U.S. Department of Education
Institute of Education Sciences
NCES 2006–026

Comparing Science Content in the National Assessment of Educational Progress (NAEP) 2000 and Trends in International Mathematics and Science Study (TIMSS) 2003 Assessments

Technical Report

March 2006

Teresa Smith Neidorf
NAEP Education Statistics Services Institute
American Institutes for Research

Marilyn Binkley
National Center for Education Statistics

Maria Stephens
American Institutes for Research

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Mark Schneider
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006

March 2006

The NCES World Wide Web Home Page is: http://nces.ed.gov
The NCES World Wide Electronic Catalog is: http://nces.ed.gov/pubsearch/

Suggested Citation
Neidorf, T.S., Binkley, M., and Stephens, M. (forthcoming). *Comparing Science Content in the National Assessment of Educational Progress (NAEP) 2000 and Trends in International Mathematics and Science Study (TIMSS) 2003 Assessments* (NCES 2006–026). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from http://nces.ed.gov/pubsearch.

For ordering information on this report, write to
U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794-1398

or call toll free 1-877-4ED-PUBS or order online at http://www.edpubs.org

Content Contacts:
Marilyn Binkley
(202) 502-7484
Marilyn.Binkley@ed.gov

Elois Scott
(202) 502-7489
Elois.Scott@ed.gov

Executive Summary

The National Center for Education Statistics (NCES) collects information on student performance in key subject areas through the National Assessment of Educational Progress (NAEP), as well as through participation and collaboration in international studies. Information from these studies is used to inform policymakers, educators, researchers, and the public about the knowledge and skills of U.S. students and how these compare with students in other countries.

This technical report describes a study that was undertaken to compare the content of two recent fourth- and eighth-grade assessments in science: the NAEP 2000 assessment and the Trends in International Mathematics and Science Study (TIMSS) 2003 assessment.[1] Its aim is to provide information that will be useful for interpreting and comparing the results from the two assessments, based on an in-depth look at the content of the respective frameworks[2] and assessment items.

The report draws upon information provided by the developers of the assessments, as well as data obtained from an expert panel convened to compare the frameworks and items from the two assessments on various dimensions.[3] The frameworks were compared with respect to

- how each assessment organizes and defines the science content, cognitive skills, and science processes to be assessed at each grade level;
- the main content areas included and the set of topics covered in each;
- how each assessment addresses scientific investigation or inquiry; and
- the types and distribution of item formats used.

Item comparisons were based on

- cross-classification of NAEP and TIMSS items to each other's assessment framework in terms of the science content covered and grade-level expectations;
- classification of all items by *cognitive domain* (*factual knowledge*, *conceptual understanding*, or *reasoning and analysis*) as defined in the TIMSS 2003 framework;
- classification of all items with respect to whether they assess scientific inquiry skills and understandings;[4] and
- comparisons of the types of item formats and tasks included.

---

[1] At the time the study was conducted, NAEP 2000 was the most recent science assessment that could be compared with TIMSS 2003. Subsequently, NAEP has conducted a science assessment in 2005, with results to be released in 2006.

[2] Assessment frameworks define what will be assessed, including the content to be covered, the types of test questions, and recommendations for how the test is administered.

[3] The panel—composed of experts in science, science education, and science assessment, with familiarity and experience with the NAEP and TIMSS assessments—is listed in appendix C.

[4] Classifications of *scientific inquiry* were based on the definitions in the TIMSS 2003 framework.

Example items are referenced throughout the report to illustrate findings from some of these comparisons.[5]

The results of this study reveal that while there is considerable overlap in the content included in the NAEP and TIMSS science assessments at fourth and eighth grade, there are some key differences between the two assessments both at the framework level and in the assessment items developed based on these frameworks. Differences exist in how each assessment defines the main science content areas, the specific topics and objectives to be assessed at each grade level, the distribution of items with respect to cognitive categories, the relative emphasis on scientific investigation or inquiry, and the type and distribution of different item formats. Some of these differences reflect the different purposes of the two assessments, with NAEP developed in the context of the U.S. system and TIMSS based on a consensus across many countries with different science curricula.

There are some similarities and differences between NAEP and TIMSS at the framework level, with NAEP defining science content in three broad *fields of science* (physical science, life science, and Earth science), and TIMSS including five *content domains* (physics, chemistry, life science, Earth science, and environmental science). Both NAEP and TIMSS include content areas related to life science and Earth science, which appear to be defined similarly based on the topic areas included in the framework. The two frameworks differ, however, in how the physical sciences are organized, with TIMSS having separate *content domains* for chemistry and physics. TIMSS also includes a separate *content domain* for environmental science, which includes topics related to environmental and resource issues that go across the *fields of science* in NAEP. The inclusion in TIMSS of separate content areas in chemistry, physics, and environmental science results in broader topic coverage in some areas. The differences at the framework level are translated into different emphases in the pool of items included in each assessment, even in the content areas of life science and Earth science where there is considerable overlap of the topic areas in the frameworks.

TIMSS reports on each of the five *content domains* separately at the eighth grade, while at fourth grade only three content areas are reported—physical science, life science, and Earth science. At the fourth grade, TIMSS items covering topics in physics and chemistry are combined in the physical science reporting category, and a few items covering environmental science topics are reported on either the life science or Earth science subscales. Similar topics are also included in the NAEP framework in the areas of life science and Earth science. Thus, there appears to be greater correspondence between NAEP and TIMSS at the fourth-grade level, at least in terms of broad content areas.

Comparing the entire set of items in the assessments overall, there are somewhat different emphases placed on each of the broad science content areas defined by either framework. While NAEP has a more even distribution of items across the three main content areas of physical science, life science, and Earth science, it does include a slightly greater proportion of items devoted to Earth science than TIMSS at either grade level. TIMSS, on the other hand, places more emphasis on life science at the fourth grade and on physical science at the eighth grade. As only TIMSS includes environmental science as a separate content area, there are more items in TIMSS than in NAEP that were classified in this area of the TIMSS framework, particularly at the eighth grade.

---

[5] Additional released item sets from each assessment are available on the NAEP and TIMSS websites: http://nces.ed.gov/nationsreportcard and http://isc.bc.edu/timss2003.

Even greater differences emerge with more detailed comparisons of the topics and specific objectives covered and the grade-level correspondence between items in each assessment and the intentions of the other assessment framework. In general, NAEP items were more frequently classified to the TIMSS framework with a higher degree of content match (e.g., to particular topics or grade-specific objectives) than TIMSS items to the NAEP framework. More than 90 percent of NAEP items (97 percent at grade 4 and 93 percent at grade 8) compared to 83 percent of TIMSS items at grade 4 and 82 percent at grade 8 were classified to one of the broad topics in the other assessment framework. The difference was even greater when considering items mapped to specific objectives, with more than 80 percent of NAEP items at either grade (83 percent at grade 4 and 82 percent at grade 8) compared to 67 percent of fourth-grade and 60 percent of eighth-grade items in TIMSS. While the majority of items in both NAEP and TIMSS were classified to topics and objectives in the other assessment framework, a substantial number were not, indicating that both NAEP and TIMSS contain items that might not be included in the other assessment at any grade level. Also, of the items that were mapped to topics and objectives in the other assessment framework, many were at a different grade level or in a different content area.

The grade-level correspondence between the NAEP and TIMSS assessments overall is quite different when comparing the fourth- and eighth-grade items. At the fourth grade, 88 percent of TIMSS items were classified at the corresponding grade level of the NAEP framework. In contrast, 61 percent of NAEP items were classified at the fourth-grade level and 35 percent were classified at the eighth-grade level according to the TIMSS framework. The correspondence between the eighth-grade assessments is greater, with more than 80 percent of items in both NAEP and TIMSS classified at the eighth-grade level of the other assessment framework. Still, there were 10 percent of eighth-grade NAEP items classified at the fourth-grade level in TIMSS and 18 percent of grade 8 TIMSS items classified at either the fourth-grade or twelfth-grade level of the NAEP framework.

There are also notable differences across the content areas in the level of grade match between the two assessments. NAEP fourth-grade items classified at the eighth-grade level in TIMSS go across all of the main content areas. TIMSS items classified as a higher grade according to the NAEP framework—either fourth-grade items classified at the eighth grade or eighth-grade items classified at the twelfth grade—come primarily from Earth science and environmental science, with some items from life science. There are no physical science items in TIMSS that were classified at a higher grade level in the NAEP framework.

Within each content area, detailed comparisons of content coverage and grade correspondence reveal some key differences. These include the following:

- Physical science: TIMSS includes a number of items at both fourth and eighth grade that do not correspond well to explicit topics in the NAEP framework, although the NAEP assessment does include some items matching these TIMSS topic areas (e.g., chemistry items related to oxidation, burning, and chemical change, and physics items related to heat conductivity and electrical circuits). NAEP and TIMSS both include items at the eighth grade that are somewhat beyond the eighth-grade specifications of the other framework (e.g., TIMSS chemistry items requiring knowledge of subatomic particles and NAEP physics items involving acceleration and momentum). TIMSS also addresses chemistry topics not included in NAEP at the eighth grade (e.g., *acids and bases*).

- Life science: TIMSS includes *human health*, a topic not included in the NAEP framework, although there are some NAEP items related to this TIMSS topic. NAEP places a greater emphasis on *ecology* and *ecosystems* in life science than TIMSS at both grade levels, although TIMSS includes additional items measuring topics related to ecosystems in its environmental science category. Also, the assessments differ with respect to what is considered fourth- and eighth-grade content in this area. Both assessments include *cells and their functions* at the eighth-grade level, but the topics included in this area are defined differently in TIMSS and NAEP.

- Earth science: There is a lower level of grade correspondence between NAEP and TIMSS in Earth science than in the physical and life sciences. About one-quarter of TIMSS fourth-grade items and almost one-half of TIMSS eighth-grade items were classified at a different grade level. Among the TIMSS eighth-grade items, 28 percent were classified at fourth grade and 21 percent were classified at twelfth grade in the NAEP framework. Based on the TIMSS framework, NAEP and TIMSS generally emphasize different aspects of Earth science at the two grade levels, with NAEP having a greater focus on *Earth in the solar system and the universe* and TIMSS on *Earth's structure and physical features* at the fourth grade. This situation is reversed at the eighth grade.

- Environmental science: While only TIMSS includes this area as a separate *content domain*, more than 70 percent of TIMSS environmental science items were classified to topics in the NAEP framework across the *fields of science,* but primarily in Earth science. Still, a number of TIMSS items in this *content domain* (29 percent at grade 4 and 30 percent at grade 8) were classified at a different grade level in the NAEP framework. Also, several items (29 percent at grade 4 and 22 percent at grade 8) were found not to match any of the NAEP topics; these items cover a range of TIMSS framework objectives related to human use of natural resources as well as global and local environmental issues due to human and natural causes.

The NAEP and TIMSS assessments have different balances of cognitive skills based on the expert panel classification of items to the TIMSS *cognitive domains* (*factual knowledge*, *conceptual understanding*, and *reasoning and analysis*). TIMSS has a larger proportion of items measuring *factual knowledge* (44 percent at fourth grade and 31 percent at eighth grade), while NAEP has more emphasis on *conceptual understanding* (53 percent of fourth-grade items and 61 percent of eighth-grade items). Both NAEP and TIMSS classify at least 20 percent of eighth-grade items as *reasoning and analysis*, while at the fourth grade there are relatively more items in this *cognitive domain* in NAEP than in TIMSS (22 percent compared to 10 percent).

While both assessment frameworks include the measurement of *scientific investigation* or *scientific inquiry*, this area receives more emphasis in NAEP than in TIMSS. Both assessments include individual items as well as longer tasks designed to measure investigation or inquiry skills. TIMSS includes "problem solving and inquiry" tasks as part of its pencil-and-paper assessment, but only NAEP includes "hands-on" tasks involving the use of materials to conduct scientific investigations. Based on the expert panel classifications, about one-quarter of NAEP items compared to about 10 percent of TIMSS items at both the fourth and eighth grades measure *scientific inquiry* skills and understandings. This difference was due at least in part to the "hands-on" tasks in NAEP, nearly all of which were classified as measuring *scientific inquiry*.

NAEP and TIMSS also differ with respect to the types and distribution of item formats. While the NAEP assessment is roughly balanced with respect to multiple-choice and constructed-response items, TIMSS has a larger proportion of multiple-choice items—about two-thirds at both grades. TIMSS includes a higher proportion of items classified as extended-response by its assessment developers; however, the definition of extended-response items differs between the two assessments and some of these items may correspond more closely to items classified as short-answer items in the NAEP assessment. There are differences between NAEP and TIMSS in the *cognitive domains* covered by multiple-choice items, with more than half of TIMSS multiple-choice items classified as *factual knowledge* compared to 60 percent of the NAEP items classified as *conceptual understanding*. The *cognitive domains* of the constructed-response items are similar for both assessments, with more than 60 percent *conceptual understanding*, at least 25 percent *reasoning and analysis,* and 10 percent or less *factual knowledge*.

The similarities and differences between NAEP and TIMSS are presented in this report for consideration when making meaningful comparisons and interpretations of the achievement results from the two assessments. While NAEP and TIMSS show similarities, particularly at the broad content area and topic level, differences are revealed by examining in more detail what the items in each assessment measure. In addition to providing information to help explain differences in achievement results from NAEP and TIMSS, this report also illustrates the complementary nature of the NAEP and TIMSS assessments. As a result of different emphases in NAEP and TIMSS, each assessment contributes more information in some content areas as well as some unique components.

Acknowledgments

Many people's contributions made this report possible, and the authors wish to thank all those who have assisted with various aspects of the report, including data analysis, reviews, and design.

Members of the expert panel listed in appendix C provided all of the item classification data used as the foundation for the results presented in this report. The authors would like to thank the panel members for their expertise and contributions to the study.

Thanks to Patrick Gonzales, Eugene Owen, and Mariann Lemke of the National Center for Education Statistics (NCES) for their input on the design and their role in reviewing the report. The authors are grateful to NCES technical reviewers, Marilyn Seastrom and Bruce Taylor, as well as to NCES program directors (Andrew Malizio and Elois Scott) and associate commissioners (Peggy Carr and Val Plisko) who provided direction and support for this publication. The authors wish to thank Steve Gorman of NCES, Barbara Lynch of Synergy Enterprises, Christine O'Sullivan of K-12 Consulting, Duc-Le To of the Institute of Education Sciences, and two anonymous reviewers for their very helpful comments.

Several people, formerly or currently from the Education Statistics Services Institute (ESSI), provided support in the form of research assistance and/or review: Ben Dalton, Kristy David, Rachel Dinkes, Aaron Douglas, Janine Emerson, J. Lane Glenn, Robin Gurley, Jamie Johnston, Dana Kelly, Geeta Kotak, David Miller, Melanie Ouellette, Lisette Partelow, Mike Planty, Anindita Sen, and Margaret Woodworth. In particular, we recognize the efforts of David Miller and Jamie Johnston in helping to shepherd the report through the final stages of the review and publication process. Thanks to Heather Block, also of ESSI, for her work on graphic design for this report, and to Brian Henigin of Westat for the final publication formatting of the report.

For permission to use secure items and publish released items, the authors would like to thank the Trends in International Mathematics and Science Study (TIMSS) and the International Association for the Evaluation of Educational Achievement (IEA). Thanks also to the staff of the Educational Testing Service for participation in the expert panel meeting and for supplying item information and review materials from the National Assessment of Educational Progress (NAEP).

Contents Page

List of Tables

Table **Page**

List of Figures

Figure **Page**

List of Exhibits

Exhibit **Page**

1. Introduction

The National Center for Education Statistics (NCES) provides information to assist policymakers, researchers, educators, and the public in obtaining a comprehensive picture of what U.S. students know and can do in key subject areas. Information on such learning outcomes can be valuable in helping to inform policy, monitor and benchmark progress, and identify factors for improving student achievement, among other benefits.

National-level data on student achievement in the United States comes primarily from two sources: the National Assessment of Educational Progress (NAEP)—otherwise known as the Nation's Report Card—and the United States' participation and collaboration in international assessments, such as the Trends in International Mathematics and Science Study (TIMSS), the Program for International Student Assessment (PISA), and the Progress in International Reading Literacy Study (PIRLS).[1] NAEP has been assessing fourth-, eighth-, and twelfth-grade students' performance in reading, mathematics, science, and other subjects at regular intervals for over 30 years. The international assessments allow the United States to benchmark its performance to other countries—in fourth- and eighth-grade mathematics and science in TIMSS, in fourth-grade reading literacy in PIRLS, and in 15 year-olds' reading, mathematical and scientific literacy in PISA. This comparative element is increasingly important as the marketplace in which U.S. students will have to compete becomes more global.

While these assessments may appear to be similar in terms of the age or grade of the students assessed and the subject matter of the assessment, each was designed to serve a different purpose and is based on a unique framework. In order to provide information to help interpret and compare results, NCES has periodically undertaken studies to examine the similarities and differences among the various assessments. The purpose of these studies is to provide an in-depth and comparative look at the frameworks and test items used in the different assessments to shed light on what each assessment can contribute to the knowledge base on student performance and help explain differences in results between the assessments. An understanding of the similarities and differences between the assessments is important to avoid misunderstanding or misinterpreting these results.

In 2003, NCES conducted two content comparison studies with NAEP—one in mathematics and one in science—following the 2003 administrations of TIMSS and PISA. This report focuses on the comparison of the science assessments—specifically NAEP 2000 and TIMSS 2003—while a companion paper (Neidorf, Binkley, Gattis, and Nohara 2006) compares the NAEP, TIMSS and PISA 2003 mathematics assessments. PISA is not included in the science comparison study because scientific literacy was not a major domain of this assessment in 2003, including only about 30 scientific literacy items. It is included in the mathematics comparison study, however, because mathematical literacy was a major domain in that year, with many more items on which to base analyses. Scientific literacy will be the major domain in 2006, when there will be a greater number of scientific literacy items and its assessment framework will be fully developed and updated. For more information on PISA see http://nces.ed.gov/surveys/PISA.

[1] TIMSS and PIRLS are conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). The Organization for Economic Cooperation and Development (OECD) is responsible for PISA.

The 2003 science and mathematics comparison studies build on several earlier studies, which were undertaken to explore the similarities and differences between NAEP and various international assessments. Such studies comparing frameworks and items are conducted periodically as NAEP and international assessments evolve, improving their frameworks and test items to reflect current research, policy, and practice.

One previous published study of mathematics and science assessments is the NAEP 2000, TIMSS 1999, and PISA 2000 mathematics and science assessments (Nohara 2001), which compares the underlying frameworks and test items from each assessment in terms of content, item format, and thinking skills required.

There also have been several studies comparing reading assessments. For example, the earliest of these compared the NAEP 1992 reading assessment and the 1991 IEA Reading Literacy Study (Binkley and Rust 1994). More recently, Binkley and Kelly (2003) examined the frameworks, passages, and items from the NAEP 2002 and the Progress in International Reading Literacy Study (PIRLS) 2001 reading assessments.

The goal of this science comparison study is to identify similarities and differences between the NAEP 2000 and TIMSS 2003 assessments based on a detailed comparison of their frameworks and items. This information may be used to help inform interpretations of students' performance in science based on the two different assessments. While there are other important aspects that might be compared, such as item difficulty, sampling, and scaling procedures, this study focuses on a comparison of the content of the assessments. This content comparison is based on the main dimensions of the assessment frameworks and focuses on a comparison of the set of assessment items as a reflection of how the frameworks are implemented. The questions driving the study are as follows:

- How do NAEP and TIMSS define the domain of science to be assessed and its content areas, in terms of both the topics that are included and the distribution of items across topics?

- How do NAEP and TIMSS define the content topics and assessment objectives appropriate for the fourth-grade and eighth-grade assessments? How do the items in each assessment compare to the grade-level expectations specified by the other framework?[2]

- How do NAEP and TIMSS compare with respect to the distribution of items across broad cognitive domains related to *factual knowledge*, *conceptual understanding*, and *reasoning and analysis*?

- How do NAEP and TIMSS define and measure scientific inquiry? To what extent does each assessment measure scientific inquiry skills?

- How do NAEP and TIMSS compare with respect to the types and distribution of item formats?

[2] The 2003 mathematics and science comparison studies are the first to compare the assessments in terms of grade level—the extent to which items from one assessment map to the same grade level of the framework of the other assessment.

To answer these questions, NCES convened an expert panel to examine the science frameworks and items for each assessment. The panel cross-classified NAEP and TIMSS fourth- and eighth-grade items to the other assessment's framework with respect to science content and grade level. They also classified all items from both assessments with respect to a common set of cognitive domain categories and a common definition of scientific inquiry. In addition to the classification data from the panel, the study drew upon information provided by the NAEP and TIMSS assessment developers that describes how each item is classified according to the main dimensions of its own framework, as well as other relevant characteristics such as item format and scoring rubrics.

Section 2 of this report opens with an overview of the NAEP and TIMSS assessments and a comparison of their respective science assessment frameworks. Section 3 reviews the methods used for this comparison study. The results of the study are then presented in two major sections. The first results section (section 4) compares the assessments overall with respect to content coverage, grade level, cognitive domain, scientific inquiry, and item format. The overall comparisons are followed (section 5) by comparisons of the two assessments in each of the main content areas of physical science, life science, Earth science, and environmental science. Section 5 provides more detailed comparisons of the extent to which items in one assessment map to the science framework of the other assessment and compares the content distribution of the items for each of the NAEP and TIMSS science subscales. The report concludes with a summary of key findings (section 6).

2. Overview of the Assessments and their Frameworks

NAEP

The National Assessment for Educational Progress (NAEP) is the United States' source for nationally representative and continuing information on what American students know and can do and is commonly known as the Nation's Report Card. NAEP policies and frameworks are established by an independent National Assessment Governing Board (NAGB), and the Department of Education's National Center for Education Statistics (NCES) administers the assessment. For over 30 years, NAEP has periodically collected and reported data on achievement in reading, mathematics, science, and other subjects for students in fourth, eighth, and twelfth grades. The comparisons in this report are based on the main NAEP science assessments conducted in 2000.[3]

The frameworks established by NAGB for all the NAEP subject areas, including science, are based on the collaborative input of a wide range of experts and involvement by participants from government, education, business, and public sectors; are informed by common curricular practices in the nation's schools; and ultimately are intended to reflect the best thinking about the knowledge, skills, and competencies needed for students to have a deep level of understanding of different subject areas at different grades.

TIMSS

The Trends in International Mathematics and Science Study (TIMSS) is the United States' source for international comparative information on mathematics and science education in the primary and middle grades. TIMSS is one of the current studies conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), which has been conducting international comparative studies since the early 1960s. TIMSS is directed by the International Study Center at Boston College and collects achievement and background data to provide information on trends in mathematics and science achievement over time as well as on the curricular, instructional, and attitudinal factors that may be related to performance. TIMSS collects data on a four-year cycle. The first administration was in 1995 (at fourth, eighth, and twelfth grades)[4], the second in 1999 (at eighth grade only), and the most recent in 2003, with 25 countries participating at fourth grade and 45 countries participating at eighth grade.

Like NAEP, the TIMSS assessment is based on collaboratively developed frameworks. In contrast to NAEP, however, the TIMSS framework development and consensus process involves science experts, education professionals, and measurement specialists not only from the United States, but from many countries.

---

[3] At the time of this study, NAEP 2000 was the most recent science assessment. A subsequent science assessment was conducted in 2005.

[4] Defined as the upper of the two grades containing the majority of 9-year-olds or 13-year-olds and the final year in secondary school. These are the fourth, eighth and twelfth grades in the U.S. and most other countries. TIMSS 1995 was also administered in third and seventh grades.

Organization of the NAEP 2000 and TIMSS 2003 Science Frameworks

Exhibits 1-A and 1-B compare schematically the organizing dimensions in the NAEP 2000 and TIMSS 2003 science frameworks. These organizing dimensions provide the basic framework for the development of the pool of items in each assessment, and the framework documents include target percentages for the distribution of assessment time across the main categories in each dimension to ensure a balanced assessment (discussed in the following sections).[5] As seen in these figures, there are some similarities as well as differences between the dimensions in the frameworks for NAEP and TIMSS.

Both the NAEP 2000 and TIMSS 2003 science frameworks are represented in exhibits 1-A and 1-B based on two main organizing dimensions—a content dimension and a cognitive dimension, as well as overarching dimensions (along the bottom) that define areas of science content knowledge—skills, and abilities that go across content and cognitive categories. In the content dimension, NAEP includes three broad *fields of science* (physical science, life science, and Earth science), while TIMSS includes five separate *content domains* (physics, chemistry, life science, Earth science, and environmental science). At the eighth grade, TIMSS reports on all five of the *content domains*, while at the fourth grade three content reporting categories are used that correspond to the NAEP *fields of science* (physics and chemistry items are combined into physical science and environmental science items are reported in either life science or Earth science).

In their main cognitive dimensions (*knowing and doing* in NAEP and *cognitive domains* in TIMSS), both assessment frameworks include three broad categories that appear to be at least partially overlapping. The NAEP *knowing and doing* categories include *conceptual understanding*, *scientific investigation*, and *practical reasoning*; while the TIMSS *cognitive domains* include *factual knowledge*, *conceptual understanding*, and *reasoning and analysis*. There are two overarching dimensions in NAEP (*nature of science* and *themes*) and one overarching dimension in TIMSS (*scientific inquiry*). These overarching dimensions include both content-related and cognitive- or skills-based components, and there is overlap between the assessment objectives defined by these overarching dimensions and those defined in the other main dimensions in the NAEP and TIMSS science frameworks. In particular, both assessments include cognitive categories related to conceptual understanding and reasoning. Both assessments also address *scientific investigation* (a *knowing and doing* category in NAEP) or *scientific inquiry* (an overarching dimension in TIMSS). There also is overlap between understandings related to NAEP's *nature of science* dimension and the *scientific inquiry* dimension in TIMSS. All items developed for NAEP and TIMSS are classified with respect to which categories in the content and cognitive dimensions they assess. Some items across these categories also are developed to assess science content knowledge, abilities and skills defined by the overarching dimensions.

The following sections describe and compare in more detail the science assessment frameworks for NAEP and TIMSS. Additional assessment framework summary documents that were used for the comparison study are found in appendixes A and B.

[5] The frameworks only provide target percentages of assessment time as guidelines for test development.

Exhibit 1-A. NAEP science framework dimensions: 2000

Fields of science	Knowing and doing
Physical science	Conceptual understanding
Life science	Scientific investigation
Earth science	Practical reasoning
Nature of science (science and technology)	
Themes Models, systems, patterns of change	

NOTE: The NAEP framework is based on two main organizing dimensions—*fields of science* and *knowing and doing*—as well as two overarching dimensions (*nature of science* and *themes*) that go across the *fields of science* and *knowing and doing* categories.
SOURCE: U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.

Exhibit 1-B. TIMSS science framework dimensions: 2003

Content domains	Cognitive domains
Physics	Factual knowledge
Chemistry	Conceptual understanding
Life science	Reasoning and analysis
Earth science	
Environmental science	
Scientific inquiry	

NOTE: The TIMSS framework is based on two main organizing dimensions—*content domains* and *cognitive domains*—as well as an overarching dimension (*scientific inquiry*) that goes across the content and cognitive domains.
SOURCE: International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003.

2.1. NAEP 2000 Science Framework

The framework for the NAEP 2000 science assessment includes a content dimension (*fields of science*) and a cognitive dimension (*knowing and doing*), as well as two overarching dimensions (*nature of science* and *themes*) (see exhibit 1-A).[6] The framework stipulates that every item developed for the assessment is given a primary classification in the two major content and cognitive dimensions according to certain distribution targets. While secondary classifications may be assigned for some items, NAEP does not use multidimensional scaling, and these secondary classifications are not used in the analysis of results. In addition, some items are also classified as addressing understandings and skills specified in the two overarching dimensions.

The first major dimension is the three content areas, *fields of science*, which are the same for fourth, eighth, and twelfth grade. They are life science, physical science, and Earth science. Within these content areas, major topics and subtopics are further identified which, with few exceptions, are the same across grades (4, 8, and 12).[7] The framework specifies that the distribution of assessment time should be approximately equal across the fields of science at the fourth and twelfth grades, and should place a slight emphasis on life science at the eighth grade (table 1).

Table 1. Target percentage of NAEP assessment time distributed across NAEP framework dimensions, by grade: 2000

NAEP framework dimensions	Grade 4	Grade 8	Grade 12
Fields of science			
Life science	33	40	33
Physical science	33	30	33
Earth science	33	30	33
Knowing and doing			
Conceptual understanding	45	45	45
Scientific investigation	45	30	30
Practical reasoning	10	25	25

NOTE: Percentages reflect the targets specified in the NAEP 2000 science framework. At the 4th and 12th grade, distributions across the three *fields of science* are approximately equal. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.

The other major organizing dimension in the NAEP science framework is the cognitive dimension, *knowing and doing science*, which defines the processes and cognitive skills the items are designed to require of students. The three categories of *knowing and doing* in NAEP are *conceptual understanding*, *scientific investigation*, and *practical reasoning*. The NAEP 2000 science framework specifies that there should be heavy emphasis on *conceptual understanding* (45 percent at each

[6] The framework for the NAEP 2000 science assessment was developed in the early 1990s and was used as the basis for the assessments in 1996, 2000, and 2005 (see NAGB (2000) for additional information). A new science framework will be developed for the 2009 assessment.
[7] A separate assessment specifications document specifies different grade-specific assessment objectives within each subtopic that are appropriate for each of the respective grades (NAGB 1994). See section 5 and appendix A for more information about the topics and objectives included in the NAEP content framework and specifications documents.

grade), which is the ability to understand basic concepts and tools of science. The category of *scientific investigation* focuses on students' use of appropriate tools and thinking skills in science to acquire new information, plan and conduct appropriate investigations, and communicate the results of investigation. These ways of *knowing and doing* science receive greater emphasis at the lowest grade (45 percent at fourth grade compared to 30 percent at eighth and twelfth grade). *Practical reasoning* also involves the application of scientific knowledge and skills but focuses more on analyzing and solving practical problems. The proportion of the assessment measuring *practical reasoning* also differs across grades (10 percent at fourth grade compared to 25 percent at eighth and twelfth grades).

In addition to the content and cognitive dimensions, the NAEP 2000 science framework includes two overarching categories, considered to be important aspects of science knowledge and skills to be captured in NAEP. The first is *nature of science*, which includes topics relating to the history of science and technology, habits of mind that characterize these fields, and methods of inquiry and problem solving. The framework specifies that at least 15 percent of the assessment should measure *nature of science*. The second category, *themes*, represents the big ideas or key concepts that transcend the scientific disciplines and enable students to consider problems with broader implications. Concepts related to *themes* are specified in each content area. The three *themes* are systems, models, and patterns of change.

In comparison to *nature of science*, a much larger proportion of the assessment is specified to assess *themes* (about one-third at the fourth grade and half at the eighth and twelfth grades), spread evenly across the three. NAEP includes both individual items that assess concepts related to *themes* as well as sets of items all related to a common theme ("theme blocks") that provide an in-depth measure of a particular theme and may include items across content areas.

The NAEP 2000 science framework specifies that multiple-choice, short-answer, and extended-response items are included in the assessment. Less than half of the assessment time should be multiple-choice, and about one-third of the constructed-response items should require an extended response. The framework also specifies that in addition to pencil-and-paper tests, about 30 percent of the assessment at each grade will be devoted to "performance" or "hands-on" tasks. For this part of the assessment, students are provided with kits and they must manipulate physical materials to solve a scientific problem or conduct an investigation involving those materials. All of the items contained in the "hands-on" tasks are also classified according to the appropriate categories in the content and cognitive dimensions as well as the overarching categories of *nature of science* and *themes*.

2.2. TIMSS 2003 Science Framework

The TIMSS 2003 science framework is based on two main organizing dimensions, a content dimension and a cognitive dimension, as well an overarching dimension of *scientific inquiry* (see exhibit 1-B earlier in this section).[8]

The TIMSS 2003 science framework specifies five *content domains*. They include life science, chemistry, physics, Earth science, and environmental science. All five *content domains* are included as separate reporting categories at the eighth grade, but at the fourth grade, there are three reporting categories (physical science, life science, and Earth science). For fourth grade, chemistry and physics items are combined on the physical science scale, and a small number of items assessing environmental science topics at the fourth grade are embedded within the life science and Earth science reporting scales. As shown in table 2, the TIMSS 2003 science framework specifies that assessment time should have greater emphasis on life science at the fourth grade (45 percent at fourth grade compared to 30 percent at eighth grade) and physical science at the eighth grade (25 percent for physics and 15 percent for chemistry at the eighth grade compared to 35 percent for physical science overall at the fourth grade). A smaller proportion of the assessments at both grades are specified for Earth science (20 percent at fourth grade and 15 percent at eighth grade) and environmental science (15 percent at eighth grade only). Within the *content domains*, the TIMSS framework further specifies topic areas and grade-specific objectives within those topic areas that are appropriate for each grade.[9]

Table 2. Target percentage of TIMSS assessment time distributed across TIMSS framework dimensions, by grade: 2003

TIMSS framework dimensions	Grade 4	Grade 8
Content domains		
Life science	45	30
Physical science	35	40
Chemistry	†	15
Physics	†	25
Earth science	20	15
Environmental science	†	15
Cognitive domains		
Factual knowledge	40	30
Conceptual understanding	35	35
Reasoning and analysis	25	35

† Not applicable. At grade 4, target percentages are defined for only three reporting categories—*life science*, *physical science*, and *Earth science*. *Physical science* is assessed as one content area that includes both *physics* and *chemistry* topics. Some topics related to *environmental science* are assessed as part of *life science* and *Earth science.*
NOTE: Percentages reflect the targets specified in the TIMSS 2003 framework.
SOURCE: International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Edition*, 2003.

[8] The TIMSS science framework was revised for 2003 from the original curriculum framework used as the basis for the 1995 and 1999 assessments. See Mullis et al. (2003) for additional information.
[9] See section 5 and appendix A for more information about the topics and objectives included in the TIMSS content framework.

On the cognitive dimension, TIMSS specifies three *cognitive domains* to describe the range of cognitive skills and abilities that the items require students to apply to answer the item successfully. The three *cognitive domains* are *factual knowledge*, *conceptual understanding*, and *reasoning and analysis*. The inclusion of items in the *factual knowledge* category provides a measure of the extent and accuracy of students' factual knowledge and their knowledge of information, tools, and procedures, which are the basis of developing scientific understanding and reasoning. The next category of *conceptual understanding* assesses students' grasp of the relationships that explain the behavior of the physical world and their ability to relate observations to more abstract or general concepts in science. The third category, *reasoning and analysis*, captures students' ability to engage in scientific reasoning for the purpose of analyzing and solving problems, developing explanations, drawing conclusions, making decisions, and extending their knowledge to new situations. The TIMSS 2003 science framework specifies an approximately equal distribution across the *cognitive domains* for the eighth grade assessment. For fourth grade, however, there is a somewhat greater emphasis on the *factual knowledge* category (40 percent) compared to *reasoning and analysis* (25 percent).

All TIMSS items are classified with respect to *content domain* (topic area and objective) and *cognitive domain* at the broad category level. Items also are classified with respect to whether they are intended to measure knowledge and skills necessary to engage in *scientific inquiry*. All items measuring *scientific inquiry* are set in a science content-based context and assess skills and abilities described in the *cognitive domain* categories.

The TIMSS 2003 science framework includes *scientific inquiry* as an overarching dimension that overlaps all the *content domains* and *cognitive domains* and has both content- and skills-based components. In TIMSS, it is expected that students at both grade levels will possess some general knowledge of the nature of science and scientific inquiry and demonstrate skills and understandings involved in the major phases of the scientific inquiry process, such as formulating questions and hypotheses; designing investigations; representing, analyzing, and interpreting data; and drawing conclusions based on evidence. The TIMSS 2003 assessment does not include "performance" or "hands-on" tasks.[10] Rather, a set of "problem solving and inquiry tasks" are included as part of the pencil-and-paper assessments. These tasks present students with an inquiry-based situation and ask a series of related questions assessing some of the skills specified in the *scientific inquiry* dimension. While these tasks are not intended to be full scientific investigations, they are designed to require a basic understanding of the processes of investigation and skills essential to the scientific inquiry process. Individual items are also included in the assessment to measure individual inquiry skills. The framework specifies that up to 15 percent of the assessment at each grade may assess *scientific inquiry*. Appendix B provides a more detailed definition of the TIMSS *cognitive domain* categories and *scientific inquiry*.

The TIMSS 2003 science framework specifies that both multiple-choice and constructed-response items requiring students to generate their own answers be included in the assessment, with up to two-thirds of the assessment time coming from multiple-choice items. About two-thirds of the constructed-response items should require a short answer, while the other third should require a more extended response.

---

[10] Performance assessment tasks were included as part of the TIMSS 1995 assessment but not repeated in the 1999 or 2003 assessments.

2.3. Comparing the NAEP and TIMSS Science Frameworks and Assessments

The NAEP and TIMSS science frameworks have some obvious similarities, including the general structure with content and cognitive dimensions, the inclusion of overarching dimensions, and the specification of curriculum-based content areas related to the physical, life, and Earth sciences. Both NAEP and TIMSS cover a broad range of science topics and skills and include similar numbers of items in their assessments.[11] However, there are differences as well, which may affect the items developed based on the respective frameworks and thus the content of each assessment overall. A first notable structural distinction is TIMSS's inclusion of a separate environmental science category and the disaggregation of physical science into the two disciplines of chemistry and physics in the eighth-grade assessment, which differs from the NAEP categorization into three main *fields of science* at both grades. While both NAEP and TIMSS are grade-based assessments including two corresponding grades (fourth and eighth), NAEP also includes a twelfth-grade assessment.[12] The set of topics and objectives included in the NAEP framework reflect this difference.

The method of specifying the science content to be assessed at each grade level is somewhat different in NAEP and TIMSS frameworks. Exhibit 2 shows the different content levels and terminology specified in each framework as well as general terminology used in this report to facilitate direct comparisons. NAEP has one additional "layer" of content specification than does TIMSS—major topic areas used to organize subtopics within each *field of science*. For example, in physical science there are three major topics in NAEP—*matter and its transformations*, *energy and its transformations*, and *motion*—with several subtopics within each.[13] The major topics and subtopics in the NAEP framework are common across grades (4, 8, and 12), and a separate assessment specifications document used by assessment developers describes specific objectives in each subtopic that are expected at each grade level (NAGB 1994). In contrast, the TIMSS framework specifies only two levels—topic areas and grade-specific objectives in each *content domain*. For this report, the NAEP subtopics (within major topics) and specific objectives are compared with the two most comparable TIMSS levels (topic area and objective).[14]

When making direct comparisons related to item content, this report uses a general terminology of content area, topic, and objective to refer to the comparable levels of specification used in the NAEP and TIMSS content framework. For the discussion of content or cognitive classifications based on a single framework (NAEP or TIMSS), the terminology from that framework is used.

---

[11] Both NAEP 2000 and TIMSS 2003 include more than 170 items at eighth grade and more than 140 items at fourth grade.

[12] Although TIMSS was administered at the twelfth grade in the 1995 survey, it was not administered at this grade level in the 2003 survey, which is the focus of this comparison study.

[13] The content area comparisons in section 5 show major topics and subtopics in each NAEP *field of science* and the topic areas within each TIMSS *content domain*.

[14] The topic/subtopic/objective structures are shown in the detailed framework summary documents in appendix A.

Exhibit 2. Terminology used in making comparisons across NAEP 2000 and TIMSS 2003 content frameworks

NAEP framework		General terminology		TIMSS framework
Field of science	⇐	Content area	⇒	Content domain
Major topic				
Subtopic (within major topic)	⇐	Topic	⇒	Topic area
Specific objective	⇐	Objective	⇒	Objective

Examples of related NAEP and TIMSS content areas, topics, and objectives

NAEP		TIMSS	
Field of science:	Physical science	Content domain:	Physics
Major topic:	Energy and its transformations		
Subtopic:	Energy sources and use, including distribution, conversion, costs, and depletion	Topic area:	Energy types, sources, and conversions
Specific objective:	Demonstrate awareness that the sun is the ultimate source of most energy we use	Objective:	Identify common energy sources and forms; know some practical uses of energy

SOURCE: U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003.

Although the TIMSS framework recognizes the importance of unified concepts and topics that bridge the domains of science, only NAEP includes some specific science *themes* (models, systems, and patterns of change) as an explicit part of the framework dimensions and assessment specifications.

While both NAEP and TIMSS frameworks define cognitive categories for assessing *conceptual understanding* and *reasoning* (defined as *practical reasoning* in NAEP and *reasoning and analysis* in TIMSS), only TIMSS includes an explicit category for *factual knowledge*. However, the NAEP definition of *conceptual understanding* does include knowledge of facts and information about procedures.

Both the NAEP and TIMSS frameworks recognize scientific investigation or inquiry as an important component of science teaching, learning, and assessment. *Scientific investigation* is included as one of the cognitive categories in NAEP; *scientific inquiry* is included as an overarching dimension in the TIMSS framework that has both content- and skill-based components. Based on the target percentages in the frameworks, more emphasis on *scientific investigation* or inquiry would be expected for NAEP, particularly at fourth grade. In addition, NAEP includes "hands-on" investigational tasks, while TIMSS includes pencil-and-paper "problem-solving and inquiry tasks"

but no performance task component. Examples of a NAEP "hands-on" task and a TIMSS "problem solving and inquiry" task from the eighth-grade assessments are included in appendix E.[15]

NAEP and TIMSS include both multiple-choice items, in which students choose the correct answer from a list of several choices (usually four), and constructed-response items, in which students generate their own answers. Some of the constructed-response items in both assessments require only a short and often objective response (e.g., an example of a mixture), while others require a more extended response. The frameworks specify that up to two-thirds of the TIMSS assessment time can be from multiple-choice items compared to half or less in NAEP. Both frameworks allow for about one-third of the constructed-response items to be extended response. Some of the extended-response items across the two assessments are similar, while others are not. A more detailed comparison of item format is included in section 4.5.

The assessment designs for NAEP and TIMSS result in each individual student taking only a portion of the total assessment items, but the testing time for individual students differs across the two assessments. NAEP requires 50 minutes at all three grades and TIMSS requires 72 minutes at fourth grade and 90 minutes at eighth grade. The hands-on tasks in NAEP require an additional 20 minutes at fourth grade or 30 minutes at eighth and twelfth grade.

Finally, the NAEP framework was developed within the specific context of the U.S. system and defines a set of achievement levels (basic, proficient, and advanced) that are used as the primary means of reporting what students should know and be able to do in science at each grade level from a national perspective. In contrast, the TIMSS framework reflects a consensus across diverse participating countries about what science topics are most appropriate and important to assess at fourth and eighth grades. In general, the topics included are in the curricula for the majority of TIMSS countries.

Some curricular differences that exist across countries are related to the structure and sequencing of science instruction (Beaton et al. 1996; Martin et al. 1997; Martin et al. 2000; Robitaille 1997). While integrated or general science is taught in many countries (including the United States), separate science courses (biology, chemistry, and physics) are taught by the eighth grade in a number of other countries. Earth science is not taught as a separate science subject in many countries, and topics related to this content area may be included in courses covering the physical and life sciences as well as in separate courses such as geography. There is considerably more emphasis placed on science instruction in the primary and middle school years in some countries than others. In particular, some countries have less formalized science instruction by the fourth grade than the United States. Some of the differences in science curricula across countries are reflected in differences between the NAEP and TIMSS frameworks and science assessments. In the following sections, more detailed analyses of the individual content areas provide some information on these possible differences.

---

[15] The example included for the NAEP "hands-on" task is from the 1996 assessment, as none of the "hands-on" tasks in the 2000 assessment were released and cannot be reproduced here. This example is included only to illustrate the nature of the "hands-on" tasks in NAEP, and this particular task is not reflected in any of the item comparisons made in this report which are based on the NAEP 2000 assessment. The NAEP hands-on tasks included in the 2000 assessment are similar in style and approach but involve different science content.

This section provided an overview of the NAEP and TIMSS assessments and a comparison of their respective science assessment frameworks. The next section reviews the methods used for this comparison study.

3. Process and Methods

To conduct comparisons of the NAEP and TIMSS assessments, NCES convened a panel of 11 experts in science, science education, and science assessment. All panel members had familiarity and experience with at least one of the assessments and their frameworks.[16] The panel met over a 2-day period to review the frameworks and classify the items from each assessment. The following two sections describe the organization of the expert panel meeting and the methods used for making the NAEP/TIMSS comparisons reported in this report. Additional methodological notes are included in appendix D.

3.1. Organization of the Expert Panel Meeting

The expert panel meeting opened with a plenary session during which the study organizers presented the goals of the study, provided an overview of the NAEP and TIMSS frameworks, and described the procedures for reviewing items. The expert panel members also had an opportunity during the opening plenary session to review, classify, and discuss several practice items in order to establish a common understanding of the classification procedures.

During the 2-day meeting, all of the NAEP and TIMSS fourth- and eighth-grade science items were reviewed, reflecting a total of about 630 items across the two assessments and grades. The items were divided into three groups according to content, with each group containing all items from both NAEP and TIMSS in the content areas of[17]

- life science;
- physical science (chemistry and physics); and
- Earth science and environmental science.

The panel also was divided into three groups, with each group responsible for reviewing and classifying all of the items in one of the content groups. Panelists and staff were assigned to subgroups according to their content area expertise and to make sure that each group contained participants familiar with each of the assessments.

The meeting concluded with a plenary session during which panelists shared their thoughts on the frameworks, items, and the study overall. While this report draws from these comments, where applicable, it reports primarily on the results from the item review and classification sessions, which were the focus of the meeting.

3.2. Methods Used for NAEP/TIMSS Comparisons

In each content area group, the panel conducted a framework-level review to familiarize the panelists with the relevant portions of the content frameworks and to uncover some of the main

[16] A list of panel members and associated staff is presented in appendix C.
[17] The division of items was based on the assessment developers' classifications by content area subscale.

similarities and differences in how the major content areas covered by each group are interpreted in the two frameworks documents. The panels then classified the items, first classifying the TIMSS items to the NAEP framework and then classifying the NAEP items to the TIMSS framework. All items were classified on the following dimensions: [18]

- **Content**: Each item was classified with respect to the content framework of the other assessment (i.e., TIMSS items to the NAEP framework and NAEP items to the TIMSS framework) by identifying the content area, topic, and objective with the best match to the item content. Some items were classified as matching the other assessment framework at only the topic or content area level. Items that could not be classified at any level were also identified.

- **Grade level**: Each item was classified with respect to the grade level corresponding to the best content match in the other framework. For TIMSS items classified to the NAEP framework, grade classification was made to grade 4, 8, or 12. However, for NAEP items classified to the TIMSS framework, grade classifications were limited to grades 4 and 8 since TIMSS does not include grade 12.

- **Cognitive domain**: All items were classified in one of the *cognitive domain* categories of *factual knowledge*, *conceptual understanding,* or *reasoning and analysis* according to the definitions in the TIMSS 2003 framework.

- **Scientific inquiry (yes/no)**: All items were classified as to whether or not they measured scientific inquiry skills as defined in the TIMSS 2003 framework.

In conducting their evaluations, panelists were given the following guidelines:

- Items should be classified to the most detailed content level possible—ideally, to the objective level. (Although panelists were allowed to make some logical inferences about what a content area, topic, or objective might include, they were instructed not to classify items further than they considered appropriate.)

- Each group should consider all content areas of the framework. The content area in one assessment may overlap with another content area in the other assessment (e.g., the best topic match for an Earth science item may be in the life science content area of the other framework).

- In cases where items appear to address multiple content areas, topics, or objectives, a primary classification for the item should be identified whenever possible. (In cases where this was not appropriate, panelists were instructed to indicate multiple or secondary classifications which were recorded.)[19]

---

[18] Additional information about the content categories and definitions of *cognitive domains* and *scientific inquiry* is provided in appendixes A and B. Considerations in selecting the methods of classification are discussed in appendix D.

[19] The results in this report are based on primary classifications in nearly all cases.

- Instances where a number of items that cannot be placed in a framework are of a similar type should be documented. These instances may indicate a potential gap in the framework to which the items are being classified.

- Grade-level classifications should be based on descriptions found in the frameworks rather than on common understandings of grade-level content (i.e., items should be placed at the grade level where they best match the descriptions in the content framework). (As with other content classifications, panelists were allowed to make some logical inferences about what a topic or objective might include at a given grade level.)[20]

- Classifications to *cognitive domain* categories and *scientific inquiry* should be based strictly on the definitions in the TIMSS 2003 framework.

Within each group, panelists classified all items individually and then discussed the classifications as a group to arrive at a group classification. In general, consensus was reached, but for some items the final classifications reflect the classifications of the majority of panelists. For *cognitive domain*, final classifications for some items reflected the fact that the panel was split between two categories or that the consensus of the group was that the item was on the borderline between two categories. To monitor consistency in the classifications of *cognitive domain* and *scientific inquiry* across the three groups, a set of common items was classified by the members of all three groups. The degree to which the three groups classified these items in the same categories on these two dimensions serves as a measure of the reliability of these classifications. The items in the reliability set were not chosen at random, but rather, were a representative set of 60 items (30 from NAEP and 30 from TIMSS) selected to cover the main categories addressed in the study (content area and grade level). Reliability items were classified at regular intervals throughout the classification process. The reliability procedure and results are described in more detail in the methodological notes (appendix D).

Expert panelists typically spent more time reviewing and classifying the items in the reliability set that were in their primary content area. Thus the classifications by the primary content area expert panel groups are the most valid and used for all of the results in the report. Results from the secondary classifications of the reliability set were used to monitor the consistency of classification and were not a complete replication of the process used by the primary group, which was most familiar with items in the respective content area.

Panelists' comments on the items were also recorded during the item review process, including observations about specific item characteristics and rationales for the classifications. In addition, general comments made by the panel about the assessments and frameworks in plenary or during the separate group discussions were recorded and used to inform the discussions in this report.

---

[20] Since the TIMSS framework contains grade-specific objectives, the grade-level classification is concurrent with an objective classification. The NAEP framework provides a single set of major topics and subtopics that apply across grades; however, the assessment specifications document provides grade-specific objectives that were used to determine the specific grade match for the TIMSS items to NAEP framework classifications. For items not classified to a grade-specific objective, the grade classification reflects the judgment of the panel of the grade at which the item is most consistent with the overall framework.

This section reviewed the methods used for this comparison study. The next section compares the assessments overall with respect to content coverage, grade level, cognitive domain, scientific inquiry, and item format.

4. Overall Comparisons

The classifications made by the expert panel as well as the information provided by each assessment provide rich data that can be organized and analyzed in numerous ways. This section compares the assessments overall with respect to content coverage, grade level, cognitive domain, scientific inquiry, and item format.

4.1. Content Coverage

Tables 3 and 4 compare the distribution of items from each assessment across the main content areas in the NAEP and TIMSS frameworks. The tables compare NAEP and TIMSS item classifications, according to their own respective frameworks, with item classifications according to the framework of the other assessment.[21] At the fourth grade, according to the NAEP framework, NAEP items are fairly evenly distributed across the three content areas of physical science, Earth science, and life science, with a slight emphasis (37 percent) on life science (table 3). When the TIMSS grade 4 items were classified according to the NAEP framework, 45 percent were classified as life science and 24 percent as Earth science. Using the TIMSS framework as the reference point, a similar picture emerges (table 4). Both NAEP and TIMSS have about one-third of their fourth-grade items classified as physical science on the TIMSS framework, while TIMSS has more items classified as life science (45 percent compared to 37 percent) and NAEP has relatively more items classified as Earth science (32 percent compared to 24 percent). NAEP and TIMSS have items distributed similarly across the TIMSS chemistry and physics *content domains*, with both assessments having somewhat more emphasis on physics. Also, a small percentage of items in each assessment (5 percent or less) were classified in the environmental science content area. There also was one NAEP fourth-grade item dealing with technology that could not be placed within the content areas as defined in the TIMSS framework.

At eighth grade, NAEP items are again fairly evenly distributed across the *fields of science* according to the NAEP framework, whereas TIMSS items show a greater emphasis on physical science (46 percent in TIMSS compared to 31 percent in NAEP) and a lesser emphasis on Earth science (24 percent in TIMSS compared to 33 percent in NAEP) (table 3). These differences in emphases are confirmed using the TIMSS framework as the classification system (table 4). The comparison based on the TIMSS framework also shows that the increased emphasis on physical science in TIMSS is distributed across both the chemistry and physics content areas. TIMSS also has a higher percentage of items classified as environmental science, which is consistent with this being an explicit content area in TIMSS but not NAEP. Most items classified as environmental science were classified as Earth science in the NAEP framework, although some TIMSS environmental science items were also classified as life or physical science in NAEP. As with the fourth grade, one NAEP eighth-grade technology item could not be placed within the content areas as defined in the TIMSS framework. Also, four items (2 percent) from each assessment were classified to multiple content areas on the other's framework.

[21] The classifications of items to their own framework were provided by the assessment developers. Cross classifications of NAEP and TIMSS items to the other's assessment framework were done by the expert panel.

Table 3. Percentage of NAEP 2000 and TIMSS 2003 science items classified to the fields of science in the NAEP science framework, by grade and survey

NAEP field of science	Grade 4		Grade 8	
	NAEP[1]	TIMSS[2]	NAEP[1]	TIMSS[2]
Total number of items	144	141	197	177
	Percentage distribution			
Physical science	31	31	31	46
Life science	37	45	36	32
Earth science	32	24	33	24
Classified to multiple fields	0	0	0	2

[1] NAEP items classified by NAEP developers.

[2] TIMSS items classified by expert panel.

NOTE: Data reflect the percentage of items classified to the NAEP content framework at any level of specificity (*field of science*, *subtopic*, or *specific objective*). Multi-part items were treated as one item for classification purposes and only contribute one to the total. However, if multi-part items covered multiple *fields of science*, then they contributed to the percentage for each *field of science*. Items classified to multiple *fields of science* were counted in each relevant category. Four TIMSS eighth-grade items were classified to multiple *fields of science*: three to Earth science and life science and one to physical science and Earth science. Detail may not sum to totals because of rounding or items classified to multiple *fields of science*.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.

Table 4. Percentage of NAEP 2000 and TIMSS 2003 science items classified to the content domains in the TIMSS science framework, by grade and survey

TIMSS content domain	Grade 4		Grade 8	
	NAEP[1]	TIMSS[2]	NAEP[1]	TIMSS[2]
Total number of items	144	141	197	177
	Percentage distribution			
Physical science	35	33	31	42
Chemistry	15	13	11	17
Physics	19	20	21	25
Life science	37	43	35	29
Earth science	24	20	27	16
Environmental science	3	5	7	13
Classified to multiple domains	1	0	2	0
Not classified to a content domain	1	0	1	0

[1] NAEP items classified by expert panel.

[2] TIMSS items classified by TIMSS developers.

NOTE: Data reflect the percentage of items classified to the TIMSS content framework at any level of specificity (*content domain*, *topic area,* or *objective*). Multi-part items were treated as one item for classification purposes and only contribute one to the total. However, if multi-part items covered multiple *content domains*, then they contributed to the percentage for each *content domain*. Items classified to multiple *content domains* were counted in each relevant category. Three eighth-grade NAEP items were classified to multiple *content domains*: two to chemistry and physics and one to life science and physics. One fourth-grade NAEP item was also classified to multiple *content domains*: chemistry and physics. Two NAEP Earth science items (one at eighth grade and one at fourth grade) were not classified to a *content domain* on the TIMSS framework. Detail may not sum to totals because of rounding or omitted items or items classified to multiple *content domains*.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Edition*, 2003.

At the broad content area level, the classification of virtually all NAEP and TIMSS items was consistent with the basic definitions of the other framework. However, a closer examination of the degree to which the NAEP and TIMSS items could be classified on the other assessment framework (i.e., at the topic or specific objective levels) reveals that the level of content match is less universal at the more detailed levels of the framework, particularly when classifying TIMSS items to the NAEP framework (table 5). About one-third of TIMSS fourth-grade items and 40 percent of TIMSS eighth-grade items could not be mapped to one of NAEP's specific objectives at any grade, and nearly 20 percent of TIMSS items in each grade could not be mapped to the NAEP framework at the topic level. Conversely, for NAEP items, the level of match to the TIMSS framework was greater, with over 80 percent classified at the specific objective level and over 90 percent classified at the topic level for both grades. This suggests that, for each assessment to varying degrees, there are specific topics and objectives represented in the items that may not be covered explicitly in the framework of the other assessment. In the case of TIMSS items, panelists' comments bear this out, as they highlighted a number of the items for which there was no explicit topic included in the NAEP framework, such as those related to chemical change and reactions, burning, heat conductivity, or electrical circuits in physical science or to human health in life science. Some NAEP items that could not be placed on the TIMSS framework related to momentum and action-reaction in the physical science content area. In some areas, the lower level of specific match between TIMSS items and the NAEP framework reflects broader content coverage in TIMSS. In other areas, the differences are more a reflection of a lack of specificity in the NAEP framework about the content to be included, as there are items in the NAEP assessment covering content similar to these TIMSS items. More information about the differences in topics covered in NAEP and TIMSS is included in the sections devoted to content area comparisons (section 5). NAEP and TIMSS items illustrating some of the findings in each content area are included in appendix E.

Table 5. Percentage of NAEP 2000 and TIMSS 2003 science items classified to the other assessment framework at the topic or objective level, by grade and survey

	Grade 4		Grade 8	
Level of content classification	NAEP items to TIMSS framework	TIMSS items to NAEP framework	NAEP items to TIMSS framework	TIMSS items to NAEP framework
Topic level	97	83	93	82
Objective level	83	67	82	60

NOTE: Data reflect the percentage of items that were classified by the expert panel to the topic and objective levels of the other assessment framework in any content area. Items classified to multiple topics or objectives are considered to match those levels of classification and are counted only once.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Edition*, 2003.

4.2. Grade Level

The cross-classification data were used to examine the extent to which items from one assessment map to the same grade level framework of the other assessment. Figures 1-A and 1-B show the percentage of items in the NAEP and TIMSS assessments overall that were classified at each grade level in the other assessment framework. For these overall comparisons, the percentages at each grade level of the other's assessment framework reflect items that were classified at the specific objective, topic, or broad content area level.[22]

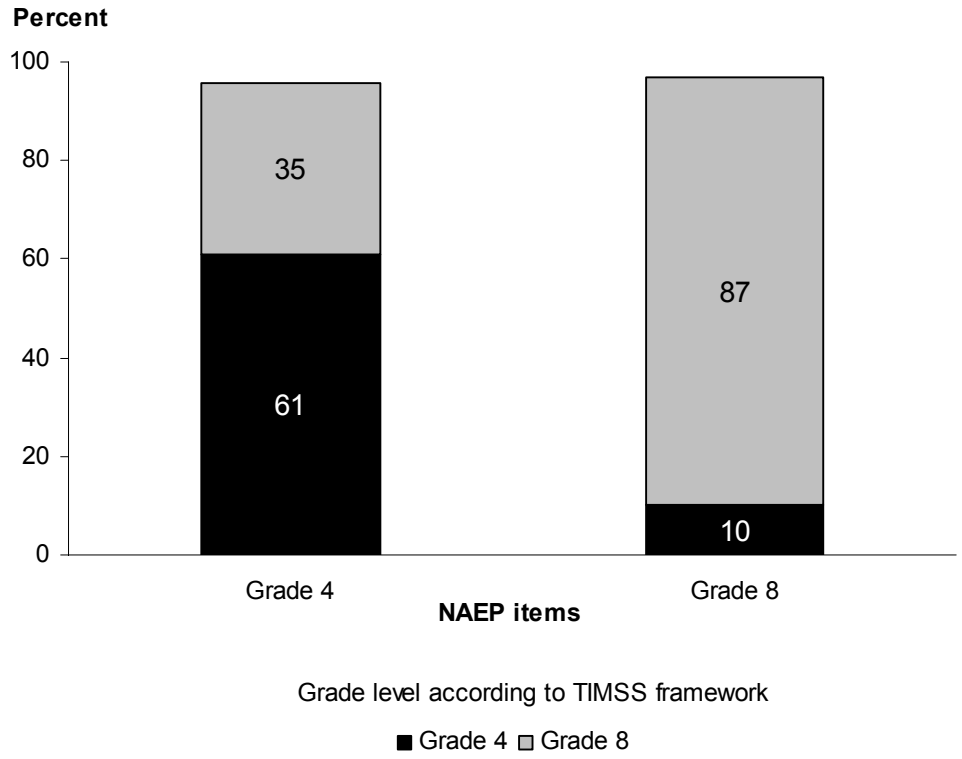
The degree of consistency regarding what is considered fourth-grade science content and what is considered eighth-grade science content varies between the two assessments and across the two grades. Sixty-one percent of NAEP fourth-grade items were classified to the TIMSS fourth-grade framework, and 35 percent were classified to the TIMSS eighth-grade framework (figure 1-A). In other words, about one-third of the items included in the NAEP fourth-grade assessment would be more likely to be included in the TIMSS eighth-grade assessment. In contrast, 88 percent of TIMSS fourth-grade items were judged to be consistent with the NAEP fourth-grade framework and the remaining 12 percent with the eighth-grade framework (figure 1-B). As is discussed in the following sections, the NAEP fourth-grade items that were classified at the eighth grade in TIMSS go across all of the main content areas of physical science, life science, and Earth science. This is consistent with the different purposes for the assessments, with TIMSS reflecting a consensus across countries including many with a less formalized science curriculum in primary school than that in the United States. It also is consistent with the relatively high performance of U.S. fourth-graders on past TIMSS science assessments (Martin et al. 1997; Robitaille 1997).

Compared to fourth grade, grade-level match appears to be more consistent for eighth grade, with 87 percent of NAEP items and 82 percent of TIMSS items classified at the corresponding grade level of the other assessment framework. The remaining TIMSS items were almost evenly split between those mapped to the fourth-grade and twelfth-grade NAEP framework. The remaining NAEP items were classified to the TIMSS fourth-grade framework. Comparisons within each content area presented in the following sections indicate that the TIMSS items classified as a higher grade level in NAEP—either fourth-grade items classified at eighth grade or eighth-grade items classified as twelfth grade—come primarily from the Earth science and environmental science categories, with some items also from life science. In contrast, there were no physical science items in TIMSS that were classified at a higher grade level in the NAEP framework. It should be noted that since the TIMSS 2003 framework includes only fourth and eighth grades, it was not possible for the panel to classify NAEP items at a grade level higher than the eighth grade on the TIMSS framework.

[22] The analyses for each of the content area comparisons in section 5 further examine the degree to which items match topics and objectives at particular grades.

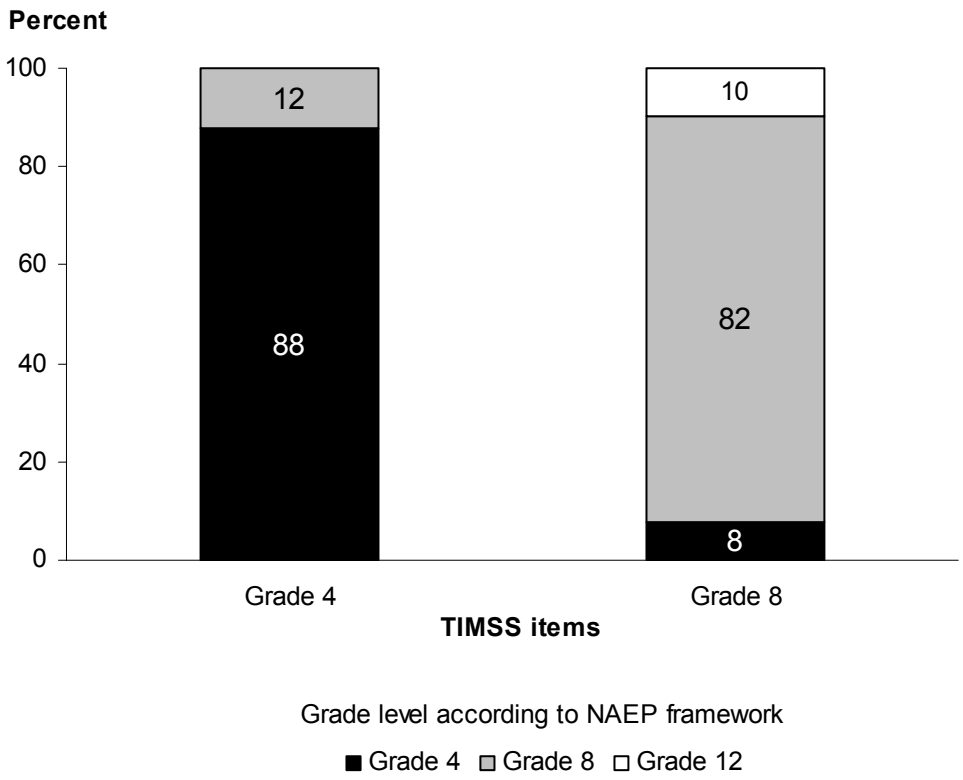
While there were few comments recorded that suggested that the NAEP items exceeded the TIMSS eighth-grade descriptions, there were some physical science items that the panel judged to be generally consistent with the topics in the TIMSS eighth-grade framework but having some item characteristics (e.g., specific concept applications or terminology) that are somewhat more advanced than the eighth-grade specifications in TIMSS. The same situation was also found with some of the TIMSS eighth-grade physical science items, which although they were classified as most consistent with the descriptions in the NAEP eighth-grade framework, had aspects that were somewhat beyond that grade level but not consistent with the twelfth-grade specifications. These items (NAEP and TIMSS), which are treated as eighth-grade classifications in the overall results shown in figure 1-A, are discussed further in the content area comparison section on physical science (section 5.1).

Figure 1-A. Percentage distribution of NAEP 2000 science items classified at each grade level according to the TIMSS 2003 science framework, by grade



NOTE: Data reflect expert panel classifications of grade level according to the TIMSS content framework at any level of specificity (*content domain*, *topic area*, or *objective*). Six NAEP items at each grade that the panel did not classify with respect to grade level on the TIMSS framework are not included.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003.

Figure 1-B. Percentage distribution of TIMSS 2003 science items classified at each grade level according to the NAEP 2000 science framework, by grade



NOTE: Data reflect expert panel classifications of grade level according to the NAEP content framework at any level of specificity (*field of science*, *subtopic*, or *specific objective*). Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.

One consideration in comparing the grade match of NAEP and TIMSS is the fact that NAEP includes a portion of items that were developed to be used at multiple grades (cross-grade items): 28 items administered at both fourth and eighth grades and 48 items administered at both eighth and twelfth grades. These cross-grade items reflect 19 percent of the items in the fourth-grade assessment and 38 percent of the eighth-grade assessment items (data not presented in tables). As shown in table 6, the inclusion of cross-grade items generally does not explain the lack of grade correspondence found for the fourth-grade NAEP items. The cross-grade items administered at fourth and eighth grades are fairly evenly distributed with respect to their classification to fourth- and eighth-grade levels in the TIMSS framework, 50 percent and 43 percent, respectively. All of the cross-grade items administered at eighth and twelfth grades are classified to the TIMSS framework at the eighth grade, which again may be due in part to the fact that there is no twelfth-grade assessment in TIMSS. Example 1 in appendix E illustrates a NAEP cross-grade item administered at fourth and eighth grades that was classified at the eighth-grade level on the TIMSS framework.

Table 6. Percentage of NAEP 2000 single-grade and cross-grade science items classified at each grade level according to the TIMSS 2003 science framework

Grade level according to the TIMSS 2003 framework	Total		NAEP item type			
			Single-grade items		Cross-grade items	
	Grade 4	Grade 8	Grade 4 only	Grade 8 only	Grades 4 and 8	Grades 8 and 12
Grade 4	61	10	64	5	50	0
Grade 8	35	87	33	92	43	100

NOTE: Data reflect expert panel classifications of grade level to the TIMSS 2003 content framework at any level of specificity (*content domain*, *topic area*, or *objective*). Single-grade items are administered at one grade level; cross-grade items are administered at more than one grade (4 and 8 or 8 and 12); totals reflect single-grade and cross-grade items included in the assessment at each grade level. Ten NAEP items (four fourth-grade, four eighth-grade, and two cross-grade) that the panel judged as having a poor fit to the TIMSS content framework were not assigned a grade level classification, and, therefore, are not included. Detail may not sum to totals because of rounding or omitted items.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003.

4.3. Cognitive Domain

All of the NAEP and TIMSS items were classified according the TIMSS 2003 definition of *cognitive domain* as a common metric for comparing the cognitive demands of the items in each assessment. Items were classified into the *cognitive domain* categories of *factual knowledge*, *conceptual understanding,* and *reasoning and analysis* according to the expert panel judgment of the primary cognitive skills and abilities required by the item. While a range of cognitive skills is defined in each of the TIMSS *cognitive domains*, there is an implicit hierarchy assumed by this classification system.

Table 7 shows the percentage distribution of items in each assessment across the *cognitive domains* at fourth and eighth grades. At fourth grade, 53 percent of the NAEP items were classified as *conceptual understanding*; in addition, 6 percent of items were classified on the borderline of this category and another. Similar percentages of items (about 20 percent) were classified in the other two categories of *factual knowledge* and *reasoning and analysis*. In comparison to NAEP, TIMSS at grade 4 is more evenly distributed between *factual knowledge* and *conceptual understanding* items,

44 percent and 39 percent, respectively. Ten percent were classified as *reasoning and analysis*. As with NAEP, there were a number of TIMSS items (6 percent) that were classified on the borderline of the *cognitive domain* categories.

At the eighth grade level, there were fewer items that the panel had difficulty classifying in a single category, with 3 percent "borderline" items for each assessment. NAEP and TIMSS both emphasize *conceptual understanding* items at the eighth grade, but TIMSS does so to a lesser degree (47 percent compared to 61 percent for NAEP). At both grades, TIMSS has a substantially greater proportion of items classified as *factual knowledge* than NAEP—44 percent at fourth grade and 31 percent at eighth grade. In contrast, NAEP has less than 20 percent at both grades. Also, the proportion of TIMSS eighth-grade items classified as *reasoning and analysis* was twice that for the fourth grade (20 percent compared to 10 percent), while in NAEP the percentage classified as *reasoning and analysis* was about one quarter at both grade levels.

For both NAEP and TIMSS, the percentage of items classified as *factual knowledge* is greater at fourth grade than at eighth grade, although this difference is greater in TIMSS than in NAEP (13 percent and 7 percent, respectively). Both framework documents suggest that some differentiation between the grades would be expected on these process-related domains because of different expectations for cognitive development.

Table 7. Percentage distribution of NAEP 2000 and TIMSS 2003 science items across TIMSS cognitive domains, by grade and survey

Cognitive domain	Grade 4		Grade 8	
	NAEP	TIMSS	NAEP	TIMSS
Classified in a single domain				
Factual knowledge	18	44	11	31
Conceptual understanding	53	39	61	47
Reasoning and analysis	22	10	25	20
Classified on the border of two domains				
Factual knowledge/conceptual understanding	3	5	2	2
Conceptual understanding/reasoning and analysis	3	1	1	1
Not classified in a cognitive domain	0	1	0	0

NOTE: Data reflect expert panel classifications of *cognitive domain* as defined in the TIMSS 2003 science framework.
Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Edition*, 2003.

Table 8 shows the distribution of items across the *cognitive domains* broken down by main content area.[23] Looking at the distributions, there are some differences across the content areas, including:

- The difference between NAEP and TIMSS in terms of the relative proportion of items in *factual knowledge* versus *reasoning and analysis* items is most pronounced in physical science. In this content area, NAEP has about one-third at both grades classified as *reasoning and analysis* (36 percent at the fourth grade and 31 percent at the eighth grade) and less than 10 percent as *factual knowledge* (9 percent at the fourth grade and 5 percent at the eighth grade). In comparison, nearly half of fourth-grade (46 percent) and a quarter of eighth-grade items in TIMSS are *factual knowledge* (26 percent) and less than 20 percent are *reasoning and analysis* (13 percent at the fourth grade and 18 percent at the eighth grade).

- While NAEP and TIMSS have the highest proportion of life science items classified as *conceptual understanding*, more NAEP items are likely to require *reasoning and analysis* and more TIMSS items are likely to require *factual knowledge*. This is true for both grades.[24]

- The profile of *cognitive domain* for the Earth science items differs from the profiles of the other content areas, particularly at the fourth grade. While NAEP fourth-grade Earth science items retain a heavy emphasis on *conceptual understanding* (61 percent), the proportion of *factual knowledge* items is greater than in other content areas. For TIMSS, there is a higher percentage of *reasoning and analysis* items in Earth Science (21 percent at fourth grade and 31 percent at eighth grade) than in the physical and life sciences. Earth science is the only content area where TIMSS has a higher proportion of *reasoning and analysis* items than NAEP.

- The TIMSS environmental science items for eighth grade showed heavier emphasis on *reasoning and analysis* compared to the other TIMSS content areas, with 39 percent classified in this *cognitive domain*. Additionally, almost half of the items were classified as *conceptual understanding*. Thirteen percent were classified as *factual knowledge*—the lowest across all the TIMSS content areas. Some of these items address topics that, in NAEP, are considered twelfth-grade topics (data not shown in table 8).

---

[23] For the data reflected in table 8 and the remaining figures in this section, items classified on the borderline between two *cognitive domains* have been counted in the "higher" category (*factual knowledge*/*conceptual understanding* as *conceptual understanding*; *conceptual understanding*/*reasoning and analysis* as *reasoning and analysis*). This procedure was based on a policy set by the expert panel and reflects an assumed hierarchy indicated in the TIMSS framework.

[24] Another difference noted for the life science items but not shown in table 8 is that a higher number of items in life science were classified on the border between two *cognitive domains*, particularly at fourth grade (18 percent of NAEP and 10 percent of TIMSS fourth-grade items were not placed in a single category).

Table 8. Percentage distribution of NAEP 2000 and TIMSS 2003 science items across TIMSS cognitive domains, by grade and content area

Content area and cognitive domain	Grade 4		Grade 8	
	NAEP	TIMSS	NAEP	TIMSS
Physical science items				
Factual knowledge	9	46	5	26
Conceptual understanding	56	41	65	56
Reasoning and analysis	36	13	31	18
Life science items[1]				
Factual knowledge	21	45	19	41
Conceptual understanding	54	48	60	49
Reasoning and analysis	24	5	21	10
Earth science items				
Factual knowledge	24	39	9	41
Conceptual understanding	61	39	65	27
Reasoning and analysis	15	21	27	31
Environmental science items				
Factual knowledge	†	†	†	13
Conceptual understanding	†	†	†	48
Reasoning and analysis	†	†	†	39

† Not applicable. TIMSS does not have a reporting category for environmental science at the fourth grade. TIMSS fourth-grade environmental science items have been included in percentages for the content area subscales in which they were reported (life science or Earth science). NAEP does not include an environmental science content area at fourth or eighth grade.
[1] One TIMSS fourth-grade life science item (2 percent) not classified by the panel with respect to *cognitive domain* is not included.
NOTE: Data reflect expert panel classifications of *cognitive domain* as defined in the TIMSS 2003 science framework.
Items classified on the borderline between two *cognitive domains* have been counted in only one category for the purposes of this table (*factual knowledge/conceptual understanding* as *conceptual understanding*; *conceptual understanding/reasoning and analysis* as *reasoning and analysis*). Detail may not sum to totals because of rounding or omitted items.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003.

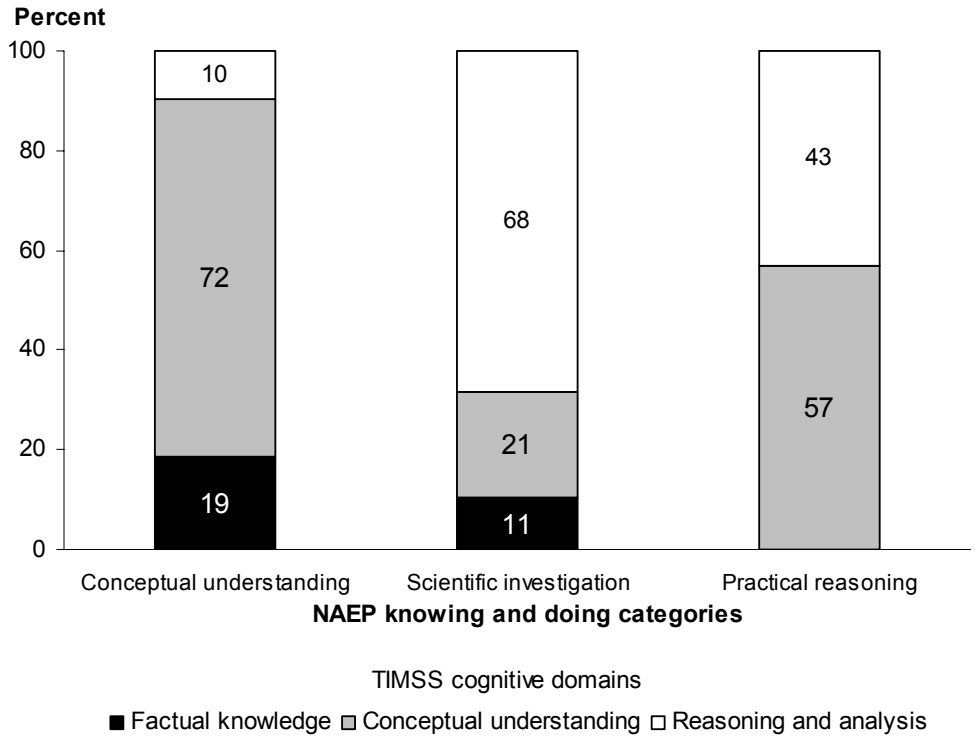
The NAEP items developed for each of the *knowing and doing* categories in the NAEP framework were compared with respect to the proportion of *factual knowledge, conceptual understanding* and *reasoning and analysis*. Figure 2 displays the percentage of items classified in each *cognitive domain* for the items from the NAEP categories of *conceptual understanding, scientific investigation*, and *practical reasoning*.

While 72 percent of the items in the NAEP *conceptual understanding* category were also classified as *conceptual understanding* in the TIMSS framework, 19 percent were classified as *factual knowledge* and 10 percent as *reasoning and analysis*. Examination of the definitions of the NAEP framework reveals that the *conceptual understanding* category does include "facts and events" and "information and procedures" which are included in the TIMSS *factual knowledge* domain. This NAEP category also includes "application of scientific knowledge in the engagement of practical tasks," which would be consistent with the TIMSS definition of *reasoning and analysis*. Thus, the NAEP category of *conceptual understanding* is defined broadly to include factual knowledge as well as conceptual understanding and reasoning abilities, and this is reflected in the items developed based on the framework.

More than two-thirds of items in the NAEP *scientific investigation* category were classified as *reasoning and analysis*, which is consistent with the fact that many *scientific inquiry* skills such as hypothesizing, designing investigations, analyzing data, and drawing conclusions are included in this TIMSS *cognitive domain*. The remaining third of NAEP *scientific investigation* items were classified to the other TIMSS *cognitive domains*. This is consistent with a NAEP definition of *scientific investigation* that includes factual knowledge and conceptual understanding about the methods and tools of science.

The NAEP *practical reasoning* items share characteristics with the TIMSS categories of *conceptual understanding* and *reasoning and analysis*, with somewhat less than half of items classified in the latter. The two TIMSS domains, *conceptual understanding* and *reasoning and analysis*, covered by the NAEP *practical reasoning* items are differentiated in terms of the complexity or familiarity of the problem situations, the level of hypothesizing, planning, decision making, and analysis required, as well as whether the problem solution requires consideration of multiple factors.

Figure 2. Percentage distribution of NAEP 2000 science items across TIMSS 2003 cognitive domains, by NAEP knowing and doing categories



NOTE: Includes both fourth and eighth grade items combined. Classifications by *knowing and doing* categories provided by NAEP assessment developers; classifications by TIMSS *cognitive domain* made by expert panel according to definitions in the TIMSS 2003 framework. Items classified on the border of two *cognitive domains* were counted in the "higher" category for the purpose of this figure (*factual knowledge/conceptual understanding* as *conceptual understanding*; *conceptual understanding/reasoning and analysis* as *reasoning and analysis*). Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003; and U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.
.

The expert panel's classifications were compared to those of the assessment developers for the TIMSS items since both were based on the *cognitive domain* categories in the TIMSS framework. As shown in table 9, there is general agreement between the classifications, although the expert panel classified more items as *factual knowledge* (37 percent versus 34 percent) and *conceptual understanding* (47 percent versus 41 percent), whereas developers classified more items as *reasoning and analysis* (25 percent versus 16 percent). The more "stringent" classifications by the panel may be due in part to the nature of the task. Whereas the experts were reviewing items *ex post facto*, developers use the definitions not as a classifying system *per se* but as a tool or guide to ensure that a range of cognitive skills are represented in assessment items being developed. Differences may also be due in part to the fact that some items may require skills from more than one of these categories, and how these items were handled may have differed. Also, the TIMSS developers have a greater familiarity with the framework definitions and how they are applied in the development of items.

Table 9. Percentage distribution of TIMSS 2003 science items across TIMSS cognitive domains: comparison between classifications by expert panel and assessment developers

TIMSS cognitive domains	Classification source	
	Classified by expert panel	Classified by TIMSS assessment developers
Factual knowledge	37	34
Conceptual understanding	47	41
Reasoning and analysis	16	25

NOTE: Includes items at both fourth and eighth grades. Items classified by the expert panel on the border of two *cognitive domains* were counted in only one category for the purpose of this table (*factual knowledge/conceptual understanding* assigned to *conceptual understanding*; *conceptual understanding/reasoning and analysis* assigned to *reasoning and analysis*). Items classified by the developers to multiple *cognitive domain* categories (e.g., if they were multi-part items) were counted in the "highest" category, assuming an implicit hierarchy of *factual knowledge*/*conceptual understanding*/*reasoning and analysis*. Classifications by *cognitive domain* were made according to the definitions in the TIMSS 2003 framework. Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003.

4.4. Scientific Inquiry

The expert panel classified items with respect to whether or not they matched the TIMSS definition for measuring *scientific inquiry* skills.[30] Figures 3 and 4 show the percentage of fourth-grade and eighth-grade items from NAEP and TIMSS that were classified by the expert panel as measuring *scientific inquiry* skills overall and broken down by main content area. For the NAEP items overall, the distribution of items is roughly the same in both grades, with nearly one-quarter of the items classified as measuring *scientific inquiry* skills. For TIMSS, a slightly higher percentage of eighth-grade items were classified as *scientific inquiry* than fourth-grade items (12 percent compared to 9 percent). The higher percentage of NAEP items classified as *scientific inquiry* is due in part to the inclusion of the hands-on performance tasks in NAEP which represent between 10 and 15 percent of the NAEP fourth- and eighth-grade assessments overall and about two-fifths of the items classified as *scientific inquiry* (data not shown). In comparison, the TIMSS assessment included two or three problem solving and inquiry tasks at each grade. About one-third of the items in the TIMSS problem solving and inquiry tasks were classified as measuring *scientific inquiry* compared to nearly all of the NAEP performance task items.[31] There was considerable discussion among the panelists about the difficulty of assessing scientific inquiry through large-scale assessments such as NAEP and TIMSS.

Looking at the results by content area, at fourth grade, the NAEP physical science items stand out as having the highest proportion of *scientific inquiry* items (36 percent), compared to 23 percent in life science and 15 percent in Earth science. This may relate to the fact that the NAEP physical science area includes more performance tasks than other content areas. At the eighth grade, there is a more even distribution of *scientific inquiry* across the content areas in NAEP than in TIMSS, but the proportion in life science is the lowest. For TIMSS at both grade levels, the Earth and environmental science content areas have more *scientific inquiry* items (close to 20 percent or more) than the life and physical science content areas in both grades, with the percentage of life science items being particularly low (4 percent or less at both grades).

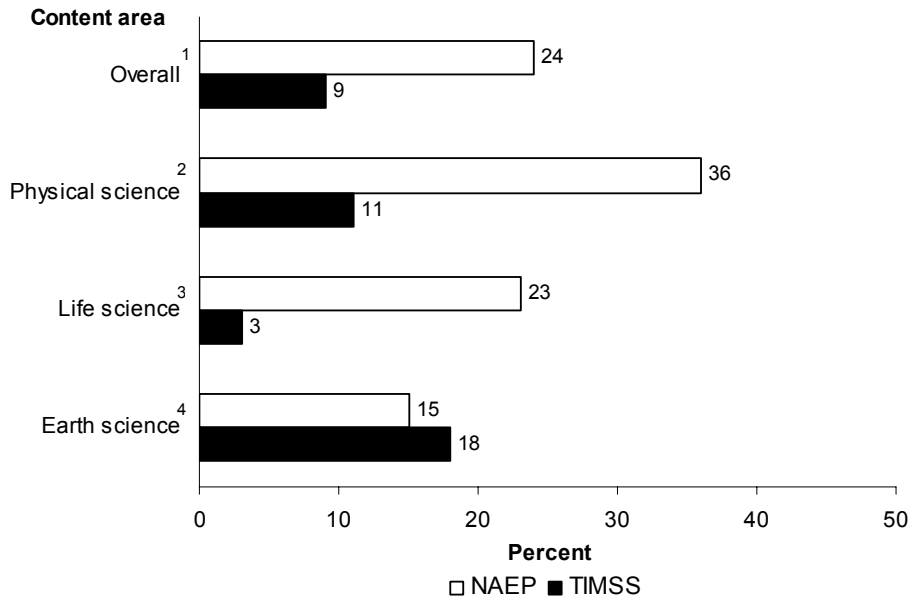
Examples 2 and 3 in appendix E show items from the NAEP and TIMSS assessments, respectively, that were classified as measuring *scientific inquiry* skills. Both of these items require students to design and plan a scientific procedure.

---

[30] See appendix B for a description of the definition of scientific inquiry.

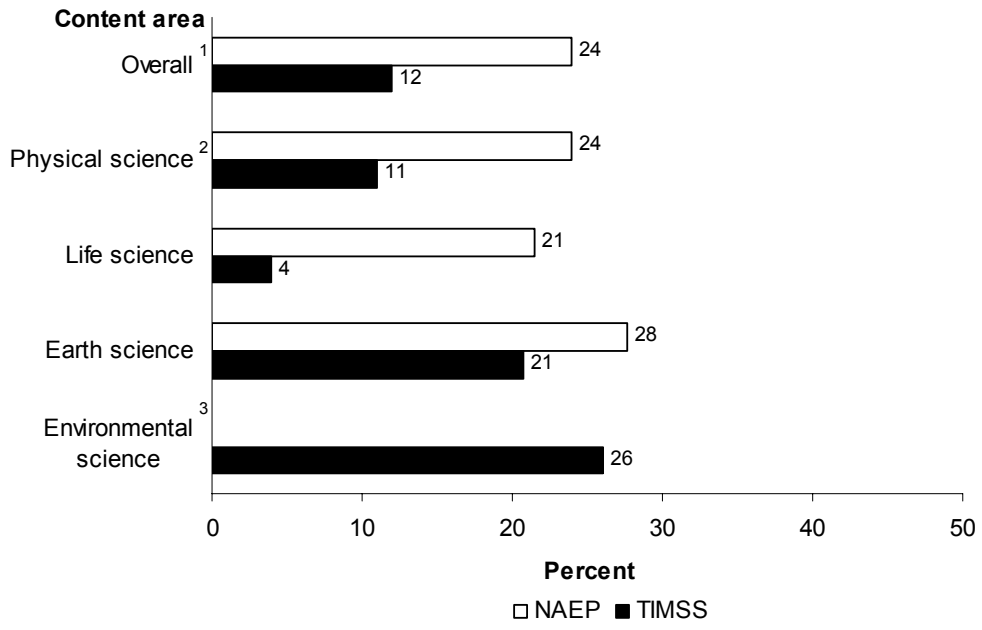
[31] Both the NAEP hands-on tasks and the TIMSS problem solving and inquiry tasks include several items that are considered separately when determining the proportion of items in each assessment classified as scientific inquiry. Example tasks are show in appendix E.

Figure 3. Percentage of NAEP 2000 and TIMSS 2003 fourth-grade items classified as measuring scientific inquiry, by science content area



[1] The overall category refers to all items combined.
[2] TIMSS physical science category combines items from both the chemistry and physics *content domains*.
[3] Two TIMSS fourth-grade items from the environmental science *content domain* are included in the life science reporting category.
[4] Five TIMSS fourth-grade items from the environmental science *content domain* are included in the Earth science reporting category.
NOTE: Data reflect expert panel classifications of *scientific inquiry* as defined in the TIMSS 2003 framework.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Edition*, 2003.
.

Figure 4. Percentage of NAEP 2000 and TIMSS 2003 eighth-grade items classified as measuring scientific inquiry, by science content area



[1] The overall category refers to all items combined.
[2] TIMSS physical science category combines items from both the chemistry and physics *content domains*.
[3] NAEP does not include an environmental science category.
NOTE: Data reflect expert panel classifications of scientific inquiry as defined in the TIMSS 2003 framework.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Edition*, 2003.

The panel's classifications of *scientific inquiry* and *cognitive domain* may be used to further compare the nature of the *scientific inquiry* items in each assessment. The TIMSS definition of *scientific inquiry* includes both content- and skills-based components, and many of the skills and abilities associated with the inquiry process (formulating hypotheses, designing investigations, collecting, interpreting and analyzing data, and drawing conclusions) overlap the *cognitive domain* of *reasoning and analysis*. In addition to these process skills, the TIMSS *scientific inquiry* dimension also includes general knowledge and understanding about the methods and nature of science. To investigate the extent to which the *scientific inquiry* items are focused on process skills, the percentage of NAEP and TIMSS *scientific inquiry* items classified in each *cognitive domain* category were computed (table 10). NAEP and TIMSS have similar profiles, with the vast majority of *scientific inquiry* items classified as *reasoning and analysis* and slightly more than 10 percent as *factual knowledge* or *conceptual understanding*.

Table 10. Percentage distribution of NAEP 2000 and TIMSS 2003 scientific inquiry items across TIMSS cognitive domains

TIMSS cognitive domains	NAEP scientific inquiry items	TIMSS scientific inquiry items[1]
Factual knowledge	5	3
Conceptual understanding	6	9
Reasoning and analysis	88	86

[1] One TIMSS item (3 percent) that the panel did not classify with respect to *cognitive domain* is not included.
NOTE: Includes both fourth- and eighth-grade items. Data reflect classifications made by the expert panel according to definitions in the TIMSS 2003 framework of *scientific inquiry* and *cognitive domains*. Items classified on the borderline between two *cognitive domains* have been counted in only one category for the purposes of this table (*factual knowledge/conceptual understanding* as *conceptual understanding*; *conceptual understanding/reasoning and analysis* as *reasoning and analysis*). Detail may not sum to totals because of rounding or omitted items.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003.

4.5. Item Format

The items in the NAEP and TIMSS assessments were also compared with respect to the types of item formats used and their proportion of the assessments. Table 11 shows the percentage distribution of NAEP and TIMSS items by item format, including multiple choice and constructed response (short answer and extended response). Items can vary in difficulty and cognitive demand regardless of format, though extended constructed-response items can be particularly important in assessing students' abilities to generate ideas and solutions and communicate their depth of scientific understanding. Including a variety of item types ensures that a range of knowledge and skills is being assessed.

Table 11. Percentage distribution of NAEP 2000 and TIMSS 2003 science items across item formats, by grade and survey

Item format	Grade 4		Grade 8	
	NAEP	TIMSS	NAEP	TIMSS
Multiple choice	49	65	48	62
Constructed response	51	35	52	38
Short answer	44	18	45	20
Extended response	7	17	7	18

NOTE: The breakdown of constructed-response items as short answer or extended response was provided by the assessment developers for the NAEP items. For the TIMSS items, the assignment was based on examination of the items and level of score points in the scoring guides in accordance with information provided by the TIMSS assessment developers—extended-response items reflect multi-part items and items that were scored with 3-level scoring rubrics. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; and International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment.

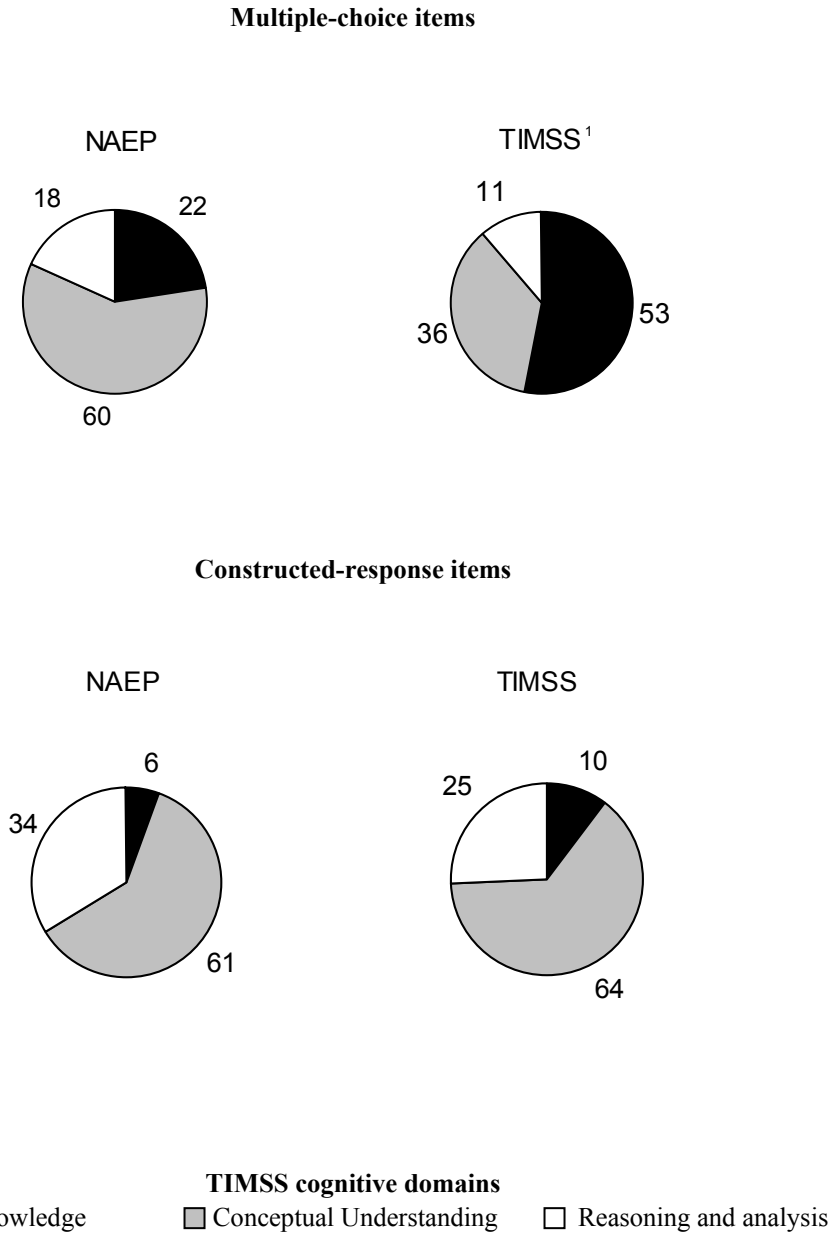
On both the fourth- and eighth-grade assessments, the NAEP items are roughly balanced between multiple-choice and constructed-response items, while nearly two-thirds of the TIMSS items are multiple-choice items. Of the constructed-response items, TIMSS has a substantially higher proportion of items defined as extended constructed response (about half). However, the definition and nature of extended-response items is not always the same across the two assessments. In both

assessments, the constructed-response items are scored with rubrics that are customized for each item. In TIMSS the short-answer items are scored with a 2-level rubric (Correct/Incorrect) and extended-response items with a 3-level rubric (Correct/Partial/Incorrect). In NAEP, the short-answer items may be scored with either a 2-level or 3-level rubric, while extended-response items are scored with a 4-level rubric (Correct/Partial/Minimal/Incorrect). Of particular note is that TIMSS includes a number of items at each grade level that ask for two responses to the same question (e.g., give two examples, write two reasons, etc.). While each response requires a short-answer and is scored with a 2-level rubric, the items as a whole are treated as having three score levels (two correct answers, one correct answer, and no correct answers), and therefore, are classified as extended response in TIMSS. Items of this type in NAEP are also scored with a 3-level rubric, but are classified as short-answer.

To compare the cognitive demand placed on students by the items of different formats, figure 5 shows the percentage of multiple-choice and constructed-response items in each assessment that were classified as *factual knowledge*, *conceptual understanding*, or *reasoning and analysis*. More than half of the multiple-choice items in TIMSS were classified as *factual knowledge* compared to 22 percent of NAEP multiple-choice items classified as such. Sixty percent of NAEP multiple-choice items were classified as measuring *conceptual understanding* compared to 36 percent of TIMSS multiple-choice items. NAEP also had a somewhat higher proportion of multiple-choice items classified as *reasoning and analysis* (18 percent compared to 11 percent in TIMSS). Example 4 in appendix E illustrates a NAEP eighth-grade multiple-choice item classified as measuring *conceptual understanding*. Relative to the multiple-choice items, the distribution of constructed-response items across the *cognitive domains* was generally more comparable between the two assessments, although a larger proportion of the NAEP items were classified as *reasoning and analysis* (34 percent compared to 25 percent) and a smaller proportion as *factual knowledge* (6 percent compared to 10 percent) and *conceptual understanding* (61 percent compared to 64 percent).

This section compared the assessments overall with respect to content coverage, grade level, cognitive domain, scientific inquiry, and item format. In the next section, more detailed comparisons of the content of the two assessments are made in each of the main content areas of physical science, life science, Earth science, and environmental science.

Figure 5. Percentage distribution of NAEP 2000 and TIMSS 2003 science items across TIMSS cognitive domains, by item format



[1] One TIMSS multiple-choice item that the panel did not classify with respect to *cognitive domain* is not included.
NOTE: Data reflect expert panel classifications of *cognitive domain* according to definitions in the TIMSS 2003 framework. Graphics for NAEP and TIMSS include both fourth- and eighth-grade items combined. Items classified on the border of two *cognitive domains* were counted in the "higher" category for the purpose of this figure (*factual knowledge/conceptual understanding* as *conceptual understanding*; *conceptual understanding/reasoning and analysis* as *reasoning and analysis*). Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003.

5. NAEP/TIMSS Comparisons by Main Content Areas

The overall comparisons section highlighted that, among other differences, NAEP and TIMSS have somewhat different emphases across the main science content areas. This section provides more detailed comparisons of the content coverage of the items for each of the NAEP and TIMSS science content area subscales. There are four sections, including:

- physical science (which includes chemistry and physics for TIMSS);
- life science;
- Earth science; and
- environmental science (which includes only TIMSS items).

Each content area section includes

- a comparison of the relevant parts of the content frameworks;[32]
- an analysis of the level of match between the items from one assessment and the topics and objectives at particular grades in the other assessment framework; and
- a comparison of how items are distributed across topics within the content areas defined by each framework.

For these analyses, the NAEP and TIMSS items are divided according to subscale and then comparisons are made within the content areas that are the same or similar across the two assessments. Grade and content classification are examined simultaneously in the analyses for this section. For each content area, this section reports on the percentage of items that were classified to the other framework at the corresponding grade level or at another grade level. For items classified at the corresponding grade level, there are three levels of content match, including:

- specific match (to a specific objective in the same content area);
- general match (at the broader topic level in the same content area but not the objective level); and
- match to another content area (at either the topic or objective level).

For items classified at another grade level in the other assessment framework, there are two types, including:

- lower grade (grade 8 items classified as grade 4 topics or subtopics); and

[32] Framework comparison exhibits and figures in this section list the topics included in each content area. For NAEP, topics are organized by major topic areas as described in section 2. Additional information about the specific objectives included for each of the topics is given in appendix A.

- higher grade (grade 4 items classified to grade 8 topics or subtopics or grade 8 TIMSS items classified to grade 12 NAEP topics or subtopics).[33]

This section also reports the percentages of items not classified to topics in the other assessment framework (i.e., those that could not be classified to a topic at a specific grade).[34] The text in this section may refer to items classified to specific objectives, but this level of detail is not shown in the tables in the report. Subtopics are listed in appendix A and example items illustrating various features that are referenced in this section are shown in appendix E.

5.1. Physical Science

TIMSS has a greater emphasis on physical science than NAEP, particularly at the eighth grade. As discussed in section 4, thirty-one percent of NAEP items at both fourth and eighth grades are from the *field* of physical science (table 3). In TIMSS, 33 percent of fourth-grade items and 42 percent of eighth-grade items are from the *content domains* of chemistry and physics (table 4). The results in the physical science section are based on 45 fourth-grade and 62 eighth-grade items in NAEP, and 46 fourth-grade and 74 eighth-grade items in TIMSS.

Framework comparison in physical science

The most visible structural difference between the NAEP and TIMSS physical science frameworks is that TIMSS includes separate content areas for chemistry and physics topics and reports achievement in these two areas on separate subscales at the eighth grade. Exhibit 3 shows a comparison of the physical science topics included in the NAEP and TIMSS science frameworks. NAEP includes topics related to chemistry in the major topic area of *matter and its transformations*, and a comparison of the frameworks at the topic level only would not necessarily reveal whether the specific content covered by the two assessments differed greatly. Closer examination of the topics and specific objectives included in each assessment, however, reveals a broader range of chemistry content covered in the TIMSS framework. Some topics within the content area of chemistry (e.g., *acids and bases* and *chemical change*) are explicit in the TIMSS framework but not the NAEP framework. As discussed in the next sections, the panel was not always able to classify some of the TIMSS chemistry items to topics within the NAEP framework.

The NAEP and TIMSS frameworks include 14 and 12 topics, respectively, in physical science, with a slight emphasis on topics relating to physics. One particular difference panelists noted between the frameworks is that the TIMSS conception of physical science explicitly includes knowledge about the *properties and uses of water*, whereas the NAEP physical science framework does not refer to water specifically but rather to properties and uses of common materials. The NAEP framework does, however, include some specific objectives in Earth science related to the special properties of water.

---

[33] Because the NAEP 2000 framework is used to guide a twelfth-grade assessment and the TIMSS 2003 framework is not, the classification of grade 8 items to the twelfth-grade level is only applicable for the classification of TIMSS items to the NAEP framework.

[34] The method for determining grade-level match in this section differs somewhat from what was used for the overall comparisons in section 4.2. Overall comparisons of grade level include items classified at any level of content match (specific objective, topic, or broad content area). In this section, grade level was not assigned unless items could be classified to at least the topic level.

Exhibit 3. Physical science topics included in the NAEP 2000 and TIMSS 2003 science frameworks

NAEP	TIMSS
Matter and its transformations	**Chemistry**
Diversity of materials: classification, types, and particulate nature of matter	Classification and composition of matter
Temperature and states of matter	Particulate structure of matter (grade 8 only)
Properties and uses of materials: modifying properties and the synthesis of materials with new properties	Properties and uses of water
Resource management (grade 12 only)	Acids and bases (grade 8 only)
Energy and its transformations	Chemical change
Forms of energy	**Physics**
Energy transformations in living systems, natural physical systems, and artificial systems constructed by humans	Physical states and changes in matter
Energy sources and use, including distribution, energy conversion, and energy costs and depletion	Energy types, sources, and conversions
Motion	Heat and temperature
Frames of reference	Light
Force and changes in position and motion	Sound and vibration (grade 8 only)
Action and reaction (grades 8 and 12 only)	Electricity and magnetism
Vibrations and waves as motion	Forces and motion
General wave behavior (grade 12 only)	
Electromagnetic radiation	
Interactions of electromagnetic radiation with matter	

NOTE: Unless otherwise noted, all topics are intended for all grades (grades 4, 8, and 12 in NAEP and grades 4 and 8 in TIMSS). The number of grade-specific objectives and level of detail varies across topics and assessments.
SOURCE: U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000; U.S. Department of Education, National Assessment Governing Board, *Science Assessment and Exercise Specifications for the 1994 National Assessment of Educational Progress*, 1994; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Edition*, 2003.

Content and grade match in physical science

At the fourth grade, there is a noticeable difference in the grade level correspondence between NAEP and TIMSS in the area of physical science. Although all the NAEP fourth-grade physical science items can be classified to topics in the TIMSS framework, these items embody some of what is considered eighth-grade content on the TIMSS framework (table 12). Whereas 80 percent of the TIMSS fourth-grade items were classified to the NAEP fourth-grade framework, nearly 40 percent of NAEP fourth-grade items were classified to the eighth-grade framework in TIMSS,

distributed across nearly all the topics. The remaining NAEP fourth-grade items were classified to physical science topics in the TIMSS fourth grade framework at varying levels of specificity. While the grade-level match of TIMSS fourth-grade items to the NAEP framework is closer, there are about 20 percent of items that were not clearly placed in NAEP framework topics at any grade. Example 5 in appendix E presents a NAEP fourth-grade physical science item placed at the eighth-grade level on the TIMSS science framework.

Table 12. Percentage of NAEP 2000 and TIMSS 2003 fourth- and eighth-grade physical science items classified to the other science assessment framework, by level of content/grade match

Level of content/grade match	Grade 4		Grade 8	
	NAEP items to TIMSS framework	TIMSS items to NAEP framework	NAEP items to TIMSS framework	TIMSS items to NAEP framework
Total number of physical science items	45	46	62	74
	Percentage distribution			
Classified as same grade	62	80	66	68
Specific match[1] in physical science	40	54	53	45
General match[2] in physical science	22	24	5	23
Match to another content area[3]	0	2	8	0
Classified as another grade[4]	38	0	13	3
Lower grade[5]	†	†	13	3
Higher grade[6]	38	0	†	0
No classification to topics[7]	0	20	21	30

† Not applicable. Grade 4 is the lowest grade in both frameworks; grade 8 is the highest grade in the TIMSS 2003 framework.

[1] Includes items that were classified to an objective at the same grade.

[2] Includes items that were classified to a topic but not to an objective at the same grade.

[3] Includes items that were classified to a topic or objective at the same grade.

[4] Includes items that were classified to a topic or objective in any content area at another grade.

[5] Includes grade 8 items classified to grade 4 topics/objectives.

[6] Includes grade 4 items classified to grade 8 topics/objectives or grade 8 TIMSS items classified to grade 12 NAEP topics/objectives.

[7] Includes items that the panel did not classify to a topic at a specific grade level.

NOTE: Data reflect the percentage of items classified by the expert panel at each level. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003; and U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.

More than half of TIMSS fourth-grade items (54 percent) and 40 percent of NAEP fourth-grade items map to one of the specific objectives in the other's fourth-grade framework in physical science. For both assessments, many of the fourth-grade items that do not have a specific match in the other's framework are items covering topics in chemistry. For example, there is no specific match for a number of TIMSS items relating to burning and chemical change. Conversely, several NAEP items from the topics of *properties and uses of new materials* and the *diversity of materials* have only a general match in the TIMSS *chemical change* topic. There are also TIMSS items related to electrical circuits and heat conductivity which were not explicitly included in the NAEP framework, although there are some items in NAEP that deal with these topics. Example 6 in

appendix E shows a TIMSS fourth-grade item involving *chemical change* that was not classified to a topic in the NAEP framework.

The proportion of TIMSS and NAEP eighth-grade physical science items classified at the corresponding grade level is similar—about two-thirds of the items are classified at the eighth-grade level of the other framework. For NAEP, however, this includes 8 percent of items that were classified to the TIMSS environmental science content area, mostly in the topic of *conservation and use of resources*. Also, 13 percent of NAEP eighth-grade items were classified to the TIMSS fourth-grade framework, concentrated in the chemistry topic relating to the *classification of matter* and spread across several topics in physics.

Both assessments have a substantial number of eighth-grade items that were not mapped to topics in the other's framework. For TIMSS, as in the fourth grade, this includes items related to burning and oxidation, heat conduction, and electrical circuits, which are topics not explicitly addressed in the NAEP framework. In addition, there are TIMSS chemistry items covering topics such as *acids and bases* that are included only in the TIMSS eighth-grade framework. Example 7 in appendix E shows a TIMSS eighth-grade item that was not classified to a topic in the NAEP framework. NAEP includes a number of items which, although the panel could assign them to a TIMSS topic, did not match well to the particular topic descriptions in the TIMSS eighth-grade framework. In many cases, this is because the panel thought the items addressed topics that were somewhat beyond the eighth-grade level descriptions in TIMSS. These included items involving acceleration, momentum, and induction as well as physics items that used terminology that was beyond the level specified for eighth-grade TIMSS, such as amplitude and frequency. While there were no TIMSS physical science items classified at the twelfth-grade level of the NAEP framework, there were some items considered to be somewhat above the eighth-grade level in the NAEP framework, in particular those dealing with the particulate structure of matter that require basic knowledge of subatomic particles, which is not included in NAEP at the eighth grade. These items were classified as "no classification to topics," since they did not clearly fit the grade-level expectations for the topics in the other assessment framework. For the overall grade-level comparisons presented in section 4.2, these items were still treated as classified at the eighth grade. In the case of NAEP items, there was no higher grade possible in the TIMSS framework. For the TIMSS items, the panel did not believe that the items warranted a classification at the twelfth-grade level, despite being somewhat beyond the specifications for eighth grade.

In terms of specificity of content match within physical science, a higher percentage of NAEP eighth-grade items can be mapped to TIMSS at the specific objective level than vice-versa (53 percent compared to 45 percent). Many of the TIMSS items that have a general match within physical science but were not classified to a specific objective in the NAEP framework are physics items related to *light* as well as items covering some chemistry topics. Some NAEP items that do not have a specific match to the TIMSS physics framework are related to momentum and action/reaction.

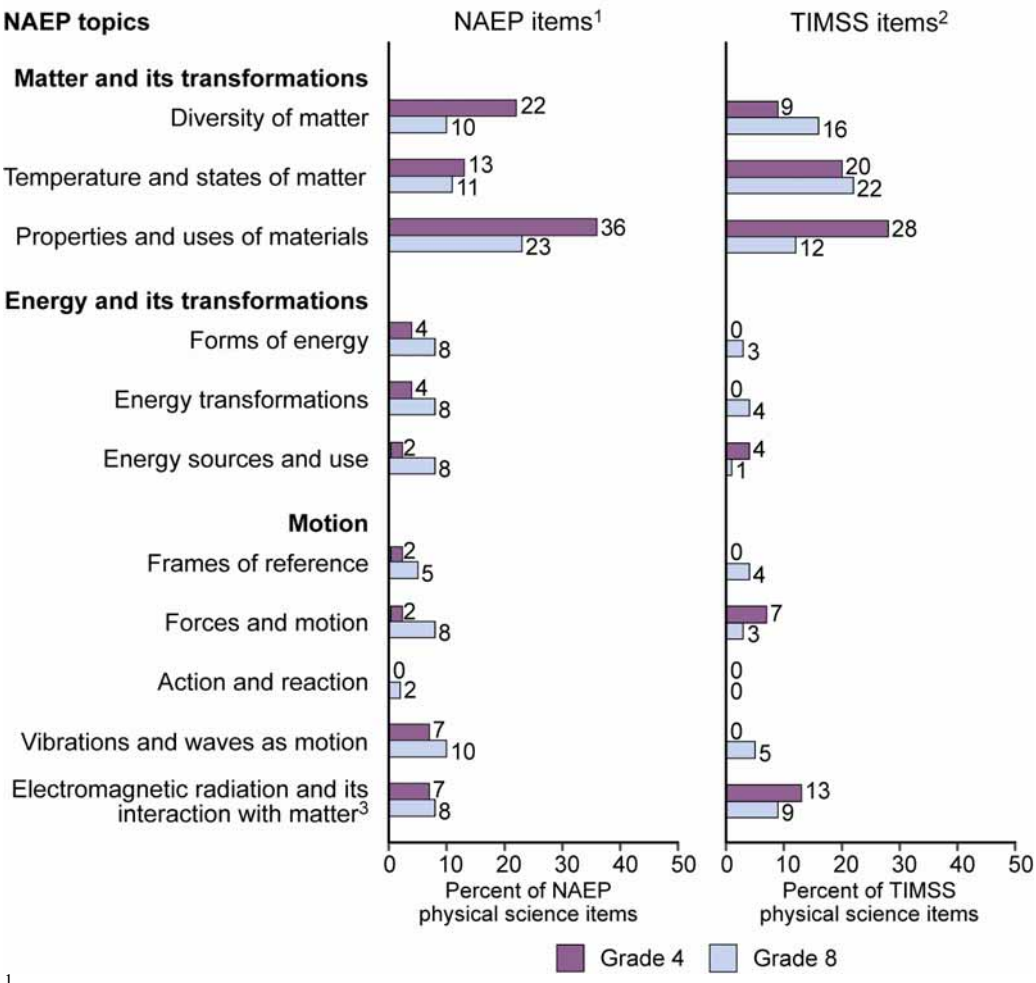
Item distribution across physical science topics

Looking at the distribution of physical science items across the topics in the respective NAEP and TIMSS frameworks shows that NAEP and TIMSS emphasize different topics (see figures 6 and 7). Using the NAEP framework as the basis for comparison, NAEP has relatively more focus at the fourth grade on the *diversity of matter* and *properties and uses of materials* (figure 6). In fact, more than one-third of NAEP items are classified to the latter topic area (36 percent). While TIMSS had almost 30 percent of items classified to the topic of *properties and uses of materials* (28 percent), it

also had a larger proportion of items (20 percent) related to *temperature and states of matter* than NAEP. NAEP had more coverage of *energy and its transformations* at fourth grade than TIMSS. A fourth-grade NAEP item from the *energy and transformation* topic area is shown in example 8 in appendix E. Within the physics topics, TIMSS has a larger proportion of items in the NAEP topic related to *electromagnetic radiation*, which focuses on basic properties and behavior of light at this grade (13 percent versus 7 percent). On the other hand, NAEP has items related to sound in the topic of *vibrations and waves as motion*, which is a topic not assessed at the fourth grade in TIMSS. This is confirmed by the distribution of items across topics in the TIMSS framework, which also shows relatively more TIMSS items in the topic of *light* (13 percent versus 9 percent) and none in the topic of *sound and vibration* (figure 7). Also, TIMSS has relatively more physical science items covering *properties and uses of water* at the fourth grade (7 percent versus 2 percent). Interestingly, NAEP has a higher percentage of fourth-grade items classified to the TIMSS topic of *chemical change* (20 percent versus 11 percent), despite the fact that the panel felt that this was an area not explicitly covered in the NAEP framework. As noted previously, the NAEP items were judged as having only a general match to this TIMSS topic.

As with the fourth-grade physical science assessments, the eighth-grade science assessments have some differences between them in the distribution of items across topics. According to the NAEP framework, there are more items in TIMSS than in NAEP that measure *temperature and states of matter*, there are more items in NAEP than in TIMSS related to *properties and uses of materials*, *vibrations and waves as motion*, and topics related to *forces and motion* (figure 6). Compared to TIMSS, NAEP also has three times the percentage of items in topics relating to *energy and transformations*. These differences in emphasis are also reflected in the corresponding topic areas in the TIMSS framework (figure 7). The comparisons based on the TIMSS framework reveal that TIMSS has a greater emphasis than NAEP on *chemical change* and *particulate structure of matter* at the eighth grade. Also, only TIMSS includes items related to *acids and bases*, a topic not included in the NAEP framework.

Figure 6. Percentage of NAEP 2000 and TIMSS 2003 physical science items classified to physical science topics in the NAEP science framework, by survey and grade



[1] NAEP items classified by NAEP developers.

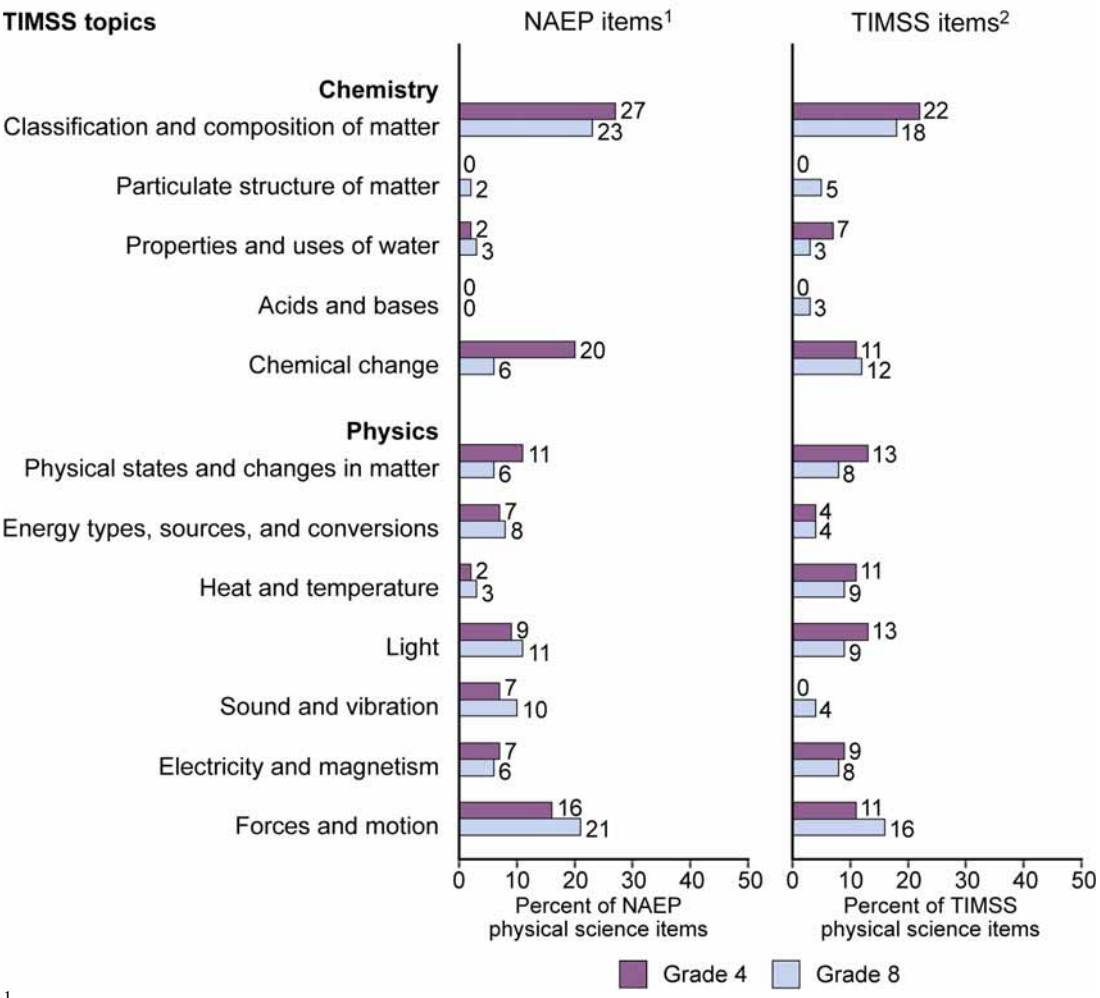
[2] TIMSS items classified by expert panel.

[3] The topic used for classification is based on the NAEP Assessment Specifications document (NAGB 1994), which combines the two framework topics related to electromagnetic radiation (*electromagnetic radiation* and *interactions of electromagnetic radiation with matter*).

NOTE: Topics may be abbreviated for graphical clarity. Two NAEP framework topics included for assessment only at the twelfth grade are not reflected in this figure, as no grade 4 or grade 8 items in either assessment were classified to these topics: *resource management* and *general wave behavior*. Percentages reflect the proportion of physical science items classified at either the topic level or the specific objective level at any grade level. Items that were classified to multiple topics were counted in all relevant topics. Bars not shown indicate that no items from that particular grade and assessment were classified to the topic.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.

Figure 7. Percentage of NAEP 2000 and TIMSS 2003 physical science items classified to chemistry and physics topics in the TIMSS science framework, by survey and grade



[1] NAEP items classified by expert panel.

[2] TIMSS items classified by TIMSS developers.

NOTE: Percentages reflect the proportion of physical science items classified at either the topic level or the specific objective level at any grade level. Items that were classified to multiple topics were counted in all relevant topics. Bars not shown indicate that no items from that particular grade and assessment were classified to the topic.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Edition*, 2003.

5.2. Life Science

TIMSS places a greater emphasis on life science at the fourth grade than does NAEP. As discussed in section 4, in NAEP, 37 percent of fourth-grade items and 36 percent of eighth-grade items were from the *field* of life science (table 3). In TIMSS, 43 percent of fourth-grade items and 29 percent of eighth-grade items were from the *content domain* of life science (table 4). The results in the life science section are based on 53 fourth-grade and 70 eighth-grade NAEP items, and 60 fourth-grade and 51 eighth-grade TIMSS items.

Framework comparison in life science

The NAEP and TIMSS frameworks for life science show some structural similarities. Exhibit 4 shows a comparison of the life science topics included in the NAEP and TIMSS science frameworks. Each includes seven topics at the fourth-grade level and eight topics at the eighth-grade level, with considerable overlap across the two frameworks at least at the topic level. Although labeled and structured somewhat differently, the topics included in both frameworks cover the broad major topic areas identified in NAEP—*change and evolution*, *cells and their functions*, *organisms*, and *ecology*. In both assessments, *cells and their functions* are not included at the fourth grade. Further examination of the specific objectives included in each framework (appendix A) reveal other differences between NAEP and TIMSS in terms of the content specified in the framework at each grade level.

One of the notable differences in the frameworks is that *human health* is an explicit topic in the TIMSS framework, whereas this does not appear to be addressed even within any of the topics in NAEP (although there is some language related to disease within the objectives under *life cycles*). Panelists noted the difficulty of mapping TIMSS *human health* items to the NAEP framework and suggested they were a special type of item that is more common in TIMSS than in NAEP. Panelists also noted the difficulty of classifying items related to the biosphere and indicated that this topic was not well specified in either framework. However, there are some specific life science objectives related to the role of organisms in the flow of energy and cycling of materials through Earth's surface in both NAEP and TIMSS topics related to *ecology* and *ecosystems*. In addition, there are some related objectives in the TIMSS environmental science content area.

Exhibit 4. Life science topics included in the NAEP 2000 and TIMSS 2003 science frameworks

NAEP	TIMSS
Change and evolution	Types, characteristics, and classification of living things
Diversity of life on Earth	Structure, function, and life processes in organisms
Genetic variation within a species	Cells and their functions (grade 8 only)
Theories of adaptation and natural selection	Development and life cycles of organisms
Changes in diversity over time (grade 12 only)	Reproduction and heredity
Cells and their functions[1]	Diversity, adaptation, and natural selection
Cells (grade 8 only)	Ecosystems
Cells as systems (grade 12 only)	Human health
Information transfer (grade 12 only)	
Energy transfer for the construction of proteins (grade 12 only)	
Communication among cells (grade 12 only)	
Organisms	
Reproduction, growth, and development	
Life cycles	
Functions and interactions of systems within organisms	
Ecology	
Interdependence of life: populations, communities, and ecosystems	

[1] Subtopics reflect those included at eighth and/or twelfth grade in the major topic of *cells and functions* in the NAEP Assessment Specifications document (NAGB 1994).
NOTE: Unless otherwise noted, all topics are intended for all grades (grades 4, 8, and 12 in NAEP and grades 4 and 8 in TIMSS). The number of grade-specific objectives and level of detail varies across topics and assessments.
SOURCE: U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progres*s, 1996, 2000; U.S. Department of Education, National Assessment Governing Board, *Science Assessment and Exercise Specifications for the 1994 National Assessment of Educational Progress, 1994*; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003.

Content and grade match in life science

Table 13 shows the level of content and grade match between NAEP and TIMSS in life science. As with physical science, NAEP and TIMSS life science items have different levels of grade match to the other's framework. Seventy-seven percent of the TIMSS fourth-grade items were classified at the fourth-grade level in the NAEP framework, and 62 percent of NAEP fourth-grade items were classified at the fourth-grade level in the TIMSS framework (table 13). On the other hand, 30 percent of NAEP fourth-grade items were considered to best match eighth-grade content according to the TIMSS framework, compared to 10 percent of TIMSS fourth-grade items classified

at the eighth-grade level in the NAEP framework. The NAEP items that were classified to the eighth-grade TIMSS framework relate almost exclusively to the topic areas of *ecosystems* and the *types, characteristics and classifications of living things*. Of the 10 percent of TIMSS items classified to the NAEP eighth-grade framework, half are in the topic of *interdependence of life*.

Most of the NAEP and TIMSS fourth-grade items that were classified at the corresponding grade level had a specific match to a life science objective in the other's framework (60 percent of NAEP items and 68 percent of TIMSS items). An additional 2 percent of NAEP items and 8 percent of TIMSS items had a general match in life science.

Notably, none of the fourth-grade life science items from either assessment were classified to other content areas, though 8 percent of NAEP and 13 percent of TIMSS items did not match topics in any area of the other's framework. Included in these are TIMSS items related to *human health* (e.g., nutritious food sources) and some NAEP items in which the life science content was considered to provide a context for the problem but was not a primary part of the knowledge that was required to answer the item (e.g., data interpretation).

Table 13. Percentage of NAEP 2000 and TIMSS 2003 fourth- and eighth-grade life science items classified to the other science assessment framework, by level of content/grade match

Level of content/grade match	Grade 4		Grade 8	
	NAEP items to TIMSS framework	TIMSS items to NAEP framework	NAEP items to TIMSS framework	TIMSS items to NAEP framework
Total number of life science items	53	60	70	51
	Percentage distribution			
Classified as same grade	62	77	71	76
Specific match[1] in life science	60	68	59	57
General match[2] in life science	2	8	10	16
Match to another content area[3]	0	0	3	4
Classified as another grade[4]	30	10	11	12
Lower grade[5]	†	†	11	0
Higher grade[6]	30	10	†	12
No classification to topics[7]	8	13	17	12

† Not applicable. Grade 4 is the lowest grade in both frameworks; grade 8 is the highest grade in the TIMSS 2003 framework

[1] Includes items that were classified to an objective at the same grade.

[2] Includes items that were classified to a topic but not to an objective at the same grade.

[3] Includes items that were classified to a topic or objective at the same grade.

[4] Includes items that were classified to a topic or objective in any content area at another grade.

[5] Includes grade 8 items classified to grade 4 topics/objectives.

[6] Includes grade 4 items classified to grade 8 topics/objectives or grade 8 TIMSS items classified to grade 12 NAEP topics/objectives.

[7] Includes items that the panel did not classify to a topic at a specific grade level.

NOTE: Data reflect the percentage of items classified by the expert panel at each level. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003; and U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.

At the eighth-grade level, 71 percent of NAEP and 76 percent of TIMSS life science items were classified at the same grade level of the other's framework. About 60 percent had a specific match to an objective in life science, and 10 percent of NAEP items and 16 percent of TIMSS items classified to the other's framework had a general match at only the topic level. Slightly more than 10 percent of both TIMSS and NAEP items were classified as a different grade level on the other's framework. The off-grade TIMSS items were classified to the NAEP twelfth-grade framework (half related to the life science topic of *interdependence of life*) and the NAEP items were classified to the TIMSS fourth-grade framework (mostly in the *ecosystem* and *structure*, *function and life processes* topics). These results, together with those above for fourth grade, indicate a lack of agreement across NAEP and TIMSS with respect to what is considered fourth-, eighth- or twelfth-grade content in the area of ecology and ecosystems. Although the content of this item may not necessarily be beyond the level of the NAEP eighth-grade life science items, the NAEP framework includes a specific objective related to the cycling of matter in ecosystems in the general topic area of *ecology* at the twelfth grade but not the eighth grade. Example 9 in appendix E shows a TIMSS eighth-grade life science item that was classified to a twelfth grade objective in the NAEP life science framework.

Seventeen percent of NAEP items and 12 percent of TIMSS items at the eighth grade have no clear match to any topic in the other's framework. As with fourth grade, these include TIMSS items in *human health* and NAEP items that are focused on data interpretations and investigations where the life science content is only a context for the item. Each assessment also has a small number of eighth-grade items that were placed in other content areas, including the Earth, environmental and physical sciences. Example 10 in appendix E shows a TIMSS eighth-grade *human health* item that was not classified to a NAEP topic.

Item distribution across life science topics

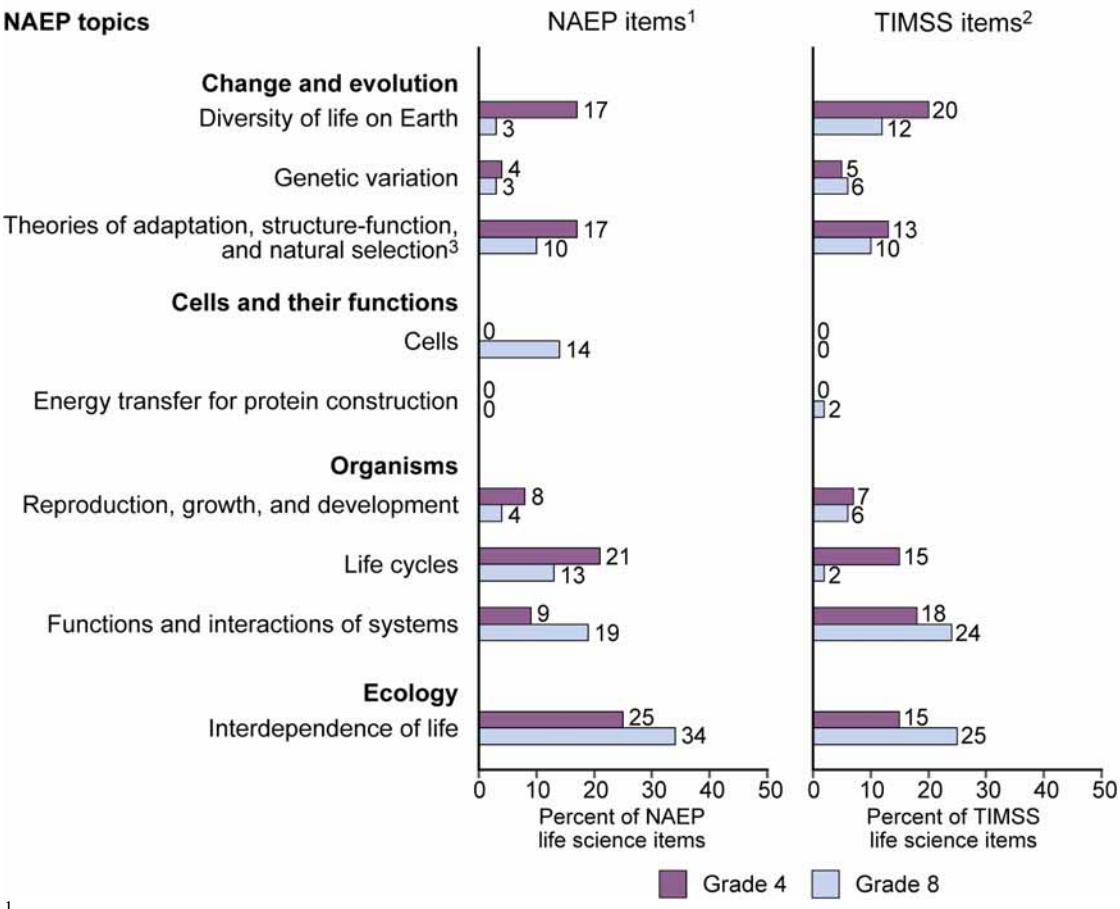
There are some similarities between the NAEP and TIMSS life science items based on their distribution across topics in the NAEP and TIMSS frameworks (see figures 8 and 9). Both NAEP and TIMSS include a larger proportion of fourth-grade than eighth-grade items covering the topics of *life cycles*; *types, characteristics and classification of living things*; and *diversity of life*. In comparison, there are higher proportions of eighth-grade items in both assessments covering the NAEP topic of *functions and interactions of systems*, and the related NAEP and TIMSS topics devoted to *ecology* and *ecosystems*. Also, neither assessment includes items related to *cells and their functions* at the fourth-grade level.

Despite these similarities, there are some important differences between the NAEP and TIMSS assessments. There is considerably more emphasis in NAEP than in TIMSS on topics related to *ecosystems* or *interdependence of life* at both fourth and eighth grades. This difference is particularly noticeable in the TIMSS framework comparisons, with more than 30 percent of NAEP items compared to less than 20 percent of TIMSS items classified to the *ecosystems* topic area at either grade level (figure 9). A NAEP eighth-grade item from the *interdependence of life* topic is shown in Example 11 in appendix E. NAEP also has more items classified to the NAEP topic related to *life cycles*, particularly at eighth grade (13 percent compared to 2 percent of TIMSS items) (figure 8). Based on the TIMSS framework, TIMSS includes relatively more items in *reproduction and heredity*. In fact, this topic (as defined by TIMSS) is addressed by 10 percent of fourth-grade items in TIMSS but none in NAEP (figure 9). In contrast, there are similar proportions of fourth-grade and eighth-grade items from each assessment classified to the closest NAEP topic of *reproduction, growth and development*. As expected, the TIMSS framework comparison shows substantially more

TIMSS items in the *human health* topic. Despite the lack of an explicit *human health* topic in the NAEP framework, NAEP still has some items devoted to *human health* as defined in TIMSS. Most of these items were from the *life cycles* topic in NAEP.

One unusual aspect of the eighth-grade comparison is the apparent discrepancy between the NAEP and TIMSS distributions related to the topic *cells and their functions*. Based on the NAEP framework, NAEP has 14 percent of items in this topic area, while TIMSS has none (figure 8). However, using the TIMSS framework as the classification system, there are relatively more TIMSS items classified to the corresponding topic than NAEP (14 percent compared to 7 percent) (figure 9). These differences are due to the differences in the definitions of what is included in the *cells and their functions* topic in the two frameworks. In TIMSS, cellular processes (respiration and photosynthesis) are included within the objectives under *cells and their functions*, while in NAEP these understandings are included in the *interdependence of life* topic. NAEP, on the other hand, includes sexual and asexual reproduction in the *cells* topic, which is addressed in the TIMSS topic of *reproduction and heredity*. Also, the basic cellular make-up of living organisms is included in the *cells and their functions* topic in TIMSS but not NAEP. In the NAEP framework, there is a twelfth-grade objective related to the cell as the fundamental unit of living organisms under the major topic of *organisms*.

Figure 8. Percentage of NAEP 2000 and TIMSS 2003 life science items classified to life science topics in the NAEP science framework, by survey and grade



[1] NAEP items classified by NAEP developers.

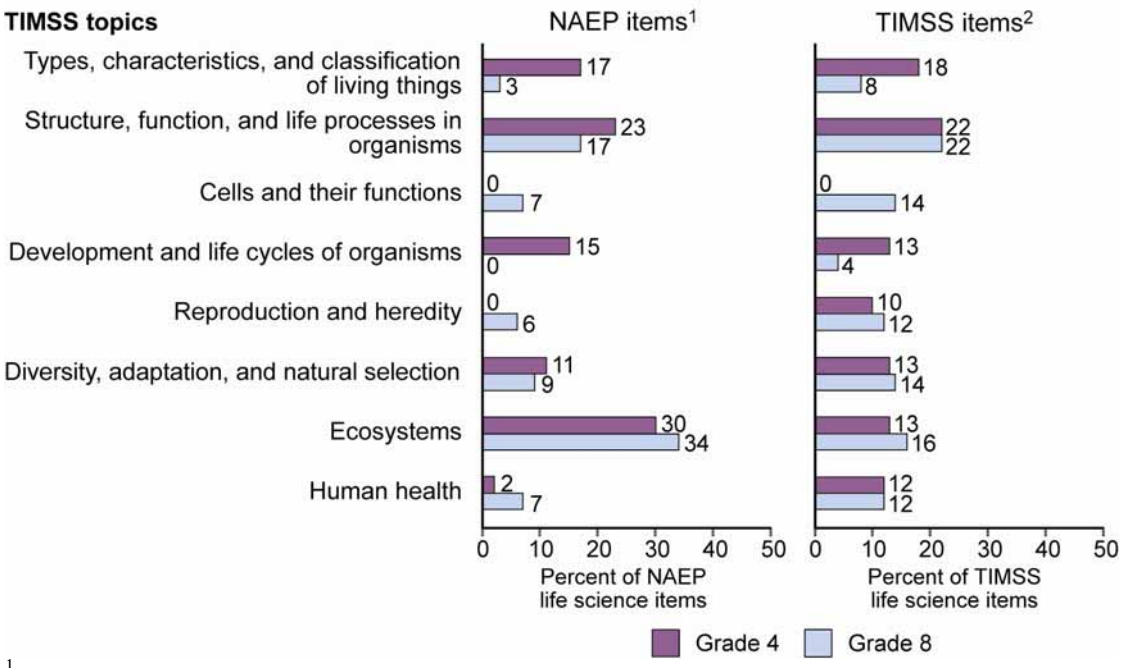
[2] TIMSS items classified by expert panel.

[3] Structure-function is included in the framework topic *of theories of adaptation and natural selection* based on subtopics in the NAEP Assessment Specifications document used for item classifications (NAGB 1994).

NOTE: Topics may be abbreviated for graphical clarity . Four NAEP framework topics included for assessment only at the twelfth grade are not reflected in this figure, as no grade 4 or grade 8 items in either assessment were classified to these topics: one from *change and evolution* (*changes in diversity over time*) and three from *cells and their functions* (*cells as systems*, *information transfer,* and *communication among cells*). Percentages reflect the proportion of life science items classified at either the topic level or the specific objective level at any grade level. Items that were classified to multiple topics were counted in all relevant topics. Bars not shown indicate that no items from that particular grade and assessment were classified to the topic.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.

Figure 9. Percentage of NAEP 2000 and TIMSS 2003 life science items classified to life science topics in the TIMSS science framework, by survey and grade



[1] NAEP items classified by expert panel.

[2] TIMSS items classified by TIMSS developers.

NOTE: Percentages reflect the proportion of life science items classified at either the topic level or the specific objective level at any grade level. Items that were classified to multiple topics were counted in all relevant topics. Bars not shown indicate that no items from that particular grade and assessment were classified to the topic.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Edition*, 2003.

5.3. Earth Science

Earth science items make up a larger proportion of the NAEP assessment than of the TIMSS assessment, at both grades. As discussed in section 4, in NAEP, 32 percent of fourth-grade items and 33 percent of eighth-grade items were from the *field* of Earth science (table 3). In TIMSS, 20 percent of fourth-grade items and 16 percent of eighth-grade items were from the *content domain* of Earth science (table 4). The results in the Earth science section are based on 46 fourth-grade and 65 eighth-grade NAEP items, and 28 fourth-grade and 29 eighth-grade TIMSS items.

Framework comparison in Earth science

The NAEP and TIMSS frameworks for Earth science are different in the breadth of coverage and the level of detail specified at the topic level (exhibit 5). The TIMSS Earth science framework has three broad topics focusing on the structure of the Earth, its processes, and its place in the solar system and universe. The NAEP framework, on the other hand, contains a larger set of topics at a more detailed level (e.g., solid earth, water, air) and also includes a separate topic on *Earth in space*. These structural differences between the two frameworks do not necessarily translate to major differences in content specified across the full set of objectives included in each. The set of topics within each sphere (lithosphere, hydrosphere, and atmosphere) in the NAEP framework contain objectives related to both structures and features and to processes. Similarly, the broad topics in the TIMSS framework contain specific objectives related to lithosphere, hydrosphere, and atmosphere. As noted in the life science section, neither NAEP nor TIMSS has an explicit topic area related to the biosphere within their Earth science frameworks.

Exhibit 5. Earth science topics included in the NAEP 2000 and TIMSS 2003 science frameworks

NAEP	TIMSS
Solid earth (lithosphere) Composition of the Earth Forces that alter Earth's surface Rocks: their formation, characteristics, and uses Soil: its changes and uses Natural resources used by humankind Forces within the Earth (grades 8 and 12 only) **Water (hydrosphere)** Water cycle Nature of the oceans and their effects on water and climate Location, distribution, and characteristics of water, and its effect and influence on human activity **Air (atmosphere)** Composition and structure of the atmosphere, including energy transfer (grades 8 and 12 only) Nature of weather Common weather hazards Air quality and climate **Earth in space** Setting of Earth in the solar system Setting and evolution of the solar system in the universe (grades 8 and 12 only) Tools and technology used to gather information about space Apparent daily motions of the Sun, Moon, planets, and stars Rotation of the Earth about its axis and the Earth's revolution around the Sun Tilt of Earth's axis that produces seasonal variations in climate Earth History: 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.	Earth's structure and physical features (lithosphere, hydrosphere, and atmosphere) Earth's processes, cycles, and history Earth in the solar system and the universe

NOTE: Unless otherwise noted, all topics are intended for all grades (grades 4, 8, and 12 in NAEP and grades 4 and 8 in TIMSS). The number of grade-specific objectives and level of detail varies across topics and assessments.
SOURCE: U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000; U.S. Department of Education, National Assessment Governing Board, *Science Assessment and Exercise Specifications for the 1994 National Assessment of Educational Progress*, 1994; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003.

Content and grade match in Earth science

Compared to the other content areas, there is less correspondence between the NAEP and TIMSS assessments with respect to what is included in Earth science, both in terms of grade level expectations and the specific content covered. At the fourth grade, the level of grade match for both NAEP and TIMSS Earth science items is not high (table 14), with less than 60 percent of Earth science items from each assessment classified to a fourth-grade topic in the other framework in Earth science or another content area. Thirty-seven percent of NAEP fourth-grade items and 25 percent of TIMSS fourth-grade items in Earth science were classified to the other assessment's eighth-grade framework, with a number of these NAEP items classified to other TIMSS content areas. Another 18 percent of TIMSS items did not have a clear match to a topic anywhere in the NAEP framework at any grade level, including some items that panelists noted were related to water but did not match the description of the NAEP topic relating to the *water cycle* very well. Example 12 in appendix E illustrates a NAEP fourth-grade Earth science item classified to an eighth-grade environmental science topic on the TIMSS science framework, while example 13 illustrates a TIMSS fourth-grade Earth science item classified at the eighth-grade level on the NAEP science framework. Example 14 presents a TIMSS fourth-grade Earth science item not classified to a topic on the NAEP science framework.

For the eighth-grade Earth science items, the patterns of grade and content match are different between the two assessments. Over 90 percent of the NAEP items were classified to the TIMSS eighth-grade framework, which is the highest percentage of NAEP items classified at the same grade level for any of the content areas. Of these, about 70 percent had a specific match to a TIMSS objective, and 14 percent were classified to topics in other content areas, primarily environmental science. Conversely, less than half of the TIMSS eighth-grade items were classified to the NAEP eighth-grade framework, which is the lowest degree of grade match of TIMSS items for any of the content areas, and only about one-third were classified to a specific objective. Similar proportions of the eighth-grade TIMSS items were classified to the fourth- and twelfth-grade NAEP framework, across multiple topics. Example 15 in appendix E shows a TIMSS eighth-grade Earth science item placed at the twelfth-grade level on the NAEP science framework.

Table 14. Percentage of NAEP 2000 and TIMSS 2003 fourth- and eighth-grade Earth science items classified to the other science assessment framework, by level of content/grade match

Level of content/grade match	Grade 4		Grade 8	
	NAEP items to TIMSS framework	TIMSS items to NAEP framework	NAEP items to TIMSS framework	TIMSS items to NAEP framework
Total number of Earth science items	46	28	65	29
	Percentage distribution			
Classified as same grade	59	57	92	48
Specific match[1] in Earth science	39	46	69	31
General match[2] in Earth science	13	11	9	14
Match to another content area[3]	7	0	14	3
Classified as another grade[4]	37	25	6	48
Lower grade[5]	†	†	6	28
Higher grade[6]	37	25	†	21
No classification to topics[7]	4	18	2	3

† Not applicable. Grade 4 is the lowest grade in both frameworks; grade 8 is the highest grade in the TIMSS 2003 framework.

[1] Includes items that were classified to an objective at the same grade.

[2] Includes items that were classified to a topic but not to an objective at the same grade.

[3] Includes items that were classified to a topic or objective at the same grade.

[4] Includes items that were classified to a topic or objective in any content area at another grade.

[5] Includes grade 8 items classified to grade 4 topics/objectives.

[6] Includes grade 4 items classified to grade 8 topics/objective or grade 8 TIMSS items classified to grade 12 NAEP topics/objectives.

[7] Includes items that the panel did not classify to a topic at a specific grade level.

NOTE: Data reflect the percentage of items classified by the expert panel at each level. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editio*n, 2003; and U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.

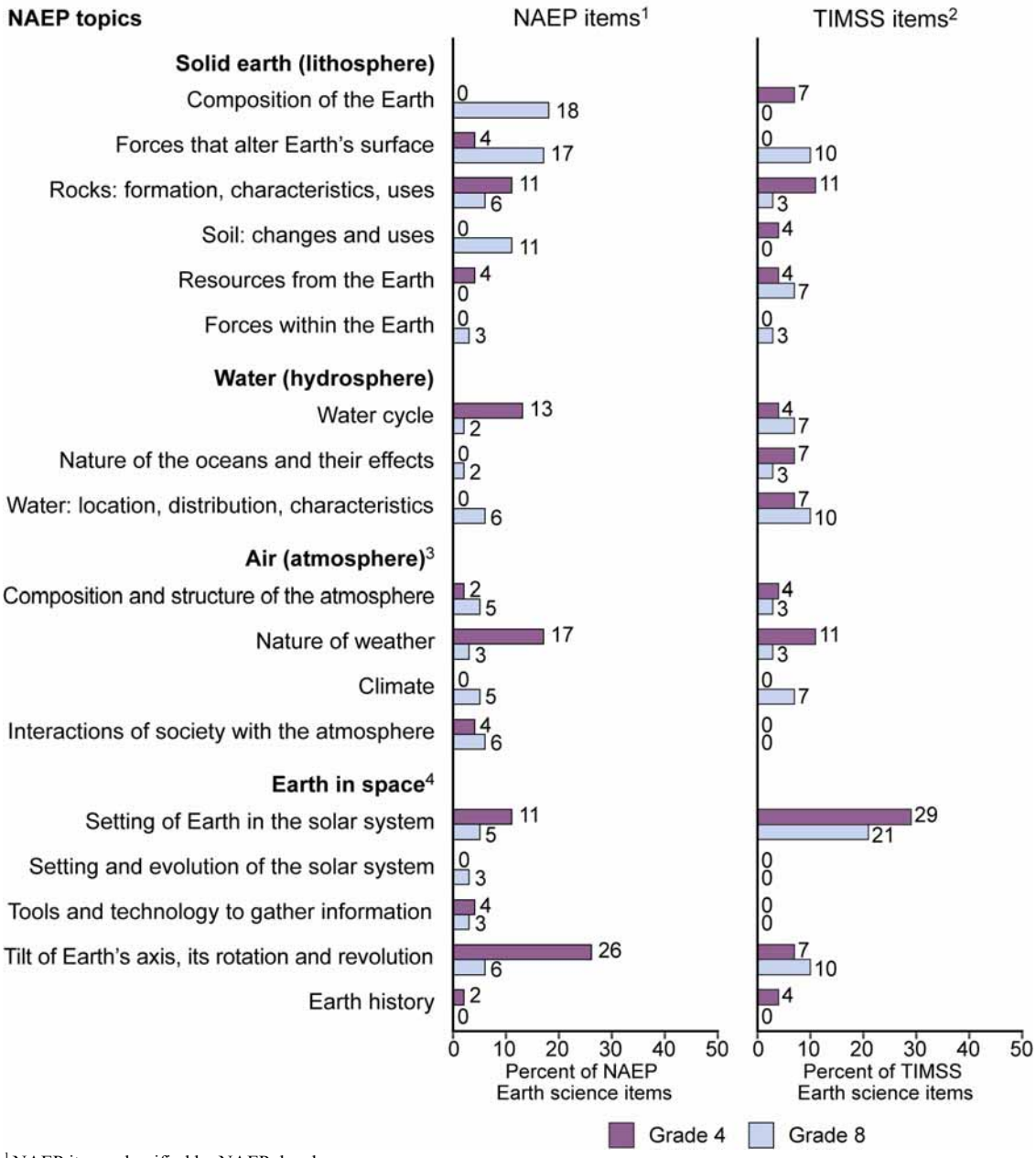
Item distribution across Earth science topics

The distribution of NAEP and TIMSS Earth science items across topics from each framework indicates that the two assessments generally emphasize different aspects of Earth science at the fourth and eighth grades (figures 10 and 11). Across the 18 topics in the NAEP Earth science framework, there are seven topics at the fourth grade and six topics at the eighth grade that are assessed by items in one assessment but not the other. In particular, items classified to NAEP topics of *soil*, *the composition of the Earth*, *the nature of the oceans and their effects*, and the *location, distribution, and characteristics of water* are included in the fourth-grade assessment in the TIMSS but not until the eighth-grade assessment in NAEP. In contrast, the topic related to *forces that alter Earth's surface* is reflected in the fourth-grade items in NAEP but only in eighth-grade items in TIMSS. Other topics related to *interactions of society with the atmosphere* and *tools and technology to gather information about space* include only NAEP items at either grade level.

Based on the NAEP framework topics, the NAEP fourth-grade assessment has relatively greater emphasis on the *water cycle*, the *nature of weather*, and the *tilt of the Earth's axis*, *its rotation and revolution* (figure 10). In contrast, TIMSS has a greater emphasis on the *setting of the Earth in the solar system*. Using the TIMSS framework as the basis for comparison, NAEP has a somewhat greater focus on *Earth in the solar system and the universe*, while TIMSS has a greater focus on *Earth's structure and physical features* (figure 11). Examples 13 and 14 in appendix E present a TIMSS fourth-grade item from each of those topics. Neither item was classified to a grade 4 topic in the NAEP framework.

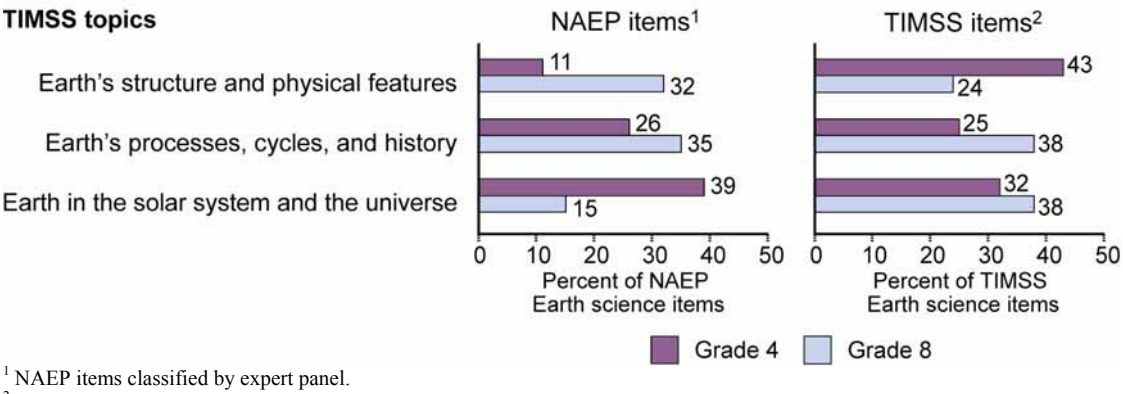
Again, there are some notable differences in the distribution of NAEP and TIMSS items across the Earth science topics in the eighth grade. While both NAEP and TIMSS have an increased emphasis on the TIMSS topic of *Earth's processes, cycles, and history* in their eighth-grade assessments compared to those at the fourth grade (figure 11), NAEP has somewhat more emphasis than TIMSS on *forces that alter the Earth's surface*, while TIMSS has relatively more on the *water cycle*, which was a focus in the NAEP fourth-grade assessment (NAEP topics in figure 10). As noted earlier, NAEP also has a number of items that pertain to the *composition of the Earth* and to *soil: its changes and uses*, topics that are not explicitly addressed by TIMSS Earth science items at the eighth grade. This is consistent with a higher proportion of the NAEP eighth-grade assessment in the TIMSS topic of *Earth's structure and physical features*. This topic area received more focus at the fourth grade in the TIMSS assessment. In comparison, TIMSS has a greater percentage of its eighth-grade assessment devoted to E*arth in the solar system* according to its own framework. Based on the NAEP framework, these topics are focused on the *setting of the Earth in the solar system* and *the tilt of Earth's axis, its rotation and revolution*, which is a focus of NAEP at the fourth grade. Based on the grade-specific objectives in the TIMSS framework, however, this topic area deals primarily with the earth/moon/sun system at the fourth grade and is substantially broadened at the eighth-grade level. Only NAEP has any eighth-grade items classified to the topic of the *setting and evolution of the solar system in the universe*.

Figure 10. Percentage of NAEP 2000 and TIMSS 2003 Earth science items classified to Earth science topics in the NAEP science framework, by survey and grade



[1] NAEP items classified by NAEP developers.
[2] TIMSS items classified by expert panel.
[3] *Air* topics listed above are slightly different from the framework version*: nature of weather* includes *common weather hazards*; the combined topic of *air quality and climate* is separated into two topics—*interactions of society with the atmosphere* and *climate*.
[4] *Earth in space* topics listed above are slightly different from the framework version: *setting of Earth in the solar system* includes *apparent daily motions of the Sun, Moon, planets, and stars*; *tilt of Earth's axis*, *its rotation and revolution* combines two topics—*rotation of the Earth about its axis and the Earth's revolution around the sun* and *tilt of earth's axis that produces seasonal variations in climate*.
NOTE: Topics may be abbreviated for graphical clarity or revised to reflect the NAEP assessment specifications document (NAGB 1994). Percentages reflect the proportion of Earth science items classified at either the topic level or the specific objective level at any grade level. Items that were classified to multiple topics were counted in all relevant topics. Bars not shown indicate that no items from that particular grade and assessment were classified to the topic.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.

Figure 11. Percentage of NAEP 2000 and TIMSS 2003 Earth science items classified to Earth science topics in the TIMSS science framework, by survey and grade



[1] NAEP items classified by expert panel.
[2] TIMSS items classified by TIMSS developers.
NOTE: Percentages reflect the proportion of Earth science items classified at either the topic level or the specific objective level at any grade level. Items that were classified to multiple topics were counted in all relevant topics.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Science Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003.

5.4. Environmental Science

Only TIMSS includes environmental science as a separate *content domain* in its framework and as a separate reporting category at the eighth-grade level. The TIMSS items that are discussed in this section for fourth grade are included in the life science and Earth science subscales for reporting purposes. This section examines the TIMSS environmental science items—7 at the fourth-grade level and 23 at the eighth-grade level.

The TIMSS environmental science framework

The TIMSS framework in environmental science focuses on three topics, including:

- changes in population (included for the eighth grade only);
- use and conservation of natural resources; and
- changes in environments.

Some of the objectives within the TIMSS environmental science topics are somewhat similar to topics and objectives included in the NAEP framework across the fields of science. These NAEP topics are predominantly in life science and Earth science, and sometimes are at the twelfth-grade level of the framework. One of the key differences, however, is that the TIMSS environmental science objectives (versus those similar objectives embedded in NAEP) focus more on the identification and possible solutions for global problems.

Content and grade match of TIMSS environmental science items to the NAEP framework

The TIMSS fourth-grade environmental science items cover topics and objectives on both the fourth- and eighth-grade NAEP framework, primarily in Earth science. Forty-three percent of the items are classified to NAEP fourth-grade topics and 29 percent to eighth-grade topics (table 15). All the items that map to the fourth-grade NAEP framework have a specific match in some content area. However, two of the TIMSS fourth-grade items (29 percent) were not classified to any NAEP topic. These items were related to human use of natural resources and the impact of human activity on the environment.

Table 15. Percentage of TIMSS 2003 fourth- and eighth-grade environmental science items classified to the NAEP 2000 science framework, by level of content/grade match

Level of content/grade match	TIMSS items to NAEP framework	
	Grade 4	Grade 8
Total number of environmental science items	7	23
	Percentage distribution	
Classified as same grade	43	48
Specific match[1] in any content area	43	22
General match[2] in any content area	0	26
Classified as another grade[3]	29	30
Lower grade[4]	†	13
Higher grade[5]	29	17
No classification to topics[6]	29	22

† Not applicable. Grade 4 is the lowest grade in the NAEP framework.

[1] Includes items that were classified to an objective at the same grade.

[2] Includes items that were classified to a topic but not to an objective at the same grade.

[3] Includes items that were classified to a topic or objective at another grade.

[4] Includes grade 8 items classified to NAEP grade 4 topics/objectives.

[5] Includes grade 4 items classified to NAEP grade 8 topics/objectives or grade 8 items classified to NAEP grade 12 topics/objectives.

[6] Includes items that the panel did not classify to a topic at a specific grade level.

NOTE: Data reflect the percentage of items classified by the expert panel at each level. Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.

Similar to the fourth-grade findings, the eighth-grade TIMSS environmental science items cross grade boundaries on the NAEP framework, with 48 percent mapping to the corresponding grade and 30 percent to the fourth- and twelfth-grade frameworks—the latter of which are primarily matched to NAEP topics in life science or physical science. The eighth-grade environmental science items show a different pattern from the fourth-grade environmental science items. The eighth-grade items are divided fairly evenly between those with a specific match and those that were classified only to the general topic level. There also are five eighth-grade items (22 percent) not classified to a NAEP topic. These items cover a range of TIMSS objectives and address both global and local environmental issues due to human or natural causes. Example 16 in appendix E illustrates a TIMSS eighth-grade environmental science item classified as a fourth-grade physical science item on the NAEP science framework.

Distribution of TIMSS environmental science items across TIMSS and NAEP science topics

When classified to their own framework, the fourth-grade TIMSS environmental science items are split four to three between topics on the *use and conservation of natural resources* and *changes in environments* (table 16). The NAEP topics to which these items map (table 17) include Earth science topics that address relationships between humans and the world—such as *oceans and their effects* and *natural resources used by humankind*—and the life science topic related to ecology (*interdependence of life*). Example 17 in appendix E presents a TIMSS fourth-grade environmental science item classified as life science on the NAEP science framework.

At the eighth-grade level, equal proportions of items (43 percent) are included in the topics covering the *use and conservation of natural resources* and *changes in environments*, and a much smaller number are in the *changes in population* topic (table 16). Almost half the TIMSS environmental science items are classified to NAEP topics in Earth science, again focused on the interaction between humans and the environment (table 17). Example 18 in appendix E presents a TIMSS eighth-grade environmental science item classified as Earth science on the NAEP framework. A number of TIMSS items at the eighth grade also were classified to life science and physical science topics in NAEP (about 20 percent in each). The items classified as physical science deal with energy sources and uses of water and some were classified at the fourth-grade level in NAEP (Example 16). The items classified in life science involve predictions related to long-term environmental effects and changes in population. A number of these items were classified at the twelfth-grade level in NAEP.

Table 16. Percentage distribution of TIMSS 2003 environmental science items across environmental science topics in the TIMSS science framework, by grade

Environmental science topic	TIMSS items to TIMSS framework	
	Grade 4	Grade 8
Changes in population	0	13
Use and conservation of natural resources	57	43
Changes in environments	43	43

NOTE: Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd Editi*on, 2003.

Table 17. Percentage of TIMSS 2003 environmental science items classified to topics in the NAEP 2000 science framework in each field of science, by grade

NAEP topic within each field of science	TIMSS items to NAEP framework	
	Grade 4	Grade 8
Earth science		
Soil: its changes and uses	14	4
Natural resources used by humankind	14	4
Nature of the oceans and their effects	14	0
Water: location, distribution, characteristics	14	22
Interactions of society with atmosphere	0	17
Life science		
Interdependence of life	14	22
Physical science		
Temperature and states of matter	0	4
Energy sources and uses of water	0	13

NOTE: Percentages reflect the proportion of TIMSS environmental science items classified by the expert panel at either the objective level (specific match) or topic level (general match) at any grade level. Items that were classified to multiple topics were counted in all relevant topics. Two fourth-grade items and five eighth-grade items that were not classified at the topic level are not included.
SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and U.S. Department of Education, National Assessment Governing Board, *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*, 1996, 2000.

While the NAEP framework does not include a separate content category related to environmental science, some of the NAEP items from other *fields of science* were found by the panel to best match this part of the TIMSS framework (3 percent at grade 4 and 7 percent at grade 8, as shown in table 4). These NAEP items (14 items at eighth grade and 5 items at fourth grade) are not reflected in the percentage of items across TIMSS topics shown in the content area sections for Earth science, life science and physical science. The fourth-grade items all came from Earth science and cover topics related to human use of natural resources as well as the impact of both human activity and natural events on the environment. In contrast, the eighth-grade items came from all three of the NAEP *fields of science* and covered topics related to human use of natural resources and impact on the environment, energy resources, and genetic engineering. All of these NAEP items at both grades were classified to environmental topics in the TIMSS framework at the eighth-grade level. One of the NAEP items classified as environmental science is shown in example 12 in appendix E.

This section provided a comparison of the two assessments in each of the main content areas of physical science, life science, Earth science, and environmental science. The last section includes a summary and conclusion of the findings of this comparison study.

6. Conclusion

The content comparisons between NAEP and TIMSS reveal some key differences in the science topics covered, grade-level correspondence, and the characteristics of the item pools on other dimensions. All of these factors together may result in differences in student performance, and it is important to consider these differences when interpreting the results from the different assessments. Differences in the science content included in each assessment can be seen at both the framework level and in the pool of items developed based on these frameworks. Even in content areas where there is considerable overlap of the frameworks, such as in life science and Earth science, a closer examination of the topics and specific objectives covered by the items in each assessment reveals some important differences as well as similarities. In comparison to NAEP, whose framework was developed in the context of the U.S. system, the TIMSS framework reflects a consensus across many countries. Some of the differences in curricula across these countries are reflected in the frameworks and the differences in content between the two assessments. In particular, the inclusion in TIMSS of separate content areas in chemistry, physics, and environmental science results in broader topic coverage in some areas. While there is a considerable overlap in the topics included in some content areas, the items included in each assessment place different emphases at the topic level. As a result, both NAEP and TIMSS assessments may each contribute more information in some areas as well as some unique components to the larger picture of what students at fourth and eighth grades know and can do in science. In addition, the "hands-on" tasks in NAEP provide further complementary information to the pencil-and-paper portions of both assessments, enabling the measurement of student performance in this area of knowing and doing science.

References

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

Beaton, A., Martin, M., Mullis, I., Gonzalez, E., Smith, T., and Kelly, D. (1996). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. International Association for the Evaluation of Educational Achievement (IEA). Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Binkley, M., and Kelly, D. (2003). *A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments* (NCES 2003–10). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Binkley, M., and Rust, K. (Eds.). (1994). *Reading Literacy in the United States: Technical Report of the U.S. Component of the IEA Reading Literacy Study* (NCES 94–259). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Martin, M., Mullis, E., Beaton, A., Gonzalez, E., Smith, T., and Kelly, D. (1997). *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study.* International Association for the Evaluation of Educational Achievement (IEA). Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Martin, M., Mullis I., Gonzalez, E., Gregory, K., Smith, T., Chrostowski, S., Garden, R., and O'Connor, K. (2000). *TIMSS 1999 International Science Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. International Association for the Evaluation of Educational Achievement (IEA). Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J., and O'Connor, K.M. (2003). *TIMSS Assessment Frameworks and Specifications 2003, 2nd Edition*. Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.

National Assessment Governing Board (NAGB). (1994). *Science Assessment and Exercise Specifications for the 1994 National Assessment of Educational Progress.* Washington, DC: Author.

National Assessment Governing Board (NAGB). (2000). *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*. Washington, DC: Author.

National Research Council (NRC). (1996). *National Science Education Standards*. Washington, DC: National Academy Press.

Neidorf, T.S., Binkley, M., Gattis, K., and Nohara, D. (forthcoming). *Comparing Mathematics Content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments* (NCES 2006-029). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Nohara, D. (2001). *A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Program for International Student Assessment (PISA)* (NCES 2001–07). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Robitaille, D. (Ed.). (1997). *National Context for Mathematics and Science Education: An Encyclopedia of the Educational Systems Participating in TIMSS.* Vancouver, BC: Pacific Educational Press.