DOCUMENT RESUME

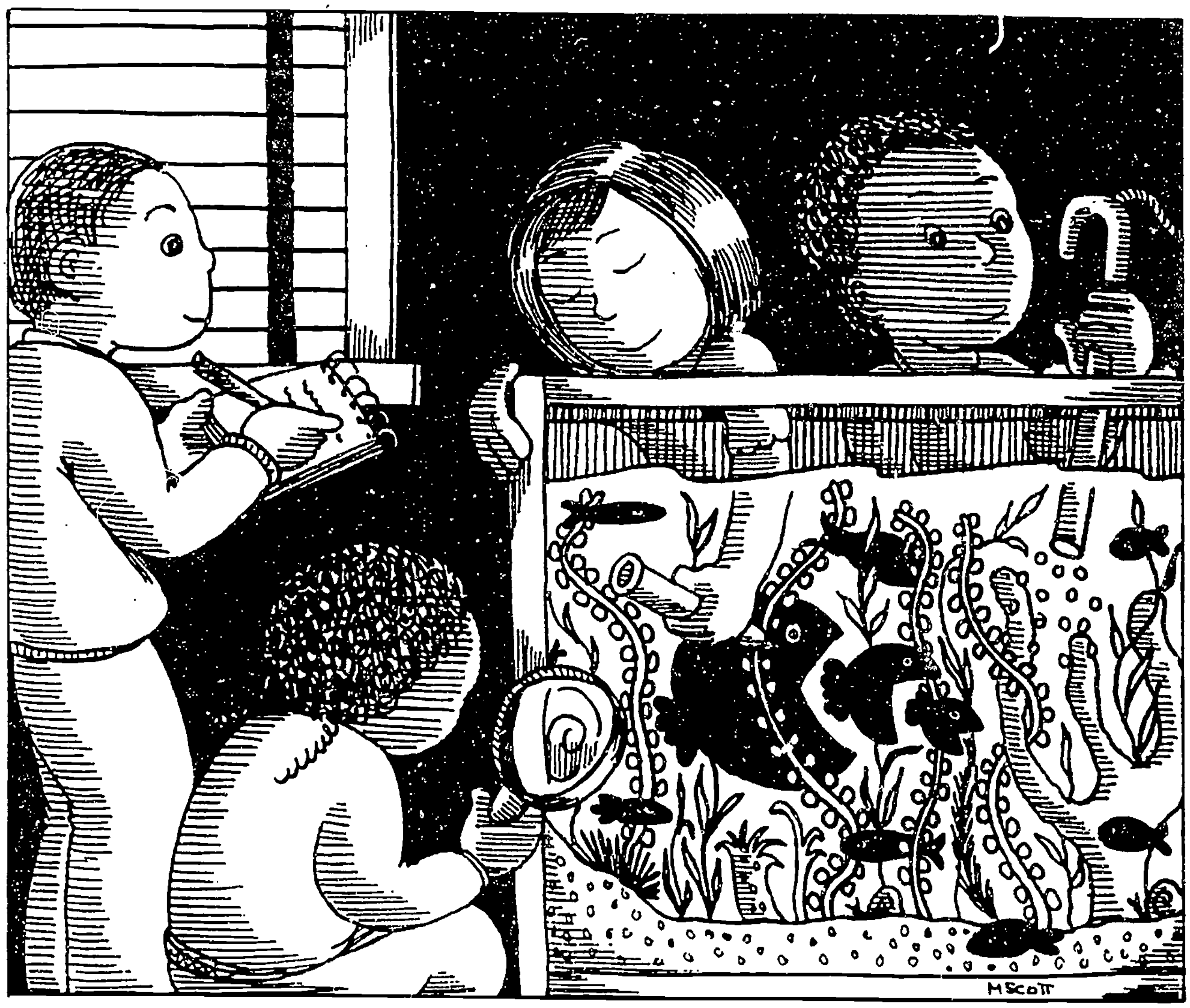
ED 374 020 SE 055 088

TITLE	Science Framework for the 1994 National Assessment of Educational Progress.
INSTITUTION	Council of Chief State School Officers, Washington, D.C.; National Assessment Governing Board, Washington, DC.
PUB DATE	94
NOTE	89p.; A product of the NAEP Science Consensus Project.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Academic Achievement; Achievement Tests; Biological Sciences; Elementary Secondary Education; Physical Sciences; *Science Education; Science Tests; *Student Evaluation
IDE TIFIERS	*National Assessment of Educational Progress; *Science Achievement

ABSTRACT

This document contains the framework and rationale for assessing science achievement of students throughout the United States in 1994. It provides a general overview of the National Assessment of Educational Progress (NAEP), describes the 1994 NAEP Science Framework adopted by the National Assessment Governing Board, and reviews the process by which the Framework was developed. The following sections are included: (1) Introduction; (2) The Framework for the 1994 NAEP Science Assessment; (3) Desired Attributes of Assessment; (4) Characteristics of Assessment Exercises and Items; and (5) Special Studies and Research. Approximately half of the document is made up of the appendices, which includes examples of themes in science and sample items of the content that students are expected to know. (ZWH)

Science Framework for the 1994 National Assessment of Educational Progress



NAEP Science Consensus Project

Council of Chief State School Officers

National Assessment Governing Board
U.S. Department of Education

What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is the only nationally representative and continuing assessment of what America's students know and can do in various subject area Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history/geography, and other fields. By making objective information on student performance available to policymakers at the national, sta and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individu students and their families.

NAEP is a congressionally mandated project of the National Center for Education Statistics, the U.S. Departme of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations. NAEP reports directly to the Commissioner, who is also responsible for providing continuing reviews, including validation studies and solicitation of public comment, on NAEP' conduct and usefulness.

In 1988, Congress created the National Assessment Governing Board (NAGB) to formulate policy guidelines fc NAEP. The board is responsible for selecting the subject areas to be assessed which may include adding to those specified by Congress; identifying appropriate achievement goals for each age and grade; developing assessment objectives; developing test specifications; designing the assessment methodology; developing guidelines and standards fc data analysis and for reporting and disseminating results; developing standards and procedures for interstate, regional, an national comparisons; improving the form and use of the National Assessment; and e uring that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias.

The National Assessment Governing Board

NOTE: Due to changes in the NAEP budget, this assessment will be given in 1995 or 1996.

TABLE OF CONTENTS

INTRODUCTION

This document contains the framework and rationale for assessing science achievement of students throughout the United States in 1994. It provides a general overview of the National Assessment of Educational Progress (NAEP), describes the 1994 NAEP Science Framework adopted by the National Assessment Governing Board, and reviews the process by which the Framework was developed.

Background

The National Assessment of Educational Progress (NAEP) 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 20 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 biannual. Originally, assessments were of students at ages 9, 13, and 17, but beginning in 1983, they have included students at grades 4, 8, and 12, as well.

NAEP reports provide descriptive information about student performance in various subjects, including basic and higher-order knowledge and 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.

In the past, only results from a national sample of students were reported for each NAEP subject assessed; but in 1987, a national study group chaired by Lamar Alexander, then Governor of Tennessee, and H. Thomas James, President Emeritus of the Spencer Foundation, recommended to the Secretary of Education that NAEP collect representative data on student achievement at the state level. For the first time in 1990, a trial state assessment was conducted for eighth-grade mathematics. The trial will continue in 1992 with state-level assessments in mathematics at grades four and eight and in reading at grade four. State-level assessments beyond 1992 are dependent upon Congressional action.

Assessment is not without critics. The Alexander-James Study Group expressed concern that "national assessments of science were sporadic and almost exclusively devoted to assessing factual information," and that they did not attempt to assess abilities to organize and transform a body of information into a coherent scientific account (*The Nation's Report Card: Improving the Assessment of Student Achievement*, 1987).

To address this criticism, the study group recommended that NAEP broaden its scope to include the collection of information on whether students are able to design, perform, and analyze experiments; on whether they have acquired complex thinking abilities essential to various fields of science; on whether they are able to integrate basic concepts of the various scientific fields; and on whether they can perceive fundamental relationships. To measure these kinds of knowledge and abilities, NAEP was urged to include open-ended items and performance tasks in its assessment techniques.

NAEP has great prestige, considerable influence, and well-publicized results. Therefore, the weight of its findings may have a substantial impact on science education, possibly influencing state curriculum frameworks and, ultimately, even what teachers teach. Therefore, the assessment should reflect consensus on priorities, best practices, and conclusions from research on science education.

A specific example of how NAEP can influence state frameworks is illustrated by the special assessment techniques that are selected, such as hands-on science. Performance, for instance, places a value on the "doing" of science. That priority actually may not be reflected in classroom practice, although it always was considered to be a standard teaching technique in science. Thus, as the NAEP science reports are circulated, states and curriculum policymakers become sensitized and may incorporate more hands-on activities into classroom lessons.

Science educators, by and large, do not quarrel with the essential concept of assessment, but there has been no formal agreement on a common framework, outcomes, or goals and objectives to assess. The 1994 NAEP Science Assessment attempts to reflect a comprehensive, contemporary view of science, so that those affected by the National Assessment are satisfied that it addresses the complex issues in science education without oversimplification. The framers of this document have attempted to tread a fine line between clear communication and technical accuracy in the hope that their efforts represent a step forward in building national consensus on the key outcomes of science education.

Development of the 1994 NAEP Science Assessment Framework

The following factors guided the process for developing consensus on the Science Assessment Framework.

- **The first factor is the general process of consensus development as is set forth in law and is evolving over time.** The process calls for active participation and broad involvement of curriculum specialists, science teachers, local science supervisors, state supervisors, administrators, parents, and representatives of scientific associations, business and industry, government officials, unions, cognitive psychologists, and science educators, as well as participation from the public and private education sectors.

- **The second factor is emphasis on the important outcomes of science education.** As much as possible, this Science Framework represents what is considered essential learning in science, setting the stage for the 1994 and subsequent assessments, and recommending innovative assessment techniques to probe the critical abilities and content areas.

- **The third factor is recognition that the various "players" in education and in industry often hold diverse and sometimes conflicting views.** Further, research and general agreement in the field is lacking. This lack of agreement on a common scope of instruction and sequence, components of scientific literacy, important outcomes of learning, and the nature of overarching themes in science hinders clear communication among science educators and the public.

The process of developing the Framework document and accompanying reports[1], occurred between October 1990 and August 1991. A Steering Committee was established of 19 members (see appendix D) recommended by the education community and private sector. Its members developed the principles that guided creation of the Framework.

A smaller Planning Committee (see appendix E), composed of practitioners, recognized experts in science education, and scientists, was established to identify goals and objectives and to produce the

---

1 *Assessment Specifications Report*, *Student Background Variables*, and *Reporting Formats*.

Framework. Together with the Steering Committee, it was also responsible for suggesting ways to assess important outcomes of science education.

The Planning Committee met monthly from November 1990 through April 1991 and was joined in the first and final meetings by the Steering Committee, which reviewed and reacted to all Framework drafts. Staff of the American Institutes for Research (AIR), a subcontractor, also attended the meetings and, on the basis of this input and with reaction and advice from the committees, AIR formulated specifications for the science assessment.

The NAGB Subject Area Committee #2 and technical staff closely monitored the consensus project work, and Board members were involved in all key phases. In addition, advice was sought from the organizations and sectors that affect and are affected by science education. For instance:

o Opportunities for public input were provided at two hearings (in Washington, D.C., and San Francisco). Opinions were expressed by representatives of institutions (e.g., public and private education and scientific and science education organizations) likely to be affected by or be concerned with the Framework and subsequent assessments.

o The Council of State Science Supervisors was kept abreast of project developments through their electronic communication system, PSInet. This was vital, because state science instruction as well as public opinion may be influenced by the 1994 assessment process. State policymakers and those with responsibility for science education leadership must become familiar with and involved in shaping the process and products of the science consensus effort. Input from the Council of State Science Supervisors was particularly important in helping the 1994 NAEP Science Assessment Project define "big ideas" or themes (see chapter 2) in science learning. For example:

o States with existing frameworks for science instruction were contacted for copies of their frameworks and assessment methods.

o Planning and Steering Committee members hosted sessions at both regional and national science education meetings to report on the development of the Science Assessment Framework.

o The revised draft of the Framework was widely circulated in June 1991 within the science education and science communities for reaction and comments.

Steering Committee Guidelines

At its early meetings, the Steering Committee drafted guidelines for the Planning Committee's work and recommended that the Framework and ensuing science assessment have the following characteristics:

1. The Framework should reflect the best thinking about the knowledge, skills, and competencies needed for a high degree of scientific understanding among all students in the United States of America. Accordingly, it should

   - Encompass knowledge and use of organized factual information, relationships among concepts, major ideas unifying the sciences, and thinking and laboratory skills; and

   - Be based on current understandings from research on teaching, learning, and student performance in science.

2. Both the Framework and the 1994 NAEP Science Assessment should

   - Address the nature and practices of knowing in science, as different from other ways of knowing;

   - Reflect the quantitative aspects of science as well as the concepts of the life, earth, and physical sciences;

   - Deal with issues raised by the role of science and technology in society;

   - Include practical problem solving that involves design, use of materials, and weighing risks in relation to benefits;

   - Take into account the developmental levels of students; and

   - Ensure that students with diverse backgrounds are assessed in ways that provide them with equal and fair opportunities to reflect their knowledge and performance.

3. Assessment formats should be employed that are consistent with the objectives being assessed. A variety of strategies for assessing student performance are advocated, including

- Performance tasks that allow students to manipulate physical objects and draw scientific understandings from the materials before them;
- Open-ended items that provide insights into students' levels of understanding and ability to communicate in the sciences, as well as their ability to generate, rather than simply recognize, information related to scientific concepts and their interconnections;
- Collections of student work over time (such as portfolios) that demonstrate what students can achieve outside the time constraints of a standardized assessment situation; and
- Multiple-choice items that probe students' conceptual understanding and ability to connect ideas in a scientifically sound way.

4. The assessment should contain a broad enough range of items at different levels of proficiency for identifying three achievement levels for each grade.

5. Information on students' demographic and other background characteristics should be collected. Additional information should be collected from students, teachers, and administrators about instructional programs and delivery systems, so that their relationships with student achievement can be ascertained and used to inform program and policy decisions.

The 1994 assessment represents a new base for trendlines, while NAEP uses previous tests for making comparisons to science achievement in past years.

CHAPTER 1

THE NATURE OF SCIENCE AND THE SCIENCE CURRICULUM

The Nature of Science

The various fields of science have their own special ways of knowing, but the essentials of the natural sciences should be defined for the purpose of planning appropriate assessments.

The natural sciences are characterized by organized explanations incorporating both theoretical and empirical elements. Scientists attempt to construct theories that encompass as much factual and conceptual knowledge, including laws and principles, as possible. The construction of theories is a process involving the consideration of factual evidence, insightful questioning, creativity, and imagination.

Atomic theory and evolution theory are good examples of modern scientific theories which are central to the sciences as ways of knowing. They are sources of new hypotheses and logical deductions that can be tested. As these theories are refined, they continue to stimulate new questions and hypotheses. As with other scientific theories, each time they come into play in experimental situations, they are again subject to testing and the possibility of refutation. Each success broadens their domain and increases their usefulness. Verification of theories does not ensure truth, but does extend their usefulness in explaining natural phenomena.

Useful theories enable an individual to make predictions under a specified range of conditions. Scientific theories must be testable, according to standards of evidence and logical argumentation set by the scientific community; they are central to the scientific way of knowing, guiding observation. When scientists observe natural phenomena, their observations are made within the context of existing scientific theories. However, non-scientists holding different views may perceive the same phenomena but, through their subjective belief systems, may arrive at different conclusions. History is replete with examples of misunderstandings and miscommunication based, at least in part, on the use of different rules of observation.

Science consists of both theoretical and experimental knowledge that is constructed by the creativity, knowledge, and world view of scientists. Experimentation plays an essential role in generating and verifying scientific information. New findings and results of scientific observation and experimentation accumulate, enlarging the base of present knowledge and laying the foundation for future learning. Usually, scientific knowledge is structured, forming a web of interrelated concepts, laws, and principles. Whereas sensory data are sometimes ephemeral, concepts are precisely formulated and are multiply connected to other concepts or sensory data within a theory. Concepts without factual content (sensory data) are empty, while sensory data without concepts are difficult to understand and open to misinterpretation.

None of the processes usually associated with science—observation, measuring, classifying, deduction, inference, etc.—is unique to science. However, in science, these processes are given meaning by the context of the subject matter under investigation. Hence, observations of a mealworm seemingly walking endlessly around the sides of a container have new meaning when mealworm anatomy, behavior, and physiology are understood. Likewise, the wiggly lines a physicist observes on a photographic negative taken of a cloud chamber remain wiggly lines to most observers, but they become important sensory data (facts) when the conditions under which the lines were produced are known. A goal of science education, therefore, is to help the student recognize the difference between personal opinion and the knowledge gained through scientific investigation and debate.

Although the scientific disciplines are alike in their reliance on evidence, use of logic, and the organization of factual information into concepts and theories, they differ with respect to what constitutes evidence, specific methods of investigation, and degree of quantification. Yet, "there are common understandings among [scientists] about what constitutes an investigation that is scientifically valid." (American Association for the Advancement of Science, 1989). This view of the nature of science has profound implications for assessment as reflected in the criteria outlined in chapter 3.

The Science Curriculum

Since a major purpose of NAEP is to illuminate education policy, assessment of student science learning must take account of the science curriculum. Hence, the 1994 NAEP Science Assessment Framework is based on a consensus regarding desirable elements of science education against which student attainment is to be measured.

In developing the Framework, the Planning Committee reviewed key blue-ribbon committee reports, examined exemplary practice, studied local and state-based innovation in the science curriculum, reviewed the science education literature, and noted innovations emerging in other countries. Recent reports by government agencies and professional societies (National Science Board, 1983; American Association for the Advancement of Science, 1989; National Science Teachers Association, 1989; National Research Council, 1990) express unanimity in their goals for science education. For example,

- Students should acquire a core of scientific understanding, including organized factual information;
- Students should acquire the ability to relate scientific concepts to each other and to problems they encounter in and out of school;
- Students should be able to apply science knowledge in practical ways;
- Students should be familiar with experimental design and have the ability to carry out scientific experiments that are developmentally appropriate; and
- Students should acquire the science knowledge and understanding that will allow them the opportunity to pursue further study in scientific fields or enter science- or technology-related careers.

There also are similarities in the reports' recommendations for science curricula and instruction needed to achieve these goals, including

- Reduction of the traditional breadth of coverage in favor of greater depth, especially in high school science;
- Emphasis on development of such thinking processes as organizing factual knowledge around major concepts, defining and solving problems, accessing information and reasoning with it, and communicating with others about scientific results and understandings;
- Multi- and interdisciplinary approaches to science teaching;
- Approaches that encourage active student engagement and active participation such as hands-on science; learning in small, cooperative groups; reflecting orally and in writing upon experience; and completing sustained projects; and
- Increased participation of under-represented populations in challenging school science.

The design of the 1994 NAEP Science Assessment Framework, while maintaining some conceptual continuity with the 1990 NAEP Science Assessment, takes account of current reforms in science education. It also is consistent with the science framework used for the 1991 International Assessment of Educational Progress and the science framework proposed for the two IEA (International Association for the Evaluation of Educational Achievement) assessments planned for 1993-94 and 1997-98. This consistency is important as the nation prepares to appraise its progress toward the National Education Goals for the year 2000.

15

CHAPTER 2

THE FRAMEWORK FOR THE 1994 NAEP SCIENCE ASSESSMENT

Introduction

It is customary to collect relevant curriculum guides. frameworks, and other course outlines to get a sense of what students are studying throughout the nation when developing a framework for an educational assessment. The union of all these "learning guides" is then used to develop some overall test specification. Such an approach tends to have an unfortunate consequence: It often leads to a broad, trivialized, lowest-common-denominator approach to assessment. However, these materials can serve another purpose, documenting trends and developments in science education throughout the country. In the 1994 NAEP Science Assessment Framework development process, reform reports in science from states and some large city science curriculum guides were gathered and used to establish consonance between the evolving framework and the most forward-looking of the reports and guides.

The traditional approach to teaching science has tended to emphasize rote memorization of facts without connection or organization. Although we must not lose sight of the need for factual knowledge as fundamental to science literacy, there are several compelling reasons why the assessment of science achievement must change in order to give more emphasis to conceptual understanding and the application of knowledge and skills. First, the expansion of scientific information has created far too many facts for anyone to memorize. It is much more efficient to store them electronically (or in other forms) and access information as it is needed. Second, isolated science facts that are unorganized and remain unused tend to be forgotten quickly and, even when remembered, form a poor basis for learning. Third, it is desirable to encourage science instruction in which science instruction is used both to deepen understanding and to address challenging problems. Science education is best served when students can understand and discuss ideas rather than simply accumulate unconnected facts.

Therefore, it is important that science assessment cover major topics like electricity and magnetism, forces and motions. life cycles. ecosystems. plate tectonics. and climatology. But. even these should be viewed as shorthand for a much richer understanding of what students should attain. Several of the current reform reports and frameworks such as *Science for All Americans* (American Association for the Advancement of Science. 1989), innovative state frameworks, and the reports of the National Center for Improving Science Education (1989. 1990) describe desired outcomes of science instruction in new ways. They advocate mastery of the traditional fundamentals, but in ways that are more likely to result in students' learning them. The approach advocated in these documents is reinforced by the findings of science educators and cognitive researchers, demonstrating that if students do not learn deeply the concepts presented to them, they end up passing through and over the K-12 curriculum without fundamentally changing the conceptual models they learn in their early years.

While it is easy to argue for a test that emphasizes conceptual understanding rather than topical listings or the recognition of definitions from a list of choices, it is more difficult to agree on how science learning should be assessed. The PSSC physics, CBA chemistry, BSCS biology, ESCP earth, and CHEMStudy chemistry courses that were developed in the late 1950s and early 1960s used deeper conceptual organizing principles for their texts and teaching materials. These programs, funded by the National Science Foundation, became major influences on the country's secondary science texts from 1960 to 1970 and helped raise standards of students' science achievement (Shymansky et al., 1983). Unfortunately however, the old-style textbooks have since returned, with ever-expanding numbers of pages to include as many newly emerging topics as possible. The effect on student learning of science has not been beneficial. For 1994, the NAEP Science Assessment Framework incorporates a balance of knowledge and skills based on current reform reports, exemplary curriculum guides, and research on the teaching and learning of science.

The Framework Elements

The Framework for the 1994 NAEP Science Assessment is organized according to two major dimensions: the major **Fields of Science** and **Knowing and Doing Science**. The Fields of Science are the earth, physical, and life sciences; Knowing and Doing Science encompasses conceptual understanding, scientific investigation, and practical reasoning. The two dimensions and sub-dimensions are explained in greater detail below and in appendix A.

The Matrix

The 1994 Science Framework is structured according to a matrix similar to that used in the 1990 NAEP assessment. The content areas are organized into the same three fields; however, there is an additional requirement for some interdisciplinary exercises that merge technology with the content areas. The "Nature of Science" (which also was part of the 1990 framework) and "Themes" (which is new to the 1994 NAEP Science Assessment) are categories that should integrate the three fields of science, rather than represent separate content. The "Knowing and Doing" dimension is a reorganization of the "Thinking Skills" in the 1990 assessment, with a clearer delineation of sub-categories, particularly with respect to "practical reasoning." Each of the elements of the 1994 Framework is addressed briefly below and in greater detail in separate appendices.

	FIELDS OF SCIENCE		
KNOWING & DOING	**Earth**	**Physical**	**Life**
Conceptual Understanding			
Scientific Investigation			
Practical Reasoning			
	Nature of Science		
	Themes Models, Systems, Patterns of Change		

With respect to the major Fields of Science, the main emphasis of the assessment should be on knowledge in the content areas.

- Distribution of content across the three science fields should be approximately equal in grades 4 and 12.

- For grade 8, the Assessment Framework will place a somewhat heavier emphasis on life science (40 percent in terms of content), with physical and earth science evenly distributed (30 percent each). This distribution for grade 8 reflects the importance of human biology for this age group, increasingly recognized both in curriculum and instruction.

- A limited number of exercises at every grade level should address technology because of its intimate relationship with science. Although not every item need do so, exercises and tasks that draw from more than one discipline at the same time are highly desirable, as they are more likely to mirror science problems occurring in the real world.

More specific guidance on the distribution of content from the three fields is given in the accompanying *Specifications* report.

The major emphasis with respect to Knowing and Doing Science should be on students' active expression of conceptual understanding.

- At each grade level, 45 percent of content should be devoted to conceptual understanding—the ability to understand basic concepts and tools used in the process of a scientific investigation.

- Scientific investigation—the ability to use the appropriate tools and thinking processes in the doing of science—should be more heavily emphasized in grade 4 (45 percent) than in grades 8 and 12 (30 percent at each of these grade levels). This is desirable because learning by doing plays a crucial role for younger students, and ways of knowing in science need to be introduced early.

- Practical reasoning involves suggesting effective solutions to everyday problems by applying scientific knowledge and skills. The ability to engage in practical reasoning is essential if one is to understand complex societal problems or be able to apply previous knowledge to an everyday problem. The proportions for practical reasoning should be 10 percent at grade 4. and 25 percent at grades 8 and 12.

- Many exercises will involve more than one of the sub-dimensions in Knowing and Doing Science.

- The percentages cited above are not to be interpreted as immutable but should serve as a general guide for test development. As new types of assessment tasks are developed, assessment of conceptual understanding may become part of exercises that measure scientific investigations and practical reasoning.

More detailed guidelines on the distribution of Knowing and Doing items are provided in the *Specifications* report.

Every question or task in the assessment should be classifiable by one or more sub-categories in each of the two major dimensions of the matrix. In addition to the two major dimensions, the Framework includes two other categories, which pertain to a limited subset of items. The first concerns students' understanding of the **Nature of Science.**

- Included in this category are the historical development of science and technology, the habits of mind that characterize these fields, and the methods of inquiry and problemsolving.

- At least 15 percent of the content should measure the Nature of Science. Within this percentage, somewhat more than half (about 60 percent) should deal with the nature of science and somewhat less than half (about 40 percent) with the nature of technology.

In specifying these percentages, it is assumed that these assessment items can be developed to do double duty in measuring knowledge of content within a field of science or an area of Knowing and Doing Science as well as the Nature of Science.

The second category, **Themes**, is new to the 1994 NAEP Science Assessment. Themes represent big ideas or key organizing concepts that pervade science. They cross the traditional science discipline boundaries and comprise a group of inquiry tools that scientists use to better investigate and understand the phenomena with which they deal. These themes include the notion of **Systems** and their application in the disciplines: **Models** and their functioning in the development of scientific understanding and its application to practical problems: and **Patterns of Change** as exemplified in natural phenomena.

- The assessment should probe, in a developmentally appropriate way, students' understanding of these organizing ideas or themes.

- In grade 4, students should build beginning notions related to systems, models, and patterns of change; about one-third of the assessment—spread evenly across the three themes—should measure themes as well as content from one or more of the fields of science.

- In grades 8 and 12, 50 percent of the assessment content should assess students' understanding of the themes, spread evenly across all three themes.

- The assessment exercises must embed a given theme in the science content that supports it, so that an understanding of the content and the theme are probed at the same time.

More detailed guidelines on assessing themes are provided in the *Specifications* report.

Understanding, doing, and using science often involve tasks that include more than one category in each dimension. Multiple-duty exercises may present some scoring challenges. Relatively simple exercises will be scorable according to one or two sub-dimensions that include, for example, a "conceptual understanding in the physical sciences." More complex items may be scorable in several sub-categories: for example, an open-ended task involving ecosystems might yield responses scorable according to "conceptual understanding in the earth sciences," "scientific investigation in the life sciences," and "systems." Such items, which contribute to more than one sub-category, may prove difficult to reproduce through similar items having the same properties in future assessments. It is important, therefore, to specify carefully the several sub-dimensions or sub-categories that each such exercise is intended to probe. Scoring rubrics for each of the sub-dimensions also must be developed.

The Fields of Science

The descriptions given here are capsule summaries of the major topic areas to be probed within each field in the 1994 NAEP Science Assessment. This content represents key elements in science that all students should be expected to know and understand. For a more complete explanation, including descriptions of subject matter knowledge to be expected at each grade level, see appendix A.

Earth Science

The 1994 NAEP Science Assessment will probe student understanding of how earth scientists depict data through maps and other means to interpret objects, their features and structures, and the events and the processes that caused them. What do students know about their own position with respect to objects and structures on, below, and above the Earth's surface? What do children know about changes in the position of objects and environments through time? What do students know about the relative movements of the Earth, Moon, Sun, and the planets? The content to be assessed in earth science in 1994 centers on objects and events that are relatively accessible or visible: Earth (lithosphere), water (hydrosphere), air (atmosphere), and the Earth in space.

With respect to earth science, the 1994 NAEP Science Assessment should center on the following concepts and topics:

Solid Earth:

- composition of the Earth
- forces that alter the Earth's surface
- rocks: their formation, characteristics and uses
- soil, its changes and uses
- natural resources used by humankind
- forces within the Earth

Water:

- the water cycle
- nature of the oceans and their effect on water and climate
- the location of water, its distribution, characteristics, and effect of and influence on human activity

Air:

- o the composition and structure of the atmosphere including energy transfer
- o the nature of weather
- o common weather hazards
- o air quality and climate

Earth in Space:

- o the setting of the Earth in the solar system
- o the setting and evolution of the solar system in the universe
- o tools and technology that are used to gather information about space
- o the apparent daily motions of the Sun, the Moon, the planets, and the stars
- o the rotation of the Earth about its axis and the Earth's revolution around the Sun
- o the tilt of the Earth's axis that produces seasonal variations in climate
- o the Earth as a unique member of the solar system that may be approximated in other galaxies in the universe and that evolved at least 4.5 billion years ago

Physical Science

The physical science component of the 1994 NAEP Science Assessment relates to basic knowledge and understanding concerning the structure of the Universe as well as the physical principles which operate within it. The assessment should probe the following major topics: matter and its transformations, energy and its transformations, and the motion of things.

The 1994 NAEP Science Assessment should center on the following physical science concepts:

Matter and its Transformations:

- o diversity of materials—classification and types, particulate nature of matter
- o temperature and states of matter
- o properties and uses of material—modifying properties, synthesis of materials with new properties
- o resource management

Energy and its Transformations:

- o forms of energy

- energy transformations in living systems, natural physical systems, and artificial systems constructed by humans
- energy sources and use, including distribution, energy conversion, and energy costs and depletion

Motion:

- an understanding of frames of reference
- force and changes in position and motion
- action and reaction
- vibrations and waves as motion
- general wave behavior
- electromagnetic radiation
- interactions of electromagnetic radiation with matter

Life Science

The fundamental goal of life science is to attempt to understand and explain the nature and function of living things. During the 20th century, the focus of biological research has changed from descriptive natural history to experimental investigation with evolution as the central, unifying theory. The major concepts to be assessed in life science are listed below:

Change and Evolution:

- diversity of life on Earth
- genetic variation within a species
- theories of adaptation and natural selection
- changes in diversity over time

Cells and their Functions:

- information transfer
- energy transfer for the construction of proteins
- communication among cells

Organisms:

- o reproduction, growth, and development
- o life cycles
- o functions and interactions of systems within organisms

Ecology:

- o the interdependence of life—populations, communities, and ecosystems

Knowing and Doing Science

In the 1990 NAEP Science Assessment, three categories were used: "knowing science," "solving problems," and "conducting inquiries." For the 1994 NAEP Science Assessment, it should be noted that not only has the "knowing science" dimension been changed, its **meaning** has been redefined as well. It has been reformulated as "conceptual understanding of science" in order to stress the connections as well as the organization of factual knowledge in science.

The sub-dimension of "scientific investigation" has been substituted for the 1990 category of "conducting inquiries". This new sub-category is intended to probe students' ability to **use** the tools of science, including both cognitive and laboratory tools. Appropriate to their age and grade level, students should be able to **acquire** new information, **plan** appropriate investigations, **use** a variety of scientific tools, and **communicate** the results of their investigations.

The 1990 performance category "solving problems" proved to be ambiguous. To emphasize the need to assess students' ability to use and apply science understanding in new, real-world applications, this category has been redefined as "practical reasoning." Practical reasoning includes **competence** in analyzing a problem, **planning** appropriate approaches, **evaluating** them, **carrying out** the required procedures for the approach(es) selected, and **evaluating** the result(s).

Each of these sub-dimensions is described below.

Conceptual Understanding

Mastery of basic scientific concepts can best be shown by a student's ability to use information in conducting a scientific investigation or engaging in practical reasoning. Optimally, essential scientific concepts involve a variety of information, including

- facts and events the student learns from science instruction and experiences with the natural environment;
- scientific concepts, principles, laws, and theories that scientists use to explain and predict observations of the natural world;
- information about procedures for conducting scientific inquiry;
- procedures for the application of scientific knowledge in the engagement of practical tasks;
- propositions about the nature, history, and philosophy of science; and
- interactions between science, technology, and society.

The goal of school science is to develop conceptual understanding. Students should acquire a data base composed of information structured in ways that will enable them to apply it efficiently in the design and execution of scientific investigations and in practical reasoning.

A challenge in the design of assessment exercises is to capture changes in the characteristics of student performance as children mature. In the primary years, when the goal of school science is to build a rich collection of information derived from examined experiences with the natural environment, the assessment of conceptual understanding will focus on the breadth of information about the natural world and the student's ability to elaborate principles from personal experiences. Does the student know the cyclical changes in the apparent shape of the Moon over time? More importantly, can the student relate how he or she knows about the changes? What evidence does the assessment exercise provide that the student's information is based on direct experience? Is there a science notebook in which the student recorded observations of the Moon over time? Does the student know that sometimes the Moon is visible during daylight hours? In the primary years, the focus should not be on explanation or prediction—rather, on knowledge obtained from rich experiences in school. Consequently, assessment exercises would not be concerned with having students explain why the Moon appears to change shape but rather with relationships between time of day, apparent positions of the Sun and the Moon, and times of moonrise and sunset.

In the middle and high school years, the emphasis should shift from richness of experience to reasonable scientific interpretation of observations. In the elementary years, the primary concern should be with how well reasoned an interpretation the student presents, not with whether it reflects the most sophisticated scientific understanding. However, at grades 8 and 12, the assessment should be concerned increasingly with the congruence of the student's interpretations with accepted interpretations, as well as with the sophistication of their reasoning in moving from observations of the natural world to explanation and prediction. Of special interest in the 1994 NAEP Science Assessment will be the extent to which the student is able to understand and use the notions of models, systems, and patterns of change.

It is important to note that many aspects of conceptual understanding as defined for the 1994 NAEP Science Assessment cannot be tested using exclusively multiple-choice items. Items of this kind are satisfactory for assessing individual parts of the information base, but they are limited in tapping highly valued aspects of conceptual understanding.

Scientific Investigation

Scientific investigation represents the activities of science that distinguish it from other ways of knowing about the world. It incorporates such previously used assessment categories as the "processes of science" and "scientific problemsolving". This category is not just another name for "the scientific method." Indeed, there is great confusion about the scientific method in the teaching of science. Real science is doing what one can in any way one can, often creative and insightful, often involving flashes of insight with little regard for a progression of steps. However, there is a familiar format and context for reporting the results of experiments. It begins with the report of the problem, continues with the hypothesis, the experimental design, the data collected, the analysis of data, and the conclusions, if any. This convention of science is often mistaken for how scientists actually work. The results must satisfy logical analysis, but the logical ordering may appear only when the report is prepared. A great disservice has been done to generations of students because well-meaning people have taught the standard method of **reporting** science as the standard method of **doing** science.

Scientific investigations must be designed at levels appropriate to the development of the students. This has important implications for assessment. Young students are limited in their ability to perceive the scale of things, both very large and very small. Students' limitations handicap them when they are forced, either by the textbook or by the curriculum, to deal with developmentally inappropriate concepts such as atoms or even cells. Young students also are developmentally limited in their ability to understand time. The distant past and the future are narrowly perceived by the egocentric student. Instruction as well as assessment must recognize where the student is and take developmental levels into account. As students develop and experience accumulates, their performance in doing scientific investigations should begin to look more and more like "real" science.

Central to the ways a scientist works is the concern for a fair test, a controlled experiment. Children seem to have an intuitive sense of what makes a fair test. What they lack is the ability to consider all the variables and the means to control the variables. It might be reasonable to consider a developmental continuum such as the following when thinking about the control of variables.

This continuum begins with the simplest type of variable—the nominal variable. Nominal variables have two or more unordered values: "This plant was watered; that plant was not." "This seed was placed in the sunlight and that one was placed in the dark." The second level of controlling variables is the ordinal variables level. These variables have a sequential order and no determined intervals (e.g., the sequential ordering of objects by relative weight). The third level of variables is called the continuous variables level. These variables have sequence and equal intervals and are on a continuous scale. "This object has a temperature of 50 degrees Celsius and that object has a temperature of 57 degrees Celsius." The fourth level of variables is called the ratio variables level. These are similar to the continuous variables but have an absolute beginning point (e.g., Kelvin temperature scale with an absolute zero point).

As students are asked to demonstrate their ability to do scientific investigations, it is important to keep in mind this sort of development in understanding and performance, not just with respect to the control of variables but also regarding the other elements of doing science. The difficulty with the assessment probe may not be the content but the level of variable imbedded in the content.

Practical Reasoning

Practical reasoning about matters with scientific content. the ability to apply one's knowledge, thought, and action to real situations (not "textbook problems") is influenced by the ability to: (1) abstract and consider hypothetical experiences; (2) consider several factors simultaneously; (3) take a depersonalized view; and (4) realize the importance of practical reasoning and life experience. These abilities develop throughout life.

One of the characteristics of young children is that they have difficulty dealing with multiple ideas simultaneously. With maturity and experience, they are better able to consider several ideas at once and weigh benefits in relation to costs and risks. The ability to abstract and to consider hypothetical situations develops as students progress in science and learn to deal with more remote phenomena and generalizations.

As they mature, students also learn to take a depersonalized view of a situation and to consider someone else's point of view. Often real-life problems involve not only the theoretical and technical elements, but also personal preferences. What will be the social impact of a new waste disposal system? What will neighbors say if a traffic light is installed? How will other students react if the lunchroom noise is diminished by staggering the lunch hour? In order to consider these questions carefully, it is necessary to understand different perspectives. Understanding the viewpoint of others grows with age and experience.

Young children also may not realize the need for scientific information in solving problems. For example, children below age 12 usually see no need to carry out measurements (Strang, 1990). Also, since young children have little responsibility for decisions affecting their lives, they may not see the need for practical reasoning. However, the more students have done or seen, the more likely it is that they can solve real-world problems. With age and experience, the possibility increases that a new situation is analogous to a previous one, and that the human, technical, and theoretical factors involved in it have been encountered before.

All these factors suggest that practical reasoning should become a major factor in science assessment at grades 8 and 12 rather than at grade 4. As students become eager to take control of their lives, wish to try out their understanding of the world, and progress in development, practical situations related to their everyday life, school, and home provide excellent exemplars to demonstrate science-related practical reasoning. For example, students might be asked to discuss problems such as noise abatement in the lunchroom, to design a simple apparatus such as a flashlight or a burglar alarm, or plan a school garden.

By grade 12, students should be able to discuss larger science- and technology-linked problems not directly related to their immediate experience. Examples of these might be waste disposal, energy uses, air quality, water pollution, noise abatement, and the trade-offs between the benefits and adverse consequences of various technologies.

The Nature of Science

Knowledge of the nature of science is central to understanding of the scientific enterprise. Yet, oftentimes this category is relegated to a discussion --or even rote memorization--of some version of the "scientific method". There was total agreement that the topic should play a prominent role in the 1994 NAEP Science Assessment. A controversy in the project committees did exist concerning whether "The Nature of Science" is sufficient unto itself, or whether the Framework should include a separate section that deals with the nature of technology? The project committees were split on this issue. It was acknowledged by all that technology is integral to the nature of science and ought to be included in the assessment of science, as long as it was clear that technology does not exist as a separate sub-category within the assessment. Technology, then, will be assessed as it relates to science and the scientific enterprise.

Science

- By grade 4, students should understand that science is trying to find out what happens in the natural world. Through careful observation of objects and events, they should be able to develop explanations for their observations. Students should also understand that different people may notice different things, and therefore may explain things differently.

- By grade 8, students should have acquired an understanding of the control of variables and the difference between showing that conditions occur together and that they are causally related. Students should grasp what makes for a good scientific explanation fitting all the relevant observations, suggesting what new observations to make, and explaining, as simply as possible, a wide variety of observations.

- By grade 12, students should demonstrate their knowledge and understanding of the following:

    - Scientific conclusions are based on logic and evidence, but there is no fixed series of steps that make up a "scientific method." Scientists try to invent explanations that are logical and fit observations, but these are subject to change based on new evidence.
    - Explanations are most believable when they also account for observations that were not known to the explainer.

- Scientists (like anyone else) tend to look for, pay attention to, and cite evidence that supports what they already believe.
- New conclusions require that scientists consider all possible objections to their own findings.
- Scientific organizations try to avoid bias and maintain quality by having scientists' reports of observations and explanations judged by other scientists before they can be published.
- Very few human problems can be solved by scientific knowledge alone. Most are too complicated and most involve values—which science has very little to say about.

Technology

Students are surrounded by and interact with the man-made world as much as with the natural world. Therefore, they must develop an understanding of what shapes the design and development of the technologies that are part of that world and their daily lives. The nature of technology, rather than being a content area, is embedded within this section because of its close association with science. The following concepts are appropriate for assessment at the given levels.

- By grade 4, students should understand that any design requires making trade-offs and that advantages and disadvantages must be weighed.

- By grade 8, students should understand that scientific knowledge often is useful in design and that much scientific investigation is done for the purpose of improving design. They should understand that there are often several ways to solve a design problem and that possible solutions should be evaluated and justified in terms of their advantages and disadvantages.

- By grade 12, students should know that scientific knowledge may help to predict consequences of one design or another, but that design decisions often depend upon human values outside science. They also should be able to apply scientific concepts to scientific, societal and/or technical concerns. They should understand that every design has limits and may fail if required to work outside of them.

The Themes

The themes are the "big ideas" of science, transcending the various scientific disciplines and enabling students to consider problems that have global implications. In order to understand the conceptual basis for the themes that have been selected, students must begin to develop an understanding of major ideas by the 4th grade; they should continue to develop their understanding through the 8th grade, and by 12th grade should have the ability to integrate their knowledge and understanding.

The review of current state frameworks carried on in the course of developing the 1994 NAEP Assessment Framework revealed that many are based in part on crosscutting themes in science. Several national organizations, including the American Association for the Advancement of Science, have issued reports that advocate the importance of common themes. The number of themes defined in these reports and state frameworks varies somewhat, but there is considerable agreement on what the common elements or big ideas of science are that should be understood by students as they complete their high school education. The decision by the 1994 NAEP Science Assessment committees to include themes underscores their emerging importance, as well as the necessity to integrate themes through programmatic threads in grades K-12. Three of the themes are common to all the documents: **Models**, **Systems**, and **Patterns of Change**. These three themes will be included in the 1994 NAEP Assessment, because they constitute major, interdisciplinary organizing principles of science. Further, they do not conflict or compete with the factual content of the various fields but, rather, augment and help organize that information into a coherent intellectual framework.

Models of objects and events in nature are ways to understand complex or abstract phenomena. Models may be first attempts to help tease out the relevant variables in order to build ever more useful representations, or they may be highly refined for predictions about the actual phenomenon. Students need to understand the limitations and simplifying assumptions that underlie the many models used in the natural sciences. A model is likely to fit data well only within a limited range of circumstances and to be misleading outside of that range.

Systems are complete, predictable cycles, structures, or processes occurring in natural phenomena, but students should understand that the idea of system is an artificial construction created by people for certain purposes—to gain a better understanding of the natural world, or to design an effective technology. The construct of a system entails identifying and defining its boundaries, identifying its component parts, the interrelations and interconnections among the component parts, and the inputs and outputs of the system.

Regardless of the topic around which the **Patterns of Change** theme is developed, students should be able to recognize patterns of similarity and difference, to recognize how these patterns change over time, to have a store of common types of patterns, and to transfer their understanding of a familiar pattern of change to a new and unfamiliar situation.

Appendix B contains more detail on each of these three themes and the developmentally appropriate expectations for students at grades 4, 8, and 12.

CHAPTER 3
DESIRED ATTRIBUTES OF ASSESSMENT

Some Assessment Issues

NAEP functions to provide information about (1) the knowledge and scientific understanding of the nation's youth and (2) the features of science education programs that relate to high levels of student achievement. Consequently, the design plan for the NAEP Science Assessment includes strategies for measurement in both of these areas. Research has produced practical and theoretical knowledge that is important to the assessment design process.

The Nature of Testing and the Nature of Knowledge and Learning

Except for early assessments, NAEP science tests have consisted primarily of short-answer, paper-and-pencil questions that were mostly multiple-choice. The previous assessments tended to focus on discrete components of science, each of which was usually learned independently of the others. Hence, tests were made up of independent items, each comparable to the others and weighted the same.

This Framework is based on a different view. It holds that scientific knowledge should be organized to provide a structure that connects and creates meaning for factual information and this organization is influenced by the context in which the knowledge is presented. Learning is perceived as an activity in which the learner interacts with the physical world, with peers and teachers, and with accepted scientific knowledge. In this view, science proficiency depends upon the ability to know and integrate facts into larger constructs and to use the tools, procedures, and reasoning processes of science for an increased understanding of the natural world.

Rather than concentrate on facts in isolation, the assessment will reflect the organization and structure of scientific knowledge and the nature of learning in science. Since scientific knowledge is expanding faster than can be accommodated by any curriculum, teachers and assessment designers must make choices about what topics, concepts, and factual information to address. Consequently, the 1994 NAEP Science Assessment Framework concentrates on assessing students' ability to relate basic facts and concepts as well as their ability to discuss and evaluate approaches to science-related problems.

The Framework also stresses that an assessment of what students know and can do must employ techniques reflective of the nature of science.

Inferring Understanding from Student Responses

Test items present students with tasks to carry out, which may range from recall of factual information to performing scientific investigations to complex reasoning. Based on analysis of the responses, experts in the field make inferences about the student's understanding—that is, the knowledge and reasoning skills that are assumed to have produced the response. The validity of these inferences is a central issue in assessment.

Responses to exercises designed to assess thinking or mental processing are generally more difficult to interpret than responses to items designed to assess factual knowledge. In practice, basing an assessment of the quality of mental processing on short responses is problematic. Often the decision about the mental processes applied are based only on the accuracy of the factual knowledge in the answer. When the answer is factually correct, the observer infers that the mental processes represent scientific reasoning (e.g., that necessary to understand information in the stem, to retrieve scientific knowledge from memory, to reason from the stem to the correct response, or to eliminate incorrect responses). But this is not necessarily a valid inference. An incorrect answer may be the result of misinformation, not flawed reasoning; a correct answer is not necessarily the product of sound reasoning. Illogical thinking or using a wrong assumption or incorrect information can produce a seemingly correct answer. Furthermore, a correct answer may not require any higher-order thinking at all; it simply may have been recalled. Only if the student response includes some indication of how it was obtained will those who score the assessment have information from which to choose among alternative interpretations.

Science entails observing objects and phenomena in the natural world and collecting and interpreting information about them. For this reason, pencil-and-paper tests have been criticized as too limited for assessing what students know and can do in science. Over the past several years, groups have developed assessment exercises that engage students in "performance tasks" using scientific equipment and materials; student responses are recorded by an observer or by the students themselves in written form. However, these exercises have limitations as well.

Indeed, in examining many examples of performance items in the course of developing the 1994 Framework with respect to science concepts and reasoning, many problems did not stand up to rigorous analysis. These exercises might have misled science teachers and educators through the faulty inferences being drawn.

The following is typical of a performance exercise that is counterproductive: The exercise requires a student to identify several unknown substances by means of indicators; but the student is given minutely detailed directions for performing each step in the process of identification. Unfortunately, even if the answers are correct, the only inference to be drawn is that the student can follow written instructions. A test item formulated with such detailed step-by-step directions reduces to zero the science understanding needed for problemsolving.

Many of the so-called performance assessment tasks that were reviewed turned out to be standard laboratory exercises which, again, were reduced to "follow-the-instructions" problems. No inferences can be drawn from such exercises about a student's knowledge of science or of its tools and procedures. In order to test higher-order thinking skills—a major goal of performance assessment—problems need to be placed in new contexts, applied to new situations, or have new elements introduced that preclude students' recalling what they have done before (Resnick, 1987).

Inferences in Assessment of Different Population Groups. Class, culture, ethnicity, gender, language ability, and access to quality instruction all may influence the manner in which science is learned and the manner in which science attitudes and knowledge are produced. Hence, individuals need opportunities to demonstrate knowledge or competencies in different contexts. NAEP should investigate ways to address this issue in the next science assessment.

Assessment techniques that show group differences are more likely to reveal problems with student learning and classroom instruction than with assessment, *per se*. However, this does not eliminate the assessment community's responsibility to the broader society; multiple assessment methods may be more effective in addressing the issues of pluralism than any one method—no matter how well it is developed.

Developments in Assessment

Currently, some state pilot-testing efforts are providing new ideas about assessment exercise and task formats. These pilots also are aimed at assessing new types of information relevant to new curriculum guides emerging in the states.

The experimental assessment work (much of it pioneered in the United Kingdom) uses new approaches including performance tasks, open-ended tasks, and new types of multiple-choice questions that are thematic or conceptual or ask students to explain their choices in short written answers. Moreover, state assessments are experimenting with collecting information on other meaningful outcomes: sustained student work, proficiency in designing and conducting experiments, and fluency of ideas critical to the natural sciences and related fields. They also are experimenting with innovative reporting approaches. The performance measures are created through consensual judgment regarding what students should know and be able to do at given grade levels or developmental stages.

Criteria for Assessing Learning and Achievement in Science

The Framework for the 1994 NAEP Science Assessment has been developed according to the following broad guidelines:

- By focusing on meaningful knowledge and skills, NAEP should be a force in fostering progress as well as measuring it, enabling more students to learn more science.
- A range of assessment means must be used, including some in which the student is required to create and construct, not only to recognize and respond.
- Assessment exercises should challenge students at developmentally appropriate levels to
  - explain commonplace natural phenomena using appropriate scientific theory, principles, and concepts;
  - plan the investigation of a novel scientific problem;
  - demonstrate understanding of the basic knowledge structures of science by using the appropriate techniques to connect concepts to each other and to the theory within which they are embedded;
  - demonstrate some understanding of pervasive crosscutting themes in science;

— solve practical problems by using the appropriate theories, principles, concepts, and techniques of science.

- The assessment must be sensitive to the need and ability of students to function in a variety of contexts.
- Assessment exercises should display a variety of formats to allow students to display the wide range of competencies expected as the outcomes of science education.
- Assessment tasks that are larger than single items, should be analyzed in multiple ways, not restricted to providing information for single scales.
- Test results should not be normalized; that is, students' outcomes should not be manipulated to fit a normal distribution curve.[2]
- The assessment must have enough questions about enough topics to explore students' knowledge in depth.

[2] Such a practice is appropriate only if it is assumed that all items are independent of others, and that they measure discrete properties of the learner which can be assumed to be randomly distributed among a total population.

CHAPTER 4

CHARACTERISTICS OF ASSESSMENT EXERCISES AND ITEMS

Types of Exercises

Innovative assessments in this and other countries use three major item types: performance exercises, open-ended pencil-and-paper exercises, and multiple-choice items probing understanding of conceptual and reasoning skills. A fourth type often added to time-limited tests is the collection and evaluation of portfolios of student work done in the course of instruction. There is also an emerging assessment technique that involves two-phase testing. The following sections discuss these exercise types and provide guidelines for the amount of assessment time to be devoted to each. Further details are provided in the *Specifications* and *Reporting Formats and Issues* documents.

Performance Exercises

In performance exercises students actually manipulate selected physical objects and try to solve a scientific problem about the objects before them. Although various types of performance tests have been piloted extensively, their standardization and administration differ widely. One method for ensuring uniform administration is the use of standardized performance test kits, with each test proctored and scored by trained personnel. Depending on the objectives established for the assessment, student answer sheets also can be used to provide responses for scoring. To measure the goals outlined for the assessment adequately, performance items generally should make up at least 30 percent of the assessment, as measured by student response time. An extra period of time (20-30 minutes) may be necessary for students assigned complex performance tasks.

The shortcomings of many performance tasks currently being used were discussed in the preceding chapter. How could a performance exercise be designed so that it meets criteria for assessing science concepts and their relations? The problem should be meaningful and not a context-free laboratory problem. Personal context, for example, is seen in the following problem: "You have just been given this new drink. It is claimed to be sugar- and calorie-free. What can you find out about the accuracy of these claims with these indicators?" Students can be given the names and procedures for the safe use of each indicator, but without any information as to their scientific use. The students would have to know, for example, that iodine solution is a test for starch and what the negative and positive reactions are. They would also have to know that if fats, protein, or carbohydrates are present, they will yield calories. They would have to plan how to conduct an investigation of the unknown in such a way as not to waste it. to be able to repeat the investigation when they believe their procedure was faulty, and to have enough solution left to replicate for verification. They would also have to design their data collecting procedures. Finally, they would need to interpret and justify their findings. The questions asked of students as part of a performance exercise need to enable the students to display understanding and to justify interpretations. Such questions as "What substance is in the unknown?" or "How far did the dye diffuse?" do not elicit such responses.

If students need additional information to carry out the task, they could be asked if they would like any other materials before they begin and why. If they request known substances for each indicator to refresh their memories, those could be provided. Such a request demonstrates one aspect of understanding the processes of science—knowing what one doesn't know and how to find out.

Scoring for such a problem could give points for science knowledge, for laboratory procedures, and (if using an observer) for a systematic approach to problemsolving in contrast to a trial and error or random approach.

Open-Ended Pencil-and-Paper Items

Open-ended items that require written responses provide particularly useful insights into students' levels of conceptual understanding. They also can be used to assess students' ability to communicate in the sciences. In addition, open-ended items, if carefully crafted, can be used to reflect students' ability to generate rather than recognize information related to scientific concepts and their interconnections. Open-ended items should make up at least 50 percent of the assessment, as measured by student response time. About one-third of the open-ended questions should consist of extended response items. Research needs to be carried out using non-written forms of responses, particularly with younger students, to explore the effects of writing ability on performance in a science assessment (see chapter 5 for a more detailed description).

Multiple-Choice Items

The 1994 NAEP Science Assessment will send important messages about the science curriculum and classroom instruction, and the use of multiple-choice items should be considered very carefully. Although they often are often over-used to test low-level recall. Balanced with other item types, however, multiple-choice items are worthwhile for getting at important facts, concepts, and deductive reasoning skills. To the extent that the assessment demonstrates an appropriate use of multiple-choice items, they should be balanced with other items types in the NAEP assessment. Multiple choice items should comprise 50 percent of the assessment, as measured by student response time.

Additional Considerations

Multiple-choice items, open-ended paper-and-pencil items, and performance tasks could produce responses less subject to faulty interpretation if students were given an opportunity to explain their responses, their reasoning processes, or their approach to a problem. Hence, the assessment should afford this opportunity; but care must be taken, particularly with fourth-graders, that language ability is not confounded with science ability. This caution also applies to the more complex multiple-choice items needed to probe conceptual understanding.

The 1994 NAEP Science Assessment, to be consonant with current reform efforts in science education, needs to probe students' depth of knowledge and scientific understanding. For this reason, it is recommended that for at least half the students sampled, the assessment include an in-depth examination involving a single problem or topic. The format could be a set of linked multiple-choice items, open-ended paper-and-pencil exercises, performance tasks, or a combination. (Examples of relevant exercises are given in chapter 6.) Pending modification after pilot testing, the suggested time to be spent by students on this type of exercise is 10 minutes for grade 4; 20 minutes for grade 8; and 30 minutes for grade 12.

Two-Phase Testing

One of the most promising testing techniques would entail the administration of a two-phase assessment. In this design, a short pretest is given to students to establish their ability level. Depending on results, students then are given different tests that match their individual competence with the difficulty level of items and exercises. The pilot test should include open-ended paper-and-pencil exercises, multiple choice, and performance tasks. This design permits a more efficient, reliable estimate of science achievement by targeting the assessment exercises to the student's level of performance. It also can be used to measure group performance.

Collection of Students' Work

Students' work collected over time—portfolios—may be more difficult to gather in a large-scale assessment program, but would be valuable in two ways. First, they can demonstrate what students know and can do under less artificially-constrained conditions than a one-shot test. Second, they reflect the content and approaches of classroom instruction. NAEP is collecting a subsample of students' work for the 1992 Writing Assessment, a field in which the evaluation of portfolios has been carefully worked out and proved quite reliable. A portfolio approach is being explored in the 1992 Reading Assessment.

A variant on portfolios is a type of assessment exercise, developed by DeLange in The Netherlands, to which students respond in two phases. In the first phase, the exercise is a timed assessment. In the second phase, students take home the exercise (usually, a reasonably complex problem) for more intensive work, including the use of resources—textbooks, their source materials and notes, instrumentation, and discussion with peers and experts, etc. An interesting finding is that in the take-home phase, female students do as well as males, although they show the usual gender differences in the timed phase.

The use of portfolios in science should be explored in the 1994 NAEP Science Assessment. A special study will be conducted for this purpose, as described in chapter 5, using a subsample of the national sample.

Pilot Testing

Multiple approaches need to be tried in the pilot testing of the assessment exercises. It would be especially useful to test the same concept(s) and performance skill(s) in different ways to see which method provides the richest but also most reliable and valid information. For example, if an open-ended question can be easily turned into a multiple-choice question without losing its intent and validity, it should be multiple-choice. Open-ended questions should tap skills and knowledge that are truly "open"—probing for the integrated application of relevant knowledge, not for recall of a series of unconnected facts.

The following additional issues need to be investigated during pilot testing:

- Scoring Rubrics. These should be developed *a priori* for open-ended questions and performance tasks, but should be modified on the basis of pilot test results. Weights should be assigned within scoring rubrics of complex items to reflect the quality of the responses. Distinct rubrics must be developed to score for multiple aspects or on items contributing to more than one component of the assessment.

- Scaffolding. To what extent should scaffolding or additional information be integrated into open-ended questions? This could be done either in the form of "hints" (to be given only after the student gets "stuck;" for example, in a computer-administered or individually timed exercise) or by informing the students how their answers will be scored. How much should students be able to learn during the test, as contrasted to what they learned in the classroom? Unfortunately, many science items in current exams require no science knowledge at all since everything is provided in the stem or in the instructions. On pencil-and-paper tests, these questions usually turn out to measure general reasoning or scale and graph reading.

- Student Self-Evaluation. Should students (perhaps in grades 8 and 12 only) be able to evaluate their own test performance? Students might be given credit for what they know they don't know, particularly if they can articulate what steps they might take to find out. Alternatively, they might be asked to indicate how well they think they did on some number of items (meta-cognitive questions).

Procedures for the Development of the Assessment

The Planning Committee responsible for developing the 1994 NAEP Science Assessment Framework has been very concerned that the nature and specifics of the Framework be faithfully mirrored in the actual assessment instrument. For this reason, a detailed review is recommended of individual items and the proposed assessment as a whole, to be conducted at the end of each of the following four stages:

- item development and selection for the pilot testing;

- analysis of the data from the pilot testing and review of the results;

- *a priori* development of the several scales at each grade level together with review of the items used for the behavioral anchoring; and
- selection of items and formulation of the assessment as a whole for the 1994 test administration.

The review committee(s) should be broadly constituted to include scientists, teachers, science educators and researchers, cognitive psychologists, psychometricians, and informed members of the lay community, including representatives from business, higher education, education governing bodies, and parent groups. The several stages of review should be carried out by the same committee; if this is not feasible, the four groups need to have overlapping membership. Members of the Planning Committee should be included throughout the entire review process.

Reporting

1. For the types of assessment exercises and assessment as a whole recommended in this Framework, the usual assumptions about independence of individual items and discrete properties of learners randomly distributed throughout the population will not hold. For this reason, items, scales, and students' test results should not be adjusted after scaling to fit a normal distribution curve. Instead, scaling should be based on *a priori* definitions of scales.

2. Information on numbers or percentages of students who have achieved some given science knowledge (Fields of Science) and science performance skills (Knowing and Doing Science) should be reported on a more detailed level than one scale or a few subscales.

3. Assessment results should be reported using within-grade rather than across-grade scaling. Assessment exercises in science must be developmentally appropriate; the corollary is that the same exercise may be eliciting very different thinking, reasoning, or performance skills at different grade levels rather than measuring growth in the same skills. Thus, a challenging reasoning problem for a fourth-grader may be a simple recall problem for the eighth-grader and therefore will not assess growth in reasoning.

4. All three types of exercises should be included in the NAEP scales, including open-ended paper-and-pencil tasks and performance exercises and multiple-choice items.

5. If the special studies recommended for the 1994 NAEP Science Assessment are carried out, results should be reported on a national level.

More detail is given in the accompanying document on *Reporting Student Achievement Data in the 1994 NAEP Science Assessment.*

CHAPTER 5
SPECIAL STUDIES AND RESEARCH

Special studies are often recommended as part of the National Assessment process because new or emerging techniques offer promise and, if they yield useful information, will make a positive contribution to the assessment process. Special studies are part of the main NAEP process and are reported along with the results from the national sample.

Two special studies should be carried out in the 1994 NAEP Science Assessment, each using a subsample of the NAEP national sample. In order of priority, they are

1. Collecting and analyzing samples of students' work (portfolios) drawn from their classroom and laboratory assignments, science projects, homework, and classroom tests; and

2. Piloting a two-stage design for identifying able students and administering to them a version of the NAEP test containing a larger number of difficult exercises.

These two studies are aimed at critical components of a valid assessment of students' knowledge and abilities in science. In each case the reason for recommending a special study rather than administration to the entire NAEP sample is the same—an adequate base of experience and research must be built to ensure reliability and validity of the recommended innovation.

Each of the two special studies is discussed briefly below with more detail provided in the *Specifications* report.

Special Studies

Collection and Analysis of Students' Work

Assessments of science achievement limited to a 50-minute period, even when extended by 20 or 30 minutes to allow for complex performance exercises, is inadequate for students to demonstrate what they know and can do. The setting and format of such an assessment is artificially constrained, and students have no opportunity to demonstrate their ability to perform sustained work. On several counts, therefore, past NAEP science assessments have failed a crucial validity criterion: that is, they have not mirrored the nature of the field. For this reason. **the special study on collecting and analyzing "Science Portfolios" is accorded the highest priority.**

A number of problems need to be resolved, however, before "science portfolios" can become a regular component of NAEP, on a state-by-state basis. These problems include formulating criteria for selecting students' work samples, creating scoring rubrics to accommodate what is likely to be a very diverse collection of materials, ensuring inter-rater reliability, and developing a method for judging the work samples that will make inclusion of science portfolios financially feasible, at least at the national level. The design of the special study must address these problems. In addition to building on NAEP experience, the design should benefit from findings of current experimentation: with portfolios in mathematics in several states, and with setting common standards among teachers in judging student work (moderating) in Great Britain. Results of the special "science portfolio" study should be included in the national report of NAEP results.

Assessing "The Best"

Past NAEP science assessments have been criticized for having too low a ceiling; that is, not including an adequate number of items at advanced levels of difficulty. As a result, NAEP tests are conjectured not to have reflected what the best-prepared students know or can do in science. This issue has become more serious since the formulation of the National Education Goals, particularly Goal 4: "By the year 2000, U.S. students will be first in the world in science and mathematics achievement." The two international assessments being planned by IEA for 1993-94 and 1997-98 will include assessment of students who are concentrating on science in high school. It would be highly desirable to have similar results from NAEP so as to have more than one study as a basis for reporting on progress toward the National Goals.

Suggestions on how to include the most competent students in the NAEP sample are to (a) over-sample magnet or specialized schools that emphasize science, particularly at the secondary level; (b) administer the NAEP 12th-grade test to students enrolled in Advanced Placement science courses; or (c) use two-phase testing to allow administration of assessments that are matched to student ability. The assessment must contain a sufficient number of challenging exercises to measure what these "best" students know and are able to do in science.

The most promising approach is the third one: administration of a version of the Mislevy/Bock two-stage design currently being pilot tested for science with 12th-graders in one state. In this design, a short pretest is given to students to establish ability level. The students are then given different tests that match their competence with the difficulty level of the items and exercises. The pilot test includes open-ended paper-and-pencil exercises as well as performance tasks. Methods have been developed for efficient assembly of alternative groupings of exercises at various levels of difficulty, matched to ability groups. The current state pilot is too limited, however, to permit conclusions about including this design in the national or state-by-state science assessment in 1994.

It is recommended, therefore, that a special study be conducted in 1994 with a subsample of the national NAEP sample to determine whether this is a useful approach to establishing the achievement and performance of the best science students, whether or not they are in special schools or enrolled in advanced science courses. Results should be included in the national report of NAEP results.

Research Studies

Assessment of Work in Groups

The assessment of science performance in the context of group work is becoming increasingly important for two reasons: (1) to meet employers' demand for employees who can work together cooperatively and solve problems as they arise, and (2) to assess certain significant learning outcomes that can be observed only in a team setting.

Group work reflects more realistically how scientific knowledge is developed and learned (see chapter 1) and sustained investigations involving group work are being developed by a compact of states with National Science Foundation funding (Baron, 1990). Such investigations may consume a class/laboratory period or be carried out over several days. At the high school level, students may be asked to participate in the evaluation of their individual performance and that of the group as a whole. An alternative for large-scale assessments might be the presentation of material on a diskette to a group of students drawn from classes or schools. Monitoring software on the diskette could provide an unobtrusive record of the group's procedures and solutions (National Center for Improving Science Education, in press).

Assessing performance of work done in groups poses obvious challenges such as judging the work of each individual, *per se*, judging his or her contribution to the group performance, and—if desired—judging the performance of the group as a whole. Research needs to be carried out on feasible approaches to assessing individual performance on science work done in groups as well as group performance. **Such research is given the highest priority.**

Expanding Assessment Time

Although NAEP is not intended as a timed assessment, review of responses clearly shows that a sizable number of students do not reach the last several test items in many blocks. This raises the question of what students might do if they had considerably more time. The question is particularly appropriate since the intent of the 1994 Framework and presumably succeeding assessments in science is to introduce exercises of much greater interest and significance to students, thereby increasing motivation and reducing the fatigue factor.

Two research studies are recommended. The first would allow students up to 2 1/2 hours to take the test. This type of study needs to oversample students from such population groups as African-Americans and Hispanics who traditionally do less well on NAEP tests. The second recommended study entails administering a set of two-phased assessment exercises, such as the DeLange mathematics assessment described earlier (a time-constrained first phase and a followup take-home phase in which students are allowed to use resource materials of their choice). This appears particularly appropriate for 12th-grade and possibly 8th-grade students. Problems to be resolved include development of appropriate exercises, relationship between the two components and responses, scoring rubrics, and comparing students' responses for the take-home component.

Non-Written Stimuli and Non-Written Responses

Reading or writing difficulties may prevent some students from demonstrating their science knowledge on more complex exercises. To allow for more adequate participation, the most straightforward addition to current NAEP procedures would be to translate each assessment into some non-English languages, particularly Spanish. This is strongly recommended for the 1994 NAEP Science Assessment, both to decrease the number of "excluded" students and to allow participation by Puerto Rico. However, the need would remain for items and exercises that do not rely on reading and writing skills, especially for students in grade 4, for students with limited English proficiency, and for those who are more competent in **doing** than in **language** about doing.

To address the reading and writing issues, a set of studies should be conducted that substitutes oral language for either the stimulus or the response, or both. One set of studies should present the exercise(s) as read by the test administrator or by audiotape, with either written or oral responses. A second set of studies should present the exercise(s) in traditional format and elicit oral responses. In eliciting oral responses, the issue of "scaffolding" or providing hints may arise, particularly in responding to an adult about a complex question. Scaffolding may be quite appropriate when exploring the respondent's thinking processes to ensure correct interpretation of the answer(s), but it may interfere with comparability to written responses. Non-language stimuii and responses also must be explored. Exercises can be presented via videodisc or live demonstration with a minimum of oral explanation, and drawings may be appropriate as a response medium for some exercises.

Curriculum-Related Assessments

A crucial issue for large-scale assessments is to ensure that conceptual understanding in science and the ability to carry out scientific investigations and engage in reasoning are probed in depth (see chapter 1). More and more, science teaching is likely to concentrate on fewer topics taken up in greater depth and this poses a dilemma for large-scale assessments like NAEP. As the science curriculum becomes more diversified, it becomes increasingly difficult to ensure that students are given an opportunity to demonstrate what they know and can do in science on a generalized test that presumes to cover the curriculum. In fact, it may be impossible to find a single in-depth assessment task for which all or even a substantial fraction of students have had the prerequisite instruction.

Research needs to address how a curriculum-related component could be included in NAEP assessments. For instance, a sample similar to the national NAEP sample could be used to collect information on the local curriculum from teachers or district science supervisors. For grades 8 and 12, students also could be asked to furnish relevant information. On this basis, one or several in-depth tasks would be selected for administration to students in each school or district, together with the core NAEP assessment. The latter could be an individual or group task, adding 25-30 minutes to the core NAEP.

The argument is not that assessments should only address what students have studied—quite the contrary. In-depth knowledge and ability to pursue a scientific question in some detail must become part of the general science proficiency and, to demonstrate such proficiency, students must be given an opportunity to show what they know and can do in an area of science that they have studied.

By providing a choice of curriculum-relevant components, the assessment can legitimize the notion that not everyone needs to possess identical knowledge and competencies. There is tension, however, between proponents of an assessment designed to allow students to best exhibit their individual abilities and proponents of an assessment that measures the knowledge and skills of the student body as a whole. For this reason, experimentation with curriculum-relevant assessment must include a core battery, to be taken by all students, based on the overall framework for the assessment.

APPENDIX A

FIELDS OF SCIENCE

FIELDS OF SCIENCE

The detailing of the fields of science below indicates themes that could be explored within major topic areas as well as brief examples of developmentally appropriate sample questions that could form the basis for assessment exercises.

Earth Science

Earth science is the study of the planet Earth's composition, process, environments, and history, focusing on the solid Earth (lithosphere) and its interactions with air (atmosphere), and water (hydrosphere). The content to be assessed in 1994 in earth science centers on objects (which includes bodies and materials, their composition, features and structures), as well as processes and events which are relatively accessible or visible. This includes objects like soil, minerals, rocks and rock outcrops, fossils, rain, clouds, the Sun and Moon, processes like erosion and deposition, and weather and climate; and events like volcanic eruptions, earthquakes and storms.

The Solar System (subset)

Earth/Moon/Sun

- Observable evidence (*Patterns of Change*)
- Description, models (*Systems, Models*)

Students should understand that the apparent daily motions of the Sun, the Moon, the planets, and the stars are due to the rotation of the Earth about its axis every 24 hours. This rotation produces the Earth's night-and-day cycle. They should also understand that the Earth's 1-year revolution around the Sun, because of the tilt of the Earth's axis, changes how directly sunlight falls on one part or another of the Earth, thus producing seasonal variations in climate. They should know that the combination of the Earth's motion and the Moon's own orbit around the Earth, once in about 28 days, results in the phases of the Moon.

Sample Questions:

Grade 4

Draw a shadow, specifying direction and length, at different times of day. Where does the sunshine fall in the classroom at different times of the year. Where is the Moon?

Grade 8

In which direction should an open-air sports field (football, tennis, baseball, etc.) be oriented, considering the position of the Sun and season in which the sport is played? Where is the Sun when we see the Moon at each of its various phases? Is the Moon full on the same day everywhere on Earth?

Grade 12

How long an overhang does one need on a house to admit maximum sunshine in winter and provide appropriate shade in summer? Why don't we spin off the rotating Earth? Why doesn't Mercury fall into the Sun? Do the horns of the Moon's crescent have the same orientation at different latitudes? Longitudes? Why?

The Earth and Forces That Shape It

With respect to the Earth, the 1994 AEP Science Assessment should center on the following concepts:

Climate (*Patterns of Change*)
Water Cycle and Ground Water (*Systems*)
Pollution Capacity (*Systems, Models*)
Interior Effects (*Models, Systems, Patterns of Change*)
- crustal plates
- rock cycle and strata

Exterior Effects (*Systems, Patterns of Change*)
- weathering
- plants, animals, civilization

Students should understand that the Earth is a unique member of the solar system but may be replicated in other galaxies in the universe; it is at least 4.5 billion years old and is a complex planet with several interacting systems; these systems have evolved through time, and changes in them occur over periods of microseconds to millions of years and vary from subatomic to astronomical. Students also should understand that the Earth's systems contain a variety of renewable and non-renewable resources that sustain life (American Geologic Institute, 1991).

Climate: Atmosphere and Hydrosphere

The atmosphere is the gaseous envelope that surrounds the Earth; it is continuously in motion, circulating in complex but regular patterns and driven by direct and stored solar energy. There are strong interactions between the atmosphere and the hydrosphere that determine weather and climate, profoundly influencing human and all other life. Desired learning goals:

Grade 4

- Students are able to communicate what is special about air. What do our senses tell us about the air? What needs air?
- Students will offer simple explanations for how the weather changes. How do we know when weather changes? How can we measure changing weather conditions?

Grade 8

- Students know about basic weather-related phenomena: tornados, hurricanes, drought, acid precipitation.
- Students know that relatively small changes in global temperatures could have dramatic effects on the Earth.
- Students can access climatological information bases, via computers and other means to extract useful information.
- Students can read a weather chart and can extract basic information.

- Students understand and can use simple instruments measuring basic phenomena related to weather (e.g., barometers—have made one, know how they work).

Grade 12

Generally, 12th-grade students are able to connect relationships between atmospheric phenomena and long-term effects. They understand that much of what determines the details of the weather depends on phenomena such as sea breezes, thunderstorms, tornadoes, and wind shear. Further, they understand how scientists can monitor atmospheric events over time (e.g., products of pollution and carbon dioxide deposition at the poles). Students are able to discuss and relate causes and possible solutions with their consequences and trade-offs.

Water Cycle and Ground Water

Desired learning goals:

Grade 4

- Students understand that water exists not only on the surface of the Earth, but beneath as well.
- Students understand that water is able to change the shape of the Earth. They can demonstrate their knowledge in two ways: by interpreting common local land features such as rivers, streams, mountain slopes, delta deposits and by designing simple models that illustrate what they are attempting to show.
- Students understand that most of the Earth's surface is covered by water.
- Students know about how waters of the Earth circulate. They can answer questions concerning where we find water, how water enters and leaves the atmosphere, and how we can use water more wisely.

Grade 8

- Students should know and demonstrate why water is special. What properties make water special? Where is water found—in the air, on earth, and under the ground?
- Students know that the oceans provide habitats for a wide variety of plant and animal life.
- Students can discuss some common problems that concern water: its availability in their areas, shortages, relationship to supply and demand, and the effects of over-population on the availability and quality of potable water.
- Students are able to relate common interactive cycles such as water cycle, nitrogen cycle, carbon cycle.
- Students are familiar with some of the ways scientists explore the water environment.

Grade 12

At grade 12, students are able to explain how water (or the lack of it) relates to their immediate state, city, or area. They can discuss national issues related to water, such as acid precipitation and the effects of global warming. They are able to understand how common cycles affect the climate of the area where they live. Students should know how to find answers to problems relating to how laws affect our use of water; and what hazards are associated with water, and how can we mitigate them.

Interaction of Earth's Systems

Desired learning goals:

Grade 4

- Students understand how the Earth relates to the Sun: periodicity, seasons, night/day.
- Students know about tides with respect to the Earth, Moon, and Sun.
- Students understand basic facts about volcanos and glaciers.
- Students understand how we can investigate rocks and minerals, what they are made of and how they form.

Grade 8

- Students are able to understand how earthquake occurrences are recorded and note some positional regularities.
- Students are able to connect short-term changes in climate with volcanic activity.
- Student. can discuss, in historical terms, that there have been changes in water levels, global temperature, and climate zones over the eons. They can advance some hypotheses as to why these changes have occurred.
- Students can identify how and where we get energy from the Earth; Earth materials that we use, and where to find them; and the advantages and disadvantages of using the Earth's resources.

Grade 12

Students in grade 12 can connect "the zone of fire" with plate tectonics. They understand how sliding plates cause sudden earth movements. They can discuss issues related to earth systems: why continents move, that they have not always been arranged the way they are now, how human activity affects earth systems. Students can access data from more than one data source to develop an environmental impact statement on their region/town/block. Students can discuss problems associated with agriculture and the lithosphere. They can relate these problems to changing atmospheric conditions.

Physical Science

The physical science component of the 1994 NAEP Science Assessment should probe the following major topics: matter and its transformations, energy and its transformations, and the motion of things. Each is detailed below, together with some example topics appropriate for assessment at the different grade levels.

Matter and Its Transformations

With respect to matter and its transformations, the 1994 NAEP Assessment should center on the following concepts:

"Many from Few"

- Diversity of materials (*Models*)
- Classifications of and types of materials (*Patterns of Change*)

- Particulate nature of matter (*Models*)

Temperature and States of Matter

Uses of Materials

- Properties and uses
- Modifying properties by mixing, processing, reacting (*Patterns of Change, Models*)
- Synthesis of materials with new properties (*Patterns of Change, Models*)

Resource Management

- Resource depletion, substitute materials (*Models*)
- Disposal and recycling (*Systems*)

Materials encountered in the physical world differ greatly in shape, density, flexibility, texture, toughness, and color; in their ability to give off, absorb, bend, or reflect light; in the form they take at different temperatures; and in many other ways. Students need to understand that these varied substances are made up of relatively few kinds of basic materials—the atomic elements—combined in various ways. Only a few of these elements are abundant in the universe.

As they advance in science, students should come to understand that the basic premise of the modern theory of matter is that materials consist of a limited number of different kinds of atoms (elements) which join together in different configurations to form substances. Thus, when substances react to form new substances, the elements composing them combine in new ways and the properties of the substances created by the new combinations may be very different from those of the old. Almost every substance can exist in a variety of states—solid, liquid, and gaseous—depending on temperature and pressure.

Patterns within the structure of matter have been elegantly described in the Periodic Table, an outstanding example of a *model* that provides a *systematic* view of *matter* and its interactions. Changes in temperature and change in state represent a category of *physical change* among substances, within the realm of matter and energy.

An understanding of the particulate nature of matter can be assessed through the identification of types of materials; for example, "mystery powders" or its equivalent at the elementary level. The effects of temperature and heat energy on systems are exemplified in how refrigerators work, how and why ice cubes melt, or how metallic fuses work. Some practical areas of human activity (e.g., cooking and much of modern chemical industry) and processes (transport of materials in biological systems) are related both to the nature of matter and the effects of external factors on its behavior. All these topics can provide rich assessment exercises.

The properties of matter determine uses to which particular materials are put by manufacturers, engineers, and others involved in technology. Materials can be physically combined or processed to serve human needs. Modern materials technology has focused increasingly on the synthesis of materials with entirely new properties. Chemical changes are typically involved, and the properties of the new materials—such as plastics and ceramics—may be entirely different from those of its constituents.

The growth of technology has led to the use of some materials from the environment—such as forests, ore deposits, and petroleum—much more rapidly than they can be replaced by natural processes. There is a continuing search for substitute materials—and in many cases they have been found or invented.

Disposal of used materials has become an increasing problem. Some used materials, such as food scraps and waste paper, can be returned safely to the environment—although as the population grows, the task becomes more difficult and expensive. Other materials, such as aluminum scrap and glass can be recycled, with resulting savings in energy and resources. Some materials, such as plastics, are not easily recycled nor do they degrade quickly when returned to the environment. Other used materials—radioactive waste being the most dramatic but not the only example—are so hazardous for such a long time that how best to dispose of them is not clear and has become the subject of widespread debate and controversy. Solving these problems of disposal will require systematic efforts that include both social and technological innovations. Assessment questions dealing with the scientific and technological issues involved in resource management are appropriate for grades 8 and 12.

Energy and Its Transformations

With respect to energy and its transformations, the 1994 NAEP Assessment should center on the following concepts:

- Forms of Energy
- Energy Transformation (qualitative) and Audits (quantitative) in
  - Living Systems
    - Plants
    - Animals
    - Protista
  - Natural Physical Systems
  - Artificial (Human Constructed) Systems
- Energy Sources and Use
  - Quantity/Kind
  - Distribution (*Patterns of Change*)
  - Energy conversions, heat gain/loss, efficiency (*Patterns of Change, Models*)
  - Slowing depletion of energy sources (conservation)
  - Costs, implications, advantage, risks, availability (*Patterns of Change*)

Students should understand that the concept of energy is central to understanding changes observed in natural and artificial systems. Observable changes occur when energy is added to a system, when energy is removed from a system, or when energy is transformed from one form to another. Energy appears in many *forms* and is categorized in different ways: Light, Heat, Sound, Kinetic and Potential (Electromagnetic, Electrical, Chemical, Gravitational, Elastic), Consumable and Renewable, Available and Unavailable Energy. Although various forms of energy appear very different, each can be *measured* in a way that makes it possible to keep track of how much of one form is transformed into another.

Students need to understand that energy is *conserved*. Within a system, whenever the energy in one place or form changes, the quantity of energy in another place or form increases or decreases by a similar quantity, but the total energy remains the same. Thus, if no energy leaks in or out across the boundaries of a system, the total energy of all the different forms in the system will not change, no matter what kinds of changes occur within the system.

Energy *transformations* usually result in producing some thermal energy (heat), which "leaks away" by radiation or conduction (i.e., from engines, electrical wires, hot-water tanks, human bodies, and stereo systems) and becomes unavailable for further transformations. Thus, the total quantity of energy available for transformation usually decreases.

Students need to develop an understanding of the more general principle that natural processes occur in the direction of increasing the *total* disorder of the system and its surroundings. Although some subsystems do increase in orderliness (such as the freezing of water to form ice), another part of the system or a connected system becomes more disordered. The cells of a human organism, for example, are always busy increasing order. as in building complex molecules and body structure. But this occurs at the cost of increasing the disorder around us even more—as in breaking down the molecular structure and order of food we eat and in warming up our surroundings.

Energy *transformations* occur naturally and in devices constructed by humans.

Naturally-occurring transformations

- Solar energy into stored energy such as starches, fats and proteins
- Solar energy into heat
- Potential energy into kinetic energy (e.g., the potential energy of roller coasters at the top converting to kinetic energy on the way to the bottom).

Transformations occurring in human artifacts

- Electric mixer converts electrical energy into mechanical energy
- Hair dryer converts electrical energy into heat energy
- Automobile converts chemical energy into mechanical energy

In the operation of these devices. as in all phenomena, the useful energy output—that is, what is available for further change—is always less than the energy input, with the difference usually appearing as heat. One goal in the design of such devices is to make them as efficient as possible; that is, to maximize the useful output for a given input and to minimize wasted heat energy.

Radiant energy from the Sun is the ultimate *source* of most of the energy we use. It becomes available to us in several ways: The energy of sunlight is captured directly in plants, which then may be eaten; it also heats the air, land, and water, causing wind and rain. For much of history, burning wood was the most common source of intense energy for cooking, for heating dwellings, and for running machines. Most of the energy used today is derived from burning fossil fuels, which contain stored solar energy that plants collected over millions of years. A new source of energy is the fission of the nuclei of heavy elements, which—compared to the burning of fossil fuels—releases an immense quantity of energy in relation to the mass of material used. In nuclear reactors, the energy generated is used mostly to heat water into steam, which drives electric generators.

Humans use energy for technological processes: transporting, manufacturing, communicating, and getting raw materials, then working them and recycling them. Students need to appreciate that different sources of energy and ways of using them involve different costs, implications, and risks. Some of the resources will continue to be available indefinitely; some can be made self-renewing, but only at a limited rate. Fuels like coal, oil, natural gas, and uranium will become more difficult to obtain as the most readily available sources become depleted. New technology may make it possible to use the remaining sources better; the ultimate limit may be prohibitive cost rather than complete disappearance.

Students should know that depletion of non-renewable energy sources can be slowed by both technical and social means. "Technical means" include maximizing the advantage that we realize from a given input of energy by means of good design of the transformation device, through insulation if we want to restrict heat flow (e.g., insulating hot-water tanks), or by doing additional work with the heat as it dissipates. "Social means" include government, which may restrict low-priority uses of energy or may establish requirements for efficiency (as in automobile engines) or for insulation (as in house construction). Individuals also may make energy efficiency a consideration in their own choice and use of technology (e.g., turning out lights and driving high-efficiency cars), either to conserve energy as a matter of principle or to reduce their personal long-term expenses. Students need to appreciate that there will always be trade-offs. For example, better insulated houses restrict ventilation and thus may increase the indoor accumulation of pollutants.

The bases for these energy-related concepts should be laid in elementary school science. The following examples illustrate appropriate activities that can be used to formulate assessment exercises for 4th grade:

1. The student gives examples from his/her own experience of heat energy and light energy changing a system.

2. The student identifies the source of energy for a familiar system (animal, plant, car, electric appliance) and describes some of the energy conversions that take place in each system.

3. The student identifies the energy stored in a stretched rubberband and in a compressed or stretched spring as potential energy. The student explains that to store potential energy in a rubberband or in a spring he/she must exert a force to stretch the rubberband, or stretch or compress a spring.

4. Given pictures of several situations, some of which depict a force being exerted or work being done and some of which do not, the student identifies those pictures in which a force is being exerted.

5. The student explains, using the words fuel and energy in context, why a candle goes out when the wax is used up.

6. The student writes a short essay on how his/her life would be different if all the coal and petroleum on Earth were used up.

The Motion of Things

With respect to this topic area, the 1994 NAEP Framework should center on the following concepts:

Reference Frames

Motion
- Force and changes in position and motion
- Action and reaction

Waves
- Vibrations and waves as motion summaries
- General wave behavior
- Electromagnetic radiation
- Effects of wavelength
- Interactions of electromagnetic radiation with matter

What students at the three grade levels should understand about these concepts is summarized below.

Reference Frames

Grade 4

Everything moves—bicycles, cars, trains; the stars, planets, and moons; the Earth and its surface, and everything on its surface; all living things and every part of living things.
Positions of things may be described, positions may change. Monitoring changes in time yields information about speed.

Grade 8

No special point in space can serve as a reference for all other motion. All motion is relative to whatever point or object we choose.

Forces and Motion

Grade 4

Changes in motion—that is, changes in speed or in direction—are due to the effects of forces.

Grade 8

Any object maintains a constant speed and direction of motion (including being at rest) unless an unbalanced outside force acts on it. When an unbalanced force does act on an object, the object's motion changes. Depending on the direction of the force relative to the direction of motion, the object may change its speed (a falling apple), or its direction of motion (the Moon in its curved orbit), or both (a fly ball). The greater the extent of the unbalanced force, the more rapidly a given object's speed or direction of motion changes. In most familiar situations, friction between surfaces brings forces into play that complicate the description of motion, although the basic principles still apply.

Grade 12

The more massive an object is, the less rapidly its speed or direction changes in response to any given force. Whenever something A exerts a force on something B, B exerts an equal force back on A, but in the direction opposite of the force exerted by A.

Vibrations and Waves

Grade 4

Some motions can be described most conveniently not in terms of forces directly but in summary descriptions of the pattern of motion, such as vibrations and waves. Vibrations may set up a traveling disturbance that spreads away from its source.

Grade 8

Vibration involves parts of a system moving back and forth in much the same place, so the motion can be summarized by how frequently it is repeated and by how far the parts of a system are displaced during the cycle. Vibration may move through a system as a wave. Wave behavior can be described in terms of speed, wavelength, and frequency. Wavelength can help determine how a wave interacts with things—how well it is transmitted, absorbed, reflected, or diffracted.

Grade 12

Apparent change in wavelength can provide information about relative motion. The ways in which shock waves of different wavelengths travel through and reflect from layers of rock are an important clue regarding the structure of the Earth's interior.

Light as Waves

Grade 4

White light is made up of all different colors of light. Things appear to have different colors because they reflect or scatter light of some colors more than others.

Grade 8

Light behaves in many ways like waves—changing direction, bouncing off surfaces, spreading out, speeding up, slowing down, changing wavelength.

Grade 12

The interaction of electromagnetic waves with matter varies greatly with wavelength. Thus, different but somewhat overlapping electromagnetic ranges have been given distinctive names: radio waves, microwaves, radiant heat or infrared radiation, visible light, ultraviolet light, X rays, and gamma rays. Materials that allow one range of wavelengths to pass through them may completely absorb others. For example, some gases in the atmosphere—including carbon dioxide and water vapor—are transparent to much of the incoming sunlight but not to the infrared radiation emitted by the warmed surface of the Earth. Consequently, heat energy is trapped in the atmosphere. The temperature of the Earth rises until its total radiation output reaches a state of balance with total radiation input from the sun.

Life Science

The fundamental goal of life science is to attempt to understand and explain the nature of life. During the 20th century, the thrust of biological research has changed its focus from descriptive natural history to experimental science, with most biological investigations conducted within the theory of evolution. The major concepts to be assessed in the life sciences, with evolution as the central, unifying theory, are listed below and developed further in the grade-level descriptions that follow:

Evolution
- The diversity of life on Earth
- Genetic variation within a species
- Adaption and natural selection
- Changes in diversity over time

Cells
- Information transfer
- Energy transfer
- Cellular communication

Organisms
- Reproduction, growth, and development
- Life cycles
- Functions and interactions of systems within organisms

Ecology
- The interdependence of life: populations, communities, and ecosystems

The three themes in the Science Framework (Patterns of Change, Systems, and Models) can be interwoven with these major concepts in the life sciences. Since evolution is the major pattern of change that occurs in the life sciences, the **Patterns of Change** theme can enhance understanding of all of the life science concepts listed above. Since the **Systems** theme can pertain to systems at the cellular, organismal, population, community, and ecosystem level, this theme also can enhance understanding of most, if not all, of the life science concepts. Although **Models** are used in the life sciences, this theme receives less emphasis in life science than in the physical and earth sciences, particularly at grade 4 and grade 8.

Students' understanding of life science concepts develops gradually as the students proceed from grade 4, to grade 8, to grade 12. A description of the developmentally appropriate concepts that should be understood at each of these grades follows:

Grade 4

Organisms

- As some animals grow, they look pretty much the same—they just increase in size. As other animals grow, they change from one form to another form that looks very different. They may change form several times before they become adults.
- Only adults can reproduce, but not all young animals survive long enough to become adults.
- Many activities go on inside the body that can not be seen—when something happens in one part of the body, it affects what goes on in other parts of the body.

Ecology

- Plants make their own food with sunlight, water, and air.
- Some animals eat plants; some of these animals are eaten by other animals.
- Plants and animals get energy and building materials from their food.

Evolution

- There are different kinds of plants and animals on Earth and in the sea.
- There also are differences between individuals of the same kind of plant or animal.
- Children of the same parents are somewhat alike and somewhat different.

Grade 8

Organisms

- Different systems of the body have different functions; however, the functioning of each system affects other systems.

- Interactions between systems are complex—these interactions maintain a fairly stable operation of the entire system that can resist disturbance from within or without.

- Interaction with other organisms (especially micro-organisms) are important to maintain health or cause disease. Avoiding or killing microorganisms can prevent disease.

Ecology

- Plants use energy in sunlight to assemble food molecules from water and carbon dioxide.

- Plants and animals break down food molecules to obtain food energy.

- The source of energy and materials for all animals is plants.

- The pattern of what-eats-what in a community can be complex.

Evolution

- The organisms that survive long enough to reproduce may be different in some ways from others in a population that do not survive long enough to reproduce; their offspring may inherit the anatomical, chemical, and/or behavioral characteristics that enabled the parents to survive.

- Gradually, over many generations, organisms with the favorable characteristics may crowd out other organisms in the population that do not have these characteristics.

- Scientists believe that these processes, operating over very long periods of time, have resulted in the diversity of organisms that can be seen on Earth today.

- Adaptation may be to either the living or non-living components of the environment.

Grade 12

Cells

- Every cell contains a recipe for running the cell, coded in DNA molecules: the code mainly specifies how to put proteins together.

- During cell reproduction, the information in the DNA code is passed on to the next generation of cells.

- Proteins control most of what goes on within cells and within the body.

- There are interactions between the cells of an organism—molecules from one cell affect what goes on inside other cells.

- In plant cells, energy from sunlight is transformed to chemical energy during photosynthesis: in plant *and* animal cells, the chemical energy stored in food molecules is released during digestion and produces heat. Some of this released energy is used to build new molecules.

Organisms

- Separate parts of the body can communicate with each other using electrical or chemical signals.

- Complex interacting systems include feedback that tends to produce cycles of activities within the body.

- In organisms that reproduce sexually, each parent passes on one-half of its DNA information to each of its offspring; therefore, half of the DNA in each cell of an organism came from one parent, and half from the other parent.

Ecology

- Interactions between living and non-living components affect how ecosystems function as a whole.

- A change in one component of an ecosystem affects other components of an ecosystem. These components in turn react in a way that will restore the ecosystem to its original condition.

- Often changes in one component of an ecosystem will have effects on the entire system that are difficult to predict.

- The size of a population and the rate of growth of a population are determined largely by the survival rate, the reproductive rate, and the death rate of the organisms in the population. Predictions about changes in the size or rate of growth of a population can be described using mathematical models.

Evolution

- Recombination and mutation are the raw materials for new traits upon which natural selection acts.
- When the environment changes, different characteristics may be important for survival—different adaptations are important for survival in different environments.
- Some descendants are so different from other descendants that they can no longer breed with each other.

Particularly at grade 12, students should be able to integrate information from different concepts within the life, physical, and earth sciences. They should understand how key concepts apply at different levels of biological organization (molecular, cellular, organism, population, community, ecosystem, and biome) and how these concepts apply to current societal problems and are significant to the development of a variety of biotechnologies. They should be able to describe common misconceptions about natural phenomena and describe how these explanations are contrary to contemporary scientific explanations. Students also should understand the effects of technologies created by humans on the life cycles of organisms and their effects on communities of plants and animals, including humans. In addition, students should have developed some familiarity with the historical development of key concepts in the life sciences.

APPENDIX B

EXAMPLES OF THEMES

Themes by Grade Level

Systems

Students should understand that systems are artificial constructions created by people for certain purposes—to gain a better understanding of the natural world or to design an effective technology.

The construct of a system entails identifying and defining its boundaries, identifying its component parts, identifying the interrelations and interconnections among the component parts, and identifying inputs and outputs of the system.

This theme should be embedded in life science learning at the three grade levels in the following ways:

Grade 4

Systems should be approached at the level of organisms. Students should have broad and rich acquaintance with structure/function relationships as a precursor to a more thorough knowledge of organ systems by grade 8. Examples of food chains and interdependencies among organisms, say, within an aquarium, are precursors to understanding complex systems.

Grade 8

Students should understand that the organism is made up of organ systems which have structure/function adaptations and interconnections among organ systems.

Interdependence of plants and animals in communities should be understood by grade 8: plants → consumers → decomposers. They should be able to explain specific examples: purple loose strife replacing cattails, the effects of the introduction of rabbits into Australia.

Disease and health should be understood in systems terms; if a part of a system is put out of kilter by disease, for example, the whole is affected. Taking drugs or smoking by an individual may impact on another system (organism); for example, secondary smoking effects on children of smoking parents, fetal damage from drugs. Measles vaccine taken by an individual, or not, impacts on the whole population of a region and further, depending on migration patterns. If a specific animal or plant population becomes unhealthy (fish poisoned, raccoons diseased, species of grass infected by virus), the food chain and, therefore, the rest of the community are affected.

Grade 12

Ecosystems should be understood in their full complexity, including interrelationships of plants and animals with each other as well as with the physical components of a system. Students also need to recognize the effects of human activity on ecosystems and the limitations on human activity imposed by natural systems.

At this level, the cell should be understood as a system itself, and also as a component of a system.

Patterns of Change

"Patterns of Change" is a particularly valuable theme in the life sciences because a conceptual understanding of patterns of change can be developed in the context of several different levels in the hierarchy of biological organization. At the *cellular and organismal* level, the primary pattern of change is the *growth and development* that occurs throughout the life of organisms. At the *population* level, the primary patterns of change are the changes in *population growth* over relatively short periods of time and the *evolutionary changes* that occur over longer periods of time. At the *community/ecosystem* level, the primary patterns of change are those that involve the non-living and living components of ecosystems during the process of *succession.* Patterns of change may be linear, or they may be cyclical; for example, many of the patterns of change that occur within cells are related to *homeostasis*, in which a change leads to feedback reactions that result in a return to conditions that existed before the change. An understanding of cyclical patterns of change can also be developed in the context of ecosystem (nutrient cycles) and in the context of organisms (life cycles).

Regardless of the context in which an understanding of the "patterns of change" theme is developed, students should be able to recognize patterns of similarity and difference, to recognize how these patterns change over time, and to transfer their understanding of a familiar pattern of change to a new, unfamiliar situation.

In order to understand the conceptual basis for the "patterns of change" theme, students must begin to develop an understanding of major ideas by the 4th grade, continue to develop their understanding through the 8th grade, and integrate their knowledge at the 12th grade.

Grade 4

Understanding Patterns of Change at the Organismal Level
- Life cycles (including growth and metamorphosis)

Understanding Patterns of Change at the Population Level
- Concept of biotic potential, birth rates, survival rates
- Diversity of many types of plants and animals (an important pre-concept for the understanding of evolution)
- Variation within species (focus on humans, dogs, cats)

Understanding Patterns of Change at the Community/Ecosystem Level
- Food chains (also important for the "systems" theme)

A more general understanding involves the notion that everything changes, sometimes quickly and sometimes slowly, and that changes may be too rapid or too slow to observe directly.

Grade 8

Understanding Patterns of Change at the Organismal Level
- Growth, development, and reproduction of the human organism
- Homeostasis of body systems

Understanding Patterns of Change at the Population Level

- Adaptation and natural selection, including learned and instinctive behavior as example
- Variation and similarity among many different organisms, including humans

Understanding Patterns of Change at the Community/Ecosystem Level

- Food webs (also part of "systems" theme)
- Environmental effects of human activity (also part of "systems" theme)

More general understandings involve the following knowledge: Changes in quantity usually have natural limits, but changes in form in which each form arises from a previous one can produce an unlimited variety. The rate of change may be as interesting as the change itself. Trends can be steady, accelerated, approach a limit gradually, or have a highest or lowest value.

Grade 12

Students should have acquired an understanding of these additional concepts and developed the ability to integrate them into the "patterns of change" theme.

Understanding Patterns of Change at the Cellular/Organismal Level

- Growth and development of cells, including an understanding of the importance of mitosis and meiosis

Understanding Patterns of Change at the Population Level

- Patterns of evolution, mechanisms for evolution, the consequences of evolution (such as speciation and diversity through time), evidence for evolution

Understanding Patterns of Change at the Community/Ecosystem Level

- Nutrient cycles and the impact of human activity on those cycles
- Succession, both natural and as a result of human disturbance

More general understandings entail knowing that trends, cycles, and randomness can occur at the same time; randomness may make it hard to see trends or cycles; on the other hand, sometimes randomness may look like a trend or cycle. Feedback in systems—often an influence that reacts against change—tends to produce cycles; changes that follow precise rules from one moment to the next may still be unpredictable in the long run: the environment in which any one change occurs is usually changing also, and they affect each other.

Models

Models of objects and events in nature are approaches to understanding. As such, they have limits, and involve simplifying assumptions but also possess generalizability and, sometimes, predictive power. Models are composed of groups of interrelated concepts, selected in an attempt to represent the interrelations of objects or events in nature or in the laboratory. Models need not be deemed correct to be useful, but may represent attempts to help tease out the relevant variables, in order to build ever more useful representations.

Models may be conceptual and consist of word descriptions or drawings. Models can also be mathematical, consisting of equations or other formal representations. Finally, physical models consist of physical objects that possess or represent some characteristics of the real thing.

The solar system often is modeled conceptually in the classroom by describing the planets as huge balls moving about an even larger sun. A mathematical model of the solar system should include quantitative descriptions of the gravitational forces between the planets and the sun as determined by their masses and distances from each other and might include the shape of a planet's orbit as being elliptical. And finally, a physical model of the solar system might consist of a series of scale-sized balls placed at appropriate distances throughout the room or hallway.

Other examples from the earth and physical sciences include models of beaches, of continental plates, and stick and ball models of molecules. Physical models, such as those of the eye, leaf, and human torso, have been used in the life sciences for decades. Experiments with animals serving as models of human beings have been used to understand the effects of medical treatments that might be useful in preventing human diseases: bacteria have been used to model population growth and decay.

Similarly, conceptual models are common in both the biological and physical sciences. The simplified treatment of photosynthesis, the stages of meiosis and mitosis, accounting for an electrical current in terms of a "water flow" analogy, and the characterization of gas molecules as bouncing balls are examples of commonly used conceptual models.

Mathematical models such as the gas laws and Newton's laws of motion are major components of the physical sciences. Some mathematical models, such as Mendel's laws have been part of the biological sciences for most of this century, whereas the Hardy-Weinberg formulation for describing ecosystems mathematically has been become part of introductory biological knowledge more recently.

Models often serve as prototypes in technology and in that case may be a full-sized representation of the final product. However, models can be used to test the workings of technology without costly investments in full-scale objects. Small boats and airplanes are tested in tanks and wind tunnels before their full-sized counterparts are built. In this way, many experiments can be tested inexpensively to optimize the design.

"Models" can be easily developed as a theme even by grade models and can be linked to the immediate experiences of children since they have grown up with a variety of toys. Children readily understand that most toys are models that look like the real objects, such as cars, airplanes, babies, and animals, but do not possess all the attributes of those objects. Many of these toys are models, sometimes scale models, of objects from the natural world. For example, models of dinosaurs enable children to develop ideas about what these creatures were like.

"Models" has been selected as a theme in this assessment because of the importance of enabling students to distinguish between the idealizations of models from the phenomena themselves. Students need to understand that the model of the human eye does not represent all aspects of human eyes as they occur in human organisms. The model is a simplification, leaving unrepresented the many important variations in human eye structure, yet the simplification has utility in illuminating some features of the eye and enables new questions about the eye to be generated.

Students need to understand the limitations and simplifying assumptions that underlie the varied models used in the natural sciences. Beliefs that models are replicas of "real" objects or events can, for example, negate the critical concept of variation that many models do not take into account. While generalized models, such as a generalized graph of growth in populations, are useful, they are not to be confused with a graph of the growth of a particular organism or population or of data from a single experiment.

Grade 4

At this level, models should be identified by students as representations of objects or events. Both conceptual and physical models can be examined by students as to how they are like and not like the object or event being represented. Examples could be models of insects, seeds, leaves, and other physical objects. These models and others in the sciences can be linked to children's experiences with scale models of cars, dinosaurs, and doll furniture.

Grade 8

Students should have knowledge of both conceptual and physical models and their uses and limitations. For example, when asked to illustrate their understanding of vertebrate structure and function with models of skeletons of different vertebrates, students need to be aware of variations in real skeletons and the generalized nature of replicas of skeletons.

Grade 12

Mathematical, physical and conceptual models should be familiar to students beyond grade 8. It is appropriate to assess students' ability to formalize the concept of models and their uses and limitations in the natural sciences and in technology.

APPENDIX C

SAMPLE ITEM CONTENT

Grade 4—Earth Science

A. Solid Earth (lithosphere)

1. Composition of the Earth

a. Students can classify substances such as soil, sand, or rock.

b. Students can identify common geographic features of landscapes.

2. Forces that alter the Earth's surface

a. Students can describe/identify/explain basic facts about major features of the Earth's surface and natural changes in those features (e.g., volcanos, glaciers).

- Students can predict the effects of weathering (e.g., rain and wind on sand piles, mud piles, or rock).

- Students can describe the relative difference in time it takes to erode a sand pile, a mud pile, and a rock pile. (*Conceptual Understanding*; *Patterns of Change*)

- Given a picture, topographical map or globe, or word description of a major Earth feature (e.g., canyon, mountain range, Great Lakes, caverns, or islands), students can identify a geologic force that contributed to producing that feature. (*Conceptual Understanding; Models*)

3. Rocks: their formation, characteristics, and uses

a. Students can identify common rocks and minerals and can explain how we can investigate what they are made of and how they form.

- Students can classify rock samples according to color, texture, or other identifying properties. (*Scientific Investigation*; *Nature of Science*)

- Students can explain that molten rock comes out of volcanos, hardens, and becomes part of the landscape. (*Conceptual Understanding*; *Patterns of Change*)

4. Soil: its changes and uses

a. Students know some facts about the composition of soil.

- Students can separate soil samples into component parts. (*Scientific Investigation*: *Nature of Science*: *Systems*)

b. Students recognize that plants grow in soil and that the soil provides both nutrients and support for the plant.

- Students can classify and relate major soil types (e.g., clay, sand, loam, subsoil) to their ability to support plant growth; that is, can identify/predict the major plant types likely to grow in those soils. (*Conceptual Understanding*; *Nature of Science*)

5. **Resources from the Earth used by humankind**

a. Students can identify Earth resources used in everyday life.

- Students can identify common uses of rock in the human environment (e.g., buildings, roads, walls). (*Practical Reasoning*; *Nature of Technology*)

b. Students can explain/identify that gasoline is processed from oil, which is pumped from the Earth. (*Practical Reasoning*; *Nature of Technology*)

Grade 8—Life Science

A. **Cells and their functions**

1. **Cells**

a. Students can describe their observations of cells under the microscope.

- Students can demonstrate the use of a microscope to examine a tissue, plant, or animal and to differentiate between plant and animal cells (e.g., students can look at an animal cell and a plant cell and notice that animal cells are flexible and plant cells are not). (*Scientific Investigation; Systems*)

- Students can look at pond water through a microscope and describe outstanding features/activities of the protista they see (e.g., locomotion, nutrition, excretion). (*Scientific Investigation*)

- Students can observe diatoms and try to distinguish as many features as possible. (*Scientific Investigation*)

b. Students can explain, in a general way, the advantages of cellular interdependence vs. independence (e.g., multicellular animals vs. single-celled animals).

c. Students can describe, in general terms, the difference between asexual and sexual reproduction in cells and the advantages and disadvantages of each. **[The stages of mitosis are not to be tested.]**

B. Organisms

1. Reproduction, growth, and development

a. Students can describe growth, development, and reproduction of the human organism.

- Students can identify the age ranges at which human beings go through common stages of development (e.g., can recognize their parents; can learn to walk, talk, socialize; can conceive or give birth). (*Conceptual Understanding; Patterns of Change*)
- Students can identify the changes human beings undergo at puberty and can explain their functions. (*Conceptual Understanding; Patterns of Change*)
- Students can, in simple terms, describe changes in human embryo development and the effects of environmental influences, such as smoking, drugs, disease, and the diet of mothers on the development of the embryo. (*Conceptual Understanding; Patterns of Change*)

2. Life cycles

a. Students can identify some major influences on the human life cycle (e.g., diet and disease).

- Students can discuss the influence of diet and food availability on human life cycles worldwide. (*Practical Reasoning; Patterns of Change*)
- Students can explain that micro-organisms can cause disease and can identify some common diseases caused by micro-organisms (e.g., bacteria, viruses, protista). [Differences between viruses and bacteria are not to be tested.] (*Conceptual Understanding*)
- Students can describe the immune system of animals as helping the animal fight disease and as controlled, in part, by the white blood cells in the body. (*Conceptual Understanding*)

3. Functions and interactions of systems within organisms

a. Students are aware that, while different systems of the body have different functions, the functioning of each system affects other systems (e.g., students can describe/identify major organ systems of the human body, state their major functions, and describe some of their interactions).

- Students can describe the primary tissues of the body and relate the special characteristics of each to its function (e.g., blood, lymph, muscle). (*Conceptual Understanding; Systems*)
- Students can distinguish cells from other structures under the microscope (e.g., can distinguish between an onion cell and a salt crystal). (*Scientific Investigation; Systems*)
- Students can describe how two or more organs of the body work together to perform a function (e.g., the heart and lungs working together in respiration). (Conceptual Understanding; Systems)

b. Students demonstrate an understanding of the functions and interactions of organ systems to maintain a stable internal environment that can resist disturbance from within or without (homeostasis).

Grade 12—Physical Science

A. Matter and its transformations

1. Diversity of matter (materials): Classification and types, particulate nature of matter, conservation of matter

a. Students can distinguish/classify objects, both regular and irregular; pure substances, both elements and compounds; and mixtures, both homogeneous (solutions, liquids, and gases) and nonhomogeneous.

b. Students can describe, measure, and compare substances in terms of mass, volume, and density/specific gravity.

- Given a substance of unknown volume or weight and appropriate laboratory equipment, students can determine its specific gravity. (*Scientific Investigation*)
- Given an irregular solid and the appropriate laboratory equipment, students can determine the density of the object. (*Scientific Investigation*)
- Students can offer a simplified distinction between weight and density. (*Conceptual Understanding*)

c. Students can identify evidence that matter is composed of tiny particles (e.g., atoms, molecules) and that the particles are in motion (kinetic molecular theory).

d. Students can define, describe, and contrast physical, chemical, and nuclear changes in molecular terms.

. Given various examples of changes in materials, students can distinguish among chemical, physical, and nuclear changes. (*Conceptual Understanding; Patterns of Change*)

e. Students can discuss the conservation of matter in physical, chemical, and nuclear changes. **[Can also be tested under temperature and states of matter, or energy and its transformations.]**

2. Temperature and states of matter (physical changes)

a. Students can discuss/identify the relation of physical states of matter to molecular energy.

. Students can associate energy states with molecular motion. (*Conceptual Understanding*)

. Students can discuss/identify the energy transfers involved in the change of phase from solid to liquid to gas and the reverse. **[Also tested under energy and its transformations.]** (*Conceptual Understanding; Patterns of Change*)

b. Students can discuss/identify the relation of physical changes in substances (i.e., melting, boiling, thermal expansion and contraction, compression and expansion under pressure, increase or decrease in density) to changes in the structural organization of the atoms or molecules of which they are composed.

. Students can explain how antifreeze solutions work to prevent freezing of the water in car radiators. (*Practical Reasoning; Systems*)

. Students can explain why NaCl is added to ice when making ice cream. (*Practical Reasoning*)

3. Properties and uses of materials: modifying properties, synthesis of materials with new properties

a. Students can relate the physical properties (e.g., compressibility, structural rigidity) of pure substances in solid, liquid, and gaseous states to the structural organization of particles in the substance and their freedom of motion.

. Students can explain/identify that the molecules in a crystal are arranged in a regular pattern which gives the crystal rigidity and causes it to take a simple geometrical shape. (*Conceptual Understanding; Models*)

b. Students can examine/utilize useful properties of materials.

- Given an unknown liquid and a universal indicator with chart, students can determine the pH of the liquid. (*Scientific Investigation*)

- Given an unknown marking pen, students can use paper chromatography to identify the brand of marking pen from among several options. (*Scientific Investigation; Nature of Science*)

c. Students can describe how common artificial materials are made, recognizing that substances can be designed to have certain properties, and that the addition of relatively small amounts of some substances can significantly alter the properties.

d. Students can describe how common artificial materials are disposed of or recycled and can discuss the technological and environmental issues involved in these processes.

4. Resource management

a. Students can discuss scientific, technological, environmental, and social issues involved in resource management.

- Students can discuss the issue of worker safety in manufacturing processes that involve poisonous chemicals. (*Practical Reasoning; Nature of Technology*)

APPENDIX D

STEERING COMMITTEE

1994 NAEP SCIENCE ASSESSMENT

STEERING COMMITTEE MEMBERSHIP

William O. Baker, Retired
AT&T Bell Laboratories
Murray Hill, New Jersey

Mary Louise Bellamy
Education Director
National Association of Biology Teachers
Reston, Virginia

Frank Betts
Director
Curriculum Technology Center
Association for Supervision
and Curriculum Development
Alexandria, Virginia

Glenn A. Crosby
Professor
Chemistry Department
Washington State University
Pullman, Washington

Hon. Wilhelmina F. Delco
Texas House of Representatives
Austin, Texas

Janice Earle
Director
Center on Educational Equity
National Association of
State Boards of Education
Alexandria, Virginia

John Fowler
Director
Triangle Coalition for
Science and Technology
College Park, Maryland

Johnnie Hamilton
Principal
Franklin Intermediate School
Chantilly, Virginia

Elam Hertzler, Retired
Secretary's Commission on
Achieving Necessary Skills
Washington, DC

Ann Kahn, Past President
PTA
Fairfax, Virginia

Douglas Lapp
Director
National Science Resources Center
Smithsonian Institution
Washington, DC

John Layman
Director
Science Teaching Center
University of Maryland
College Park, Maryland

Harold Pratt
Executive Director
Science and Technology Management
Jefferson County Schools
Golden, Colorado

Judith Torney-Purta
Department of Human Development
University of Maryland
College Park, Maryland

Douglas Reynolds
Chief
Bureau of Science Education
State Department of Education
Albany, New York

Bella Rosenberg
Assistant to the President
American Federation of Teachers
Washington, DC

Jane Sisk
Biology Teacher
Calloway County High School
Murray, Kentucky

Gerald Difford
Executive Director
Colorado Association of Schools
Englewood, Colorado

William B. Campbell
Executive Director
National Industry Council
for Science Education
College Park, Maryland

APPENDIX E

PLANNING COMMITTEE

and

CONSENSUS PROJECT STAFF

1994 NAEP SCIENCE ASSESSMENT

PLANNING COMMITTEE MEMBERSHIP

Andrew Ahlgren
Associate Director
Project 2061
American Association for
the Advancement of Science
Washington, DC

Bill Aldridge
Executive Director
National Science
Teachers' Association
Washington, DC

J. Myron Atkin
Professor
School of Education
Stanford University
Stanford, California

Joan Boykoff Baron
Assessment Coordinator
Connecticut Common Core of Learning
Bureau of Evaluation
and Student Assessment
Connecticut Department of Education
Hartford, Connecticut

Audrey Champagne
Professor
School of Education
State University of New York
Albany, New York

Sally Crissman
Lower School Head /
Science Teacher
Shady Hill School
Cambridge, Massachusetts

Edmund W. Gordon
Professor
Department of Psychology
Yale University
New Haven, Connecticut

Henry Heikkinen
Director, M.A.S.T. Center
University of Northern Colorado
Greeley, Colorado

George Hein
Professor
Lesley College
Cambridge, Massachusetts

Joseph L. Premo
Science Consultant
Minneapolis Public Schools
Minneapolis, Minnesota

James Robinson, Retired
Curriculum and Evaluation
Boulder Valley Schools
Boulder, Colorado

Thomas P. Sachse
Manager, Math and Science Unit
California Department of Education
Sacramento, California

Gary E. Skaggs
Test Development Analyst
Office of Research and Evaluation
Fairfax County Public Schools
Falls Church, Virginia

1994 NAEP Consensus Project Staff

Ramsay W. Selden, Director
State Education Assessment Center
Council of Chief State School Officers

Richard C. Clark
Consensus Coordinator

Senta A. Raizen
National Center for Improving
Science Education

Julia H. Mitchell
American Institutes for Research

Bonnie L. Verrico
Administrative Assistant

86

References

Aldridge, B.G.
1989 *Essential Changes in Secondary School Science: Scope, Sequence and Coordination.* Washington. D.C.: National Science Teachers Association.

American Association for the Advancement of Science
1989 *Science for All Amcricans: A Project 2061 Report On Literacy Goals in Science, Mathematics, and Technology.* Washington, D.C.: American Association for the Advancement of Science.

American Geologic Institute
1991 *Earth Science Education for the 21st Century: A Planning Guide.* Alexandria, Virginia: American Geological Institute.

Bybee. Rodger W., et al.
1989 *Science and Technology Education for The Elementary Years: Frameworks for Curriculum and Instruction.* Washington, D.C.: The National Center for Improving Science Education.

1990 *Science and Technology Education for the Middle Years: Frameworks for Curriculum and Instruction.* Washington, D.C.: The National Center for Improving Science Education.

California Department of Education
1990 *Science Framework for California Public Schools: Kindergarten Through Grade Twelve.* Sacramento, Calif.: California Department of Education.

Davis, F.
1990 "Assessing Science Education: A Case for Multiple Perspectives," In G.E. Hein, ed., *The Assessment of Hands-On Elementary Science Programs*, Grand Forks, North Dakota: North Dakota Study Group, 1990.

Frank, Phillip
1957 *Philosophy of Science, the Link Between Science and Philosophy.* A Spectrum Book. Englewood Cliffs, N.J.: Prentice-Hall.
Belmont, Calif.: Wadsworth.

National Center for Improving Science Education
in press *The High Stakes of High School Science.* Washington, D.C.: The National Center for Improving Science Education.

National Research Council
1990 *Fulfilling the Promise: Biology Education in the Nation's Schools.* Washington, D.C.: National Academy Press.

National Science Board Commission on Precollege Education in Mathematics, Science and Technology, The
1983 *Educating Americans for the 21st Century.* Washington, D.C.: National Science Foundation.

Resnick, L.
1987 *Education and Learning to Think.* Washington, D.C.: National Academy Press.

Shymansky, James A., William C. Kyle, Jr., and Jennifer M. Alport
1983 "The Effects of New Science Curricula on Student Performance." *Journal of Research in Science Teaching.* 20(5): 387-404.

Stiggins, Richard J.
1987 "Design and Development of Performance Assessment." *Educationcl Measurement: Issues and Practices,* 33-42, Fall).

Strang, J.
1990 *Measurement in School Science.* London: APU.

Acknowledgments

The publication of this Science Assessment Framework has been made possible with help from the following individuals, in addition to the project committees and staff listed in the appendices: Gene Owen and Gary Phillips of the National Center for Education Statistics; staff of the Publications Branch, Office of Educational Research and Improvement; and Margaret Scott, who provided the cover illustration.

Special thanks go to Carl Moser and the members of the National Assessment Governing Board Subject Area Committee #2 for their policy guidance during this Science Consensus Project.

Mary Crovo
Assistant Director for Test Development
National Assessment Governing Board

89