	*PISA*
• LIFE SCIENCE	• LIVING SYSTEMS
• PHYSICAL SCIENCE Chemistry Physics	• PHYSICAL SYSTEMS
• EARTH SCIENCE	• EARTH and SPACE SYSTEMS
• ENVIRONMENTAL SCIENCE	

At the broad level, content domains for the two international assessments generally are similar. It seems appropriate for this review to examine the conceptual understandings for the two assessments. The following Tables (3 through 5) describe the understandings for the domains of life, physical, and Earth sciences. To the degree possible, parallel understandings are identified.

Table 3. Life Science Understandings for TIMSS and PISA

*TIMSS**	*PISA*
• Types, characteristics, and classification of living thing	
• Structure, function, and life processes in organisms	
• Cells and their functions	• Cells (e.g., structures and function, DNA, plant and animal)
• Development and life cycles of organisms	
• Reproduction and heredity	
• Diversity, adaptation, and natural selection	• Populations (e.g., species, evolution, biodiversity, genetic variation)
• Ecosystems	• Ecosystems (e.g., food chains, matter and energy flow)
• Human health	• Humans (e.g., health, nutrition, subsystems [i.e., digestion, respiration, circulation, excretion, and their relationships], disease, reproduction)
• Changes in population ***** • Use and conservation of natural resources***** • Changes in environments*****	• Biosphere (e.g., ecosystem services, sustainability)

*****Environmental Science for TIMSS 2003 is included in this Figure.

Table 4. Physical Science Understandings for TIMSS and PISA

TIMSS	*PISA*
CHEMISTRY	
• Classification and composition of matter • Particulate structure of matter • Properties and uses of water	• structure and properties of matter (e.g., thermal and electrical conductivity)
• Acids and bases • Chemical change	• chemical changes of matter (e.g., reactions, energy transfer, acids/bases)
PHYSICS	
• Physical states and changes in matter	• physical changes of matter (e.g., states of matter, elements, bonds)
• Energy types, sources, and conversions	• energy and its transformation (e.g., conservation, dissipation, chemical reactions)
• Heat and temperature • Light • Sound and vibration • Electricity and magnetism	• interactions of energy and matter (e.g., light and radio waves, sound and seismic waves)
• Forces and motion	• motions and forces (e.g., velocity, friction)

Table 5. Earth Science Understandings for TIMSS and PISA

TIMSS	*PISA*
• Earth's structure and physical features (lithosphere, hydrosphere and atmosphere)	• structures of the Earth systems (e.g., lithosphere, atmosphere, hydrosphere)
• Earth's processes, cycles, and history	• change in Earth systems (e.g., plate tectonics, geochemical cycles, constructive and destructive forces) • Earth's history (e.g., fossils, origin and evolution) • energy in the Earth systems (e.g., sources, global climate
• Earth in the solar system and the universe	• Earth in space (e.g., gravity, solar systems)

PISA 2006 also will include knowledge *about* science. The primary categories of these understandings include: scientific inquiry, scientific explanation, and science and technology in society. For PISA 2006 this category will have equal "weight" with students' understanding of scientific knowledge (see Table 6).

Table 6. PISA 2006 Knowledge About Science

Scientific Enquiry
• origin (scientific questions). • purpose (e.g., to produce evidence that helps answer scientific questions, current ideas/models/theories guide enquiries). • observations and experiments (e.g., different questions suggest different scientific investigations, current scientific knowledge). • data (e.g., quantitative [measurements], qualitative [observations]). • measurement (e.g., inherent uncertainty, replicability, variation, accuracy/precision in equipment and procedures). • characteristics of results (e.g., empirical, tentative, testable, falsifiable, self-correcting).
Scientific Explanations
• types (e.g., hypothesis, theory, model, law). • formation (e.g., extant knowledge and new evidence, creativity and imagination, logic). • rules (e.g., logically consistent, based on evidence, based on historical and current knowledge). • outcomes (e.g., new knowledge, new methods, new technologies, new investigations).
Science and Technology in Society
• role of science (e.g., understand the natural world, answers questions) and role of science-based technology (e.g., attempts to solve human problems, develop artifacts, design processes, human adaptation [non-biological]). • relationships between science and technology (e.g., science often advances due to new technologies, advances in scientific knowledge can advance technology). • risks (e.g., may create new problems, knowledge is often not public, benefits versus costs, unintended consequences). • influence (e.g., science and technology influence society through their knowledge, procedures, products, and world views). • challenges (e.g., societal issues and aspirations often inspire questions for scientific research and problems for technological innovations). • limits (e.g., science cannot answer all questions and technology cannot solve all societal problems or meet all human aspirations).

For TIMSS the assessment of scientific understandings and abilities are joined under the category of cognitive domains. Those domains include: factual knowledge, conceptual understanding, and reasoning and analysis. The TIMSS Framework

> … is based on the idea of science as a process used to learn about the physical world that involves observation, description, investigation, and explanation of natural phenomena.
> As such, it includes both demonstration of content knowledge and the ability to apply and communicate understanding of concepts in solving problems, developing explanations, and conducting and reporting results of investigations.
> (Mullis, et al., 2001, p. 61)

PISA 2006 gives priority to scientific competencies. The primary categories of competencies include: identifying scientific questions, explaining phenomenon scientifically, and using scientific evidence. Within these three major categories for PISA one can identify most of the cognitive domains of TIMSS; however, the contexts of the two assessments vary considerably. TIMSS has a school science curriculum emphasis and requirements grounded in understandings of scientific disciplines and processes, while PISA has a "literacy" emphasis of application of knowledge and cognitive abilities to situations of life and living.

> ***Issue 5:*** *Which of the science understandings assessed by TIMSS and PISA should be contained in the NAEP Science Framework?*
>
> ***Recommendation:*** *Frameworks for both TIMSS 2003 and PISA 2006 should inform decisions about the NAEP Science Framework. Consideration should be given to the conventional content domains, which include: Life, Earth and Physical Sciences and the more specific content classifications, especially for the upper grade levels of NAEP, 8th and 12th. In addition, consideration should be given to knowledge about science as represented in the PISA 2006 Science Framework.*
>
> ***Issue 6:*** *Which of the scientific abilities assessed by TIMSS and PISA should be contained in the NAEP Science Framework?*
>
> ***Recommendation:*** *The hierarchy presented in TIMSS (i.e., factual knowledge, conceptual understanding, and reasoning and analysis) should be given consideration. As a complement to this view, the organizing competencies (i.e., identifying scientific questions, explaining phenomena scientifically, and using scientific evidence) should be considered as an organizing framework for assessing cognitive abilities.*

> ***Issue 7:*** *On what basis should decisions be made about the inclusion of content from TIMSS and PISA in the NAEP Science Framework?*
>
> ***Recommendation:*** *Primary consideration should be given to content that is common to TIMSS and PISA.*
>
> ***Issue 8:*** *What terms should be used for topics and abilities in the NAEP Science Framework?*
>
> ***Recommendation:*** *Consideration should be given to the terms used in earlier NAEP Frameworks and the common terms used in NAEP and TIMSS (i.e., Life Science, Earth Science, and Physical Science). These terms also are generally consistent with PISA Science Content. In addition, consideration should be given to the Nature of Science category or, for PISA, the Knowledge about Science. For cognitive abilities, consideration should be given to the creation of new categories that represent a synthesis of cognitive domains from TIMSS and competencies from PISA.*

Influence of state standards and assessments—A first look at science education in the fifty states gives an impression of commonality. Forty-nine states have developed science content standards and most of these states have a form of statewide assessment in place. This commonality extends to the traditional content of science. All states show agreement on the inclusion of many familiar concepts in biological, physical, Earth, and space science. Inquiry may be included as part of policies guiding science instruction, and may be an expectation for what students should learn, but may not be measured on state assessments. However, beyond these initial areas of agreement, the focus of state standards varies greatly. For example, the Michigan Curriculum Framework in Science (developed with close attention to the *Benchmarks for Science Literacy* (AAAS, 1993) has been reinforced by a statewide testing program that has compatible goals and structure. More recently, incentives for high achievement on the testing program include scholarships for students funded by the State of Michigan. In contrast, California has modified its science standards and its testing program over the same period. The state has moved from curricular guidelines compatible with the *National Science Education Standards* and Project 2061 recommendations to a more traditional set of guidelines. A traditional model of testing was recently adopted after installing a performance-based approach. These changes have sent confusing signals to teachers, students, parents, and administrators. Other states appear to lie between these two in their implementation of the recommendations that are underscored in the *National Science Education Standards* and Project 2061 documents.

> ***Issue 9:*** *How should state standards and assessment frameworks influence the Framework Committees' decisions?*

> ***Recommendation:*** *Informed decisions require members of the Framework Committees to be knowledgeable about what expectations for science literacy state standards and assessments have in common with the National Science Education Standards, Benchmarks for Science Literacy, and the TIMSS and PISA Frameworks.*

Framework Content Specific Issues

Inquiry—Inquiry is a valued component of science literacy. It is mentioned in standards and in assessment frameworks. It is a complex and multifaceted concept. Inquiry defines the disciplines of the natural sciences. It encompasses the methods used by scientists to study the natural world. The product of inquiry in the natural sciences is enhanced understanding of the natural world. As practiced by natural scientists, inquiry is characterized not so much by processes as by its philosophical foundations, (especially its epistemology or perspectives on what constitutes scientific knowledge).

Inquiry is an approach used by scientifically-literate adults to pursue questions about the natural world. It is an instructional strategy and a valued outcome of science education. Teachers use inquiry as instructional strategy to develop students' understanding of the natural world. Inquiry is also an expectation for student achievement including:

- understanding inquiry as it is practiced by scientists;
- understanding inquiry as the method they use to investigate their natural world; and
- being able to inquire at a level of competence appropriate to their age.

Unpacking inquiry into its component abilities such as observing, controlling variables, hypothesizing, thinking critically, and developing well-reasoned arguments illustrates further its complexity. While there is little question that inquiry will be measured in the science assessments, the challenge is the decision about the elements of inquiry that will be measured and how they will be measured (Baxter & Glaser, 1998; Champagne & Newell, 1994; Champagne, Kouba, & Hurley, 2000; Glaser & Baxter, 1999; Duschl, 2003).

> ***Issue 10:*** *A concept as complex as inquiry presents challenges to developing unambiguous specification as a construct that will be measured.*
>
> ***Recommendation:*** *The NAEP Science Framework must communicate unambiguously those elements of inquiry that will be measured in the assessment.*
>
> ***Issue 11:*** *What views of science inquiry should be included in the Science Framework?*

> ***Recommendation:*** *Framework Committee members' decisions regarding inquiry in the Science Framework should be informed by inquiry in U.S. national and state standards, and in the TIMSS and PISA Frameworks.*

Expectations of the public, post-secondary education, the private sector, and the military—The value stakeholders place on science education is well documented as is concern that far too many high school graduates have not met the expectations necessary for success in post-secondary education, the private sector or the military (National Research Council (2001a); National Association of System Heads, 2002; Congressional Commission on the Advancement of Women and Minorities in Science, Engineering and Technology Development in their 2000 report). Policy analysts and economists call for education to give more attention to the requirements of the workplace (Coplin, 2004) while higher education administrators decry the need of entering freshman for remedial courses (American Diploma Project, 2004). Descriptions of the science required for existing and future jobs are scant. However reports such *Report for America 2000 (The Secretary's Commission on Achieving Necessary Skills* (U.S. Department of Labor, 1991) provide some insights. The skills associated with science are those necessary to 1) "put knowledge to work," 2) acquire and use information (including collecting, evaluating, interpreting and communicating data), 3) master complex systems, and 4) work with various technologies.

The disconnect between what students learn in K-12 science and the expectations and demands of post-secondary education is documented in reports such as A*ligning K-12 and Postsecondary Expectations: State Policy in Transition* (U.S. Department of Education, Office of Vocational and Adult Education, 2002*)* which illustrate the disparity between what students must do in science to earn a high school diploma and what colleges/universities require in science for admission. That the science expectations of these stakeholders should be measured is clear. The question is how should the expectations of stakeholders be represented in the content of the Science Framework?

> ***Issue 12:*** *To what extent should the understandings and abilities necessary for entry from high school into post-secondary education, the workforce, and the military influence the Science Framework?*

> ***Recommendation:*** *Expectations of various stakeholders should be given consideration by the Framework Committees.*

Influence of the Cognitive and Measurement Sciences

Knowing What Students Know: the Science and Design of Educational Assessment, (Pelligrino, 2001) and *How People Learn: Brain, Mind, Experience, and School* (Bransford, Brown, & Cocking, 2000) are reports of the National Research Council that bring current developments in the cognitive and measurement sciences to the design of assessments. The reports point out that assessments have different purposes (e.g., accountability and informing classroom practices) and different loci of control (e.g., international, federal, state, district, and classrooms). As a consequence, assessments have certain characteristics that are unique to purpose and locus of control; however, models of cognition are common to all well-designed assessments.

Models of cognition—Models of cognition derive from contemporary theories of knowing and include "the way knowledge is represented, organized, and processed in the mind." (NRC, p.3) The report characterizes assessment as reasoning from data. Data result from students' interaction with items and tasks. Items and tasks are posed and students respond to the tasks in ways that are observable. The observations of students' responses (performances, or behaviors) are interpreted using a model of cognition. The report suggests that "[A]ssessments of academic achievement need to consider carefully the knowledge and skills required to understand and answer a question or solve a problem, including the context in which it is presented" (Pelligrino, 2001, pp. 4-5).

Understanding and ability describe two types of student achievement. That the two are sides of the same coin is illustrated by the components of cognitive models that are used to interpret students' responses to tasks designed to elicit understanding and ability. A cognitive model developed by Shavelson and his colleagues (see Table 7) proposes four types of knowledge that allow the interpretation of students' science responses to items and tasks (Li & Shavelson, 2001; Shavelson & Ruiz-Primo, 1999). The knowledge types are declarative, procedural, schematic and strategic. If understanding atomic-molecular theory is an expectation for K-12 science, then understanding can be defined in terms of the types of knowledge contained in the Shavelson model and students' performance on tasks interpreted in terms of the four types of knowledge

Table 7. Shavelson Cognitive Model

Shavelson Cognitive Model

- Declarative Knowledge is "knowing that," including scientific definitions and facts, mostly in the forms of terms, statements, descriptions, or data.
- Procedural Knowledge is "knowing how" to do science, e.g., how to design a study that manipulates one relevant variable and controls others; how to follow a series of if-then production rules or a sequence of operations (measurements or procedures) to achieve a particular goal.
- Schematic Knowledge is "knowing why," e.g., how principles, schemes, and mental models that are based on scientifically justifiable "theory" or "conceptions" explain the physical world.
- Strategic Knowledge is "knowing when, where, and how" to apply domain-specific knowledge and strategies to solve a unique scientific problems or approach a new situation.

(CAESL Assessment Model p. 11)

While the components of the Shavelson cognitive model are different types of knowledge, mental processes (information processing) also are components of cognitive models. Reference to the knowledge and processes that comprise cognitive models appear in some science standards and assessment frameworks, For instance, the TIMSS Science Framework identifies Science Cognitive Domains: factual knowledge; conceptual understanding; reasoning and analysis (Mullis, et al., 2001, p. 61) and are assessed across age and natural science domains. The correspondence of the components of the TIMSS cognitive domains and the components of cognitive models requires explication so that the cognitive research can inform the conceptualization and design of the NAEP Science Framework.

> ***Issue 13:*** *Which cognitive dimensions should be a part of the NAEP Science Framework?*
>
> ***Recommendation:*** *Interpretation of students' test performance in terms of cognitive models cuts across the natural science disciplines. The abilities of inquiry add a useful dimension to the information gained from the assessment and should be given serious consideration for inclusion in the Framework.*

Naïve conceptions—Over the past quarter century, much information has been recorded about common naïve conceptions and misconceptions that children (and adults) harbor about science principles and processes (Harvard-Smithsonian Center for Astrophysics, Science Education Department, 1987). These conceptions are resistant to change even when instruction is exemplary. Examples of naïve conceptions prevalent in the general public:

- Substances (water, for instance) expand when they are heated because the molecules get bigger.
- Phases of the moon are the result of the Earth's shadow.
- An object in motion (a baseball just hit by a bat, for instance) continues in motion until the force of the bat is used up.
- Characteristics acquired by parents (dogs' clipped ears or tails, for instance) can be inherited by their offspring.

> ***Issue 14:*** *Items assessing naïve conceptions are so difficult that they seldom appear on operational tests. However, changing these conceptions is an important goal of science education.*
>
> ***Recommendation:*** *In the design of the NAEP Science Framework, attention should be given to the body of research on naïve conceptions and its implications for testing.*

Learning trajectories—Inquiry, evolution, and kinetic molecular theory are three "big ideas" of the natural sciences. Students' understanding of these ideas develops across the K-12 experience. However, the big ideas typically appear explicitly in standards, test frameworks, and assessments only late in the K-12 sequence. So, for instance, "understanding kinetic molecular theory" typically appears in grade 9-12 standards and on assessments administered at grade 12. However, components essential to the understanding the theory develop early in school science. Children observe water "disappearing," from a pan being heated on the stove and water droplets "appearing" on the outside of glasses of ice water. They notice the relationships between warm and cold and the behavior of water. They develop models of water, warmth, and cold that they use to make sense of their observations. They reason that the water on the outside of the glass came from inside the glass. But their reasoning is challenged by the observation that droplets don't form on a glass of water that is room temperature. Does the water really disappear? If so, where did the water droplets come from when a cover is put on the pot, and why doesn't the water continue disappearing when the cover is on?

These observations, models of matter, warmth and cold, are foundations of the sophisticated understandings of kinetic-molecular theory. Water is composed of molecules, they are in motion, and some have sufficient energy to escape from the surface of the water. This model of matter allows us to explain the observation that water evaporates from open containers. Understanding temperature as a measure of the average kinetic energy of the molecules, provides a model for explaining why the rate at which water evaporates is temperature dependent. The higher the temperature of water the greater the rate of evaporation. This simple description illustrates that at different points along the learning continuum the understandings and abilities that need to be assessed are fundamentally different. Assessments that trace the trajectories of the development of big ideas over the K-12 years would provide valuable information about students' progress[6] (Roberts, Wilson, & Draney, June 1997; Wilson, 2000).

> ***Issue 15:*** *A single framework for multiple assessments administered at different grades does not distinguish the fundamentally different content that is to be measured across grade levels.*
>
> ***Recommendation:*** *Consider developing the Framework so that the knowledge, understandings and abilities that are components of the big ideas are assessed across grade level, and students' progress toward achieving them can be monitored.*

Evolving nature of science inquiry —Science inquiry is an object of study by researchers in many disciplines including historians, philosophers, sociologists,

[6] Two papers were written for the National Research Council's Committee on Test Design for K-12 Science Achievement that discuss learning theory and psychometrics of tracing learning trajectories. These are not currently available but should be available when the Framework Committees begin their work. The papers: M.D. Reckase and J. Martineau, The Vertical Scaling of Science Achievement Tests and C. Anderson, C. Smith, M. Wiser, J. Krajcik, Implications of Research on Children's Learning for Assessment

cognitive psychologists, linguists, and others (Kuhn, 1996; Kuhn, 2000; Longino, 2002; Longino, 1990; Magnani, Nersessian, & Thagard, 1999; Pickering, ed.,1992; Pintrich, Marx, & Boyle, 1993).

Their evolving descriptions of the nature of science inquiry will have implications for the practice of science education and expectations for students' learning about the nature of science inquiry. The evolving perspectives imply expectations focusing on social, conceptual, and epistemic understanding and abilities, specifically the abilities to communicate scientific ideas, to reason scientifically, and to assess the epistemic status that can be attached to scientific claims, theories, and models.

> ***Issue 16:*** *The nature of scientific inquiry has and will continue to evolve as humans continue to learn how to learn. Normative features of scientific inquiry have extended beyond pure experimentation and now include theory building and revision, and model building and revision.*
>
> ***Recommendation:*** *The Framework Committees should become familiar with the research on the evolving nature of science inquiry and consider its implications for the Science Framework.*

Construct under-representation —Messick (1993) defines construct under-representation as the extent to which a test may fail to measure important dimensions of the intended construct. A construct is the “it” that is to be measured. In the context of Science NAEP the “it” is science literacy. Science literacy, in turn, is comprised of other constructs including principles of chemistry, physics, life and Earth sciences; inquiry; history and philosophy of science; and engineering design. Each of these constructs can be further analyzed into component constructs (e. g., see the analysis of inquiry). Construct under-representation characterizes a measurement (test) that does not fully sample the content (facts and principles) or evoke the reasoning processes that comprise the construct to be measured. To omit a facet of science literacy in the Science Framework would be to under-represent it. However, to attempt to measure too many facets would be to create construct-irrelevant variance in estimating what students are learning in their classes. (Shavelson, personal communication)

Coherence—Coherence is a critical assessment principle identified in *Knowing What Students Know* (NRC, 2001). It calls for an assessment system composed of standards and assessments all of which have their foundations in the same expectations of what students should understand and be able to do. Maintaining coherence of standards and assessments is a responsibility each state must meet. What are the challenges of meeting this responsibility and how does Science NAEP influence the challenge?

Perspectives on expectations for what students should know and be able to do come to states from national standards, international assessment frameworks, the private sector, post-secondary education, and the military. Of these, the Science NAEP will be a major reference point. Inevitably, states will compare their expectations for what students should understand and be able to do against the NAEP Science Framework and any information available about the form and content of the Science Assessments developed from the Framework. In those states where international comparisons are important to the private sector, the frameworks for these assessments will be scrutinized. To the degree that national standards, international assessment frameworks, and the Science NAEP are consistent in their attention to the facets of science literacy, they will contribute to coherence in states' assessment systems. However, lack of coherence can be a motivation for change in assessment systems. At issue is how should information about variability and commonality across states and school districts influence decisions about content to be included in the NAEP Science Framework?

> ***Issue 17:*** *Coherence is a desirable attribute of the nation's assessment system and of each state's assessment system. The NAEP Science Framework will play an important role in establishing coherence. The NAEP Science Framework will contribute to system coherence to the degree its contents are consistent with perspectives on science literacy common to national, state, and international standards and frameworks.*
>
> ***Recommendation:*** *Framework Committee members must be well informed about the commonalities and differences among the views of science literacy contained in U.S. and international documents and keep the importance of coherence in mind as they select the Framework content.*

The Framework's Form

In addition to making recommendations about the content of the Framework, the Committees are charged with making recommendations about the representation of the Framework content. The form of the Framework must follow its function, to effectively communicate the content that will be assessed.

The current NAEP Science Framework represented the content as a matrix (see Figure 2). Two foundational strands, themes and the nature of science, are at the base of the matrix. The content in the matrix is organized around two primary dimensions each of which has three components, Knowing and Doing (Conceptual Understanding, Scientific Investigation, Practical Reasoning) and Fields of Science (Earth, Physical, Life). Frameworks for TIMSS and PISA organize the content differently and use different words to describe the dimensions of content. A primary consideration of Framework Committee members in making decisions about the Framework's dimensions and form must be that it communicates unambiguously to many audiences.

Figure 2. The 1996 – 2005 NAEP Science Framework Matrix

Knowing And Doing	**Fields of Science: Earth**	**Physical**	**Life**
Conceptual Understanding			
Scientific Investigation			
Practical Reasoning			
	Nature of Science		
	Themes Systems, Models, Patterns of Change		

In addition, the Framework must communicate the content that will be assessed at each grade level. Consequently, the correspondence of the Framework's form to the assessment it will be used to design is another matter for the Committees' consideration. Previously a single Science Framework was used to design three assessments (grade 4, grade 8, and grade 12). Developments in cognitive and measurement theories especially the importance of tracing trajectories of students' learning of inquiry and the big ideas of science motivate the consideration of how the form of the framework will communicate significant differences in the content that will be assessed at each grade level. Can a single framework design communicate the differences effectively or should consideration be given to individual framework designs for each grade assessed by NAEP?

> ***Issue 18:*** *What form should the Framework take?*
>
> ***Recommendation:*** *The Framework Committees must carefully review the effectiveness of extant state science assessment frameworks, to communicate the content of science to be assessed at three grade levels, (4, 8, and 12) as they develop recommendations for the NAEP Science Framework.*

CONCLUDING REMARKS

The Framework Committees are charged with recommending the science content that will be assessed by NAEP. Committee members face the challenging task of making choices of the content from the rich array of understandings and abilities that comprise science literacy as it is described by organizations representing U.S. and international science education communities. The choices will be influenced by the Committee members' professional judgment of what science education should be. The members must allow their professional judgment to be challenged by reason.

Should *all* the content proposed by the U.S. and international science and science education communities be a part of the Framework? What principles will be applied to determine if the proposed content is too broad? If the content is deemed too broad, what principles should be applied to make the decisions about the content to be excluded? Will the principled choices be based on the content common across content proposed by the different organizations? Or on demands of the public, higher education, the private sector, and the military? Or on the requirements of active citizenship, personal well being, productivity in the workplace, preparation for post secondary education? How will these potentially conflicting possibilities be prioritized?

The decision principles must derive from the current state of science education and the future promise of developments in cognitive and measurement theory. All decision principles must be moderated by the purpose of the National Assessment of Education Progress and conscientious consideration of the effects of a Framework too far ahead of U.S. science education.

REFERENCES

American Association for the Advancement of Science. (1989). *Science for all Americans*. Washington, DC: American Association for the Advancement of Science.

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

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

American Diploma Project (*Ready or Not: Creating a High School Diploma That Counts* (www.americandiplomaproject.org)

Baxter, G.P., and Glaser, R. (1998). Investigating the cognitive complexity of science assessments. *Educational Measurement: Research and Practice, 17*(3), 37-45.

Bransford, J.D., Brown, A.L., & Cocking, R.R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Champagne, A. B. & Newell, S. (1994). Directions for Research and Development: Alternative Methods of Assessing Scientific Literacy. *Journal of Research in Science Teaching*, 29, 841-860.

Champagne, A. B., Kouba, V. L. & Hurley, M., (2000). Assessing Inquiry in J.Minstrell and E.H. Van Zee (Eds.) *Inquiring into inquiry learning and teaching in science.* Washington, D.C.: American Association for the Advancement of Science. Congressional Commission on the Advancement of Women and Minorities in Science, Engineering and Technology Development in their 2000 Report *Land of Plenty: Diversity as America's Competitive Edge in Science, Engineering and Technology*

Coplin, B. Education reform should keep the workplace in mind. (October, 21, 2004). USA Today.

Duschl, R. (2003). Assessment of inquiry. In J.M. Atkin & J. Coffey (Eds.) *Everyday assessment in the science classroom*. Washington, DC: NSTA Press.

Glaser, R., and Baxter, G.P. (1999). *Assessing active knowledge*. Paper presented at the 1999 CRESST Conference, Benchmarks for Accountability: Are We There Yet?, September 16-17, UCLA, Los Angeles, CA.

Harvard-Smithsonian Center for Astrophysics, Science Education Department. (1987). *A private universe* (Video). Cambridge, MA: Science Media Group.

Hollweg, K. S., and Hill, D. (Eds.). (2003). What Is the Influence of the National Science Education Standards? Reviewing the Evidence, A Workshop Summary. Washington, DC: National Academies Press.

Kuhn, T. (1996). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press. (4th ed)

---------(2000). *The Road since Structure: Philosophical Essays, 1970-1993 (with an autobiographical interview).* Edited by James Conant and John Haugeland. Chicago: University of Chicago Press.

Li, M. & Shavelson, R. J. (2001). *Examining the linkage between science achievement and assessment.* Paper presented at the annual meeting of the American Educational Research Association, Seattle: WA.

Longino, H. (2002). *The Fate of Knowledge,* Princeton: Princeton University Press

_________(1990). *Science as Social Knowledge*. Princeton, NJ: Princeton University Press.

Magnani, L, N. Nersessian, & P. Thagard, eds. (1999). *Model-based Reasoning in Scientific Discovery.* New York: Kluwer Academic/Plenum Publishers.

Messick, S. (1993). Trait equivalence as construct validity of score interpretation across multiple methods of measurement. In R. Bennet & W. Ward (Eds.), *Construction versus choice in cognitive measurement: lssues in constructed response, performance testing, portfolio assessment* (pp. 61-73). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mislevy, R.J. (1993). Foundations of a new test theory. In N. Frederiksen, R.J. Mislevy, and I.I. Bejar (Eds.), *Test theory for a new generation of tests*. Hillsdale, NJ: Erlbaum.

Mislevy, R. (2003). *Overview of the principled assessment design for inquiry (PADI) project*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois.

Mullis, V. S., Martin, M.O., Smith, T. A. Garden, R.A., Gonnzalez, E.J., et al. (2001). *TIMSS assessment frameworks and specifications*. The International Association for the Evolution of Educational Achievement. Boston Massachusetts: The International Study Center, Lynch School of Education, Boston College.

National Assessment Governing Board. (2004). *Science Framework for the 2005 National Assessment of Educational Progress.* Washington, DC: Author.

National Association of System Heads, 2002. A*ligning K-12 and postsecondary expectations: State policy in transition.* Washington, DC: Author.

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

National Research Council. (2000) *Inquiry and the national science education standards. Washington D.C.: National Academy Press.*

National Research Council. (1999a). *Grading the nation's report card: Evaluating NEAP and transforming the assessment of educational progress*. Committee on the Evaluation of National and State Assessments of Educational Progress. J.W. Pellegrino, L.R. Jones, and K.J. Mitchell, (Eds.). Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2001a). *Building a workforce for the information economy*. Committee on Workforce Needs in Information Technology. Board on Testing and Assessment; Board of Science, Technology, and Economic Policy; and Office of Scientific and Engineering Personnel. Washington, DC: National Academy Press.

National Research Council (2001b). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Neidorf, T. S., Marilyn Binkley, M. & Stephens, M. (in press). *A content comparison of the NAEP 2000 and TIMSS 2003science assessments*. National Center for Education Statistics, Institute of Education Sciences. Washington, DC: U.S. Department of Education

Pea, R. (1993), 'Learning Scientific Concepts Through Material and Social Activities: Conversational Analysis Meets Conceptual Change', *Educational Psychologist* 28, 265--277.

Pickering, Andrew, ed. (1992). *Science as Practice and Culture*. Chicago: University of Chicago Press.

Pintrich, P., Marx, R. & Boyle, R. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research, 63*(2), 167-199.

Popham, W.J. (2000). *Modern educational measurement: Practical guidelines for educational leaders*. Needham, MA: Allyn and Bacon.

Posner, G., Strike, K., Hewson, P. & Gertzog, W. (1982). Accommodation of a Scientific Conception Toward a Theory of Conceptual Change. *Science Education, 66*(2), 211-227.

Quellmalz, E., & Haertel, G. (in press). *Breaking the mold: Technology-based assessment in the 21st Century*, Center for Technology in Learning, SRI International. (http://pals.sri.com/)

Roberts, L., Wilson, M., and Draney, K. (1997, June) *The SEPEP assessment system: An overview* Bear Report Series, SA-97-1. University of California, Berkley.

Schauble, L., Glaser, R., Duschl, R., Schultz, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *The Journal of the Learning Sciences, 4*(2) 131-166.

Schmidt, W.H., McKnight, C.C., and Raizen, S.A. (1997). *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Shavelson, R.J., and Ruiz-Primo, M.A. (2000). On the psychometrics of assessing science understanding. In J.J. Mintz, J.H. Wandersee, and J.D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp.303-341). San Diego, CA: Academic Press.

Shepard, L.A. (2000). The role of assessment in a learning culture. Presidential Address presented at the annual meeting of the American Educational Research Association, New Orleans, April 26.

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

Science & Engineering Indicators-2004, National Science Board, National Science Foundation.

Sinatra, G. & Pintrich, P. eds., (2003). *Intentional conceptual change*. Mahwah, NJ, USA: Erlbaum Publishers.

U. S. Department of Labor. (June, 1991). *What work requires of schools: A SCANS report for America 2000, the Secretary's Commission on achieving necessary skills.* Washington, DC: Author.

U. S. Department of Elementary and Secondary Education. (April 28, 2004). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001*. Washington, DC: U. S. Department of Elementary and Secondary Education.

U.S. Commission on National Security/21st Century. (2001). *Road Map for National Security: Imperative for Change.*

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