NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

U.S. Department of Education
Institute of Education Sciences
NCES 2006–029

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

Technical Report

May 2006

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

Marilyn Binkley
National Center for Education Statistics

Kim Gattis
NAEP Education Statistics Services Institute
American Institutes for Research

David Nohara
Independent Consultant

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

May 2006

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

Suggested Citation
Neidorf, T.S., Binkley, M., Gattis, K., and Nohara, D. (2006). *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. 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 in international studies of student achievement. 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 three mathematics assessments conducted in 2003: the NAEP fourth- and eighth-grade assessments; the Trends in International Mathematics and Science Study (TIMSS), which also assessed mathematics at the fourth- and eighth-grade levels; and the Program for International Student Assessment (PISA), which assessed the mathematical literacy of 15-year-old students. Its aim is to provide information useful for interpreting and comparing the results from the three assessments, based on an in-depth look at the content of the respective frameworks[1] 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 three assessments on various dimensions.[2] The frameworks were compared with respect to

- how each assessment organizes and defines the mathematics content and process skills to be assessed at each grade (or age) level;
- the main content areas included and the set of topics covered in each; and
- other aspects, such as item format and calculator policy.

Item comparisons were based on

- cross-classification of NAEP and TIMSS items to each other's assessment framework in terms of the mathematics content covered and grade-level expectations;
- classification of PISA items to the NAEP framework on these same dimensions;
- classification of all items with respect to their level of *mathematical complexity*;[3] and
- comparisons based on other framework dimensions related to cognitive processes, item formats, and item contexts.

[1] 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.
[2] The panel members—experts in mathematics, mathematics education, and mathematics assessment, with familiarity and experience across the three assessments—are listed in appendix C.
[3] *Mathematical complexity* reflects the demands on thinking that an item makes, assuming that a student is familiar with the mathematics content of the task. The classifications in this report are based on the definitions in the NAEP 2005 framework for three levels of *mathematical complexity*—low, moderate, and high—that form an ordered description of the demands an item may make on a student (described in appendix B).

While comparisons between NAEP, TIMSS, and PISA were focused on the common classification systems based on the NAEP framework, the study also included a limited comparison between PISA items and NAEP eighth- and twelfth-grade *problem solving* items in light of the dimensions in the PISA framework. Example items are referenced throughout the report to illustrate some key similarities and differences.[4]

The results of this study indicate that although the NAEP, TIMSS, and PISA 2003 mathematics frameworks address many similar topics and require students to use a range of cognitive skills and processes, it cannot be assumed that they measure the same content in the same way. A hypothetical student who takes all three assessments might indeed perform equally well on them, but depending on the curriculum they have been exposed to and their skill and experience in various types of mathematical thinking, other students might exhibit quite different levels of performance across the three assessments. For NAEP and TIMSS, this is also true within each of the five corresponding content areas related to *number*, *measurement*, *geometry*, *data*, and *algebra*.

At the overall level, there is apparent agreement between the NAEP and TIMSS frameworks on the general boundaries and basic organization of mathematics content across the fourth and eighth grades, with nearly all items from each assessment being placed in one of the major content areas of the other assessment framework at the broadest level. Furthermore, both NAEP and TIMSS place similar emphases on each of the five major content areas, as evidenced by similar distributions of items across the main content areas of both frameworks at both the fourth- and eighth-grade levels. These types of comparisons, however, do not consider the grade level correspondence or the level of content match based on the distribution of items across the specific set of topics and subtopics included at each grade level in each of the assessments.

Despite the similarity between NAEP and TIMSS at the broadest content area level, there are differences between the two assessments when considering more detailed comparisons of the mathematics content covered and the grade level correspondence between the items in each assessment and the intentions of the other assessment framework. Differences between the NAEP and TIMSS assessments emerge with more detailed content analyses that consider the level of content match to specific topics and subtopics in the other assessment framework, with 20 percent of fourth-grade and about 15 percent of eighth-grade items from both assessments not classified to specific subtopics in the other assessment framework at any grade level. This finding indicates that both assessments contain items that might not be included in the other assessment and supports the general claim that NAEP and TIMSS do not necessarily assess the same mathematics content.

Most NAEP and TIMSS items were placed at the same grade on the other assessment framework, but this was not always within the corresponding content area. The overall grade-level correspondence between the NAEP items and the TIMSS framework, 86 percent at fourth grade and 73 percent at eighth grade, was lower than that between the TIMSS items and the NAEP framework (at least 90 percent). This is related at least in part to the inclusion of cross-grade items in NAEP that were administered at multiple grade levels. There are notable differences across content areas in the level of grade match between the two assessments. In the TIMSS assessment, *measurement* and *geometry* account for most of the items classified at different grade levels to the NAEP framework (10 percent or more). In the NAEP assessment, the content area of *data analysis, statistics and probability* has the largest percentage of fourth-grade items classified at a higher grade level (almost

[4] Additional released item sets from each assessment are also available on the NAEP, TIMSS, and PISA websites: http://nces.ed.gov/nationsreportcard, http://isc.bc.edu/timss2003.html, and http://www.pisa.oecd.org.

half), most notably from items covering probability topics which are not assessed in TIMSS until the eighth grade. In the NAEP eighth-grade assessment, between 10 and 43 percent of items in all content areas were classified at the fourth-grade level in TIMSS, with the largest proportion of items in *measurement* and *geometry and spatial sense* (37 and 43 percent, respectively) and the smallest in *data analysis, statistics, and probability* (10 percent).

Within each content area, detailed comparisons of content coverage and grade correspondence indicate that NAEP and TIMSS items are not necessarily measuring the same content. Some of the differences in topic and subtopic emphasis or grade level match in each content area include the following:

- *Number*: At the eighth grade, TIMSS has a relatively larger emphasis on ratio, proportion, and percent compared to whole numbers in NAEP. There is a somewhat greater emphasis on computation in TIMSS at both fourth and eighth grades. Only the NAEP eighth-grade assessment includes scientific notation (data not shown).

- *Measurement*: A larger proportion of NAEP fourth-grade items involve the selection and use of appropriate measurement instruments and units. While TIMSS has a greater emphasis at both grades on problems involving properties (area, perimeter, volume, surface area) of two- and three-dimensional shapes, a number of fourth-grade TIMSS items were classified to the NAEP eighth-grade framework (16 percent). In each assessment, at least 25 percent of eighth-grade items was classified at the lower grade level of the other assessment. In addition, there is an overlap of NAEP *measurement* items with topics in the TIMSS *geometry* framework.

- *Geometry*: A larger proportion of NAEP items involve two- and three-dimensional shapes, while TIMSS has a greater emphasis on congruence and similarity. There are differences in the nature of problem-solving items (TIMSS with more application of geometric properties and NAEP with more use of geometric models). Forty-three percent of NAEP eighth-grade items were classified to the TIMSS fourth-grade framework, while 13 percent of TIMSS eighth-grade items were classified to the NAEP twelfth-grade framework.

- *Data*: NAEP includes probability items in the fourth-grade assessment, while TIMSS does not include this topic until the eighth grade. In TIMSS, there is a greater emphasis on reading and interpreting data in tables and graphs at the fourth grade. In NAEP, there is a higher proportion of eighth-grade items involving the organization and display of data.

- *Algebra*: TIMSS has a greater emphasis on algebraic expressions and operations at eighth grade. Some of the eighth-grade NAEP items involving patterns, equations, and functions were classified to the fourth-grade TIMSS framework (18 percent). There is an overlap of NAEP *algebra and functions* topics involving the use of number lines and coordinate systems with the TIMSS *geometry* framework, and some TIMSS eighth-grade items were classified to the NAEP twelfth-grade framework (5 percent).

NAEP and TIMSS appear to be quite similar overall in terms of the distribution of items across the low, moderate, and high *mathematical complexity* levels. Sixty-four percent of fourth-grade items and more than half of eighth-grade items were classified at the low complexity level and less than 5 percent were classified at the high complexity level at both grade levels in NAEP and

TIMSS. The content areas with the highest proportion of items (more than 60 percent) classified at the moderate or high complexity level are *algebra and functions* in fourth-grade NAEP, *data analysis, statistics, and probability* in eighth-grade NAEP, and *measurement* in eighth-grade TIMSS.

PISA stands apart from NAEP and TIMSS in a number of important areas, including the organization of its mathematics content framework (which is based on *overarching ideas*), its focus on problem solving in real-world applications, and the fact that it samples students based on age (15-year-olds) rather than grade level. Interestingly, PISA items, which are distinct from NAEP and TIMSS items in numerous ways, do have a relatively high degree of content match to NAEP subtopics from a purely mathematics content perspective (more than 90 percent classified to a NAEP subtopic). Grade-level analyses based on classifications to the NAEP content framework also indicate that although the target population of PISA is somewhat older than the students taking the NAEP and TIMSS eighth-grade assessments, the mathematics content of most of the PISA items (85 percent) are at the eighth grade level.

The different nature of PISA makes it complementary to both NAEP and TIMSS. The mathematics topics addressed may not necessarily be substantially different, although PISA places greater emphasis on data analysis and less on algebra than do either NAEP or TIMSS, but it is in how that content is presented that makes PISA different. In terms of item type and level of *mathematical complexity*, PISA is quite different from NAEP and TIMSS. Not only does PISA use multiple-choice items to a far lesser degree, but it also contains a substantially higher proportion of items (71 percent) classified at the two upper levels of *mathematical complexity* (moderate and high).

Differences in the demands that the problem-solving items place on students' mathematical thinking skills are also found when comparing PISA items and NAEP eighth- and twelfth-grade *problem solving* items with respect to the PISA *competency clusters*.[5] From the perspective of the PISA framework, the mathematical thinking skills required of the NAEP *problem solving* items are focused more on *reproduction* and much less on *reflection* than PISA. This is consistent with their different purposes—NAEP being more closely aligned with curriculum-based mathematics outcomes at fourth, eighth and twelfth grades and PISA assessing the preparedness of 15-year-olds to be able to apply mathematics to solve novel, real-world problems. The *situations or contexts*[6] involved in the NAEP *problem solving* items also differed from PISA, with NAEP having a relatively higher proportion of items focused on *educational/occupational* and *scientific* contexts and lower proportions involving *personal* and *public* contexts than PISA. A number of the NAEP *problem solving* items investigated were judged by the panel as not appropriate for the PISA assessment (due to contexts or mathematical applications that were not authentic) or requiring revisions related to the level of instructions, general formatting, and sequencing in order to be included in the PISA assessment.

This report illustrates the complementary nature of the assessments, as there are certainly cases, especially looking within content areas, where results from NAEP, TIMSS, or PISA might be more informative than the others regarding a specific topic or skill. However, as scores are not reported at the topic or subtopic level, the ability to use assessment results to make statements about these student skills or abilities is limited to performance on individual items.

[5] The PISA framework defines three *competency clusters*—*reproduction*, *connections*, and *reflection*—to describe the mathematical cognitive processes required by its mathematics items. These are described in section 2.3.
[6] The PISA framework includes a situations or contexts dimension with four categories—*personal*, *educational/occupational*, *public*, and *scientific*.

For all three assessments, when reviewing results, it is important to look beyond the overall scores and content area subscales and examine in detail what each assessment measures. This study has yielded data that can be used to make informed readings of results. While there is no single factor that may be related to differences in student performance, the numerous differences noted here, whether dramatic or more minor, may have a substantial effect overall. As each assessment program continues, this type of research can continue, not only to help explain differences in student scores, but also to understand the complementary nature of the three assessments.

This report provides a first-level comparison of items in each assessment in terms of the coverage of broad content areas and distribution across mathematics topics as defined in the frameworks. All items in each assessment were considered in order to make overall comparisons of content coverage and grade-level expectations as well as distributions with respect to three broad levels of mathematical complexity. In addition, the types of item classifications conducted within the time constraints of this study permit comparisons at the mathematics topic level for each content area. While this method provides a broad view of some of the similarities and differences between the assessments, it is limited in terms of the types of comparisons that are provided at the item level. More in depth analyses of the exact nature of the items from each assessment within topics would reveal other important differences related to difficulty, scope, depth, complexity, and other item attributes. These types of more focused comparisons were outside the scope of this study, but may be important to include in future comparative studies of the assessments.

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 Shelley Burns, 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 Lisa Bridges of the Institute of Education Sciences, Christine Brusselmans-Dehairs of Ghent University, Jeff Haberstroh of the Educational Testing Service, Marcie Dingle of Synergy Enterprises, 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, Aaron Douglas, J. Lane Glenn, Jamie Johnston, Juan Carlos Guzmán, Dana Kelly, Janine Emerson, David Miller, Melanie Ouellette, Lisette Partelow, Anindita Sen, and Margaret Woodworth. The authors wish to thank ESSI technical reviewers, Lauren Gilbertson and Zeyu Xu, for their hard work and expertise. Thanks also to David Miller in helping to shepherd the report through the final stages of the review and publication process, 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), as well as the Program for International Student Assessment (PISA) and the Organization for Economic Cooperation and Development (OECD). 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

Researchers, policymakers, educators, and members of the general public interested in the achievement of U.S. students currently have available several major sources of national-level data: results from the U.S. Department of Education's National Assessment of Educational Progress (NAEP) and U.S. results from various international assessments, such as the Progress in International Reading Literacy Study (PIRLS), the Program for International Student Assessment (PISA), and the Trends in International Mathematics and Science Study (TIMSS). NAEP administers periodic assessments in reading, mathematics, science, and other subjects at the fourth, eighth, and twelfth grades; TIMSS assesses mathematics and science at fourth and eighth grade; and PIRLS is a reading literacy assessment administered to fourth-grade students. In comparison, PISA primarily assesses the literacy[1] of 15-year-old students in reading, mathematics and science. In cases where the different assessments address the same subject areas (e.g., mathematics, reading, science) at the same or similar grade levels, the opportunity exists to measure U.S. student achievement using multiple instruments. Comparing results across assessments can be useful not only for interpreting the results, but also for developing a more complete picture of student achievement than would be possible with the results of just one assessment.

In order to provide useful guidance for comparing the results of the different assessments, the U.S. Department of Education's National Center for Education Statistics (NCES) has periodically conducted studies comparing various assessments in terms of their underlying frameworks, items, and other related features. In 2003, NCES conducted two comparison studies—one in mathematics and one in science—following the 2003 administrations of TIMSS and PISA. This report focuses on a comparison of the mathematics assessments—NAEP 2003, TIMSS 2003, and PISA 2003—while a companion report (Neidorf, Binkley, and Stephens 2006) compares the NAEP 2000 and TIMSS 2003 science assessments.

The 2003 mathematics and science comparison studies build on several earlier studies, which were also 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.

Previous published studies of mathematics and science assessments included comparisons of the TIMSS 1995 and NAEP 1996 mathematics assessments (McLaughlin, Dossey, and Stancavage 1997) and the NAEP 2000, TIMSS 1999, and PISA 2000 mathematics and science assessments (Nohara 2001). Both studies compared 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. 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, reading passages and items from the NAEP 2002 and PIRLS 2001 reading assessments.

[1] PISA uses the terminology of "literacy" in each subject area to denote its broad focus on application of knowledge and skills; that is, PISA seeks to ask if the 15-year-olds are mathematically literate, or to what extent they can apply mathematical knowledge and skills to a range of different situations they may encounter in their lives.

The goal of this mathematics comparison study is to identify similarities and differences between the 2003 NAEP, TIMSS, and PISA assessments based on a detailed comparison of their frameworks and items. This information may be used to help inform interpretations of student performance in mathematics on the three 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 main questions driving the study are as follows:

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

- How do NAEP, TIMSS, and PISA define the content and process skills appropriate for the assessments at different grade or age levels? How do the items in each assessment compare to the grade-level expectations specified by the other frameworks?[2]

- How do the items in the NAEP, TIMSS, and PISA assessments compare with respect to the level of *mathematical complexity* demanded of students?

- How do NAEP, TIMSS, and PISA compare with respect to the types and distribution of item formats used? How do the items in the different assessments compare in terms of their problem-solving contexts?

To answer these questions, NCES convened an expert panel (appendix C) to examine the mathematics frameworks and items for each assessment. The panel cross-classified NAEP and TIMSS fourth- and eighth-grade items to each other's assessment frameworks with respect to mathematics content and grade level. PISA items also were classified to the NAEP framework on the same dimensions. The panel classified the items from all three assessments with respect to a common definition of *mathematical complexity* level based on the NAEP 2005 framework.[3] A limited comparison was also made between PISA items and NAEP eighth- and twelfth-grade *problem solving* items. Although TIMSS and PISA were not compared directly, this approach permits the comparison of NAEP, TIMSS, and PISA through the common classification systems based on the NAEP framework. In addition to the classification data from the panel, the study draws upon information provided by the NAEP, TIMSS, and PISA 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.

---

[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.

[3] The rationale for using this dimension from the NAEP 2005 framework as the basis for item classifications is described in appendix D.

Section 2 of this report presents an overview of the NAEP, TIMSS, and PISA assessments and a comparison of their respective mathematics assessment frameworks. Section 3 reviews the methods used for this comparison study. The results of the study are then presented in three major sections. The first, section 4, compares the NAEP, TIMSS, and PISA assessments overall with respect to content coverage, grade level, *mathematical complexity* level, and item format. The overall comparisons are followed by comparisons of the NAEP and TIMSS assessments with respect to each of the following main content areas (section 5): *number*, *measurement*, *geometry*, *data*, and *algebra*. This section provides more detailed comparisons of the extent to which items in one assessment map to the mathematics framework of the other assessment. It compares the content distribution of the items for each of the NAEP and TIMSS mathematics subscales. Section 6 contains additional comparisons made between the NAEP and PISA assessments, including detail on the mathematics topics covered by the PISA items and how NAEP eighth- and twelfth-grade *problem solving* items compare to those included in the PISA assessment. The report concludes with a summary of key findings (section 7).

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 well 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 assessments. 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 assessments conducted in 2003 at the fourth- and eighth-grade levels and in 2000 at the twelfth-grade level.[1]

The frameworks established by NAGB for all the NAEP subject areas, including mathematics, are based on the collaborative input of a wide range of experts and involvement by participants from government, education, business, and public sectors. They 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 at different grades and in different subject areas.

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 elementary 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, and is directed by the International Study Center at Boston College. TIMSS 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 4-year cycle, with the first administration in 1995 (at fourth, eighth, and twelfth grades),[2] the second in 1999 (at eighth grade only), and the most recent in 2003 (at fourth and eighth grades), with about 50 countries participating.

Like NAEP, the TIMSS assessments are based on collaboratively developed frameworks. In contrast to NAEP, however, the framework development and consensus process involved mathematics experts, education professionals, and measurement specialists from many countries.

---

[1] At the time this study was conducted, NAEP 2000 was the most recent mathematics assessment at grade 12, and NAEP 2003 (which did not include grade 12) was the most recent mathematics assessment at grades 4 and 8. NAEP long-term trend assessments in mathematics were also administered in 2003–04 but were not included in this study. Later, in 2005, NAEP conducted a mathematics assessment at fourth, eighth, and twelfth grades.

[2] 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.

PISA

The Program for International Student Assessment (PISA) is conducted by the Organization for Economic Cooperation and Development (OECD). The main objective of PISA is to provide regular, policy-relevant data on the "yield" of education systems, and so targets students at an age that is near the end of compulsory schooling in most countries (15-year-olds). PISA focuses on literacy—the ability to use and apply knowledge and skills to real-world situations encountered in adult life—in the key subject areas of reading, mathematics, and science. PISA is, thus, the United States' source of comparative information on the reading, mathematical, and scientific literacy skills of students in the upper grades, and it provides benchmarks to international performance levels based on other OECD countries. The frameworks guiding the PISA assessments reflect a consensus across the OECD countries regarding the skills and abilities that demonstrate literacy in these key areas.

A key design feature of PISA is its cycle of rotating emphasis among the three key assessment areas every three years. Each subject area is assessed in each data collection, but the design distinguishes between major and minor domains. When a subject is the major domain it comprises a relatively greater share of the total assessment time, with a larger number of items and an assessment framework that is more fully developed and updated. Reading literacy was the major domain in the first PISA assessment in 2000 (32 countries), mathematical literacy was the major domain in the most recent 2003 assessment (41 countries), and scientific literacy will be the major domain in the next assessment in 2006.[3]

Organization of the NAEP, TIMSS, and PISA 2003 Mathematics Frameworks

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. Exhibits 1-A, 1-B, and 1-C compare schematically the organizing dimensions in the NAEP, TIMSS, and PISA 2003 mathematics frameworks. These organizing dimensions provide the basic framework for the development of the pool of items in each assessment, and the frameworks include target percentages for the distribution of the assessments across the main categories in each dimension to ensure a balanced assessment (discussed in the following sections).[4] As seen in these exhibits, there are some basic organizational differences between the frameworks, especially between PISA and NAEP or TIMSS.

Both the NAEP and TIMSS 2003 mathematics frameworks represented in exhibits 1-A and 1-B are based on two main organizing dimensions—a content dimension and a cognitive dimension—as well as an overarching dimension (along the bottom) that defines processes that go across the content and cognitive categories. Both NAEP and TIMSS include five similarly labeled categories in the content dimension (*content strands* in NAEP and *content domains* in TIMSS) that correspond to major mathematics curricular areas related to *number*, *measurement*, *geometry*, *data*, and *algebra*. In the main cognitive dimensions (*mathematical abilities* in NAEP and *cognitive domains* in TIMSS), NAEP has three broad categories (*conceptual understanding*, *procedural knowledge*, and *problem solving*), while TIMSS has four (*knowing facts and procedures*, *using concepts*, *solving routine problems*, and *reasoning*). There is overlap between the categories defined in the cognitive dimensions in NAEP and TIMSS as well as the processes defined by the overarching dimensions in each assessment (*mathematical power* in NAEP and *communicating mathematically* in

---

[3] The 2003 PISA assessment also included an additional component assessing cross-disciplinary problem solving. Items from this separate component were not included in this comparison study.
[4] The frameworks only provide target percentages of items or assessment time as guidelines for test development.

TIMSS). All items developed for NAEP and TIMSS are classified with respect to which categories in the two main dimensions they assess. The overarching dimensions are also considered as items are developed.

In contrast to NAEP and TIMSS, the PISA mathematical literacy assessment framework includes three main dimensions as shown in exhibit 1-C. Like NAEP and TIMSS, there is one dimension related to mathematics content (*overarching ideas*); the four *overarching ideas* in PISA, however, do not directly correspond to the main content categories in NAEP and TIMSS. Also like NAEP and TIMSS, PISA includes a cognitive dimension (*competency clusters*). In addition to these two dimensions, the PISA framework includes a third main dimension related to the *situations or contexts* in which the application of mathematics concepts is required. The *situations or contexts* dimension does not have an analogue in the NAEP and TIMSS frameworks. All items developed for PISA are classified with respect to the main categories in each of its three dimensions.

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

Exhibit 1-A. NAEP mathematics framework dimensions: 2003

Content strands	Mathematical abilities
Number sense, properties, and operations Measurement Geometry and spatial sense Data analysis, statistics, and probability Algebra and functions	Procedural knowledge Conceptual understanding Problem solving
Mathematical power (Reasoning, connections, communication)	

NOTE: The NAEP framework is based on two main organizing dimensions—*content strands* and *mathematical abilities*—as well as an overarching dimension (*mathematical power*) that defines processes that go across the content and abilities categories.
SOURCE: U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002.

Exhibit 1-B. TIMSS mathematics framework dimensions: 2003

Content domains	Cognitive domains
Number Measurement Geometry Data Algebra	Knowing facts and procedures Using concepts Solving routine problems Reasoning
Communicating mathematically	

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

Exhibit 1-C. PISA mathematical literacy framework dimensions: 2003

Overarching ideas	Competency clusters	Situations or contexts
Change and relationship Quantity Space and shape Uncertainty	Reproduction Connections Reflection	Personal Educational/occupational Public Scientific

NOTE: The PISA framework is based on three main organizing dimensions—*overarching ideas* (content), *competency clusters*, and *situations or contexts*.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*, 2003.

2.1. NAEP 2003 Mathematics Framework

The framework for the NAEP 2003 mathematics assessment is based on two major organizing dimensions—*content strands* and *mathematical abilities*—as well as an overarching dimension of *mathematical power* (exhibit 1-A).[5] The framework stipulates that every item in the assessment is given a primary classification in the two major dimensions according to certain distribution targets. The NAEP 2003 framework allows secondary classification of items to more than one content category. However, NAEP does not use multidimensional scaling, and these secondary classifications are not used in the analysis of results.

The first major dimension is defined by five broad *content strands*, which are the same for fourth, eighth, and twelfth grades. They are *number sense, properties, and operations*; *measurement*; *geometry and spatial sense*; *data analysis*, *statistics, and probability*; and *algebra and functions*. Each *content strand* is further defined by topics and subtopics.[6] The framework document indicates which topics and subtopics are intended for each grade level (4, 8, and 12). While many are intended for all three grades (4, 8, and 12), there are some topics and subtopics that should be assessed only at a specified grade level(s). In particular, a number of topics and subtopics are not intended to be assessed either at grade 4 or at grade 8. Others may only be introduced at a simple level at a lower grade(s) (such as by using a manipulative or pictorial model).

The NAEP 2003 framework specifies target percentages for the distribution of items across the *content strands.* As shown in table 1, the framework specifies that a large proportion of the fourth-grade assessment be devoted to *number sense, properties, and operations* (at least 40 percent) and that moderate emphasis be placed on *measurement* (20 percent)*, geometry and spatial sense* (15 percent) and *algebra and functions* (15 percent). The least emphasis at fourth grade is placed on *data analysis, statistics, and probability* (10 percent). At the higher grades there is a more even distribution across the *content strands*. The greatest emphasis is placed on *algebra and functions* for both eighth and twelfth grades (25 percent). At eighth grade, 25 percent of the assessment is also focused on *number sense, properties, and operations*.

[5] The mathematics framework for the NAEP 2003 assessment was developed in the early 1990s and was used as the basis for the assessments in 1996, 2000, and 2003 (NAGB 2002). The framework was updated for the 2005 assessment (NAGB 2004).
[6] See section 5 and appendix A for more information about the topics and subtopics included in the NAEP content framework.

Table 1. Target percentage of NAEP items distributed across NAEP framework dimensions, by grade: 2003

NAEP framework dimensions	Grade 4	Grade 8	Grade 12
Content strand			
Number sense, properties, and operations	40	25	20
Measurement	20	15	15
Geometry and spatial sense	15	20	20
Data analysis, statistics, and probability	10	15	20
Algebra and functions	15	25	25
Mathematical abilities			
Conceptual understanding	33	33	33
Procedural knowledge	33	33	33
Problem solving	33	33	33

NOTE: Percentages reflect the minimum target percentages specified in the NAEP 2003 mathematics framework. At all grades, distributions across *mathematical abilities* are approximately equal. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002.

The framework also specifies cognitive abilities to ensure a balanced assessment that demonstrates mathematical thinking in various situations, which are described by a combination of *mathematical abilities* and *mathematical power*.[7] The cognitive dimension of *mathematical abilities* addresses the aspects of knowing and doing mathematics and is defined by three broad categories. These three categories are *conceptual understanding*, *procedural knowledge*, and *problem solving* and the framework specifies that there should be approximately equal emphasis on each of the *mathematical abilities*.

Mathematical power is defined as consisting of three overall cognitive processes in mathematics. These processes are *reasoning*, *connections*, and *communication*. *Mathematical power* as conceived in the framework reflects cognitive processes within a broader context of reasoning and with connections across the scope of mathematical content and thinking. Communication is a unifying thread, and the framework emphasizes the inclusion of extended constructed-response items as a way for students to provide meaningful responses to mathematics tasks and demonstrate their ability to communicate mathematically. While *mathematical power* is to be used as a foundation when developing items in each of the content and cognitive dimensions, no target percentages are specified for the individual process categories in this dimension of the framework.

Approximately one-third of the NAEP 2003 mathematics assessment comprises items for which students are permitted to use calculators.[8] The NAEP framework specifies that calculators be provided for the assessment. These are four-function calculators at fourth grade and scientific calculators at eighth and twelfth grades. The NAEP framework also specifies that manipulatives (e.g., rulers, protractors, and geometric shapes) be used for some portion of tasks in the mathematics assessment.

[7] The *mathematical abilities* and *mathematical power* dimensions are replaced in the NAEP 2005 framework with a single system of three levels of *mathematical complexity.* As discussed in forthcoming sections, the 2005 *levels of mathematical complexity* were used in this study as a means of comparing the cognitive demands of items across assessments.
[8] Calculators are only available to students while they are working on item blocks designed for use with calculators.

The NAEP framework specifies that multiple-choice, short-answer, and extended-response items be included in the assessment. The framework used for the 1996, 2000, and 2003 assessments does not specify exact proportions to be devoted to each of these item types, but does indicate an increasing emphasis on extended-response items and more balance between short-answer and multiple-choice items than in previous frameworks.

2.2. TIMSS 2003 Mathematics Framework

The TIMSS 2003 framework is based on two main organizing dimensions, *content domains* and *cognitive domains*, as well as an overarching dimension of *communicating mathematically* (exhibit 1-B).[9] All TIMSS items are classified with respect to *content domain* and *cognitive domain*.[10] There are five broad *content domains* assessed at both fourth and eighth grades (Mullis et al. 2003). These include *number*, *measurement*, *geometry*, *data*, and *algebra*. Within the *content domains*, the TIMSS framework further specifies main topic areas and grade-specific objectives within those topics that are appropriate for assessment at each grade.[11]

The TIMSS 2003 framework specifies target percentages for the distribution of assessment time across the content domains (table 2). The framework emphasizes the *number content domain* at fourth grade (40 percent of assessment time), with *measurement* as the domain with the next highest level of emphasis (20 percent). Since algebra is not taught as a formal subject in primary school across TIMSS countries, the *algebra* content area (patterns, equations and relationships) represents a relatively low proportion of the fourth grade assessment (15 percent). *Geometry* reflects the same proportion (15 percent). The *content domain* with the least emphasis at the fourth grade is *data* (10 percent). At eighth grade the largest percentage of assessment time is still *number* (30 percent), but there is an increased coverage of *algebra* (25 percent) and an equal distribution of assessment time across the other *content domains* of *measurement*, *geometry* and *data* (15 percent to each).

[9] The TIMSS mathematics 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.
[10] While items developed for TIMSS may address more than one category, only primary classifications were used during test development to ensure framework coverage.
[11] See section 5 and appendix A for more information about the topics and objectives included in the TIMSS framework.

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

TIMSS framework dimensions	Grade 4	Grade 8
Content domain		
Number	40	30
Measurement	20	15
Geometry	15	15
Data	10	15
Algebra	15	25
Cognitive domains		
Knowing facts and procedures	20	15
Using concepts	20	20
Solving routine problems	40	40
Reasoning	20	25

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 ed.,* 2003.

On the cognitive dimension, TIMSS specifies four broad *cognitive domains* to describe the range of skills and abilities that students apply in responding to items in the assessment. These include *knowing facts and procedures, using concepts, solving routine problems,* and *reasoning.* Whereas the definitions of the *cognitive domains* are the same for both fourth and eighth grades, the distribution of assessment time specified in the framework differs somewhat across grades. Both fourth- and eighth-grade assessments include the same emphasis on *solving routine problems* (40 percent) and *using concepts* (20 percent) (table 2). The two grades differ, however, with respect to the distribution across the *knowing facts and procedures* and *reasoning* categories. At the fourth grade, equal emphasis is placed on each of these (20 percent), while at eighth grade a greater emphasis is placed on *reasoning* (25 percent) than *knowing facts and procedures* (15 percent).

The TIMSS framework also specifies *communicating mathematically* as an overarching dimension that is to be demonstrated through description and explanation. Adequate levels of constructed-response items at both grades are included to measure students' ability to communicate mathematically across a wide range of content and processes. Some portion of the items in each of the content and cognitive categories measure abilities related to the overarching dimension of *communicating mathematically*, but a target percentage is not specified in the framework.

TIMSS permitted the use of calculators at the eighth grade for the first time in the 2003 assessment. Students in the eighth grade were permitted to use calculators on about half of the assessment at the discretion of each participating country.[12] Because calculators were not permitted in the previous TIMSS assessments, calculators were permitted only for the new items in 2003 and not the trend items carried over from the 1995 and 1999 assessments. Calculators were not permitted at fourth grade in any TIMSS assessment. The TIMSS assessment also includes some extended problem-solving and inquiry tasks that involve the use of manipulatives such as rulers or geometric shapes.

[12] Calculators can be used only during the second half of the testing session. TIMSS does not provide calculators, but students may use their own calculators or those provided by their schools. In the United States, simple-function calculators were provided to eighth-grade students for the TIMSS 2003 assessment.

The TIMSS 2003 framework also specifies that both multiple choice and constructed-response items (requiring students to provide a written response) 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 require a short answer, while the other third require a more extended response.

2.3. PISA 2003 Mathematical Literacy Framework

The PISA 2003 mathematical literacy framework includes three major dimensions related to mathematical content (*overarching ideas*), mathematical cognitive processes (*competency clusters*) and *situations or contexts* (exhibit 1-C).[13] All items in PISA are classified with respect to each of these three main dimensions, and the framework includes specifications for the distribution of the assessment across the categories in each (table 3).

The mathematics content to be assessed in PISA is defined by four *overarching ideas*, which include *quantity*, *space and shape*, *change and relationships*, and *uncertainty*. Collectively, they are intended to cover most of the mathematics domains that students are typically exposed to through their mathematics school curriculum, but these particular conceptual areas also were chosen because they encompass a set of phenomena and concepts within a broad range of situations that students are likely to encounter outside of school. *Quantity* includes the topics of number sense, meaning of operations, mental arithmetic, and estimation. *Space and shape* covers recognizing shapes and patterns, understanding dynamic changes to shapes, similarities and differences, and 2- and 3-dimensional representations and relationships between them. *Change and relationships* refers to functional thinking and covers different types of growth (i.e., linear, exponential, periodic, logistic) and the relationships between them. *Uncertainty* includes such topics as data collection, analysis, and representation; probability; and inference.

In the process dimension, PISA defines mathematical competencies important for mathematical literacy and describes the cognitive activities that these competencies encompass according to three *competency clusters* including *reproduction, connections,* and *reflection*. For each *competency cluster*, the framework addresses the abilities and skills associated with eight competencies: thinking and reasoning; argumentation; communication; modeling; problem posing and solving; representation; using symbolic, formal, and technical language and operations; and use of aids and tools.

The competencies demanded by items in the *reproduction* cluster involve "reproduction of practiced knowledge" and may require students to perform routine operations. The *connections* cluster involves the demonstration of competencies in problem-solving situations that are not routine but still familiar. Items in this cluster usually require evidence of integration of different mathematical concepts or making connections across *overarching ideas*. The competencies in the *reflection* cluster require students to plan solution strategies and implement them. Items in this cluster contain more elements and involve settings that are more "original" (less familiar) than in the other two categories. The framework specifies that approximately half of the assessment should be devoted to problems measuring competencies in the *connections* cluster and the remaining portion should be divided equally between the *reproduction* and *reflection* clusters.

[13] The same basic framework was used as the basis for the assessment of mathematical literacy in 2000. However, the framework was updated and expanded in 2003 when mathematical literacy became the major domain (OECD 2003).

Table 3. Target percentage of PISA assessment distributed across PISA framework dimensions: 2003

PISA framework dimensions	
Overarching ideas	
Change and relationships	25
Quantity	25
Space and shape	25
Uncertainty	25
Competency clusters	
Reproduction	25
Connections	50
Reflection	25
Situations or contexts	
Personal	25
Educational/occupational	25
Public	25
Scientific	25

NOTE: Percentages reflect the targets specified in the PISA 2003 framework. The PISA framework specifies that assessment time be distributed as evenly as possible across the four *overarching ideas* and across the four *situations or contexts*. The target percentages across the three *competency clusters* reflect the proportion of items in each cluster specified by the PISA framework.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*, 2003.

The last dimension, *situations or contexts*, is an important aspect of the PISA framework. PISA places a heavy emphasis on authentic contexts for the use of mathematics and tasks that might be encountered in real-world situations. Four types of *situations or contexts* are defined and used in developing the problems in PISA. These include *personal*, *educational/occupational*, *public*, and *scientific*.

These *situations or contexts* types are based on a model of "distance" from the students' individual world. The *personal* category reflects situations that are within the immediate realm of students' personal experience and interest. The *educational/occupational* category includes problems encountered in students' school or work life involving the application of mathematics. Items in the *public* category involve problem-solving situations that students might encounter in the local community or as a functioning member of society at large. The *scientific* category involves more hypothetical scenarios or scientific applications of mathematics. The PISA framework specifies that the assessment should be balanced with respect to the proportion of problems from each of these types of *situations or contexts*.

The PISA assessment is organized into a set of tasks designed to be authentic problem-solving *situations or contexts* that involve the application of mathematics concepts from the *overarching ideas* and embody the mathematical processes in the *competency clusters*. In general, the tasks include some stimulus information, an introduction, and a series of related items, although there also are some individual questions. PISA places the decision to use calculators at the discretion of participating countries, with the intention of mirroring common instructional practices.[14] The framework specifies a range of format types, including multiple-choice, closed constructed-response, and open constructed-response, and that about equal numbers of each of these item types be included in the mathematical literacy assessment.[15]

2.4. Comparing the NAEP, TIMSS, and PISA Mathematics Frameworks and Assessments

There are a number of similarities between the frameworks and general design aspects of all three assessments. All three frameworks encompass a broad range of mathematics content knowledge and skills and include classification systems for describing the cognitive skills students use to respond to items. There are nevertheless differences, especially between the framework of PISA and the frameworks of NAEP and TIMSS. Whereas the NAEP and TIMSS frameworks are tied closely to the organizational structures used in traditional school curricula, the PISA framework is based on a situation- and phenomena-based approach that results in noticeably different categories for describing content. Both NAEP and TIMSS are grade-based assessments, with NAEP and TIMSS assessing fourth- and eighth- grade students and NAEP also assessing twelfth grade students.[16] In comparison, PISA defines an age-based population of students in secondary school (15-year-olds). Although the target populations for the eighth-grade NAEP and TIMSS assessments and the twelfth-grade NAEP assessment are reasonably close in age to PISA's target population of 15-year-olds, there is still a difference of roughly two grade levels. Furthermore, even though mathematical literacy was the major domain in the 2003 administration of PISA, it nevertheless included a much smaller number of items than either NAEP or TIMSS, meaning that assessment items may not as fully reflect the breadth and depth of the framework.[17] PISA is also unique in its extensive use of sets of two or more items based on a common stimulus. Although both NAEP and TIMSS do include such item sets, they are not as numerous as in PISA.

The NAEP and TIMSS frameworks are quite similar with respect to the broad structure of their content dimensions, but there are differences at the finer levels. Both are organized into five major areas of mathematics related to *number*, *measurement*, *geometry*, *data*, and *algebra*, although the terminology used for these content areas are not exactly the same and the content dimensions are defined somewhat differently based on the topics included in each. In particular, NAEP and TIMSS differ in terms of how they specify what is to be assessed at each grade level. The NAEP framework includes a list of topics and subtopics within each *content strand* and indicates which are intended to be included at each grade level (grades 4, 8, and 12). Although this indicates grade-level appropriateness of each topic and subtopic, the topics and subtopics themselves are the same for all grades in which they are included. As a result, there are few grade-specific subtopics specified in the NAEP framework. A separate NAEP assessment specifications document provides illustrative examples of appropriate items that might be developed for each topic at each grade level for use by

---

[14] In PISA in the United States, the decision to use calculators was left to schools based on school, district, or state policy.
[15] Although only three basic item formats are described in the PISA framework (multiple-choice, closed constructed-response, and open constructed-response), two additional types (complex multiple-choice and short response) were included in the item classification information provided by the assessment developer. These item types are discussed further in section 4.4.
[16] Although TIMSS was administered at the 12th 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.
[17] As shown in section 4, there are 85 items in the PISA 2003 mathematics literacy assessment compared to at least 180 items in the eighth grade assessments in NAEP 2003 and TIMSS 2003 mathematics assessments.

the assessment developers (NAGB 1992). In contrast, the TIMSS framework includes a list of main topics within each *content domain* and grade-specific assessment objectives for each of the main topics. In most cases, the set of specific objectives in TIMSS is unique for each grade level (grades 4 and 8).

When making direct comparisons related to item content, this report uses a general terminology of content area, topic, and subtopic to refer to the comparable levels of specification used in the NAEP and TIMSS content framework (exhibit 2). For the discussion of content, cognitive or other classifications based on a single framework (NAEP, TIMSS or PISA), the terminology from that framework is used.

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

NAEP framework		**General terminology**		**TIMSS framework**
Content strand	⇐	Content area	⇒	Content domain
Topic	⇐	Topic	⇒	Topic area
Subtopic	⇐	Subtopic	⇒	Objective

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

NAEP		**TIMSS**	
Content strand:	Data analysis, statistics, and probability	Content domain:	Data
Topic:	Read, interpret, and make predictions using tables and graphs	Topic area:	Data interpretation
Subtopic:	Read and interpret data	Objective:	Draw conclusions from data displays

SOURCE: U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.,* 2003.

Since the purpose of PISA is to measure literacy rather than specific educational outcomes that are closely linked to curriculum-based content areas, its frameworks are structured rather differently than those of NAEP or TIMSS. First, PISA defines the mathematics content in terms of broad *overarching ideas* that go across the content areas defined in NAEP and TIMSS. Also, PISA includes a dimension not found in NAEP or TIMSS related to the *situation or context* of the item and emphasizes the use of “authentic” tasks and the application of mathematics to solve real-world problems. While NAEP and TIMSS do not have the same focus as PISA, there is still considerable language in the NAEP and TIMSS assessment frameworks about the application of mathematical knowledge and skills to problem-solving situations.

All three assessments structure their cognitive dimensions differently, but there is considerable overlap in the specific process skills, abilities, and competencies that are deemed important to be included in each assessment to demonstrate performance in mathematics. In particular, reasoning and communication are explicitly emphasized in all three assessment frameworks in the main cognitive or overarching dimensions. Making mathematical connections is also emphasized as one of the categories within the NAEP *mathematical power* dimension and PISA *competency clusters*. While mathematical connections is not an explicit category in TIMSS, it is included in the abilities to be demonstrated by items assessing the *reasoning* cognitive domain.

NAEP, TIMSS, and PISA all include both multiple-choice items, in which students choose the correct answer from a list of four or five choices, and constructed-response items, in which students generate their own answers. All frameworks allow for both short-answer and more extended constructed-response items, but the exact definition of these may vary across assessments. Some differences in item formats are discussed in the overall results section (section 4.4). The TIMSS framework specifies that one-third or more of assessment time be devoted to constructed-response items, and that about one-third of these items require an extended response. The PISA framework specifies an approximately equal distribution of items across the main format types of multiple-choice, closed constructed-response, and open constructed-response. The NAEP framework does not provide specific targets for proportions of items but emphasizes the importance of extended constructed-response items and a balance between multiple-choice and short-answer items.

All three assessment frameworks outline policies for calculator use. The NAEP 2003 framework permits calculators to be used by students in fourth, eighth, and twelfth grades on some portion of mathematics items (about one-third). In TIMSS, calculators were not allowed during the fourth-grade assessment. However, beginning with 2003, calculators were permitted but not required for newly developed eighth-grade assessment materials. In both TIMSS and PISA, participating countries could decide whether or not students were allowed to use calculators. In NAEP, calculators are provided—four-function calculators in fourth grade and scientific calculators in eighth and twelfth grades.[18] In TIMSS, the United States allowed students to use simple-function calculators that were provided with the test. In PISA, in the United States, the decision to allow calculators was left to schools based on school, district, or state policy.

Both the NAEP and TIMSS frameworks include the use of manipulatives (e.g., rulers, cardboard geometric shapes) in some tasks, and the assessments include small numbers of items involving them. In PISA, items using manipulatives are neither specified in the framework nor reflected in the assessment.

The assessment designs for NAEP, TIMSS, and PISA result in each individual student taking only a portion of the total assessment items, but the testing time for individual students differs across the three assessments. NAEP requires 50 minutes at all three grades, TIMSS requires 72 minutes at fourth grade and 90 minutes at eighth grade, and PISA requires two hours of testing time.

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 intended to provide descriptions of what students should know and be able to do in mathematics at each grade level from a national perspective. In contrast, the TIMSS and PISA frameworks reflect a consensus across diverse participating countries about the mathematics content and processes that should be assessed. For TIMSS, the framework reflects a consensus about what mathematics topics are most appropriate and important to assess at fourth and eighth grade; in general, the topics included are in the curricula for a majority of TIMSS countries. The PISA framework reflects a consensus across the OECD

[18] This NAEP policy is being revised for the 2005 assessment in which twelfth grade students will be permitted to use their own calculators.

countries about what knowledge, skills, and abilities reflect mathematical literacy and preparedness for adult life. Some of the differences in mathematics curricula and emphases across countries are reflected in differences between the frameworks and the items included in the assessments. The results presented in the following sections that compare the assessments overall and in each of the mathematics content areas provide some information on these possible differences.

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

3. Process and Methods

To conduct comparisons of the NAEP, TIMSS, and PISA 2003 assessments, NCES convened a panel of 11 experts in mathematics, mathematics education, and mathematics assessment. All panel members had familiarity and experience with at least one of the three assessments and their frameworks.[19] The panel met over a 3-day period to review the frameworks and classify the items from each assessment. The following three sections describe the organization of the expert panel meeting and the methods used for making the NAEP/TIMSS and NAEP/PISA 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, TIMSS, and PISA frameworks and assessments, 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.

The first two days of the expert panel meeting were devoted to NAEP/TIMSS comparisons. All of the NAEP and TIMSS fourth and eighth grade mathematics items were reviewed, reflecting a total of about 650 items across the two assessments and grades. The items were divided into three groups according to content, with each group containing items from both NAEP and TIMSS in the content areas of[20]

- *number*;
- *measurement* and *geometry*; and
- *data* and *algebra*.

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 to make sure that each group contained participants with expertise in each of the assessments. This division of items and panelists ensured a balance across the groups with respect to the coverage of assessments and grades as well as the number of items to be reviewed.

NAEP/PISA comparisons were conducted on the third day of the meeting. These comparisons involved a subset of the full panel, including participants with expertise in both NAEP and PISA and representatives from each of the original content area groups. After an initial orientation and plenary discussion of the PISA framework, the panel was divided into two groups to review and classify items. One group focused on reviewing the 85 PISA mathematical literacy items and the other group reviewed a set of 79 NAEP problem-solving items from the eighth- and twelfth-grade mathematics assessments.[21]

---

[19] A list of panel members and associated staff is presented in appendix C.
[20] The division of items was based on the assessment developers' classifications by content area subscale.
[21] The NAEP items selected for comparison with PISA were the items from the 2003 eighth-grade assessment and the 2000 twelfth-grade assessment that were extended constructed-response and/or were classified in the *problem solving* category of *mathematical abilities*.

Both components of the study 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 describes primarily 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 these portions of the content frameworks and to uncover some of the main 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:[22]

- **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 subtopic 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.[23]

- ***Mathematical complexity* level** (low, moderate, or high): All items were classified with respect to *mathematical complexity* level as defined in the NAEP 2005 framework. Items in the *low complexity* category rely heavily on recall and recognition of previously learned concepts and principles. They may require students to carry out a mechanical or stated procedure or recognize an example of a concept. Items in the *moderate complexity* category involve more flexibility of thinking and choice among alternatives and often require more than a single step. These items include those that require students to use informal methods of reasoning and problem-solving strategies such as comparing figures or statements. In the *high complexity* category, items make heavy demands on students to engage in more abstract reasoning, planning, analysis, judgment, and creative thought. Items may require students to generalize a pattern or to describe, compare, and contrast solution methods.[24]

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

[22] Additional information about the content categories and definitions of levels of mathematical complexity is provided in appendixes A and B.
[23] The grade-level classifications are based on the content descriptions in the NAEP and TIMSS frameworks and in the NAEP specifications document (provided to panelists for reference). The classifications reflect the judgment of this particular expert panel and their knowledge of the NAEP assessment.
[24] A more complete description of the levels of *mathematical complexity* is included in appendix B. The considerations in selecting this dimension from the NAEP 2005 framework are discussed in appendix D.

- Items should be classified to the most detailed content level possible—ideally, to the subtopic level. (Although panelists were allowed to make some logical inferences about what a content area, topic, or subtopic might include, they were instructed not to classify items further than they believed was 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 a *geometry* item may be in the *measurement* content area of the other framework).

- In cases where items appear to address multiple content areas, topics, or subtopics, a primary classification for the item should be identified whenever possible. (In cases where this was not appropriate, panelists indicated multiple or secondary classifications which were recorded. 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 indicated. These instances may indicate a potential gap in the framework to which the item is 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 subtopic might include at a given grade level).[25]

- Classifications to *mathematical complexity* level should be based strictly on the descriptions found in the NAEP 2005 framework and not on more general conceptions of complexity.

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. To monitor consistency in the classifications of *mathematical complexity* level 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 this dimension 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

---

[25] Since the TIMSS framework contains grade-specific objectives, the grade-level classification is concurrent with a classification to an objective. For items classified to a topic or content domain, but not 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. Since the NAEP framework provides a set of topics and subtopics that usually apply to more than one grade, the grade classification reflects a judgment of grade-level correspondence in terms of the panelists' knowledge of the NAEP assessment. To assist in this process, the panel also consulted the grade-specific item examples in the NAEP assessment specifications document (NAGB 1992).

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, rationales for the classifications, and judgments about whether items exceeded the grade-level descriptions in the framework. 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.

3.3. Methods Used for NAEP/PISA Comparisons

As noted in the previous section, the NAEP/PISA comparisons were accomplished by two subgroups of the panelists retained for this component of the comparison study conducted on the third day of the expert panel meeting.

The first group reviewed and classified the 85 PISA items to the NAEP framework for content, grade level, and level of *mathematical complexity*. This group included representatives from each of the previous content groups to ensure a consistent method of classification as that used for the TIMSS items for any part of the NAEP content framework.

The second group classified 79 NAEP *problem solving* and extended-response items from the eighth- and twelfth-grade assessments. These items were selected because of their potential alignment with PISA's emphasis on problem solving. They were drawn from both the eighth-grade and twelfth-grade assessments since PISA's 15-year old target population falls between these two grade levels. The group classified the set of NAEP items with respect to the main dimensions of the PISA framework, including the following:

- *overarching idea*;
- *competency cluster*; and
- *situation or context*.

A set of example items from the PISA framework were used to illustrate the classification procedures. The group also commented on whether or not the NAEP items might appear on the PISA assessment and, if not, how they were different from PISA items that might assess comparable mathematics content and processes.

This section reviewed the methods used for this comparison study. The next section compares the assessments overall with respect to content coverage, grade level, levels of mathematic complexity, 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, *mathematical complexity* level, and item format.

4.1. Content Coverage

Tables 4 and 5 compare the distribution of items from each assessment across the main content areas as defined 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.[26] Generally speaking, the three assessments appear to share similar boundaries for the definition of mathematics content, with nearly all items from each assessment classified as being consistent with the definitions of the five content areas in both the NAEP and TIMSS frameworks. There were only a few items (2 percent of NAEP items at the fourth grade, 1 percent of NAEP items at the eighth grade, and 1 percent of TIMSS items at the eighth grade) that were not classified by the panel at the broad content area level to the other assessment's framework. These classifications, however, do not consider grade level correspondence, which is discussed in the next section.

Using the five *strands* of the NAEP framework to compare NAEP and TIMSS at both the fourth- and eighth-grade levels, the two assessments have very similar distributions of items across the five strands (table 4). At the fourth-grade level on both assessments, the highest percentage of items was classified to *number sense, properties, and operations* (40 and 42 percent, respectively) and the area with the lowest percent was *data analysis, statistics, and probability* (10 percent for both assessments). A similar pattern is true at the eighth-grade level as well, although the distribution of items was more balanced. The largest difference between the two assessments at eighth grade is in *number sense, properties, and operations*, where one-third of TIMSS items were classified compared to about one-quarter of NAEP items.

The distribution of NAEP and TIMSS items across the five *content domains* of the TIMSS framework is very similar to that based on the corresponding content areas of the NAEP framework, giving at least partial support to the idea that the broad content areas defined in the NAEP and TIMSS framework are similar (table 5). While the NAEP and TIMSS assessments appear to place similar emphases on these broad content areas, there are substantial differences between the assessments noted when the item content is examined more closely, as discussed in the following sections.

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

Table 4. Percentage of NAEP, TIMSS, and PISA mathematics items classified to the content strands in the NAEP 2003 mathematics framework, by grade/age and survey: 2003

NAEP content strand	Grade 4 NAEP[1]	Grade 4 TIMSS[2]	Grade 8 NAEP[1]	Grade 8 TIMSS[2]	15-year-olds PISA[2]
Total number of items	181	145	197	180	85
			Percentage distribution		
Number sense, properties, and operations	42	40	26	33	22
Measurement	18	19	15	14	18
Geometry and spatial sense	15	15	19	18	12
Data analysis, statistics, and probability	10	10	15	11	40
Algebra and functions	14	15	25	23	11
Classified to multiple strands	0	0	0	0	2
Not classified to a content strand	0	0	0	1	0

[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 (*content strand*, *topic*, or *subtopic*). Multi-part items were treated as one item for classification purposes and only contribute one to the total. Items classified to multiple *content strands* were counted in each relevant category. Two PISA items were classified to multiple *content strands*: one to *measurement* and *data analysis, statistics, and probability* and one to *algebra and functions* and *geometry and spatial sense*. One TIMSS *data* item at eighth grade was not classified to a NAEP *content strand*. Detail may not sum to totals because of rounding, omitted items, or items classified to multiple *content strands*.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment; and U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress,* 2002.

Table 5. Percentage of NAEP and TIMSS mathematics items classified to the content domains in the TIMSS 2003 mathematics framework, by grade and survey: 2003

TIMSS content domain	Grade 4 NAEP[1]	Grade 4 TIMSS[2]	Grade 8 NAEP[1]	Grade 8 TIMSS[2]
Total number of items	181	145	197	180
		Percentage distribution		
Number	44	38	29	31
Measurement	16	21	12	16
Geometry	17	14	24	17
Data	10	10	15	13
Algebra	12	16	19	23
Not classified to a content domain	2	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. Five NAEP items were not classified to *content domains* on the TIMSS 2003 framework. These items included three fourth-grade items in *algebra*, *measurement*, and *geometry* and two eighth-grade items in *algebra*. 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) 2003 Mathematics 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 ed*., 2003.

PISA has a different balance across the main content areas than NAEP or TIMSS, with a much higher percentage of items classified to *data analysis, statistics, and probability* on the NAEP framework (40 percent, compared to 15 and 11 percent on the NAEP and TIMSS eighth-grade assessments, respectively). PISA also has relatively fewer items classified to *algebra and functions* (11 percent, compared to about one-quarter for NAEP and TIMSS at the eighth grade) and *number sense, properties, and operations* (22 percent, compared to 26 and 33 percent for the eighth-grade NAEP and TIMSS assessments, respectively).

At the broad content area level, virtually all NAEP, TIMSS, and PISA items were classified as being consistent with the basic definitions of the content areas in the other assessments when the grade level correspondence to the other framework is not considered. At the topic level, there was also a high level of content match for all three assessments (table 6). For all assessments (and grades), at least 95 percent of items were classified at the topic level on the other assessment frameworks. However, at the finest level of classification (either the subtopic level or the topic level, when no subtopics existed), the match between items and frameworks was not as universal. Still, classification at this level was relatively high, with about 80 percent of fourth-grade items and 85 percent of eighth-grade items for both NAEP and TIMSS being classified at the subtopic level on the other assessment frameworks. The level of content match for the PISA items was even higher, with 92 percent of items classified at the subtopic level to the NAEP framework. These findings indicate that there is a generally high level of agreement between the three assessments regarding the general definitions of the mathematics content areas; the level of agreement decreases at more specific levels of classification.

Table 6. Percentage of NAEP, TIMSS, and PISA mathematics items classified to other assessment framework at the topic or subtopic level, by grade/age and survey: 2003

	Grade 4		Grade 8		15-year-olds
Level of content classification	NAEP items to TIMSS framework	TIMSS items to NAEP framework	NAEP items to TIMSS framework	TIMSS items to NAEP framework	PISA items to NAEP framework
Topic level	96	98	98	99	95
Subtopic level[1]	80	80	84	85	92

[1]Includes items classified to a subtopic or to a topic, when no subtopics exist in the NAEP framework.
NOTE: Data reflect the percentage of items that were classified by the expert panel to the topic and subtopic levels of the other assessment framework in any content area at any grade. Items classified to multiple topics or subtopics 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) 2003 Mathematics Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment; U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress,* 2002; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed*., 2003.

4.2. Grade Level

The cross-classification data were used to examine the extent to which items from each assessment map to the other assessment frameworks at corresponding grade levels. Figures 1-A and 1-B show the percentage of items in the NAEP, TIMSS, and PISA assessments overall that were classified at each grade level in the other assessment frameworks. For these overall comparisons, the percentages classified at each grade level of the other assessment frameworks reflect items that were classified at the subtopic, topic, or broad content area level.[27]

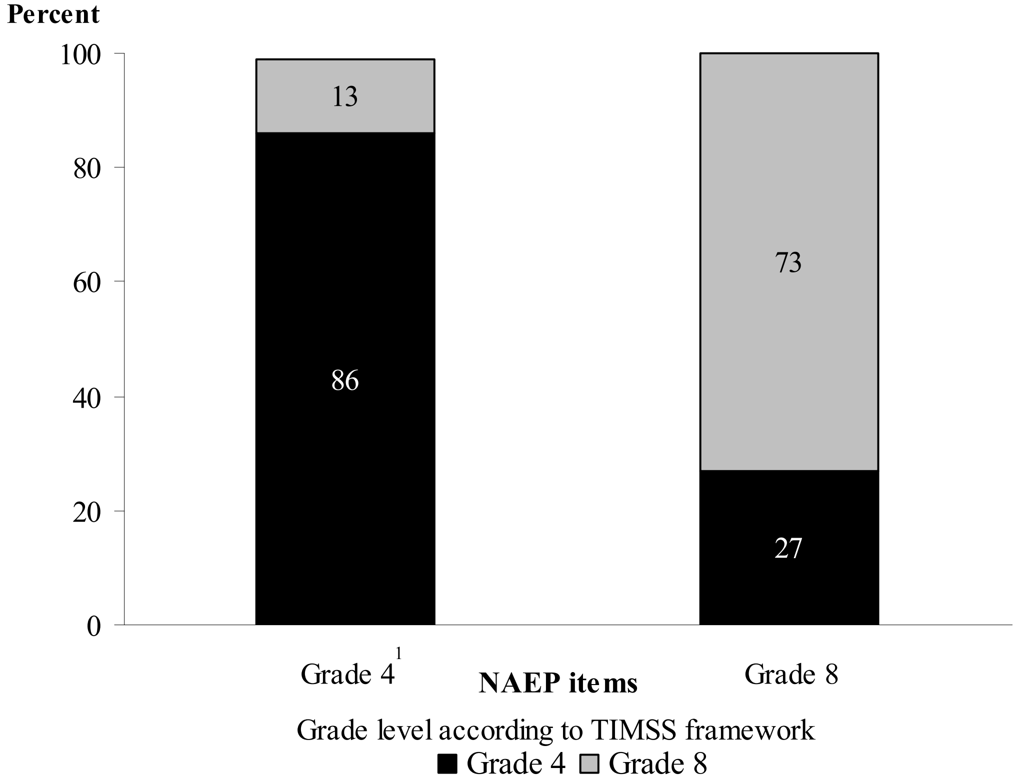
Comparing NAEP and TIMSS, there appears to be a moderately high level of consistency regarding what is considered to be fourth-grade content and what is considered to be eighth-grade content. On both the fourth- and eighth-grade assessments, most NAEP items were placed at the same grade level using the definitions and criteria of the TIMSS framework, and vice versa (figure 1-A). Eighty-six percent of all fourth-grade NAEP items were classified at the fourth grade and 73 percent of all eighth-grade NAEP items were classified at the eighth grade according to the TIMSS framework. Ninety-four percent of fourth-grade TIMSS items were classified at the fourth grade and 90 percent of eighth-grade TIMSS items were classified at the eighth grade on the NAEP framework (figure 1-B). Of the remaining TIMSS eighth-grade items, 6 percent were classified at the fourth grade and 3 percent at the twelfth grade according to the NAEP framework. It should be noted that since the TIMSS 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. At the same time, there were no comments recorded that suggested that any of the NAEP items exceeded the TIMSS eighth-grade descriptions.

Although PISA items are not designed to meet criteria for any specific grade, panelists did examine how they correspond to grade level(s) according to the NAEP framework. Most PISA items, 85 percent, were classified as being consistent with the eighth-grade NAEP framework, similar to the results for the TIMSS eighth-grade assessment (figure 1-B). Twelve percent of PISA items were classified at the fourth grade and four percent at the twelfth grade. Thus, according to the definitions

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

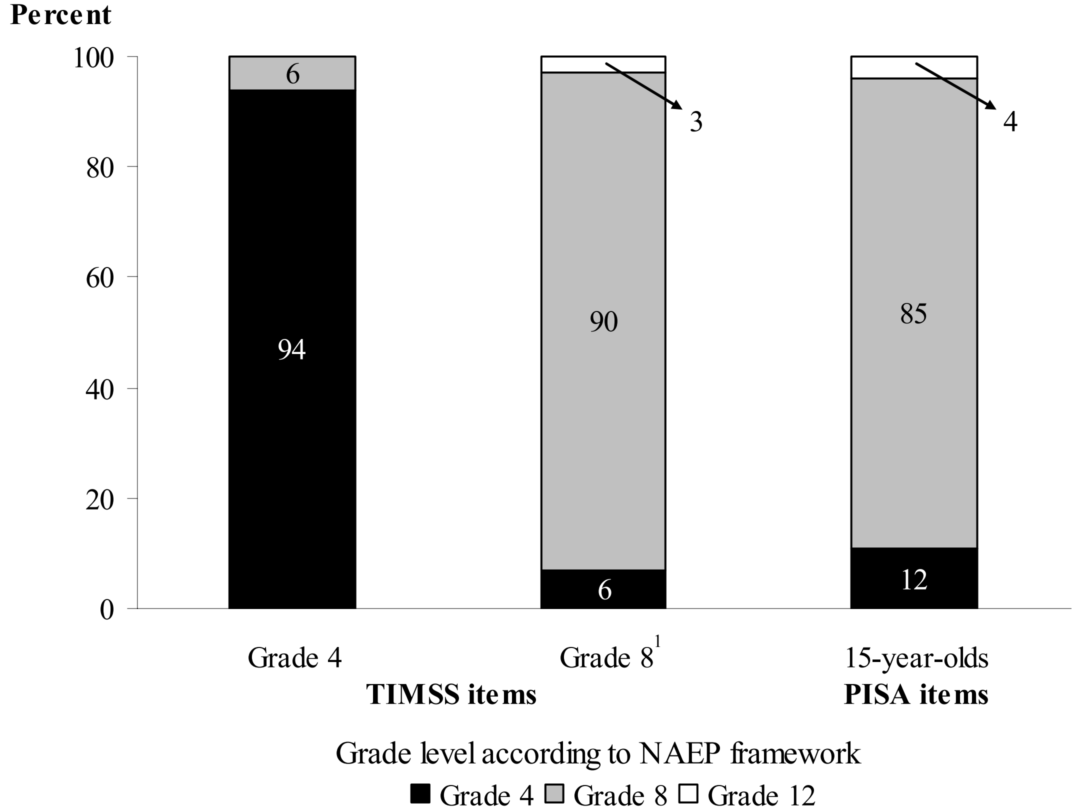
of the NAEP framework, the mathematics content of the PISA assessment is predominantly at the eighth-grade level.

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



[1]Two NAEP fourth-grade items that the panel did not classify with respect to grade level on the TIMSS assessment framework are not included.
NOTE: Data reflect expert panel classifications of grade level to the TIMSS content framework. 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) 2003 Mathematics Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.*, 2003.

Figure 1-B. Percentage distribution of TIMSS and PISA mathematics items classified at each grade level according to the NAEP mathematics framework, by grade/age: 2003



[1] One TIMSS eighth-grade item that the panel did not classify with respect to grade level on the NAEP assessment framework is not included.
NOTE: Data reflect expert panel classifications of grade level to the NAEP content framework . Detail may not sum to totals because of rounding or omitted items.
SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment; and U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress,* 2002.

The fact that grade-level agreement is less for the comparison between NAEP items and the TIMSS framework than for the TIMSS items to NAEP framework comparison may be related at least in part to the presence of NAEP cross-grade items that were developed to be used at multiple grades. Eighteen percent of fourth-grade NAEP items also appeared on the eighth-grade NAEP assessment and an additional 10 percent appeared on the eighth- and twelfth-grade NAEP assessments (data not shown). This represented 16 and 10 percent, respectively, of the eighth-grade assessment items. An additional 18 percent of eighth-grade items were administered at both the eighth and twelfth grades (but not the fourth grade). Assuming that cross-grade items are designed to be appropriate for all the grade levels at which they are administered, the presence of cross-grade items may affect the grade level correspondence to the TIMSS 2003 framework (TIMSS does not include cross-grade items). As reflected in table 7, most of the cross-grade items were classified by the panel at the lower or lowest grade level according to the TIMSS framework. For example, about 80 percent of NAEP items administered at grades 4 and 8 or grades 4, 8, and 12 were classified at the fourth grade level according to the TIMSS framework. As a result, the grade level match for all fourth-grade items, including both single-grade and cross-grade items, is similar to that for the single-grade items administered only at the fourth grade (86 percent compared to 88 percent). The difference is much greater at the eighth grade, where 93 percent of the NAEP single-grade items were classified at the corresponding grade level according to the TIMSS framework, compared to 73 percent for the NAEP eighth-grade assessment overall when the cross-grade items are included.

Table 7. Percentage of NAEP single-grade and cross-grade mathematics items classified at each grade level according to the TIMSS mathematics framework: 2003

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 4, 8, and 12	Grades 8 and 12
Grade 4	86	27	88	7	84	79	11
Grade 8	13	73	11	93	16	21	89

NOTE: Data reflect expert panel classifications of grade level to the TIMSS 2003 content framework. Single-grade items are administered at one grade level; cross-grade items are administered at more than one grade (4 and 8; 4, 8, and 12; or 8 and 12); totals reflect single-grade and cross-grade items included in the assessment at each grade level. Two NAEP fourth-grade items that the expert panel did not classify with respect to grade level 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) 2003 Mathematics Assessment; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.*, 2003.

4.3. Levels of Mathematical Complexity

Comparison of Mathematical Complexity Levels for NAEP, TIMSS, and PISA Assessments

All items from all three assessments were classified based on the three levels of *mathematical complexity* defined in the NAEP 2005 framework (NAGB 2004) and briefly described earlier in this report (see section 3.2). The *mathematical complexity* dimension of the NAEP 2005 framework replaces the previous NAEP cognitive dimensions of *mathematical abilities* and *mathematical power*. *Mathematical complexity* reflects the demand placed on students by the items and focuses on the properties of items rather than inferred abilities of students. *Mathematical complexity* is not necessarily related to item difficulty, which is based on actual student performance. *Mathematical complexity* should also be independent of curriculum, meaning it is determined assuming that students are familiar with the mathematical content of the item. The three levels—low, moderate, and high—are used to describe an increasing level of complexity of steps and processes required of students in order to succeed on an item and are based in part on the degree to which items require flexible or abstract thinking. A more complete description of the levels of *mathematical complexity* is included in appendix B.

Table 8 shows the percentage distribution of items in all three assessments classified by the expert panel at the three levels of *mathematical complexity*. It should be noted that although the classifications of *mathematical complexity* level are based on definitions in the NAEP 2005 framework, the NAEP 2003 items were not originally developed using this new dimension specified in the revised NAEP framework. Overall, NAEP and TIMSS exhibited very similar profiles of *mathematical complexity* level. At the fourth-grade level, the percentages of items classified in the three levels of *mathematical complexity* were nearly identical. At the eighth-grade level, a slightly higher percentage of TIMSS items were placed in the moderate level than NAEP items (46 percent compared to 39 percent), and a correspondingly lower percentage of items were placed in the low level (51 percent compared to 57 percent). For both NAEP and TIMSS, only a few items at each grade level (less than 5 percent) were classified at the high complexity level.

Table 8. Percentage distribution of NAEP, TIMSS, and PISA mathematics items across levels of mathematical complexity, by grade/age and survey: 2003

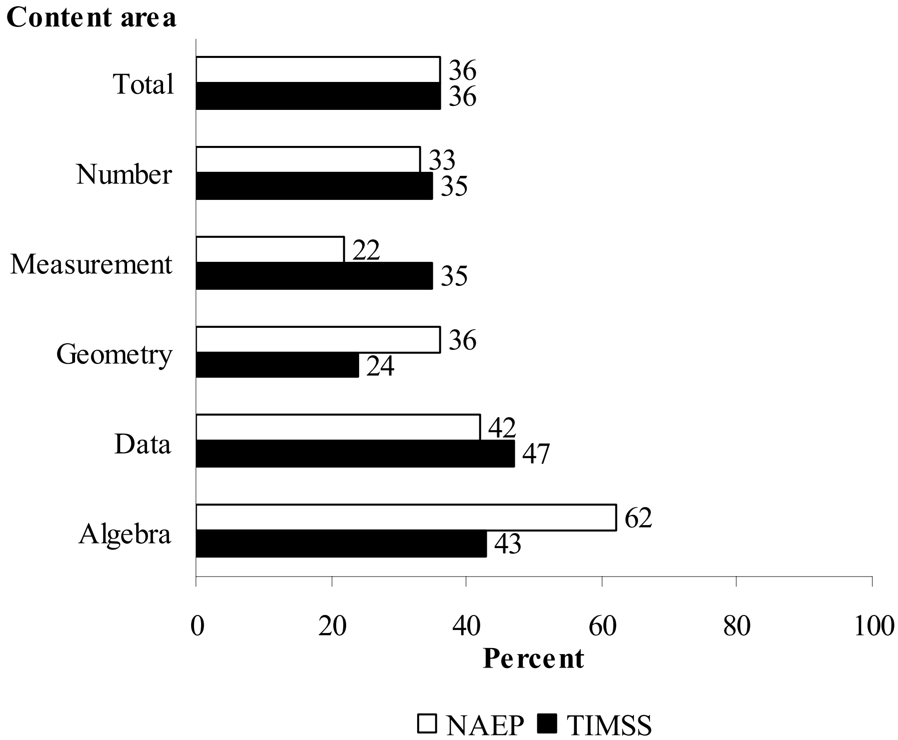
Mathematical complexity level	Grade 4		Grade 8		15-year-olds
	NAEP	TIMSS	NAEP	TIMSS	PISA
Low	64	64	57	51	29
Moderate	33	34	39	46	64
High	3	2	4	3	7

NOTE: Data reflect expert panel classifications of *mathematical complexity* level as defined in the NAEP 2005 mathematics framework. Levels of *mathematical complexity* (low, moderate, and high) form an ordered description of the cognitive demands of the item. 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) 2003 Mathematics Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment; and U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2005 National Assessment of Educational Progress*, 2004.

Relative to both NAEP and TIMSS, PISA has a much higher proportion of items at the moderate level than the low level (64 percent compared to 29 percent). Although a relatively higher percentage of PISA items were placed in the high complexity level than in NAEP or TIMSS, these items nevertheless made up only a small percentage of all PISA items (7 percent).

The percentage of items at the moderate or high complexity level in the NAEP and TIMSS assessments varied across the main content areas (figure 2). At the fourth-grade level, the *data* and *algebra* content areas had the highest proportion of items at the moderate or high complexity level for both NAEP and TIMSS (more than 40 percent). NAEP, however, had a much higher percentage in *algebra* (62 percent compared to 43 percent in TIMSS).

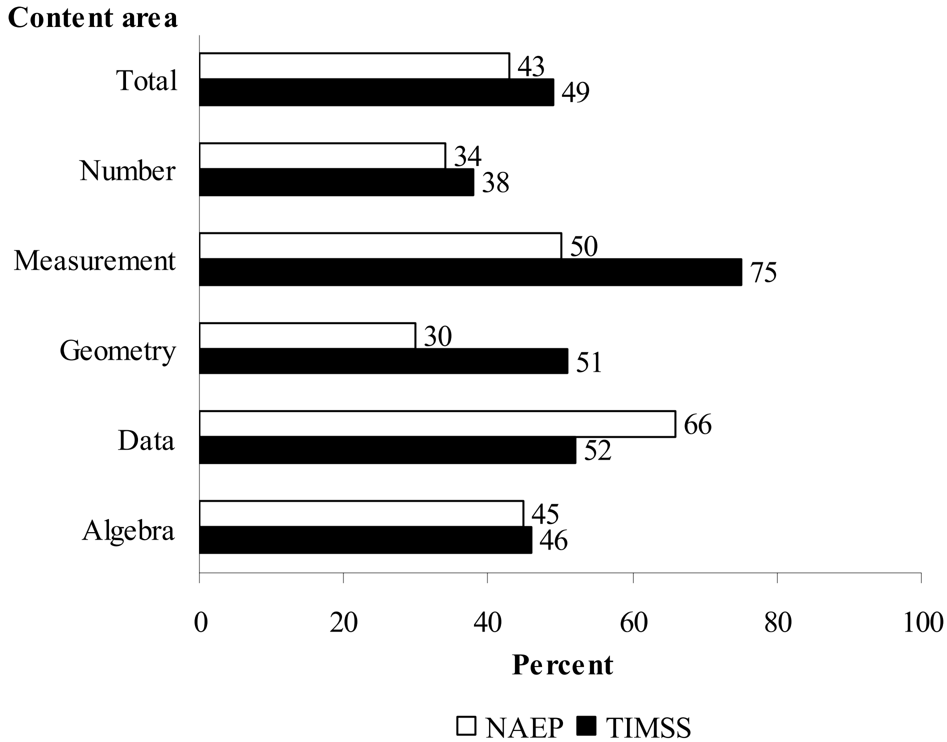
Figure 2. Percentage of NAEP and TIMSS fourth-grade items classified as moderate or high mathematical complexity level, by mathematics content area: 2003



[1] The total category refers to all items combined.
NOTE: Data reflect expert panel classifications of *mathematical complexity* level as defined in the NAEP 2005 mathematics framework.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics 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, *Mathematics Framework for the 2005 National Assessment of Educational Progress*, 2004.

At the eighth-grade level, relatively large differences between the NAEP and TIMSS items are found across the content areas of *measurement*, *geometry,* and *data* (figure 3). For TIMSS, there were a high proportion of items of moderate or high complexity level in *measurement*, with 75 percent of items compared to 50 percent of NAEP items. *Geometry* was also an area of higher complexity in TIMSS, with about half of items of moderate or high complexity level compared to only 30 percent of *geometry* items in NAEP. For NAEP, the content area with the highest complexity level was *data*, with 66 percent of NAEP items classified at the moderate or high level compared to 52 percent of TIMSS items. In none of the content areas did the percentage of items classified at the high complexity level exceed 10 percent for either assessment at fourth or eighth grade.

Figure 3. Percentage of NAEP and TIMSS eighth-grade items classified as moderate or high mathematical complexity level, by mathematics content area: 2003



[1] The total category refers to all items combined.
NOTE: Data reflect expert panel classifications of *mathematical complexity* level as defined in the NAEP 2005 mathematics framework.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics 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, *Mathematics Framework for the 2005 National Assessment of Educational Progress*, 2004.

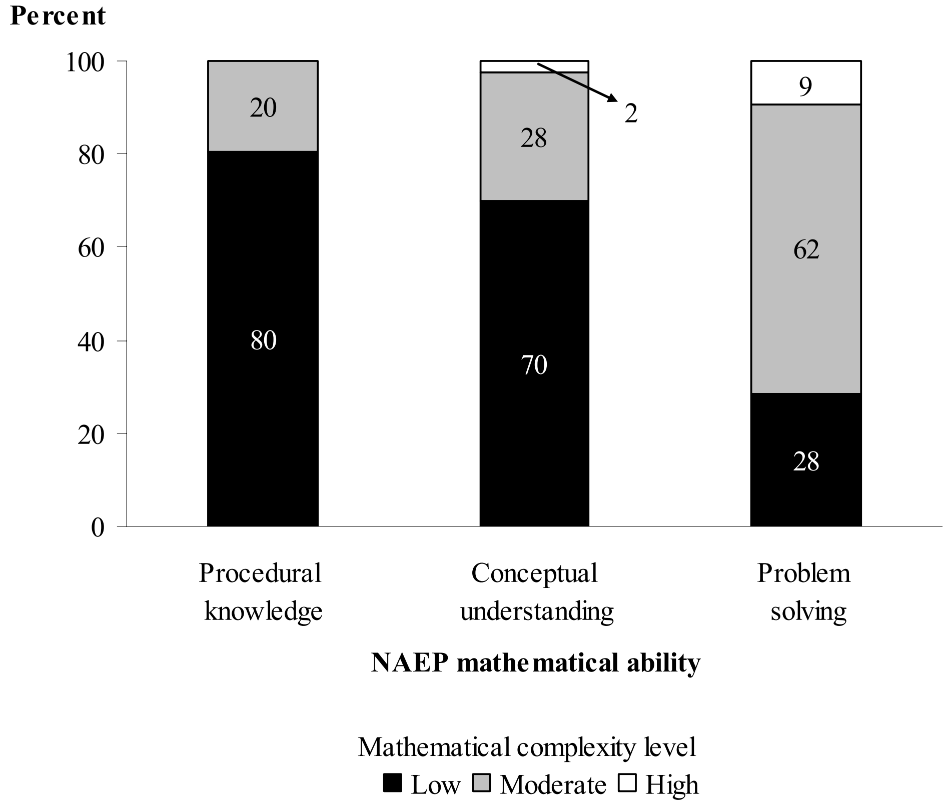
The frameworks for all three assessments have dimensions that describe and classify items according to cognitive abilities they require of students. Although a great deal can be learned from the framework documents about the meanings and intentions of each category and the ways in which they might influence item development, the classifications of all items to a common system of describing cognitive demand, the level of *mathematical complexity* in the NAEP 2005 framework, provides an additional means of examining these systems. Comparing the classification of items on the cognitive dimensions of each framework to the classifications for levels of complexity may reveal, for example, the degree to which each system is related to *mathematical complexity* level and the degree to which the systems represent a hierarchy of cognitive skills. To relate the expert panel classifications of *mathematical complexity* level to the intentions of the original framework, the NAEP, TIMSS and PISA items developed for each of the original cognitive categories were compared with respect to the proportion of low, moderate, and high complexity level. The results of these cross-classifications by cognitive categories are shown in the following sections.

Cross-classification by NAEP Mathematical Abilities Dimension

Figure 4 displays the percentage of items from each of the NAEP 2003 *mathematical abilities* categories (*procedural knowledge*, *conceptual understanding*, and *problem solving*) classified at each level of *mathematical complexity* according to the NAEP 2005 framework.[28] Not surprisingly, most of the *procedural knowledge* items were classified at the low complexity level (80 percent). The majority of items in the *conceptual understanding* category (70 percent) also were classified at the low complexity level. While a higher proportion of items were classified at the moderate complexity level in the *conceptual understanding* category, these first two *mathematical ability* categories do not reflect a clear hierarchy of complexity. Example 1 in appendix E shows a NAEP *conceptual understanding* item classified at the low *mathematical complexity* level. Many items from the *conceptual understanding* category tested student knowledge of a definition or concept with no additional steps or analysis required. These types of items are consistent with the definition of low complexity level. In comparison, most *problem solving* items were classified at the moderate complexity level (62 percent), although 28 percent were classified at the low complexity level and a number of these were word problems (figure 4). Example 2 in appendix E shows a NAEP *problem solving* item classified at the low *mathematical complexity* level. Most of the NAEP word problems were from the *problem solving* category, but many of these items were classified by the panel at the low complexity level because the underlying mathematics problem was fairly obvious and the solution required little abstract or flexible thought. Few items even in the *problem solving* category were classified at the high complexity level (9 percent).

[28] Because not all NAEP items are given classifications for mathematical power, this section examines only classifications for mathematical ability.

Figure 4. Percentage distribution of fourth- and eighth-grade NAEP mathematics items across mathematical complexity levels, by NAEP mathematical ability category: 2003

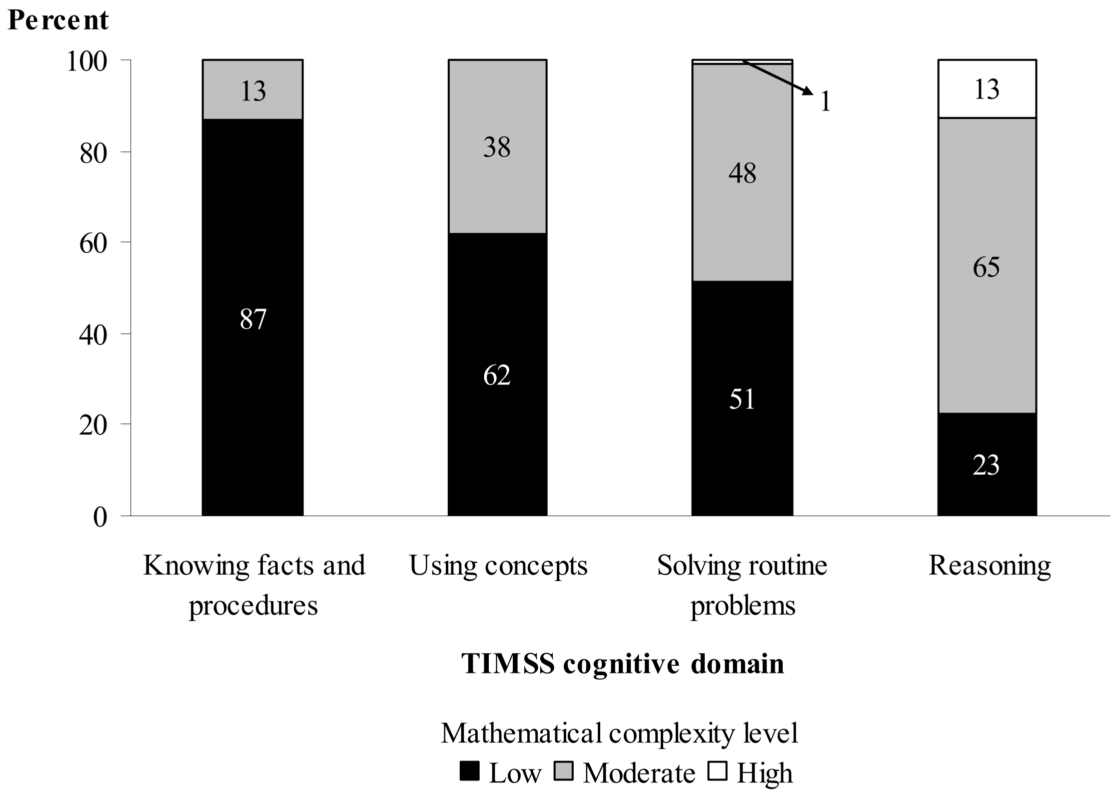


NOTE: Data reflect both fourth- and eighth-grade items combined. Classifications by *mathematical abilities* were provided by NAEP assessment developers; classifications by *mathematical complexity* level made by expert panel according to definitions in the NAEP 2005 framework. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment; and U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2005 National Assessment of Educational Progress*, 2004.

Cross-classification by TIMSS Cognitive Domains

Figure 5 displays the percentage of items from each of the TIMSS 2003 *cognitive domain* categories (*knowing facts and procedures*, *understanding concepts*, *solving routine problems*, and *reasoning*) classified at each level of *mathematical complexity* according to the NAEP 2005 framework. Similar to the NAEP *procedural knowledge* category, a very high percentage of TIMSS *knowing facts and procedures* items were classified at the low complexity level (87 percent). Also similar to the classification of NAEP *conceptual understanding* items, 62 percent of TIMSS *understanding concepts* items were classified at the low complexity level for similar reasons, most notably that recalling or using a concept does not necessarily require abstract or flexible thought. Items developed for the category of *solving routine problems* in TIMSS were about evenly distributed across the low and moderate complexity levels (51 and 48 percent, respectively), with only one item at the high complexity level (1 percent). The *mathematical complexity* level distribution for TIMSS *reasoning* items is similar to that for the NAEP *problem solving* items, with 23 percent low, 65 percent moderate, and 13 percent at the high complexity level. A TIMSS *reasoning* item classified at the low *mathematical complexity* level is illustrated by Example 3 in appendix E. A review of items indicates that most items representing non-routine situations were classified as *reasoning* in TIMSS. Using the different classification system, however, the panel judged some of these items as not requiring much abstract or flexible thinking and classified them as low or moderate with respect to the level of *mathematical complexity*.

Figure 5. Percentage distribution of fourth- and eighth-grade TIMSS mathematics items across mathematical complexity levels, by TIMSS cognitive domain category: 2003

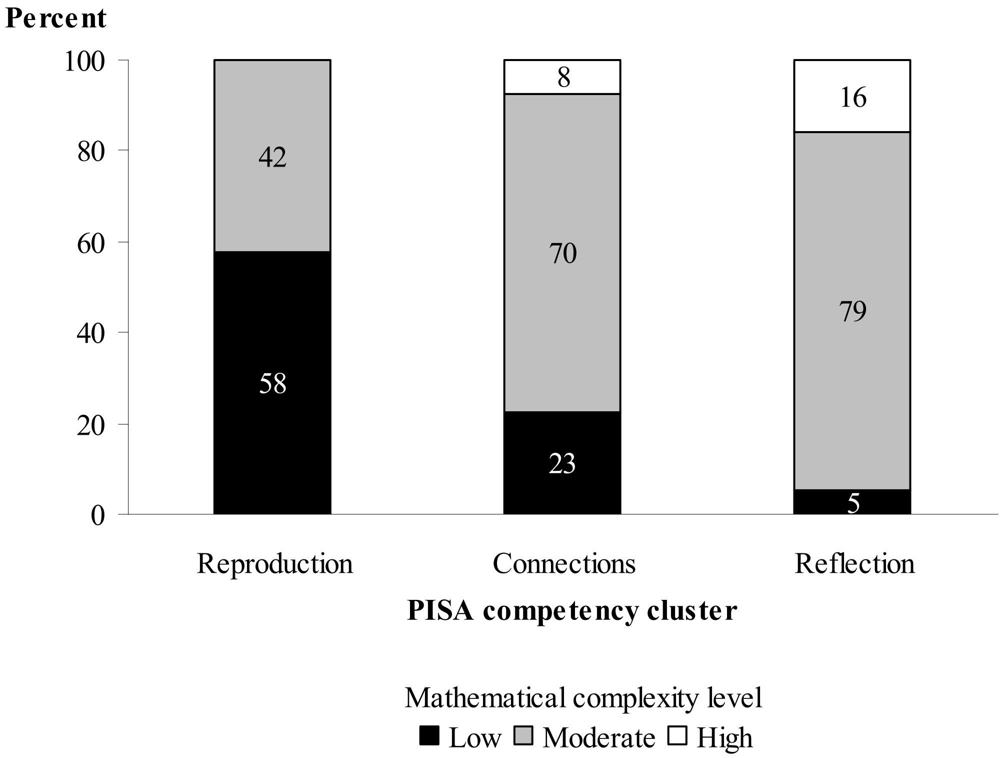


NOTE: Data reflect fourth- and eighth-grade items combined. Classifications by *cognitive domains* were provided by TIMSS assessment developers; classifications by *mathematical complexity* level were made by expert panel according to definitions in the NAEP 2005 framework. Multi-part items were counted in the reasoning category if any of the item parts were classified as reasoning; otherwise, multi-part items were counted in the category corresponding to the most frequent category across item subparts. 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; U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2005 National Assessment of Educational Progress*, 2004; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.*, 2003.

Cross-classification by PISA Competency Clusters

Figure 6 displays the percentage of items from each PISA 2003 *competency cluster* (*reproduction*, *connections*, and *reflections*) classified at each level of *mathematical complexity* according to the NAEP 2005 framework. Many items in all PISA *competency clusters* were found to be at the moderate complexity level. That 42 percent of PISA *reproduction* items were classified at the moderate complexity level may at first appear inconsistent with the description of the *reproductions* cluster in the PISA framework: "The competencies in this cluster essentially involve reproduction of practiced knowledge…." (OECD 2003). However, the framework presents a fairly broad interpretation of this phrase, providing illustrations of "the reproduction of practiced knowledge" in settings such as "thinking and reasoning," "modeling," and "problem posing and solving." Example 4 in appendix E presents a PISA *reproduction* item classified at the moderate *mathematical complexity* level. A high degree of correspondence is seen between the definitions provided in the PISA framework and the percentage of low complexity level items found across the series of competency clusters from *reproduction* to *connections* to *reflection*. The trend is less compelling when considering the differentiation between the moderate and high complexity levels. While there is the highest proportion of high complexity level items in the PISA *reflections* category than for any other set of items across the three assessments, this still represents less than 20 percent of items.

Figure 6. Percentage distribution of PISA mathematics items across mathematical complexity levels, by PISA competency cluster: 2003



NOTE: Classifications by *competency cluster* were provided by PISA assessment developers; classifications by *mathematical complexity* level were made by expert panel according to definitions in the NAEP 2005 framework. Detail may not sum to totals because of rounding.
SOURCE: Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment; U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2005 National Assessment of Educational Progress*, 2004; and Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*, 2003.

4.4. Item Format

The items in the NAEP, TIMSS, and PISA assessments were also compared with respect to the types of item formats used and their proportion on the assessments. Table 9 shows the percentage distribution of NAEP, TIMSS, and PISA items by item format, including the main formats of multiple choice and constructed response as well as different types of constructed-response items defined by each assessment. Items vary in difficulty level and cognitive demand regardless of format, though constructed-response items can be particularly important in assessing students' abilities to communicate their mathematical understanding and explain their solutions. Including a variety of item types ensures that a range of knowledge and skills is being assessed. Examples are included in appendix E to illustrate some of the item formats across the three assessments: a short constructed-response and an extended constructed-response item from the NAEP assessment (Examples 5 and 6), an extended constructed-response item from the TIMSS assessment (Example 7), a complex multiple-choice item and an open constructed-response item from the PISA assessment (Examples 8 and 9).

At both grades, the distribution of NAEP and TIMSS items by item format is similar, with more than 60 percent multiple-choice items, although at eighth grade TIMSS has a slightly higher proportion of multiple choice (71 percent). At both the fourth- and eighth-grade levels, the TIMSS assessment identified a relatively higher proportion of items as extended constructed response than the NAEP assessments. However, the definition and nature of extended-response items may not be 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 may have up to five score levels (extended/satisfactory/partial/minimal/incorrect). The criteria that differentiate each response level vary by item within and across assessments.

PISA relies on multiple-choice items far less than do NAEP or TIMSS. Only one-third of PISA items use a multiple-choice format, compared to approximately two thirds of items on NAEP and TIMSS. In PISA, this item format category includes traditional multiple-choice items as well as "complex multiple choice" items that require students to answer a series of multiple-choice or true-false questions based on the same information (see Example 8). In most cases, students must answer all questions correctly in order to receive credit for the item, while a few items allow partial credit for answering one or more, but not all, questions correctly. There also are three types of constructed response items identified in PISA: short response, closed response, and open response. Short-response items and closed constructed-response items require students to write or otherwise indicate the answer to the question but not show their work. These two item types are similar although the closed constructed-response items are more constrained to a specific set of possible answers. These types of items are scored dichotomously and do not allow partial credit. They are, thus, different from the short constructed-response items in NAEP, which sometimes allow for partial credit and often require students to write brief explanations or show their work. The PISA open constructed-response items may require students to show their work, and there may be numerous ways in which students may receive credit. Partial credit rubrics are often used for these items, but only with three levels (full/partial/no credit). In this respect, they are more similar to the extended constructed-response items in TIMSS and some of the short-answer items in NAEP.

Table 9. Percentage distribution of NAEP, TIMSS, and PISA mathematics items across item formats, by grade/age and survey: 2003

	Grade 4		Grade 8		15-year -olds
Item format	NAEP	TIMSS	NAEP	TIMSS	PISA
Multiple choice	63	63	65	71	33[1]
Constructed response	37	37	35	29	67
Short answer (short response)	33	28	30	16	27
Extended response	4	10	5	13	—
Closed response	—	—	—	—	15
Open response	—	—	—	—	25

— Not available. These constructed-response item format classifications were not included in the information provided by the assessment developers.

[1] PISA also includes "complex multiple-choice" items (13 percent), which are reflected in the percentage of multiple-choice items.

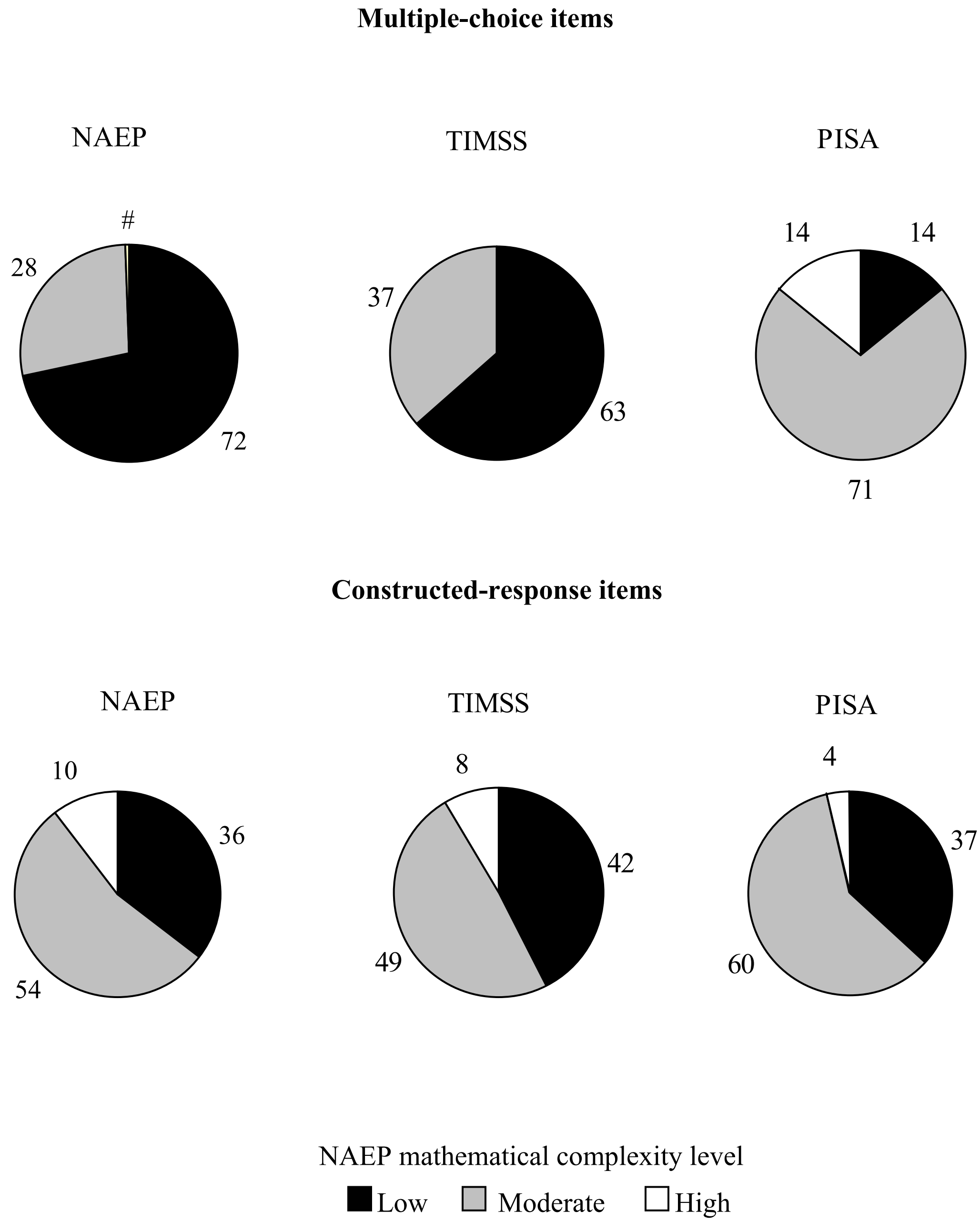
NOTE: The breakdown of constructed-response items was provided by the assessment developers for the NAEP and PISA items. For the TIMSS items, the assignment was based on examination of the items and the level of score points in the scoring guides in accordance with information provided by the TIMSS assessment 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) 2003 Mathematics Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; and Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment.

To compare the cognitive demand placed on students by the items of different formats, figure 7 shows the percentage of multiple-choice and constructed-response items in each assessment that were classified at the low, moderate, or high *mathematical complexity* level. NAEP and TIMSS show similar profiles, with multiple-choice items being composed primarily of low complexity level items (72 percent in NAEP and 63 percent in TIMSS) and essentially no high complexity level items. The complexity profiles for constructed-response items are also quite similar between NAEP and TIMSS, with about half of items classified at the moderate complexity level and 8 percent of TIMSS items and 10 percent of NAEP items classified at the high complexity level.

Once again, PISA stands alone in terms of the *mathematical complexity* level of its multiple-choice items. In PISA, the multiple-choice items (including both the more traditional multiple-choice items as well as the complex multiple-choice items) are predominantly moderate complexity level (71 percent) compared to about one-third in NAEP and TIMSS. In addition, this type of item format in PISA also includes several high complexity level items (14 percent), which were not found for NAEP or TIMSS. In fact, in PISA a larger proportion of multiple-choice items than constructed-response items are at the high complexity level. This reflects the fact that none of the short-response or closed constructed-response items were classified at the high complexity level. Among constructed-response items, the profile for PISA is similar to that for NAEP and TIMSS, but there are still a higher proportion of items at the moderate complexity level than found in NAEP and TIMSS (60 percent compared to about half in NAEP and TIMSS).

Figure 7. Percentage distribution of NAEP, TIMSS, and PISA mathematics items across levels of mathematical complexity, by item format: 2003



Rounds to zero.
NOTE: Data reflect expert panel classifications of *mathematical complexity* level according to definitions in the NAEP 2005 framework. Data for NAEP and TIMSS reflect both fourth- and eighth-grade items combined. The graphic for NAEP does not show one multiple-choice item classified at the high complexity level (< 0.5 percent). 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) 2003 Mathematics Assessment; International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment; Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment; and U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2005 National Assessment of Educational Progress*, 2004.

This section compared the assessments overall with respect to content coverage, grade level, levels of mathematic complexity, and item format. In the next section, more detailed comparisons of the content of the NAEP and TIMSS assessments are made in each of the main content areas of *number*, *measurement*, *geometry*, *data*, and *algebra*.

5. NAEP/TIMSS Comparisons by Main Content Areas

The overall comparisons section highlighted that NAEP and TIMSS have similar emphases across the broad mathematics content areas. This section provides more detailed comparisons of the content coverage of the items for each of the NAEP and TIMSS mathematics content area subscales. There are five content area sections presenting comparisons for items in each of the NAEP *content strands* with those in the corresponding TIMSS *content domains*. These five content-area sections include:[1]

- *number*;
- *measurement*;
- *geometry*;
- *data*; and
- *algebra*.

Each content area section includes

- a comparison of the relevant parts of the content frameworks;[2]
- an analysis of the level of match between the items from one assessment and the topics and subtopics included at particular grades in the other assessment framework; and
- a comparison of how items are distributed across topics within these content areas as 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. Content and grade classification are examined simultaneously in the analyses for this section. For each content area, the report reports 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, as follows:

- specific match (to a specific subtopic in the same content area);[3]
- general match (at the broader topic level in the same content area but not the subtopic level); and

---

[1] Although NAEP and TIMSS use somewhat different labels to refer to each of these content areas, the one-word TIMSS categories are used in the discussions in this section for the sake of convenience.
[2] Framework comparison tables in this section list the topics included in the content area. Additional information about the specific subtopics included for each of the main topics is given in appendix A.
[3] Specific match also applies to items classified to a topic in the NAEP framework for which no subtopics are included.

- match to another content area (at either the topic or subtopic level).

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

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

- 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)[4]

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 or subtopic at a specific grade).[5] The text in this section may refer to items classified to specific subtopics, but this level of detail is not shown in the tables in the report. Subtopics are shown in appendix A and example items illustrating various features that are referenced in this section are shown in appendix E.

5.1. Number

The *number* content area receives greater emphasis at fourth grade than at eighth grade for both the NAEP and TIMSS assessments. As discussed in section 4, in NAEP, 42 percent of fourth-grade items and 26 percent of eighth-grade items are from the *content strand* of *number sense, properties, and operations* (table 4). In TIMSS, 38 percent of fourth-grade items and 31 percent of eighth-grade items are from the *content domain* of *number* (table 5). The results in the *number* section are based on 76 fourth-grade and 51 eighth-grade items in NAEP, and 55 fourth-grade and 56 eighth-grade items in TIMSS.

Framework comparison in number

The framework structures used in NAEP and TIMSS to organize content in this area are quite different. Exhibit 3 shows a comparison of the *number* topics included in the NAEP and TIMSS mathematics frameworks. The NAEP framework is organized around skills, such as *represent numbers and operations in a variety of equivalent forms using models, diagrams and symbols*, and *compute with numbers*. The TIMSS framework, on the other hand, is organized by types of numbers, for example, *whole numbers*, and *fractions and decimals*. When considering all subtopics (appendix A), there are fewer differences between the content specified in NAEP and TIMSS. NAEP subtopics often specify types of numbers and TIMSS subtopics specify types of skills. The correspondence is not complete, however, leaving some ambiguity regarding whether a topic or subtopic not mentioned specifically in one framework might still be implied. For example, one TIMSS subtopic is “solve problems with fractions.” Similar content might be implied in the NAEP subtopic of “solve application problems involving numbers and operations, using exact answers or estimates, as appropriate,” although fractions are not mentioned specifically.

---

[4] Because the NAEP 2003 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.

[5] 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 (subtopic, topic, or broad content area). In this section, grade level was not assigned unless items could be classified to at least the topic level.

One difference resulting from the different organizational approaches is that the NAEP framework contains more detail about the aspects of computation to be emphasized. Whereas both frameworks address computation, in TIMSS it is at the subtopic level, under two separate topics, *whole numbers* and *fractions and decimals*. In NAEP, it is at the topic level (*compute with numbers*) and is further specified by four subtopics. One thing noted by the panel related to computations was that the NAEP framework does not explicitly include non-contextualized computation that is not placed in a problem-solving or application setting. Although there is a main topic in NAEP entitled *compute with numbers*, the subtopic of "apply basic properties of operations" was interpreted by the panel as requiring an application context. In contrast, in the TIMSS framework, there are explicit subtopics related to the mechanics and properties of computation within the main topics by number type (e.g., "compute with whole numbers;" and "add and subtract fractions or decimals"). This difference is primarily an issue with the specificity and interpretation of the framework, however, since both assessments include computation items.

One similarity between the frameworks is that both focus on representation and computation involving whole numbers, fractions, and decimals at the fourth-grade level, with ratios, proportions, and percents and working with integers not emphasized until the eighth-grade level.[6]

Exhibit 3. Number topics included in the NAEP and TIMSS mathematics frameworks: 2003

NAEP ***Number sense, properties, and operations topics***	**TIMSS** ***Number topics***
Relate counting, grouping and place value	Whole numbers
Represent numbers and operations in a variety of equivalent forms using models, diagrams and symbols	Fractions and decimals
Compute with numbers (that is, add, subtract, multiply, divide)	Integers (grade 8 only)
Use computations and estimation in applications	Ratio, proportion, and percent
Apply ratios and proportional thinking in a variety of situations	
Use elementary number theory	

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 subtopics or objectives at each grade and level of detail varies across topics and assessments.
SOURCE: U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.*, 2003.

Content and grade match in number

With some exceptions, NAEP and TIMSS appear to share similar definitions of the content area of *number* at both the fourth and eighth grades, since, for each assessment most items were classified at either the topic or subtopic level on the framework of the other assessment at the same grade level (table 10). In NAEP, 79 percent of fourth-grade and 73 percent of eighth-grade items were classified at the same grade level with either a specific or general match to TIMSS *number* topics. There was an even higher correspondence between the TIMSS items and *number* topics in

[6] Although the word *integer* does not appear in the NAEP framework topics, it does appear in the grade-level item illustrations given in the assessment specifications document (NAGB 1992).

the NAEP framework, with 98 percent of fourth-grade and 96 percent of eighth-grade items having a specific or general match to *number* topics in the NAEP framework.

The 22 percent of TIMSS fourth-grade items with a general match were all classified to the NAEP topic of *compute with numbers*. Panelists noted that these items were of two types, which they described as, "basic computation" and "select an operation or procedure to solve a problem." Examples of each of these types of TIMSS items are shown in Examples 10 and 11, respectively. As defined by the panel, "basic computation" items were items that required students to perform a computation in a non-contextualized situation and involved only the mechanics and properties of computation. Panelists also identified several items of these types on the TIMSS eighth-grade assessment as well. As noted previously, the panel interpreted the subtopic within the NAEP *compute with numbers* topic as requiring an application context, so any purely computational items in TIMSS were classified at only the general topic level in the NAEP framework.

An examination of the classification of NAEP items to TIMSS subtopics (data not shown) revealed that NAEP also includes computation items, although there were not as many of these types of items as in TIMSS. The TIMSS framework includes three fourth-grade subtopics within the topics of *whole numbers* and *fractions and decimals* that address only computation. Although relatively more TIMSS items were placed in these subtopics, some NAEP items were also classified to the computation subtopics in TIMSS. Almost all of these NAEP items came from a single NAEP subtopic, *apply basic properties of operations*. The appearance of the computation items in NAEP indicates that the strict interpretation of this subtopic by the panel to mean that an application setting is required was not applied in the same fashion by the NAEP assessment developers. An item from the NAEP assessment that was described by the panel as "basic computation" is shown in Example 12 in appendix E.

At both grade levels, all or almost all TIMSS items were classified at the same grade level on the NAEP framework (table 10). On the other hand, 16 percent of fourth-grade NAEP items were classified as eighth-grade items on the TIMSS framework, almost all as *fractions and decimals* spread across various subtopics. An example of a NAEP fourth-grade *number* item placed at the eighth-grade level on the TIMSS mathematics framework is illustrated by Example 13. Twenty-seven percent of eighth-grade NAEP items were placed at the fourth-grade level on the TIMSS framework, all but one of which were cross-grade items administered at both fourth and eighth grades. Most of these items were placed in the TIMSS topics of *whole numbers* and *fractions and decimals* and came from several different NAEP topics and subtopics.

Although all but 4 percent of eighth-grade TIMSS items were placed at the eighth-grade level on the NAEP framework, based on additional comments made by the panel roughly one third of TIMSS items were noted as being slightly below the specifications for the eighth grade. Therefore, while these items were judged to correspond most closely with the general descriptions of the eighth-grade level in the NAEP framework, they were found to have some characteristics consistent with a somewhat lower level than eighth grade (but not consistent with the specifications at the fourth-grade level).

A few eighth-grade NAEP items were identified that included the use of scientific notation, which was not addressed in the TIMSS framework subtopics, nor reflected in the TIMSS items (data not shown). Also, there were a number of NAEP items that include mathematical operations involving money. Some of these items had a general match to TIMSS, since they were consistent with the descriptions of the broader topics. Some items in the fourth grade that require knowledge of the value of specific U.S. coins, however, were classified as not having a match in the TIMSS framework topics, as items of this type would not be included in an international assessment like TIMSS. In TIMSS and PISA, any items with money contexts use a common fictitious currency such as *zeds* (see example 7 in appendix E).

Table 10. Percentage of NAEP and TIMSS fourth- and eighth-grade number items classified to the other mathematics assessment framework, by level of content/grade match: 2003

Level of content/grade match	Grade 4: NAEP items to TIMSS framework	Grade 4: TIMSS items to NAEP framework	Grade 8: NAEP items to TIMSS framework	Grade 8: TIMSS items to NAEP framework
Total number of number items	76	55	51	56
	Percentage distribution			
Classified as same grade	79	100	73	96
Specific match[1] in number	71	76	47	80
General match[2] in number	8	22	25	16
Match to another content area[3]	0	2	0	0
Classified as another grade[4]	16	0	27	4
Lower grade[5]	†	†	27	4
Higher grade[6]	16	0	†	0
No classification to topics[7]	5	0	0	0

† 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 at the subtopic level at the same grade (and items classified to NAEP framework topics for which no subtopics are included).
[2] Includes items that were classified to a topic but not to a subtopic at the same grade.
[3] Includes items that were classified to a topic or a subtopic in a different content area at the same grade.
[4] Includes items that were classified to a topic or subtopic in any content area at another grade.
[5] Includes grade 8 items classified to grade 4 topics/subtopics.
[6] Includes grade 4 items classified to grade 8 topics/subtopics or grade 8 TIMSS items classified to grade 12 NAEP topics/subtopics.
[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) 2003 Mathematics 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, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.*, 2003.

Item distribution across number topics

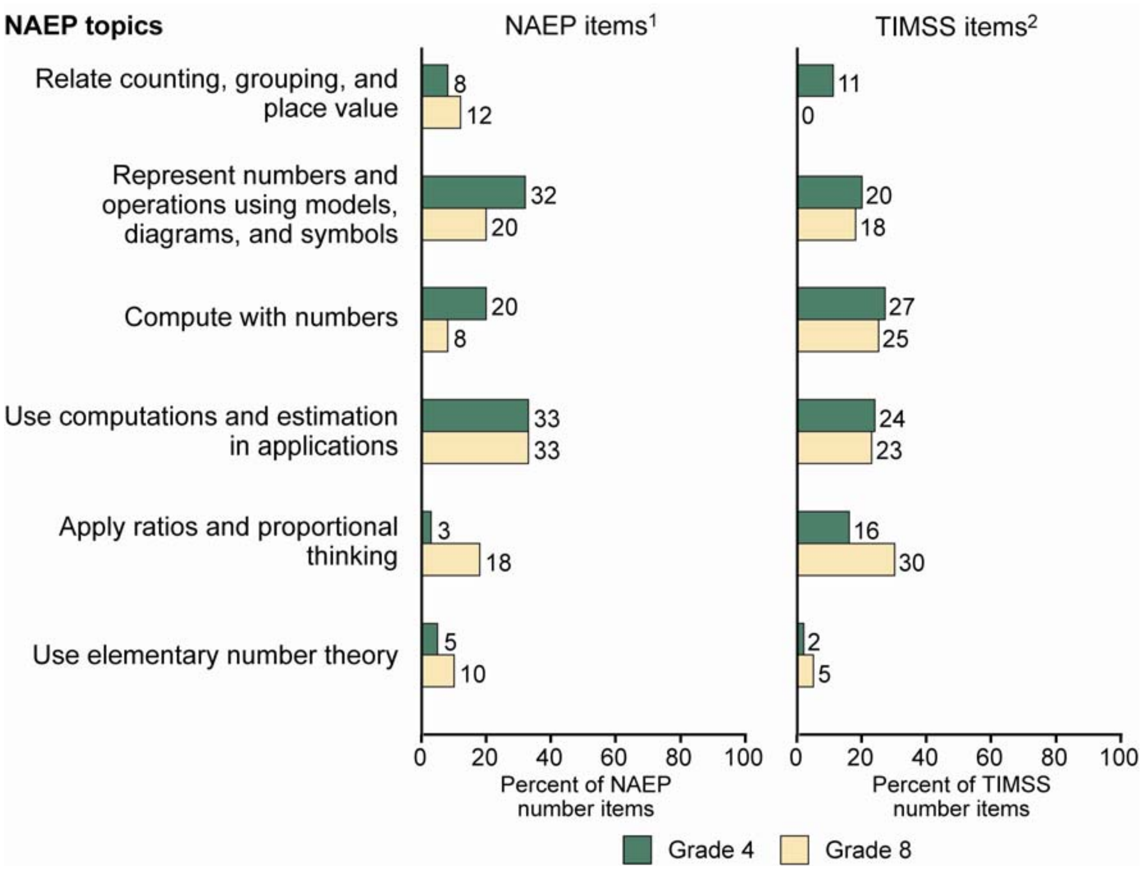
Figures 8 and 9 show the percentage of NAEP and TIMSS number items across the topics included in the *number* frameworks for each assessment. The distribution of items across NAEP topics indicates a greater emphasis in NAEP on the application of computation (as opposed to the mechanics and properties of computation) than in TIMSS. At the fourth-grade level, whereas TIMSS includes a higher percentage of items classified to the NAEP framework as *compute with numbers* (27 percent compared to 20 percent of NAEP items), NAEP includes a higher percentage of items classified as *use computations and estimation in applications* (33 percent compared to 24 percent). This pattern is found at the eighth-grade level as well—8 percent of NAEP items compared to 25 percent of TIMSS items classified to the *compute with numbers* topic and 33 percent of NAEP items compared to 23 percent of TIMSS items classified to the NAEP topic of *use computations and estimation in applications*. However, this finding is contradicted by the fact that almost half of all items on both eighth-grade assessments were classified to TIMSS subtopics that dealt with problem solving across the main topic areas, such as solving problems involving decimals, fractions, integers, etc. (data not shown). This points to some lack of specificity in both frameworks in distinguishing between pure computation and application or problem-solving contexts.

Another difference is that TIMSS includes a higher percentage of items involving ratios and proportions. This is supported by classifications on both frameworks—the NAEP topic of *apply ratios and proportional thinking* and the TIMSS topic of *ratio, proportion, and percent*. Considering both classification systems, TIMSS has 13 and 16 percent, respectively, of items in this topic area at the fourth grade compared to 3 and 5 percent in NAEP at the fourth grade. At eighth grade, TIMSS has about 30 percent of items focused on *ratio, proportion, and percent* compared to about 18 percent in NAEP. The NAEP eighth-grade assessment has a higher proportion of items classified as *whole numbers* on the TIMSS framework (37 percent compared to 14 percent of TIMSS).

Other notable differences based on the NAEP framework include a higher percentage of fourth-grade NAEP items classified as *represent numbers and operations using models, diagrams, and symbols* (32 percent compared to 20 percent in TIMSS). Also, at the eighth grade, 12 percent of NAEP items were classified to the topic of *relate counting, grouping, and place value* (half of which addressed scientific notation); in TIMSS there were items classified to this topic at fourth grade but none at eighth grade, and no items involving scientific notation.

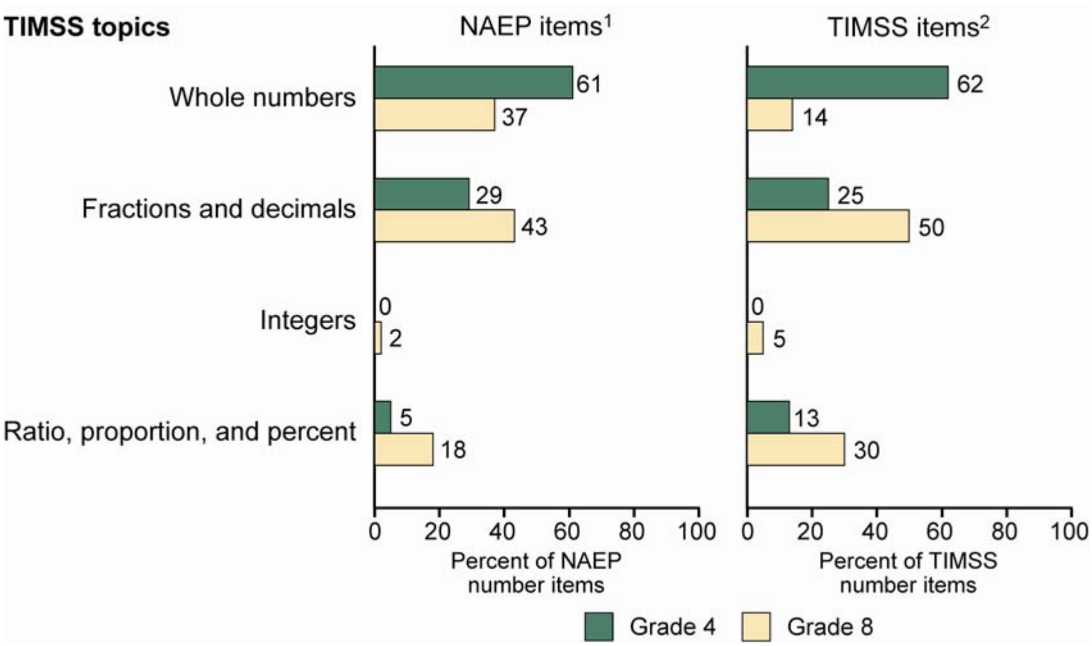
In both NAEP and TIMSS, the fourth-grade assessments emphasize items involving *whole numbers* (more than 60 percent of items), while the eighth-grade assessments are focused on items involving *fractions and decimals* (43 and 50 percent of items, respectively). A small percentage of items (5 percent or less) involving *integers* were in included in both assessments at the eighth grade but not the fourth grade.

Figure 8. Percentage of NAEP and TIMSS number items classified to number sense, properties, and operations topics in the NAEP mathematics framework, by survey and grade: 2003



[1] NAEP items classified by NAEP developers.
[2] TIMSS items classified by expert panel.
NOTE: Topics may be abbreviated for graphical clarity. Percentages reflect the proportion of *number* items classified at either the topic level or the subtopic 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) 2003 Mathematics 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, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002.

Figure 9. Percentage of NAEP and TIMSS number items classified to number topics in the TIMSS mathematics framework, by survey and grade: 2003



[1] NAEP items classified by expert panel.
[2] TIMSS items classified by TIMSS developers.
NOTE: Percentages reflect the proportion of *number* items classified at either the topic level or the subtopic 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) 2003 Mathematics 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 ed.*, 2003.

5.2. Measurement

Measurement items contribute similar proportions of the NAEP and TIMSS assessments at both grade levels. As discussed in section 4, in NAEP, 18 percent of fourth-grade items and 15 percent of eighth-grade items are from the *content strand* of *measurement* (table 4). In TIMSS, 21 percent of fourth-grade items and 16 percent of eighth-grade items are from the *content domain* of *measurement* (table 5). The results in the *measurement* section are based on 32 fourth-grade and 30 eighth-grade items in NAEP, and 31 fourth-grade and 28 eighth-grade items in TIMSS.

Framework comparison in measurement

Similar to the *number* frameworks, the NAEP and TIMSS *measurement* frameworks appear quite different at the topic level. However, when all subtopics included in NAEP and TIMSS are considered (appendix A), there is considerable overlap. Exhibit 4 shows a comparison of the *measurement* topics included in the NAEP and TIMSS mathematics frameworks. The TIMSS framework includes two broad topics with a set of eight subtopics across these two topics at both the fourth- and eighth-grade levels. The NAEP framework includes 10 topics, only three of which include subtopics. In addition, only five of the NAEP topics are intended to be included in the fourth-grade assessment. Comparing the topics and subtopics at a general level, both frameworks appear to address the major content areas typically associated with measurement, including measuring and computing attributes of figures (e.g., length, area, volume, perimeter); selecting appropriate tools, units, and methods; and converting units of measure.

Content and grade match in measurement

The level of content and grade match for NAEP and TIMSS *measurement* items is shown in table 11. Fourth-grade NAEP *measurement* items had a closer match to TIMSS *measurement* topics and subtopics than fourth-grade TIMSS items did to NAEP, with 85 percent of NAEP items having either a specific or general match compared to 74 percent of TIMSS items. One reason for this difference is that the fourth-grade TIMSS assessment contained a number of items (16 percent) classified by the panel as being consistent with the eighth-grade NAEP framework. Most of these were nevertheless classified as *measurement* items.

In contrast, at the eighth-grade level, there was a closer match between the TIMSS items and the NAEP framework than between the NAEP items and the TIMSS framework. More than two-thirds of TIMSS items had either a specific or general match to the NAEP framework, whereas less than half of NAEP items did so to the TIMSS framework. A substantial number of items on both the TIMSS and NAEP eighth-grade assessments (25 and 37 percent) were classified as having a better match to the descriptions at the fourth-grade level of the other assessment framework. Example 14 in appendix E illustrates a TIMSS eighth-grade *measurement* item placed at the fourth-grade level on the NAEP mathematics framework.

NAEP contained a larger percentage of items that, although classified at the corresponding grade level, were classified to another content area (13 percent of fourth-grade and 17 percent of eighth-grade items in NAEP compared to about 5 percent of TIMSS items). Most of these NAEP items involved measurements of angles or properties of geometric shapes and were classified to *geometry* topics in TIMSS.

In *measurement* as well as in *number*, there was a case of similar items appearing on both assessments but apparently serving different purposes. Not classified to an appropriate NAEP subtopic were several TIMSS fourth-grade *measurement* items that required students to perform calculations with time and temperature. The NAEP fourth-grade assessment does include items of this type, but they were classified across various NAEP topics and subtopics, meaning that they served a purpose other than assessing students' ability to perform calculations with these types of measures.

Exhibit 4. Measurement topics included in the NAEP and TIMSS mathematics frameworks: 2003

NAEP ***Measurement topics***	**TIMSS** ***Measurement topics***
Estimate the size of an object or compare objects with respect to given attributes (such as length, area, capacity, volume, weight/mass)	Attributes and units
Select and use appropriate measurement instruments (for example, manipulatives such as a ruler, meter stick, protractor, thermometer, scales for weight or mass, gauges)	Tools, techniques, and formulas
Select and use appropriate units of measurement	
Estimate, calculate (using basic principles or formulas), or compare perimeter, area, volume, and surface area in meaningful contexts to solve mathematical and real-world problems	
Apply given measurement formulas for perimeter, area, volume, and surface area in problem settings (grades 8 and 12 only)	
Convert from one measurement to another within the same system (customary or metric) (grades 8 and 12 only)	
Determine precision, accuracy, and error (grades 8 and 12 only)	
Make and read scale drawings (grades 8 and 12 only)	
Select appropriate methods of measurement (such as direct or indirect)	
Apply the concept of rate to measurement situations (grades 8 and 12 only)	

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 subtopics or objectives at each grade and level of detail varies across topics and assessments.
SOURCE: U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.*, 2003.

Table 11. Percentage of NAEP and TIMSS fourth- and eighth-grade measurement items classified to the other mathematics assessment framework, by level of content/grade match: 2003

Level of content/grade match	Grade 4: NAEP items to TIMSS framework	Grade 4: TIMSS items to NAEP framework	Grade 8: NAEP items to TIMSS framework	Grade 8: TIMSS items to NAEP framework
Total number of measurement items	32	31	30	28
	Percentage distribution			
Classified as same grade	97	81	63	71
Specific match[1] in measurement	66	58	33	61
General match[2] in measurement	19	16	13	7
Match to another content area[3]	13	6	17	4
Classified as another grade[4]	0	16	37	25
Lower grade[5]	†	†	37	25
Higher grade[6]	0	16	†	0
No classification to topics[7]	3	3	0	4

† 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 at the subtopic level at the same grade (and items classified to NAEP framework topics for which no subtopics are included).
[2] Includes items that were classified to a topic but not to a subtopic at the same grade.
[3] Includes items that were classified to a topic or a subtopic in a different content area at the same grade.
[4] Includes items that were classified to a topic or subtopic in any content area at another grade.
[5] Includes grade 8 items classified to grade 4 topics/subtopics.
[6] Includes grade 4 items classified to grade 8 topics/subtopics or grade 8 TIMSS items classified to grade 12 NAEP topics/subtopics.
[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) 2003 Mathematics 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, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.*, 2003.

Item distribution across measurement topics

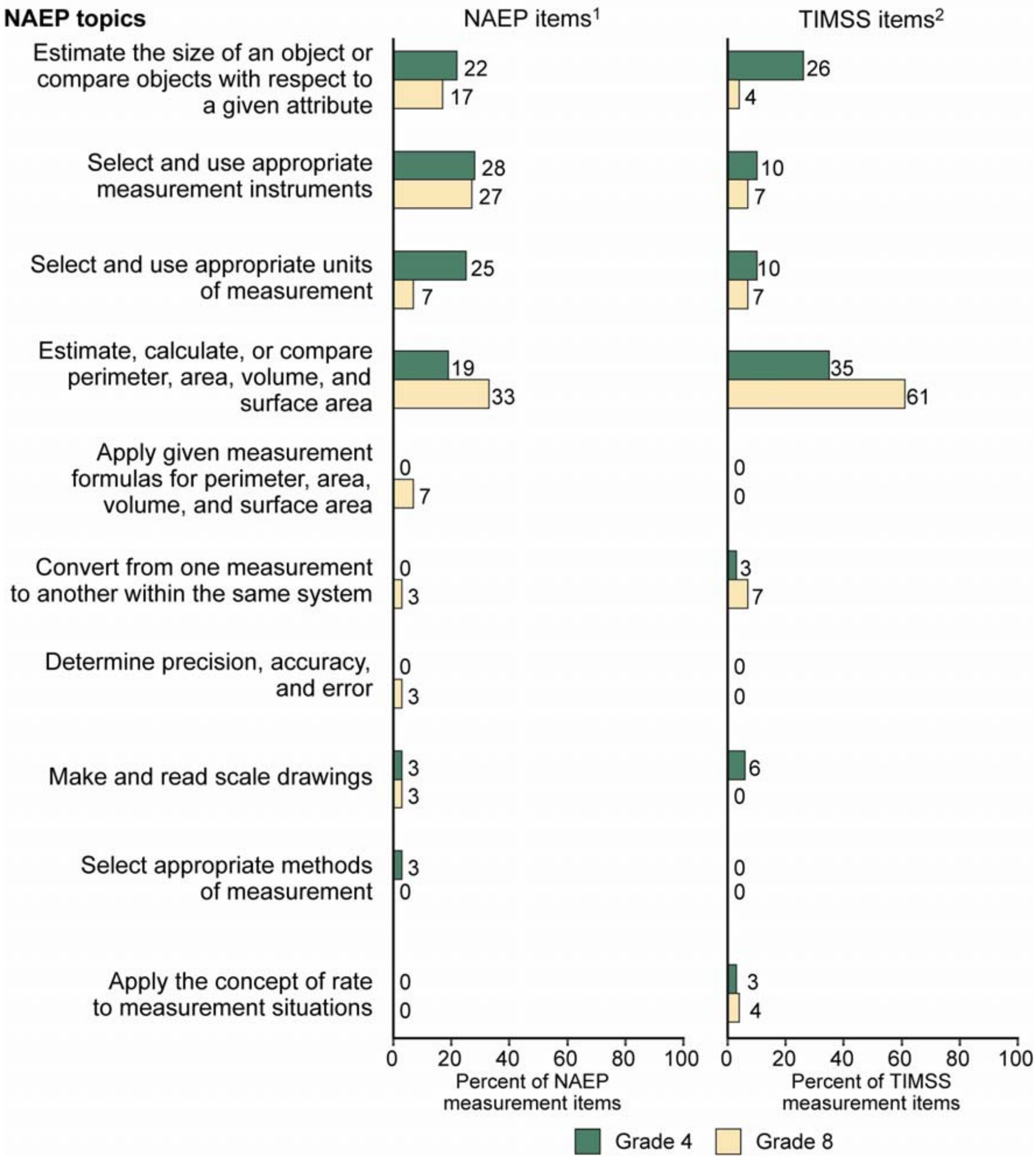
The percentage of *measurement* items across the NAEP and TIMSS *measurement* topics is shown in figures 10 and 11. On the NAEP *measurement* framework, relatively more fourth-grade NAEP items were placed in the topics of *select and use appropriate measurement instruments* (28 percent) and *select and use appropriate units of measurement* (25 percent) than were fourth-grade TIMSS items (10 percent for both topics). A further indication of fourth-grade NAEP items' relative emphasis on selecting and using appropriate units is that on the TIMSS framework, 38 percent of NAEP items (compared to 29 percent in TIMSS) were classified to the topic of *attributes and units*, with a large number of these items classified to the subtopic of *select appropriate standard units to measure length, area, etc*.

Looking at the two main topics in the TIMSS framework, although more items on both fourth-grade assessments were placed in the topic of *tools, techniques, and formulas* than in *attributes and units*, the emphasis on the former topic is more pronounced in TIMSS than in NAEP. In both assessments, there is still a greater proportion of items in *attributes and units* at the fourth grade than at the eighth grade. Within the *tools, techniques, and formulas* topic, the subtopic with the greatest number of items on both assessments was *compute measurements in simple problem situations*, although there were more TIMSS items addressing this subtopic than NAEP items.

A higher percentage of TIMSS items at both the fourth and eighth grades were classified to the NAEP topic *estimate, calculate, or compare perimeter, area, volume, and surface area,* with 35 percent of TIMSS items compared to 19 percent of NAEP items at fourth grade. At eighth grade, this topic includes the highest proportion of *measurement* items for both NAEP and TIMSS, but the percentage share is much smaller in NAEP (33 percent compared to 61 percent in TIMSS). This potential difference in emphasis is in part supported by the fact that within the fairly broad TIMSS topic of *tools, techniques, and formulas*, close to a third of TIMSS items came from the TIMSS subtopic of *select and use appropriate measurement formulas for perimeter of a rectangle, circumference of a circle, areas of plane figures, surface area and volume of rectangular solid,* compared to very few NAEP items classified by the panel in this subtopic (data not shown). Classification to specific *measurement* subtopics in NAEP indicate that most of the items in this area in both assessments at either grade are related to problems involving properties of two-dimensional shapes rather than three-dimensional objects.

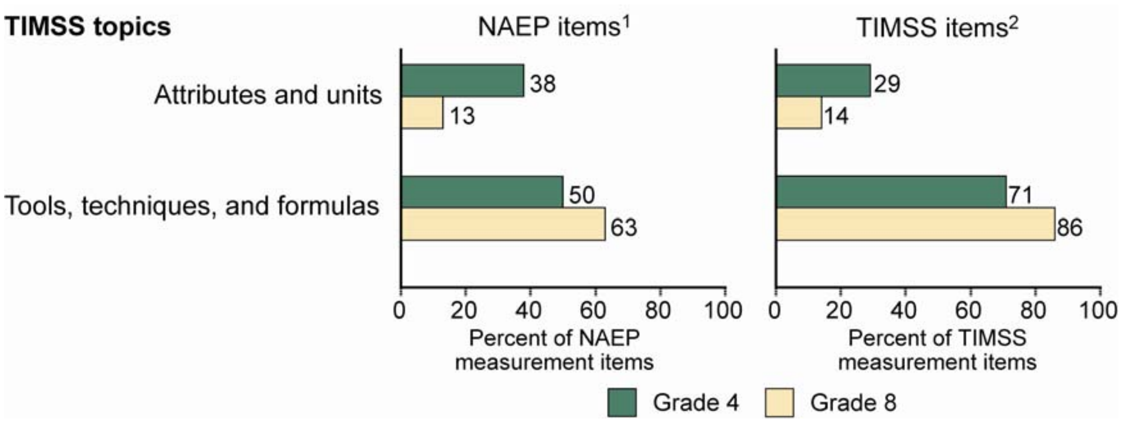
Not surprisingly, there were no items or very few items in either fourth-grade assessment classified to the NAEP topics that were intended for inclusion only at the two higher grades (eighth or twelfth) such as *convert from one measurement to another*, *determine precision, accuracy and error*, and *apply the concept of rate to measurement situations*. Although these NAEP topics were included in the framework at eighth grade, there were also very few eighth-grade items classified to these topics.

Figure 10. Percentage of NAEP and TIMSS measurement items classified to measurement topics in the NAEP mathematics framework, by survey and grade: 2003



[1] NAEP items classified by NAEP developers.
[2] TIMSS items classified by expert panel.
NOTE: Topics may be abbreviated for graphical clarity. Percentages reflect the proportion of *measurement* items classified at either the topic level or the subtopic 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) 2003 Mathematics 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, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002.

Figure 11. Percentage of NAEP and TIMSS measurement items classified to measurement topics in the TIMSS mathematics framework, by survey and grade: 2003



[1] NAEP items classified by expert panel.
[2] TIMSS items classified by TIMSS developers.
NOTE: Percentages reflect the proportion of *measurement* items classified at either the topic level or the subtopic 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) 2003 Mathematics 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 ed.*, 2003.

5.3. Geometry

Similar proportions of the fourth- and eighth-grade items in NAEP and TIMSS are in the area of *geometry*. As discussed in section 4, in NAEP, 15 percent of fourth-grade items and 19 percent of eighth-grade items are from the *content strand* of *geometry and spatial sense* (table 4). In TIMSS, 14 percent of fourth-grade items and 17 percent of eighth-grade items are from the *content domain* of *geometry* (table 5). The results in the *geometry* section are based on 28 fourth-grade and 37 eighth-grade items in NAEP, and 21 fourth-grade and 31 eighth-grade items in TIMSS.

Framework comparison in geometry

Although the NAEP and TIMSS *geometry* frameworks are organized somewhat differently, it is difficult to identify clear cases of skills or knowledge addressed in one framework but not the other based on a comparison at the topic level. Exhibit 5 shows a comparison of the *geometry* topics included in the NAEP and TIMSS mathematics frameworks. The NAEP framework consists of nine topics, six of which are further described by subtopics. Only six topics are intended to be addressed at the fourth-grade level. Most NAEP topics address skills and knowledge relevant to a variety of types of geometric figures, with little specification of types of figures (e.g., lines, angles, quadrilaterals). In contrast, of the five *geometry* topics of the TIMSS framework, two relate to specific types of geometric figures (*lines and angles* and *two- and three-dimensional shapes*); a third (*congruence and similarity*) includes subtopics related almost exclusively to triangles.

One potential difference between the frameworks is the relative lack of topics and subtopics in the TIMSS framework that deal with logic and reasoning, ones that would correspond to the NAEP topic of *establish and explain relationships involving geometric concepts* and its related subtopics (*make conjectures*, *validate and justify conclusions and generalizations* and *use informal induction and deduction*). Although the NAEP framework topics and subtopics indicate this difference from TIMSS, it does not appear to exist when the frameworks are implemented in the assessments, as only one NAEP item was classified to this topic by the assessment developers.

Exhibit 5. Geometry topics included in the NAEP and TIMSS mathematics frameworks: 2003

NAEP *Geometry and spatial sense topics*	**TIMSS** *Geometry topics*
Describe, visualize, draw, and construct geometric figures	Lines and angles
Investigate and predict results of combining, subdividing, and changing shapes (such as paper folding, dissecting, tilting, rearranging pieces of solids)	Two- and three-dimensional shapes
Identify the relationship (congruence, similarity) between a figure and its image under a transformation	Congruence and similarity
Describe the intersection of two or more geometric figures (grades 8 and 12 only)	Locations and spatial relationships
Classify figures in terms of congruence and similarity, and informally apply these relationships using proportional reasoning where appropriate (grades 8 and 12 only)	Symmetry and transformations
Apply geometric properties and relationships in solving problems	
Establish and explain relationships involving geometric concepts	
Represent problem situations with geometric models and apply properties of figures in meaningful contexts to mathematical and real-world problems	
Represent geometric figures and properties algebraically using coordinates and vectors (grades 8 and 12 only)	

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 subtopics or objectives at each grade and level of detail varies across topics and assessments.
SOURCE: U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.*, 2003.

Content and grade match in geometry

The degree to which NAEP and TIMSS *geometry* items were matched to topics and subtopics in the other assessment's framework is shown in table 12. At the fourth-grade level, most NAEP and TIMSS *geometry* items were classified to fourth-grade *geometry* topics or subtopics on the other assessment's framework. Although the percentage with either a specific or general match was higher for NAEP than for TIMSS items (82 percent compared to 76 percent), a higher percentage of TIMSS items had a specific match to the NAEP framework than did NAEP items to the TIMSS framework (57 percent compared to 46 percent). Example 15 in appendix E illustrates a fourth-grade TIMSS item that the panel classified in *geometry and spatial sense* but without a good match to a particular NAEP topic. Example 16 in appendix E shows a fourth-grade NAEP item with a general match to the TIMSS topic of *two- and three-dimensional shapes* but not classified to a specific subtopic in the TIMSS *geometry* framework.

At the eighth grade, the level of match of NAEP items to the TIMSS framework was less precise than the match of TIMSS items to the NAEP framework: 57 percent of NAEP items had a general or specific match to eighth-grade TIMSS *geometry* topics or subtopics, compared to 80 percent of TIMSS items. This is the result of a substantial number of eighth-grade NAEP items being classified to fourth-grade TIMSS topics and subtopics, 43 percent. Most of these items were cross-grade items administered at both fourth and eighth grades. In contrast, a number of TIMSS eighth-grade items (13 percent) were classified to topics in the NAEP framework at the twelfth-grade level. Nearly all of the off-grade items were, however, classified to *geometry* topics. A NAEP *geometry* item administered at the fourth, eighth and twelfth grades that was classified at the fourth-grade level on the TIMSS mathematics framework is shown in Example 17. A TIMSS eighth-grade item that was classified as most consistent with the twelfth-grade level of the NAEP framework and specifications document is shown in example 18 in appendix E.

A review of items and their classifications revealed no single reason based on obvious differences between the frameworks that might explain the lack of general or specific match. The items from one assessment that did not match well to the other assessment framework—either because of content or grade classification—came from a variety of original content classifications.

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

	Grade 4		Grade 8	
Level of content/grade match	NAEP items to TIMSS framework	TIMSS items to NAEP framework	NAEP items to TIMSS framework	TIMSS items to NAEP framework
Total number of geometry items	28	21	37	31
	Percentage distribution			
Classified as same grade	89	81	57	84
Specific match[1] in geometry	46	57	41	45
General match[2] in geometry	36	19	16	35
Match to another content area[3]	7	5	0	3
Classified as another grade[4]	4	10	43	16
Lower grade[5]	†	†	43	3
Higher grade[6]	4	10	†	13
No classification to topics[7]	4	10	0	0

† 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 at the subtopic level at the same grade (and items classified to NAEP framework topics for which no subtopics are included).
[2] Includes items that were classified to a topic but not to a subtopic at the same grade.
[3] Includes items that were classified to a topic or a subtopic in a different content area at the same grade.
[4] Includes items that were classified to a topic or subtopic in any content area at another grade.
[5] Includes grade 8 items classified to grade 4 topics/subtopics.
[6] Includes grade 4 items classified to grade 8 topics/subtopics or grade 8 TIMSS items classified to grade 12 NAEP topics/subtopics.
[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) 2003 Mathematics 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, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.*, 2003.

Item distribution across geometry topics

The distribution of items across NAEP and TIMSS *geometry* topics is shown in figures 12 and 13. At the fourth-grade level, on the NAEP framework, the topic in which the highest percentage of both NAEP and TIMSS items were classified was *describe, visualize, draw, and construct geometric figures*. NAEP included relatively more of these items than TIMSS (43 percent compared to 29 percent). Typically, items in this topic required students to use their knowledge of the properties of figures to either answer questions about them or draw geometric figures based on given characteristics.

Another difference between NAEP and TIMSS at the fourth-grade level is that TIMSS includes items related to congruence and similarity whereas NAEP does not. Fourteen percent of TIMSS items were classified to the TIMSS topic of *congruence and similarity*. These same items were classified to the NAEP framework as *classify figures in terms of congruence and apply proportional reasoning*. In contrast, no fourth-grade NAEP items were placed in either the TIMSS topic of *congruence and similarity* or the corresponding NAEP topic which was intended only for assessment at the eighth or twelfth grade.[7]

Using the TIMSS framework as a reference, the distributions of both NAEP and TIMSS eighth-grade items is roughly similar to that for fourth grade. The highest percentage of items on both assessments are classified to the topic of *two- and three-dimensional shapes*, though the relative proportions are greater for the NAEP assessment. Using the NAEP framework, on the other hand, reveals some differences. Although at eighth grade there is still an emphasis with NAEP for items on *describe, visualize, draw, and construct geometric figures,* the percentage of items classified to that topic is lower (22 percent) than at the fourth grade and is closer to the percentage of TIMSS eighth-grade items classified to that topic (16 percent).

Another notable difference is that at the eighth-grade level, a higher percentage of TIMSS items were classified to the NAEP topic of *apply geometric properties and relationships in solving problems*, with more than 40 percent of TIMSS items classified to this topic, compared to 16 percent of NAEP eighth-grade items. In comparison, 5 percent or less of fourth-grade items in either assessment were classified to this topic. Example 19 in appendix E shows an eighth-grade TIMSS item involving the measure of angles in a hexagon that was classified to the topic of *apply geometric properties and relationships in solving problems* in the NAEP mathematics framework.

It should not be concluded, however, that TIMSS places a greater emphasis on problem solving than NAEP, since 11 percent of eighth-grade NAEP items were classified to another NAEP topic that specifically addressed problem solving, *represent problem situations with geometric models and apply properties of figures*. No TIMSS items were placed in this topic.

A review of eighth-grade items indicates that the emphasis on problem solving in TIMSS is connected to the inclusion of numerous items that involved finding angle measures, for which the most appropriate topic was one that included the phrase *solving problems*. All but one of the TIMSS items classified to the NAEP framework as *apply geometric properties and relationships in solving problems* required students to find angle measures by relying on, among other things, their

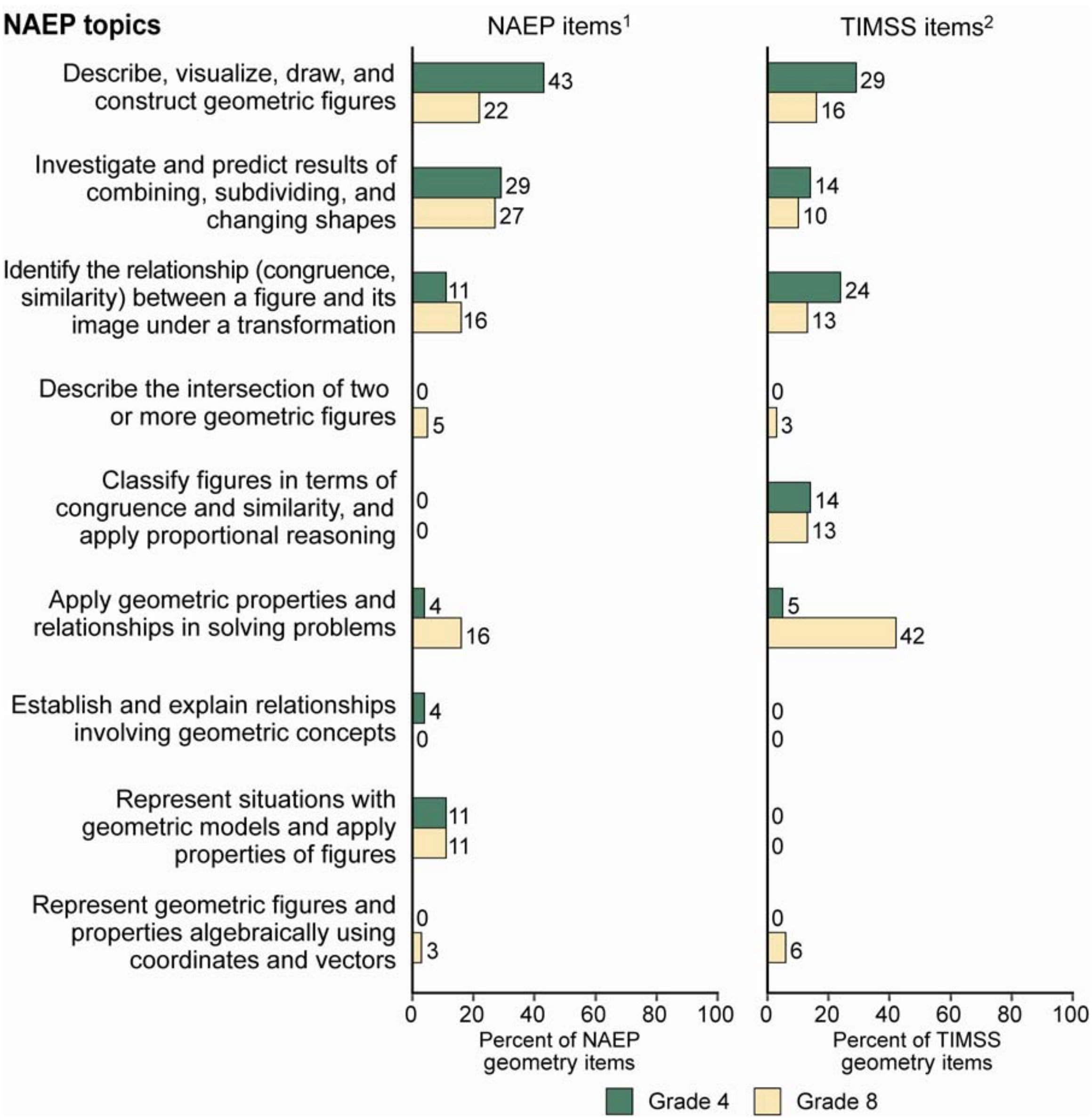
---

[7] By its name, the NAEP topic of *identify the relationship (congruence, similarity) between a figure and its image under a transformation*, would appear to include items related to congruence and similarity, but in fact none of the NAEP or TIMSS items classified in it dealt with congruence or similarity. Instead, most focused on the transformation aspect of the topic.

knowledge of the properties of angles and geometric figures and algebraic skills (data not shown). In contrast, only one NAEP item involved finding angle measures. This difference does appear to be grounded in the frameworks, as the TIMSS items dealing with angle measures could not be placed in any specific NAEP subtopic.

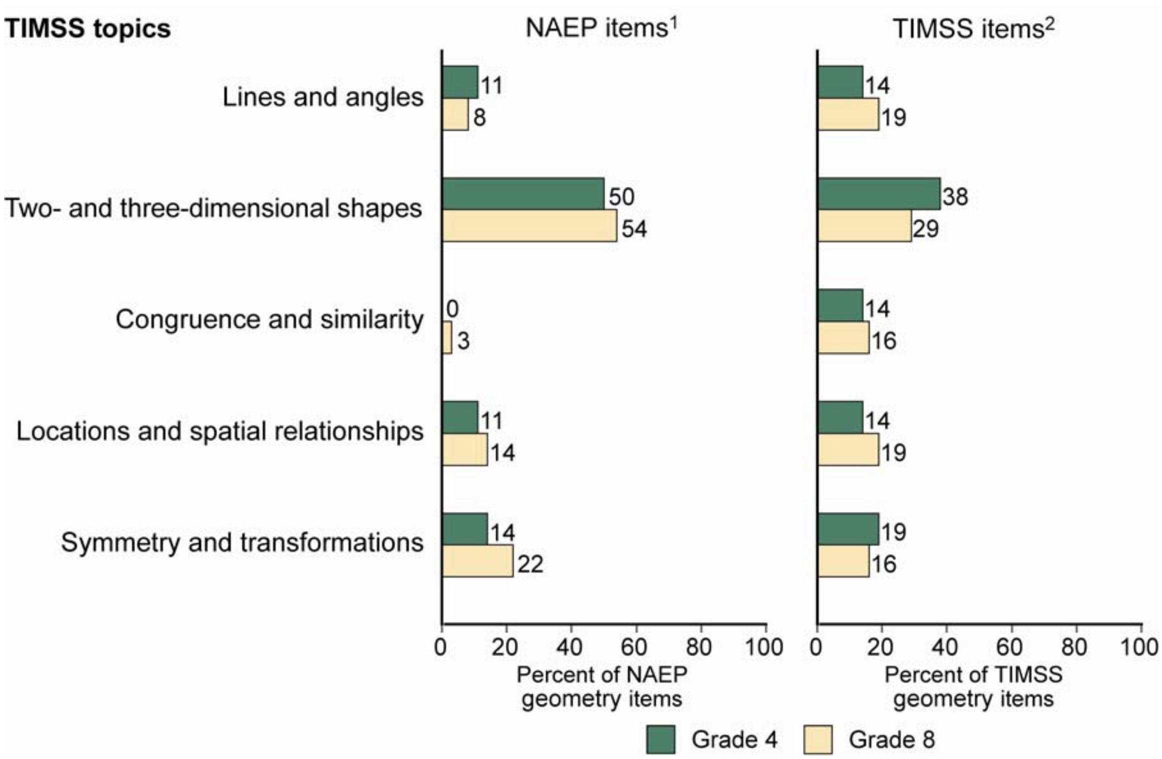
At both the eighth- and the fourth-grade levels, relatively more NAEP items than TIMSS items were classified to the NAEP topic of *investigate and predict results of combining, subdividing, and changing shapes* and to the TIMSS topic of *two- and three-dimensional shapes*. Neither assessment had separate geometry subtopics that facilitated the classification of items on the basis of the use of two- versus three-dimensional shapes. However, further review of the items indicated that there was no substantial difference between the NAEP and TIMSS assessments with respect to their relative emphasis. Both assessments included a relatively small proportion of *geometry* items that involved properties, spatial relationships, or transformations involving three-dimensional objects. This was also true in the *measurement* content area where there were many more items involving measurements of area and perimeter of two-dimensional shapes than volume and surface area of three-dimensional shapes. Both assessments also included some items that required students to recognize the relationship between two-dimensional and three-dimensional shapes (e.g., use of perspective, results of folding, two-dimensional nets of boxes).

Figure 12. Percentage of NAEP and TIMSS geometry items classified to geometry and spatial sense topics in the NAEP mathematics framework, by survey and grade: 2003



[1] NAEP items classified by NAEP developers.
[2] TIMSS items classified by expert panel.
NOTE: Topics may be abbreviated for graphical clarity. Percentages reflect the proportion of *geometry* items classified at either the topic level or the subtopic 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) 2003 Mathematics 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, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002.

Figure 13. Percentage of NAEP and TIMSS geometry items classified to geometry topics in the TIMSS mathematics framework, by survey and grade: 2003



[1] NAEP items classified by expert panel.
[2] TIMSS items classified by TIMSS developers.
NOTE: Percentages reflect the proportion of *geometry* items classified at either the topic level or the subtopic 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) 2003 Mathematics 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 ed.,* 2003.

5.4. Data

The *data* content area reflects the lowest proportion of both NAEP and TIMSS items at either grade. As discussed in section 4, in NAEP, 10 percent of fourth-grade items and 15 percent of eighth-grade items are from the *content strand* of *data analysis, statistics, and probability* (table 4). In TIMSS, 10 percent of fourth-grade items and 13 percent of eighth-grade items are from the *content domain* of *data* (table 5). The results in the *data* section are based on 19 fourth-grade and 30 eighth-grade items in NAEP, and 15 fourth-grade and 23 eighth-grade items in TIMSS.

Framework comparison in data

Exhibit 6 shows a comparison of the *data* topics included in the NAEP and TIMSS mathematics frameworks. Although the NAEP framework contains more topics than the TIMSS framework (11 compared to 4), in both frameworks there is a clear distinction between topics related to data (e.g., data collection, organization, presentation, and interpretation) and those related to probability. The NAEP framework includes three topics related to probability (*determine the number of ways an event can occur*, *determine the probability of a simple event*, and *apply the basic concept of probability to real-world situations*), while the TIMSS framework includes a single topic of *uncertainty and probability*. The remaining three TIMSS topics cover the *collection*, *organization*, *representation*, and *interpretation of data*. The same could be said of the remaining eight NAEP topics, but it should be noted that three of these topics do not have obvious analogues in the TIMSS framework, either at the topic or subtopic level (*understand and reason about the use and misuse of statistics in our society*, *fit a line or curve to a set of data*, and *design a statistical experiment*).

It is also important to remember that of the numerous NAEP topics and subtopics, only a few are intended to be assessed at the fourth-grade level, and a larger but still limited set is intended for assessment at the eighth-grade level. Two of the topics in the NAEP framework are included at the twelfth grade only—*use measure of central tendency, correlation, dispersion and shapes of distributions to describe statistical relationships* and *fit a line or curve to a set of data and use this line or curve to make predictions about the data, using frequency distributions where appropriate*. In TIMSS, the topic of *uncertainty and probability* is not intended to be addressed at the fourth-grade level.

In the NAEP framework, there appears to be some overlap of subtopics across the topics related to *data*, which could complicate attempts to analyze the distribution across topics. For example, the NAEP topic of *organize and display data and make inferences* includes subtopics and grade-level illustrations that also may correspond to another NAEP topic, *read, interpret, and make predictions using tables and graphs*.

Exhibit 6. Data topics included in the NAEP and TIMSS mathematics frameworks: 2003

NAEP *Data analysis, statistics, and probability topics*	TIMSS *Data topics*
Read, interpret, and make predictions using tables and graphs Organize and display data and make inferences Understand and apply sampling, randomness, and bias in data collection (grades 8 and 12 only) Describe measures of central tendency and dispersion in real-world situations Use measure of central tendency, correlation, dispersion and shapes of distributions to describe statistical relationships (grade 12 only) Understand and reason about the use and misuse of statistics in our society Fit a line or curve to a set of data and use this line or curve to make predictions about the data, using frequency distributions where appropriate (grade 12 only) Design a statistical experiment to study a problem and communicate the outcomes (grades 8 and 12 only) Use basic concepts, trees, and formulas for combinations, permutations, and other counting techniques to determine the number of ways an event can occur (grades 8 and 12 only) Determine the probability of a simple event Apply the basic concept of probability to real-world situations	Data collection and organization Data representation Data interpretation Uncertainty and probability (grade 8 only)

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 subtopics or objectives at each grade and level of detail varies across topics and assessments.
SOURCE: U.S. Department of Education, National Assessment Governing Board, Mathematics *Framework for the 2003 National Assessment of Educational Progress*, 2002; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.,* 2003.

Content and grade match in data

Table 13 shows the results of the level of content and grade match analyses for the NAEP and TIMSS *data* items. The most striking results from the cross-classification of items to the two frameworks are that 47 percent of fourth-grade NAEP *data* items were classified to topics at the eighth-grade level on the TIMSS framework, whereas all fourth-grade TIMSS items had a specific match to the NAEP fourth-grade framework. Most of the NAEP items that were classified at the eighth-grade level on the TIMSS framework deal with probability, a topic that is not intended to be addressed at the fourth-grade level in TIMSS. An additional 11 percent of NAEP items were not given a grade classification at the topic level on the TIMSS framework.

Of the fourth-grade NAEP items that were placed at the fourth-grade level on the TIMSS framework, most had a specific match (37 percent of all items, compared to 5 percent that had a general match).

Similar to the fourth-grade items, eighth-grade TIMSS *data* items were classified to the NAEP framework more precisely than NAEP items to the TIMSS framework, although the contrast was not as great: for TIMSS, more than 80 percent of items had a specific match to the NAEP framework compared to the 67 percent of NAEP items with a specific match to the TIMSS framework.

Unlike the fourth-grade assessment, where the high percentage of NAEP items dealing with probability resulted in a relatively high percentage of items classified to a higher grade level, at the eighth-grade level, there is no single reason for the lack of specific match between NAEP items and the TIMSS framework. It is worth noting, however, that some of the NAEP items that were not given a subtopic classification on the TIMSS framework required students to compute either the median or the mode of a single dataset. The related TIMSS subtopics were strictly interpreted by the panel to be limited to comparisons of measures of central tendency across data sets. Other NAEP items were found to deal with combinations, a common area in probability, but not one mentioned explicitly in the TIMSS subtopics in *uncertainty and probability*.

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

	Grade 4		Grade 8	
Level of content/grade match	NAEP items to TIMSS framework	TIMSS items to NAEP framework	NAEP items to TIMSS framework	TIMSS items to NAEP framework
Total number of data items	19	15	30	23
		Percentage distribution		
Classified as same grade	42	100	83	91
Specific match[1] in data	37	100	67	83
General match[2] in data	5	0	17	0
Match to another content area[3]	0	0	0	9
Classified as another grade[4]	47	0	10	4
Lower grade[5]	†	†	10	4
Higher grade[6]	47	0	†	0
No classification to topics[7]	11	0	7	4

† 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 at the subtopic level at the same grade (and items classified to NAEP framework topics for which no subtopics are included).
[2] Includes items that were classified to a topic but not to a subtopic at the same grade.
[3] Includes items that were classified to a topic or a subtopic in a different content area at the same grade.
[4] Includes items that were classified to a topic or subtopic in any content area at another grade.
[5] Includes grade 8 items classified to grade 4 topics/subtopics.
[6] Includes grade 4 items classified to grade 8 topics/subtopics or grade 8 TIMSS items classified to grade 12 NAEP topics/subtopics.
[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) 2003 Mathematics 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, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.,* 2003.

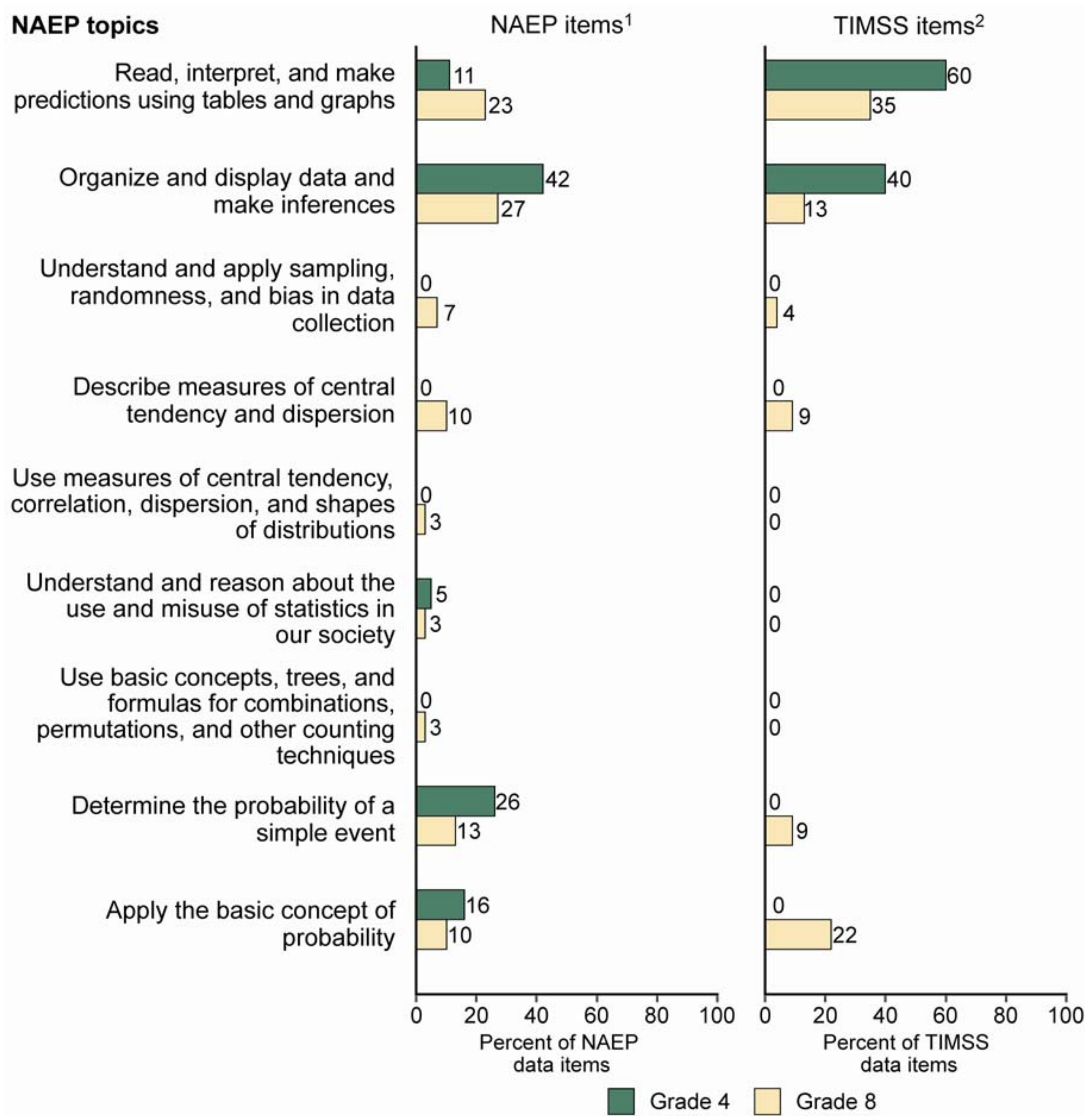
Item distribution across data topics

Figures 14 and 15 show the distribution of NAEP and TIMSS *data* items across the *data* topics included in the two frameworks. As noted above, the most notable difference between NAEP and TIMSS fourth-grade *data* items is that NAEP includes several items dealing with probability (about 40 percent), whereas TIMSS does not include any at the fourth grade. These NAEP items are clearly identifiable both in their classifications to the NAEP framework (in the topics of *determine the probability of a simple event* and *apply the basic concept of probability to real-world situations*) and in their cross-classification to the TIMSS framework, in the topic of *uncertainty and probability*.

More fourth-grade TIMSS items were classified to the NAEP topic of *read, interpret, and make predictions using tables and graphs* than were NAEP items (60 percent compared to 11 percent). Within this topic, both NAEP and TIMSS include a few items that require students to perform some sort of computation based on information provided in a table or graph. However, most of the TIMSS items classified to this topic required students to either answer a straightforward question based on a given graph or table or choose from a set of tables or graphs the one that best represents a set of data. Most of these items were originally classified to the TIMSS framework in the topic of *data representation*, which helps explain the higher percentage of TIMSS items classified to that TIMSS topic. A TIMSS *data* item of this nature is demonstrated by Example 20 in appendix E.

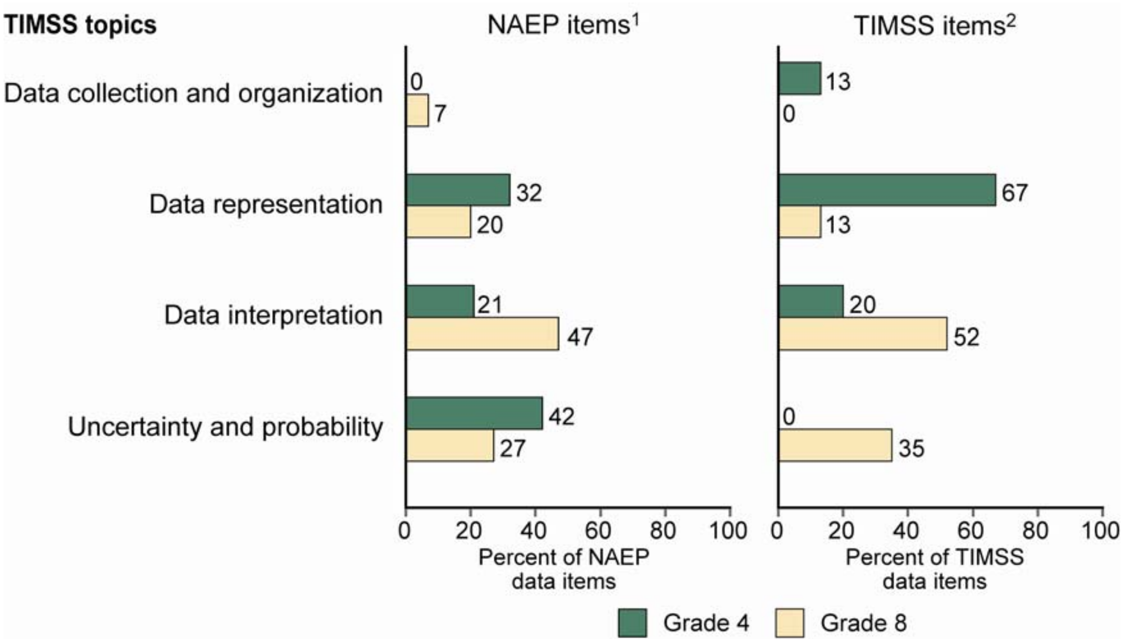
For NAEP and especially TIMSS, the relative emphasis on *data representation* at the fourth grade shifts to an increased emphasis on *data interpretation* at the eighth grade. Compared to fourth grade, at the eighth-grade, there were less dramatic differences between the distribution of NAEP and TIMSS *data* items across both the NAEP and TIMSS frameworks, with a large concentration of items in topics related to the display or interpretation of data. At the eighth grade, NAEP and TIMSS had more comparable numbers of items covering topics related to probability. One of the differences between the two eighth-grade assessments is the higher percentage of NAEP items classified to the NAEP topic of *organize and display data and make inferences* (27 percent compared to 13 percent in TIMSS).

Figure 14. Percentage of NAEP and TIMSS data items classified to data analysis, statistics, and probability topics in the NAEP mathematics framework, by survey and grade: 2003



[1] NAEP items classified by NAEP developers.
[2] TIMSS items classified by expert panel.
NOTE: Topics may be abbreviated for graphical clarity. Two NAEP framework topics included for assessment at eighth and/or twelfth grade(s) only are not reflected in this figure, as no grade 4 or grade 8 items in either assessment were classified to these topics: *fit a line or curve to a set of data and use this linear curve to make predictions about the data, using frequency distributions where appropriate* and *design a statistical experiment to study a problem and communicate the outcomes*. Percentages reflect the proportion of *data* items classified at either the topic level or the subtopic 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) 2003 Mathematics 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, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002.

Figure 15. Percentage of NAEP and TIMSS data items classified to data topics in the TIMSS mathematics framework, by survey and grade: 2003



[1] NAEP items classified by expert panel.
[2] TIMSS items classified by TIMSS developers.
NOTE: Percentages reflect the proportion of *data* items classified at either the topic level or the subtopic 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) 2003 Mathematics 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 ed.,* 2003.

5.5. Algebra

Algebra items make up a larger proportion of the eighth-grade than the fourth-grade items for both NAEP and TIMSS. As discussed in section 4, in NAEP, 14 percent of fourth-grade items and 25 percent of eighth-grade items are from the *content strand* of *algebra and functions* (table 4). In TIMSS, 16 percent of fourth-grade items and 23 percent of eighth-grade items are from the *content domain* of *algebra* (table 5). The results in the *algebra* section are based on 26 fourth-grade and 49 eighth-grade items in NAEP, and 23 fourth-grade and 42 eighth-grade items in TIMSS.

Framework comparison in algebra

The NAEP and TIMSS *algebra* frameworks are compared in exhibit 7. Exhibit 7 shows a comparison of the *algebra* topics included in the NAEP and TIMSS mathematics frameworks. The NAEP framework includes several topics and subtopics included exclusively or primarily for the twelfth-grade assessment that are not addressed in the TIMSS framework. Among these are the topics of *solve polynomial equations with real and complex roots using a variety of algebraic and graphical methods and using appropriate tools*, *apply function concepts to model and deal with real-world situations*, and *use trigonometry*. As in the other content areas, the TIMSS framework includes a few broad topics with a number of grade-specific subtopics within each. Only the *algebraic expressions* topic is not to be assessed at the fourth grade in TIMSS.

Considering only the NAEP topics and subtopics included for the fourth and eighth grades, there is rough agreement between the two frameworks regarding *algebra* content, with a few exceptions. The topics related to patterns appear to be covered in both frameworks, in TIMSS under the topic of *patterns* and in NAEP under the topic of *describe, extend, interpolate, transform, and create a wide variety of patterns and functional relationships*. However, although functions and functional thinking are addressed by both frameworks, the correspondence between NAEP and TIMSS topics and subtopics is not obvious.

There is also some correspondence between the TIMSS topics of *algebraic expressions* and *equations and formulas* and the NAEP topics of *represent and describe solutions to linear equations and inequalities* and *interpret contextual situations and perform algebraic operations on real numbers and algebraic expressions to solve mathematical and real-world problems*. In TIMSS, however, the subtopics indicate a focus on simplifying and evaluating algebraic expressions whereas the subtopics in the NAEP framework indicate a greater focus on problem solving.

One clear difference is that the NAEP framework includes an explicit *algebra* topic and related subtopics (*make conjectures, validate and justify conclusions and generalizations, use formal induction, and deduction*) that deal with *mathematical reasoning* whereas the TIMSS *algebra content domain* does not. Rather, the TIMSS framework specifies related abilities within its *cognitive domain of reasoning* that goes across all *content domains*. Another difference is that the NAEP subtopic of *identify or graph sets of points on a number line or in a rectangular coordinate system* appears more similar to a subtopic of the TIMSS *geometry* topic of *locations and spatial relationships* (*locate points using number lines, coordinate grids, maps*) than any TIMSS topic or subtopic found in *algebra*.

Exhibit 7. Algebra topics included in the NAEP and TIMSS mathematics frameworks: 2003

NAEP ***Algebra and functions topics***	**TIMSS** ***Algebra topics***
Describe, extend, interpolate, transform, and create a wide variety of patterns and functional relationships	Patterns
Use multiple representations for situations to translate among diagrams, models, and symbolic expressions	Algebraic expressions (grade 8 only)
Use number lines and rectangular coordinate systems as representational tools	Relationships
Represent and describe solutions to linear equations and inequalities to solve mathematical and real-world problems	Equations and formulas
Interpret contextual situations and perform algebraic operations on real numbers and algebraic expressions to solve mathematical and real-world problems (grades 8 and 12 only)	
Solve systems of equations and inequalities using appropriate methods (grades 8 and 12 only)	
Use mathematical reasoning	
Represent problem situations with discrete structures (grades 8 and 12 only)	
Solve polynomial equations with real and complex roots using a variety of algebraic and graphical methods using appropriate tools (grade 12 only)	
Approximate solutions of equations (bisection, sign changes, successive approximations) (grades 8 and 12 only)	
Use appropriate notation and terminology to describe functions and their properties (grade 12 only)	
Compare and apply the numerical, symbolic, and graphical properties of a variety of functions and families of functions, examining general parameters and their effect on curve shape (grades 8 and 12 only)	
Apply function concepts to model and deal with real-world situations (grades 8 and 12 only)	
Use trigonometry (grade 12 only)	

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 subtopics or objectives at each grade and level of detail varies across topics and assessments.
SOURCE: U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.,* 2003.

Content and grade match in algebra

As seen in table 14, nearly all fourth-grade level NAEP and TIMSS *algebra* items were classified to fourth-grade topics in the other assessment framework (96 percent for both). Of these items, the majority had a specific match to the other assessment's framework, although the percentage of NAEP items with a specific match to the TIMSS framework was higher than the percentage of TIMSS items with a specific match to the NAEP framework (81 percent compared to 65 percent).

At the eighth-grade level, the classification of NAEP items to the TIMSS framework was less precise than that for the TIMSS items to the NAEP framework. While 95 percent of eighth-grade TIMSS items were classified at the eighth-grade level on the NAEP framework (93 percent with a specific match to *algebra* subtopics), 78 percent of eighth-grade NAEP items were classified to eighth-grade topics in TIMSS (57 percent had a specific match to algebra subtopics). The lower level of match between NAEP items and the TIMSS framework is due in large measure to the 20 percent of items classified in *content domains* other than *algebra*. This indicates that the NAEP framework and NAEP items reflect a broader conception of *algebra* than the TIMSS framework or items, including topics that in TIMSS would be found in other *content domains*. Example 21 in appendix E shows an eighth-grade NAEP *algebra* item that was classified to the *number content domain* in TIMSS. This item was from the NAEP *algebra* subtopic of *perform basic operations using appropriate tools, on real numbers in meaningful contexts (including grouping and order of multiple operations involving basic operations, exponents, and roots)*. In addition, nearly 20 percent of NAEP eighth-grade items were classified to fourth-grade TIMSS topics and subtopics. Most were classified to either of the two *algebra* topics of *patterns* and *equations and formulas*. Five percent of TIMSS eighth-grade *algebra* items were classified at the twelfth-grade level according to the NAEP framework.

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

	Grade 4		Grade 8	
Level of content/grade match	NAEP items to TIMSS framework	TIMSS items to NAEP framework	NAEP items to TIMSS framework	TIMSS items to NAEP framework
Total number of algebra items	26	23	49	42
	Percentage distribution			
Classified as same grade	96	96	78	95
Specific match[1] in algebra	81	65	57	93
General match[2] in algebra	0	17	0	0
Match to another content area[3]	15	13	20	2
Classified as another grade[4]	0	4	18	5
Lower grade[5]	†	†	18	0
Higher grade[6]	0	4	†	5
No classification to topics[7]	4	0	4	0

† 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 at the subtopic level at the same grade (and items classified to NAEP framework topics for which no subtopics are included).
[2] Includes items that were classified to a topic but not to a subtopic at the same grade.
[3] Includes items that were classified to a topic or a subtopic in a different content area at the same grade.
[4] Includes items that were classified to a topic or subtopic in any content area at another grade.
[5] Includes grade 8 items classified to grade 4 topics/subtopics.
[6] Includes grade 4 items classified to grade 8 topics/subtopics or grade 8 TIMSS items classified to grade 12 NAEP topics/subtopics.
[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) 2003 Mathematics 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, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002; and International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.,* 2003.

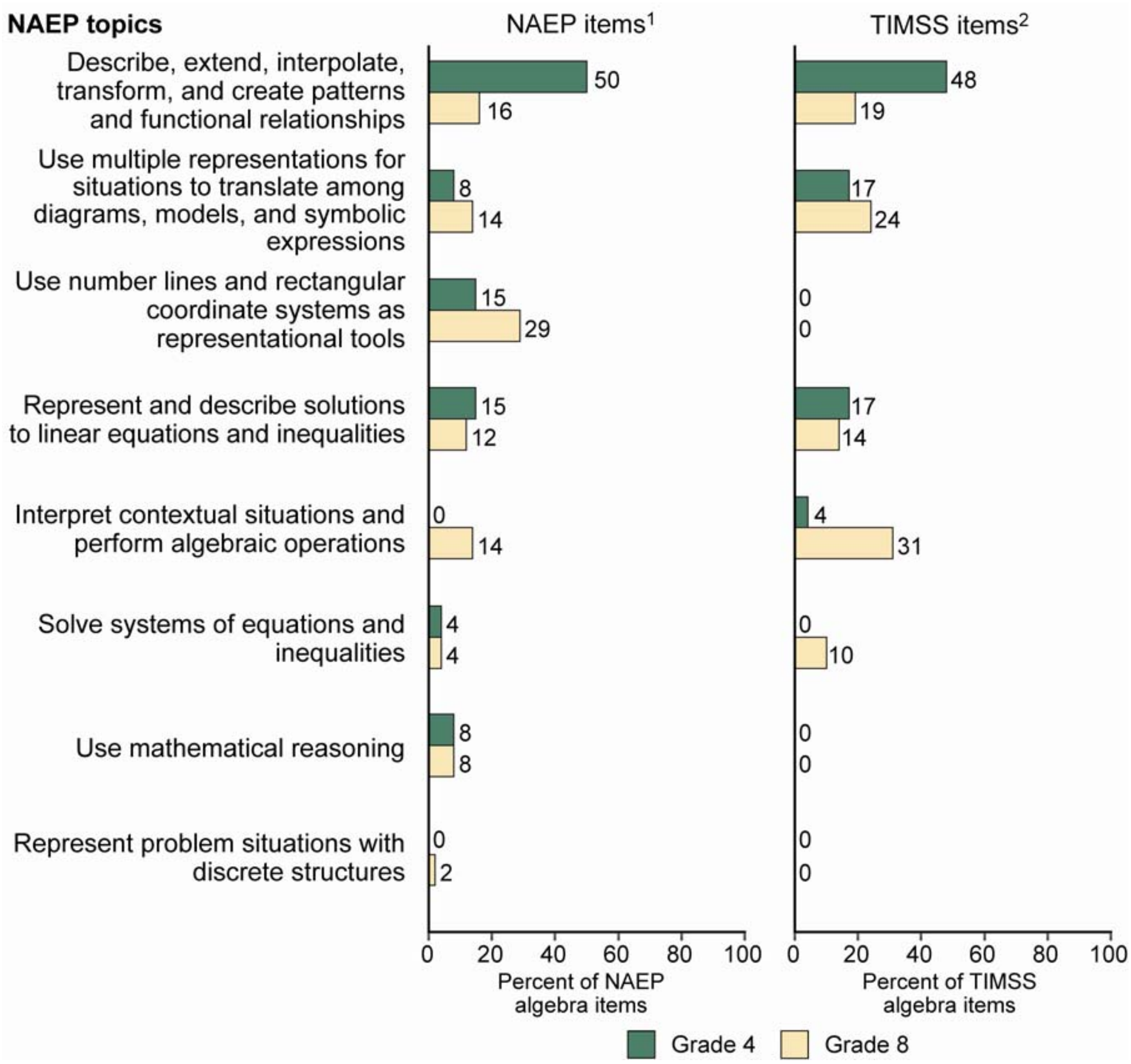
Item distribution across algebra topics

Figures 16 and 17 show the distribution of *algebra* items from each assessment across *algebra* topics in the NAEP and TIMSS framework, respectively. At the fourth-grade level, NAEP and TIMSS items had somewhat similar distributions across NAEP topics, with approximately half of the items on both assessments classified to the topic of *describe, extend, interpolate, transform, and create patterns and functional relationships*. This emphasis on patterns is reflected in the TIMSS framework as well, with 39 percent of items on both assessments classified to the topic of *patterns* and an additional 17 percent of TIMSS items classified to the topic of *relationships*. No fourth-grade NAEP items were classified to the TIMSS topic of *relationships*. Although it is possible that some NAEP items did in fact address this topic but were classified elsewhere, the fact that the panel did place NAEP items in this topic at the eighth-grade level indicates that they would have done so at the fourth-grade level as well had they found any items that matched this TIMSS topic. Neither NAEP nor TIMSS had any fourth-grade items classified to the TIMSS topic of *algebraic expressions*, which is consistent with the definition in the TIMSS framework of this as a topic appropriate for the eighth grade only.

One notable difference at both the fourth- and eighth-grade levels is that sizeable percentages of NAEP items (15 percent at the fourth-grade level and 29 percent at the eighth-grade level) were classified to the NAEP topic of *use number lines and rectangular coordinate systems as representational tools*. In contrast, no TIMSS *algebra* items at either grade level were placed in this topic. This difference at the item level reflects one of the differences in the *algebra* frameworks discussed above, that this NAEP topic is perhaps more closely aligned with TIMSS topics and subtopics in the *geometry content domain*. In fact, when the panel placed these NAEP items on the TIMSS framework, they placed a number of the fourth-grade items in TIMSS *geometry* topics and some of the eighth-grade items in a single subtopic in the TIMSS *geometry* framework (*locate points using number lines, coordinate grids, maps*), which is reflected in the 15–20 percent of NAEP *algebra* items matched to another content area in table 11.

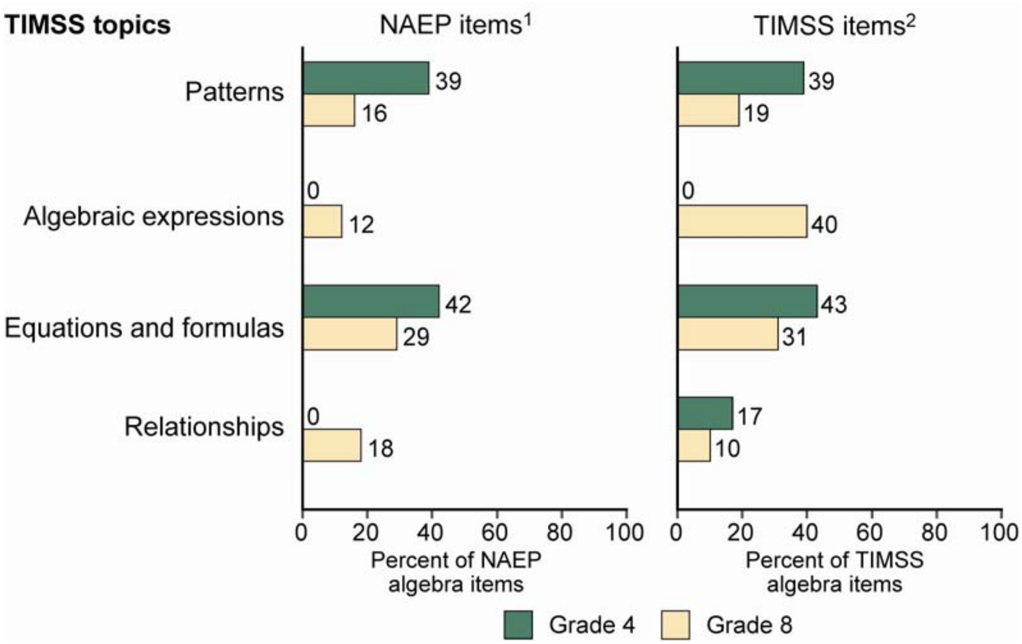
Another difference at the eighth-grade level is that 40 percent of TIMSS items were classified to the TIMSS topic of *algebraic expressions*, compared to 12 percent of NAEP items. All but one of these TIMSS items were classified to the NAEP topics of either *use multiple representations for situations to translate among diagrams, models, and symbolic expressions* or *interpret contextual situations and perform algebraic operations*, which might explain the relatively higher percentages of TIMSS items placed in those NAEP topics.

Figure 16. Percentage of NAEP and TIMSS algebra items classified to algebra and functions topics in the NAEP mathematics framework, by survey and grade: 2003



[1] NAEP items classified by NAEP developers.
[2] TIMSS items classified by expert panel.
NOTE: Topics may be abbreviated for graphical clarity. Six NAEP framework topics included for assessment at eighth and/or twelfth grade(s) only are not reflected in this figure, as no grade 4 or grade 8 items in either assessment were classified to these topics: *solve polynomial equations with real and complex roots using a variety of algebraic and graphical methods and using appropriate tools*; *approximate solutions of equation (bisection, sign changes, successive approximations)s*; *use appropriate notation and terminology to describe functions and their properties*; *compare and apply the numerical, symbolic, and graphical properties of a variety of functions and families of functions, examining general parameters and their effect on curve shape*; *apply function concepts to model and deal with real-world situations*; and *use trigonometry*. Percentages reflect the proportion of *algebra* items classified at either the topic level or the subtopic 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) 2003 Mathematics 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, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002.

Figure 17. Percentage of NAEP and TIMSS algebra items classified to algebra topics in the TIMSS mathematics framework, by survey and grade: 2003



[1] NAEP items classified by expert panel.
[2] TIMSS items classified by TIMSS developers.
NOTE: Percentages reflect the proportion of *algebra* items classified at either the topic level or the subtopic 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) 2003 Mathematics 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 ed.,* 2003.

This section compared the content of the NAEP and TIMSS assessments in each of the main content areas of *number*, *measurement*, *geometry*, *data*, and *algebra*. The next section compares PISA items with NAEP *problem solving* items from the fourth- and eighth-grade assessments. Comparisons are made with respect to content coverage, *competency cluster* and *situations or contexts* dimensions from the PISA framework, and general item characteristics.

6. NAEP/PISA Comparisons

As described in the methods section, a subset of the panel conducted a comparison of PISA items and a specially selected set of NAEP items. The selected set of NAEP items were drawn from the 2003 eighth-grade and the 2000 twelfth-grade NAEP assessments.[8] They included all items from the *mathematical abilities* category of *problem solving* and all items requiring an extended response. All of the extended-response items were from the NAEP *problem solving* category except one item that was classified as *conceptual understanding*. The NAEP and PISA items were compared on several dimensions of the other assessment's framework. Some of the results relating to PISA items are presented in the earlier section on overall comparisons (section 4). PISA items were not included in the sections devoted to each of the content areas (section 5) because PISA is not designed to correspond as closely to the curriculum-based content areas defined by NAEP and TIMSS. This section presents some additional comparisons between NAEP and PISA items related to mathematical content covered in terms of both the NAEP *content strands* and the PISA *overarching ideas* as well as comparisons based on the *competency clusters* and *situations or contexts* dimensions defined in the PISA framework.[9]

It is important to emphasize that the NAEP items included in these comparisons are not a complete set of items for either eighth or twelfth grade, but rather are selected items across the two grade levels that are closest in age to the PISA target population (15-year-olds). Therefore, comparisons made between NAEP and PISA items in this section should not be interpreted as representative of NAEP overall. Instead, the purpose of these comparisons is to compare PISA items with NAEP items that were designed to measure students' problem-solving abilities. Unlike the previous sections, this section includes comparisons based on the dimensions in the PISA framework, which, because of its emphasis on the application of knowledge and skills to real-life problem solving, may provide an additional perspective for examining the NAEP items.

6.1. Content Comparisons Based on NAEP and PISA Frameworks

Although the correspondence between the content dimensions defined in the NAEP and PISA frameworks is not as strong as between NAEP and TIMSS, there was nevertheless considerable overlap of the items from each assessment with the other assessment framework from a strictly mathematics content perspective. In fact, as discussed in the section on overall comparisons (tables 4 and 7), all of the PISA items were classified to at least the broad *content strand* level, with 92 percent classified to a specific subtopic in the NAEP framework. As displayed in tables 15 and 16, there is a strong correspondence between the PISA *overarching idea* of *uncertainty* and the NAEP *content strand* of *data analysis, statistics, and probability*. Ninety-five percent of PISA *uncertainty* items and 81 percent of NAEP *data analysis, statistics, and probability* items were classified to the corresponding content area in the other framework. There also appears to be a strong correspondence between *space and shape* in PISA and *measurement* and *geometry and spatial sense* in NAEP. Most PISA items from *space and shape* (90 percent) and NAEP items from *geometry and spatial sense* (95 percent) and *measurement* (67 percent) were cross-classified to these corresponding content areas across NAEP and PISA frameworks. Even for the *overarching idea*s of *change and relationships* and *quantity*, which have less direct correspondence with a particular NAEP *content strand*, PISA items in these areas still mapped to NAEP topics across the *content strands*.

---

[8] The 2000 assessment was the most recent twelfth-grade NAEP assessment at the time of the expert panel meeting.
[9] For all comparisons based on the PISA framework dimensions, NAEP item classifications were made by the expert panel, while those for PISA items reflect classification information provided by the PISA assessment developers.

Additional information about the distribution of PISA items across NAEP mathematics topics is provided in the supplementary data in appendix D (table D-3). While the *uncertainty* items in PISA are classified across six different topics in the NAEP *content strand* of *data analysis, statistics, and probability*, the majority of items relate to three topics: *read, interpret, and make predictions using tables and graphs*; *describe measures of central tendency and dispersion*; and *understand and reason about the use and misuse of statistics in our society*. The PISA items in *space and shape* cover a range of *measurement* and *geometry* topics in NAEP, with particular focus on the *measurement* topic of *estimate, calculate or compare perimeter, area, volume, and surface area.* Items in *change and relationships* have the greatest focus on the topic *read, interpret, and make predictions using tables and graphs.* The *quantity* items cover a couple of topics in each of the NAEP *content strands* except *geometry*. Not surprisingly, a number of these PISA items were classified as *use computations and estimation in applications*.

In general, the NAEP items were more difficult to match to the PISA framework than vice-versa. While all PISA items were classified to a NAEP *content strand* and most to a specific NAEP subtopic, a number of NAEP items could not be placed in any PISA *overarching idea* (table 16). These items came from three NAEP *content strands*: *number sense, properties, and operations* (12 percent); *geometry and spatial sense* (5 percent), and *algebra and functions* (11 percent). Two of these items are related to logic, a topic not found on the PISA framework, and the other items covered a range of topics in NAEP. There also were NAEP items in all content areas except *geometry* that were classified to multiple *overarching ideas* in PISA.

Table 15. Percentage distribution of PISA mathematics items across NAEP mathematics content strands, by PISA overarching idea: 2003

NAEP content strand	Total	PISA overarching idea: Change and relationships	Quantity	Space and shape	Uncertainty
Number of PISA items	85	22	23	20	20
		Percentage distribution			
Number sense, properties, and operations	22	14	57	10	5
Measurement	16	9	13	45	0
Geometry and spatial sense	11	0	0	45	0
Data analysis, statistics, and probability	39	41	22	0	95
Algebra and functions	9	27	9	0	0
Classified to multiple strands	2	9	0	0	0

NOTE: Percentages reflect the proportion of PISA items classified by the expert panel to the NAEP 2003 mathematics *content strands* at any level of specificity (*content strand*, *topic*, or *subtopic*) and at any grade level. Of the items classified to multiple strands, one item was classified to both *geometry and spatial sense* and to *algebra and functions*, and one was classified to *measurement* and to *data analysis, statistics, and probability*. Detail may not sum to totals because of rounding.
SOURCE: Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment; U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002; and Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*, 2003.

Table 16. Percentage distribution of NAEP 2003 eighth-grade and NAEP 2000 twelfth-grade mathematics problem solving items across PISA overarching ideas, by NAEP content strand: 2000 and 2003

		NAEP content strand				
PISA overarching idea	Total	Number sense, properties, and operations	Measurement	Geometry and spatial sense	Data analysis, statistics, and probability	Algebra and functions
Number of NAEP problem solving items[1]	79	17	9	19	16	18
		Percentage distribution				
Change and relationships	28	47	0	0	13	67
Quantity	9	35	11	0	0	0
Space and shape	32	0	67	95	0	11
Uncertainty	16	0	0	0	81	0
Classified to multiple overarching ideas	8	6	22	0	6	11
Not classified to a PISA overarching idea	6	12	0	5	0	11

[1] NAEP data are based on a selected set of 79 mathematics items from the NAEP 2003 grade 8 and NAEP 2000 grade 12 assessments. This selected set of items consists of all extended-response items and all items from the *problem solving* category of the NAEP 2003 framework. All extended-response items are from the *problem solving* category except one item from the *conceptual understanding* category.
NOTE: Percentages reflect the proportion of NAEP items classified by the expert panel to each *overarching idea* category in the PISA 2003 mathematical literacy framework. Of the items classified to multiple *overarching ideas*, five items were classified to both *change and relationships* and *quantity*, and one item was classified to both *space and shape* and *quantity*. 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) 2003 Mathematics Assessment; U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Mathematics Assessment; U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002; and Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills,* 2003.

6.2. PISA Competency Clusters

As described in the overview section (section 2), the PISA framework defines three *competency clusters*. The *reproduction* cluster involves the reproduction of practiced knowledge, including the recall of facts, execution of routine procedures, and use of standard solution and thinking strategies. The *connections* cluster involves familiar settings similar to the *reproduction* cluster, but goes beyond it by involving solution strategies that are not routine. The *reflection* cluster requires more planning and original thought than do the other two clusters.

The fact that the NAEP items that were compared to the PISA framework were those from the *problem solving* category does not necessarily mean that all would be classified to a particular PISA cluster. All three clusters can involve problem solving and are distinguished more by the balance between the demand for recall of facts or procedures versus more creative thinking and solution strategies. In fact, only 4 percent of the NAEP *problem solving* items overall were classified to PISA's *reflection* cluster, compared to 22 percent of PISA items (table 17). The remaining NAEP items were split between the *reproduction* and the *connections* clusters. When considering the overall set of the NAEP *problem solving* items for both eighth and twelfth grades, a similar proportion of NAEP and PISA items are in the *connections* category (44 and 47 percent), while a greater proportion of NAEP items were classified as *reproduction* (46 percent compared to 31 percent for PISA). A breakdown by grade level for the NAEP items reveals a greater percentage in eighth-grade classified to the *reproduction* cluster, and less to the *connections* cluster compared to twelfth grade. Five percent or less of items from both grades was classified to the *reflection* cluster. Examples 22 and 23 in appendix E illustrate NAEP *problem solving* items classified to the PISA *competency clusters* of *reproduction* and *connections*, respectively. Both of these items are cross-grade items administered at both the eighth and twelfth grades.

Table 17. Percentage distribution of PISA 2003 mathematics items and NAEP 2003 eighth-grade and NAEP 2000 twelfth-grade mathematics problem solving items classified to PISA competency clusters, by grade/age: 2000 and 2003

PISA competency cluster	PISA[1]	NAEP[2]		
	15-year-olds	Total[3]	Grade 8	Grade 12
Reproduction	31	46	58	39
Connections	47	44	36	48
Reflection	22	4	2	5

[1] PISA items classified by PISA developers.
[2] NAEP items classified by expert panel.
[3] NAEP eighth- and twelfth-grade items combined.
NOTE: NAEP data are based on a selected set of 79 mathematics items from the NAEP 2003 grade 8 and NAEP 2000 grade 12 assessments. This selected set of items consists of all extended-response items and all items from the *problem solving* category of the NAEP 2003 framework. All extended-response items are from the *problem solving* category except one item from the *conceptual understanding* category. Percentages reflect the proportion of items classified to each *competency cluster* in the PISA 2003 mathematical literacy framework. Five NAEP items that the panel did not classify to the PISA framework are not included. NAEP cross-grade items administered at grade 8 and grade 12 are reflected in the percentages for both grades, but only counted once in the total column. 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) 2003 Mathematics Assessment; U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Mathematics Assessment; Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment; and Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*, 2003.

Table 18 displays the distribution of NAEP *problem solving* items across PISA *competency clusters* by NAEP *content strand*. These results show that all of the NAEP items classified to the *reflection* cluster came from the *content strands* of *geometry and spatial sense* and *data analysis, statistics, and probability*. In addition, the *data analysis, statistics, and probability content strand* also had a substantially lower percent of *reproduction* items than the set of NAEP items overall (25 percent compared to 46 percent).

Table 18. Percentage distribution of NAEP 2003 eighth-grade and NAEP 2000 twelfth-grade mathematics problem solving items across PISA competency clusters, by NAEP content strand: 2000 and 2003

PISA competency cluster	Total	NAEP content strand: Number sense, properties, and operations	Measurement	Geometry and spatial sense	Data analysis, statistics, and probability	Algebra and functions
Number of NAEP problem solving items[1]	79	17	9	19	16	18
		Percentage distribution				
Reproduction	46	53	56	53	25	44
Connections	44	35	44	37	63	44
Reflection	4	0	0	5	13	0

[1] NAEP data are based on a selected set of 79 mathematics items from the NAEP 2003 grade 8 and NAEP 2000 grade 12 assessments. This selected set of items consists of all extended-response items and all items from the *problem solving* category of the NAEP 2003 framework. All extended-response items are from the *problem solving* category except one item from the *conceptual understanding* category.
NOTE: Percentages reflect the proportion of NAEP items classified by the expert panel to each *competency cluster* in the PISA 2003 mathematical literacy framework. Five NAEP items that the panel did not classify to the PISA framework 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) 2003 Mathematics Assessment; U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Mathematics Assessment; and Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*, 2003.

6.3. PISA Situations or Contexts

The *situations or contexts* dimension of the PISA mathematical literacy framework describes the particular context or situation in which a student is engaged in the application of mathematics. These situations are considered to be part of a students' everyday experience within which mathematical tasks are presented and there is a need for mathematical problem solving (OECD 2003). The PISA 2003 framework outlines an ordered taxonomy of four *situations or contexts* based on the "distance" they are from students' experience as explained by the framework. The four *situations or contexts*, starting with the closest to the student, are as follows: *personal, educational/occupational, public,* and *scientific*.

Table 19 compares the distribution of the PISA items and NAEP *problem solving* items across the PISA *situations or contexts* categories. For the set of NAEP items classified by the panel, there is a smaller percentage of items in the *personal* category compared to the PISA items (8 percent compared to 21 percent). There also is a smaller percentage of NAEP items addressing *public situations* (22 percent compared to 34 percent in PISA) and a somewhat greater percentage addressing *educational/occupational situations* (30 percent compared to 24 percent in PISA). A distinction can also be made with regards to the *scientific context*, with 34 percent of the NAEP items compared to 20 percent of PISA items. Consistent with its design, PISA has a reasonably balanced representation across the *situations or contexts*, with a slight emphasis on the *public* category.

Table 19. Percentage distribution of PISA 2003 mathematics items and NAEP 2003 eighth-grade and NAEP 2000 twelfth-grade mathematics problem solving items across PISA situations or contexts categories: 2000 and 2003

PISA situations or contexts category	PISA[1]	NAEP[2]
Personal	21	8
Educational/occupational	24	30
Public	34	22
Scientific	20	34

[1] PISA items classified by PISA developers.
[2] NAEP items classified by expert panel.
NOTE: NAEP data are based on a selected set of 79 mathematics items from the NAEP 2003 grade 8 and NAEP 2000 grade 12 assessments. This selected set of items consists of all extended-response items and all items from the *problem solving* category of the NAEP 2003 framework. All extended-response items are from the *problem solving* category except one item from the *conceptual understanding* category. Percentages reflect the proportion of items classified to each *situations or contexts* category in the PISA 2003 mathematical literacy framework. Five NAEP items that the panel did not classify to the PISA framework, and one PISA item, for which a *situations or contexts* classification was not provided by the assessment developers, 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) 2003 Mathematics Assessment; U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2000 Mathematics Assessment; Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment; and Organization for Economic Cooperation and Development (OECD), *Program for International Student Assessment (PISA) 2003 Assessment Framework*, 2003.

6.4. Comparing General Characteristics of NAEP and PISA Items

In addition to classifying NAEP *problem solving* items to the PISA framework dimensions, the panel also evaluated these items with respect to whether or not they could appear on the PISA assessment and, if not, in what way the characteristics of the NAEP items were different from those that might appear on PISA. While some of the NAEP *problem solving* items reviewed by the panel were judged as likely to appear in PISA as they were, a substantial number were identified as not appropriate for PISA. Others were identified as requiring revision in order to be included in the PISA assessment. Some of the general characteristics noted by the panel that distinguished the NAEP and PISA items included the authenticity of the *problem solving context*, the nature of the mathematical application, the level of instructions given, and the general formatting and sequencing of items. The panel noted that scaffolding—breaking an item into component parts and presenting the parts as a series, in steps of increasing complexity to engage students in the task—is used more in NAEP than in PISA.

In general, the NAEP items are presented in isolation, while PISA typically presents a series of items related to a particular *problem solving situation*. In some cases, the panel noted that individual NAEP items might be appropriate for PISA if included as part of a series of items in a larger task. Also, there were other NAEP items that presented a series of questions all within one item, but these were all scored together on a single rubric. In contrast, the panel noted that PISA would present these as separate questions within a larger task, and each would be scored separately.

Some example items are included in appendix E to illustrate some of these general item characteristics noted in comparing NAEP with PISA. Example 23 shows a NAEP *problem solving* item that the panel judged as being appropriate for the PISA assessment. Example 24 illustrates a NAEP *problem solving* item that the panel believed would not appear in the PISA assessment because it is based on a contrived situation. Example 25 was identified as a NAEP problem that might appear on PISA after revision. The panel noted that the presentation of this item would be revised for PISA by breaking it into a series of questions to be assessed and scored separately.

Example 26 shows a PISA graphical interpretation item, which asks students to draw conclusions based on the data in the graph.

This section compared the PISA assessment with NAEP fourth- and twelfth-grade *problem solving* items. The last section includes a summary and conclusion of the findings of this comparison study of NAEP , TIMSS, and PISA.

7. Conclusion

In summary, the content comparisons between NAEP, TIMSS, and PISA reveal some key differences in the mathematics 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 useful to consider these differences when interpreting the results from the three assessments. In addition, the PISA assessment, with its focus on problem solving and the application of mathematics in real-world situations, provides additional information on student performance that is complementary to that obtained from NAEP and TIMSS.

With respect to NAEP and TIMSS, a comparison of the frameworks revealed considerable agreement on the general boundaries and basic organization of mathematics content, with both assessments including five main content areas corresponding to traditional mathematics curricular areas related to *number*, *measurement*, *geometry*, *data*, and *algebra*. Both NAEP and TIMSS frameworks also include dimensions that define a range of cognitive skills and processes that overlap across the two assessments. Despite some apparent similarities at the broadest level, a closer examination of the items in each assessment reveals different emphases placed at the topic and subtopic level, as well as some differences in grade level expectations across mathematics topics. 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 mathematics.

PISA stands apart from NAEP and TIMSS in a number of important areas including a mathematics framework organized around *overarching ideas* rather than curriculum-based content areas, a focus on problem solving in real-world applications, and sampling an age-based population of secondary school students (15-year-olds). Based on the results from this study, PISA also includes larger proportions of constructed-response items and items classified at higher levels of *mathematical complexity* than either NAEP or TIMSS. Although PISA is distinct from NAEP and TIMSS in numerous ways, there are still some similarities when the PISA items are directly compared with the other assessments. The mathematics content covered by most PISA items is consistent with topics included in the NAEP eighth-grade mathematics framework.

This report provides a first-level comparison of items in each assessment in terms of the coverage of broad content areas and distribution across mathematics topics as defined in the frameworks. All items in each assessment were considered in order to make overall comparisons of content coverage and grade-level expectations as well as distributions with respect to three broad levels of mathematical complexity. In addition, the types of item classifications conducted within the time constraints of this study permit comparisons at the mathematics topic level for each content area. While this method provides a broad view of some of the similarities and differences between the assessments, it is limited in terms of the types of comparisons that are provided at the item level. More in depth analyses of the exact nature of the items from each assessment within topics would reveal other important differences related to difficulty, scope, depth, complexity, and other item attributes. These types of more focused comparisons were outside the scope of this study, but may be important to include in future comparative studies of the assessments.

References

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.

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 ed*. Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.

National Assessment Governing Board (NAGB). (1992). *1996 Mathematics Specifications for the National Assessment of Educational Progress (NAEP)*. Washington, DC: Author.

National Assessment Governing Board (NAGB). (2002). *Mathematics Framework for the 2003 National Assessment of Educational Progress.* Washington, DC: Author.

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

Neidorf, T.S., Binkley, M., and Stephens, M. (2006). *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.

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 Programme 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.

Organization for Economic Cooperation and Development (OECD). (2003). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills.* Paris: Author.

Appendix A: Content Framework Summary Documents

This appendix presents information about the NAEP 2003 and TIMSS 2003 mathematics content frameworks used for item classifications at the expert panel meeting.

Exhibit A-1 is the summary document that was used by the expert panel for the classification of items to the *content strands*, *topics*, and *subtopics* in the NAEP 2003 mathematics framework.

Exhibit A-2 is the summary document that was used by the expert panel for the classification of items to the *content domains*, *topic areas*, and *objectives* in the TIMSS 2003 mathematics framework.

These summary documents are based on the NAEP 2003 and TIMSS 2003 framework and assessment specifications documents, but have been reformatted and adapted slightly to facilitate the classification process.

Framework summary documents were not prepared for PISA; expert panel content classifications based on the PISA framework were made only at the *overarching idea* level, and there are no topics or subtopics specified in the PISA framework. For more information about the PISA framework, see OECD (2003).

Exhibit A-1. NAEP mathematics framework and specifications summary: 2003

A	NUMBER SENSE, PROPERTIES AND OPERATIONS	Grade(s)		
A1	**Relate counting, grouping and place value**			
1a:	Use place value to model and describe whole numbers and decimals	4	8	12
1b:	Use scientific notation in meaningful contexts	•	8	12
A2	**Represent numbers and operations in a variety of equivalent forms using models, diagrams and symbols**			
2a:	Model numbers using set models such as counters	4	•	•
2b:	Model numbers using number lines	4	8	•
2c:	Use two- and three-dimensional region models to describe numbers	4	8	12
2d:	Use other models appropriate to a given situation (e.g., draw diagrams to represent a number or an operation; write a number sentence to fit a situation or describe a situation to fit a number sentence; interpret calculator or computer displays)	4	8	12
2e:	Read, write, rename, order, and compare numbers	4	8	12
A3	**Compute with numbers (i.e., add, subtract, multiply, divide)**			
3a:	Apply basic properties of operations	4	8	12
3b:	Describe effect of operations on size and order of numbers	4	8	12
3c:	Describe features of algorithms (e.g., regrouping with or without manipulative, partial products)	4	8	12
3d:	Select appropriate computation method (e.g., pencil and paper, calculator, mental arithmetic)	4	8	12
A4	**Use computations and estimation in applications**			
4a:	Round whole numbers, decimals, and fractions in meaningful contexts	4	8	12
4b:	Make estimates appropriate to a given situation			
	i. Know when to estimate	4	8	12
	ii. Select appropriate type of estimate (overestimate, underestimate, range of estimate)	4	8	12
	iii. Describe order of magnitude (estimation related to place value; scientific notation)	4	8	12
4c:	Select appropriate method of estimation (e.g., front-end, rounding)	4	8	12
4d:	Solve application problems involving numbers and operations, using exact answers or estimates, as appropriate	4	8	12
4e:	Interpret round-off errors using calculators/computers (i.e., truncation)	•	(8)	12
4f:	Verify solutions and determine the reasonableness of results			
	i. In real-world settings	4	8	12
	ii. In abstract settings	•	•	12

See notes at end of exhibit.

Exhibit A-1. NAEP mathematics framework and specifications summary: 2003—Continued

A	NUMBER SENSE, PROPERTIES AND OPERATIONS	Grade(s)		
A5	**Apply ratios and proportional thinking in a variety of situations**			
5a:	Use ratios to describe situations	(4)	8	12
5b:	Use proportions to model problems	•	8	12
5c:	Use proportional thinking to solve problems (including rates, scaling, and similarity)	•	8	12
5d:	Understand the meaning of percent (including percents greater than 100 and less than 1)	(4)	8	12
5e:	Solve problems involving percentages	•	8	12
A6	**Use elementary number theory**			
6a:	Describe odd and even numbers and their characteristics	4	8	12
6b:	Describe number patterns	(4)	8	12
6c:	Use factors and multiples to model and solve problems	•	8	12
6d:	Describe prime numbers	•	8	12
6e:	Use divisibility and remainders in problem settings (including simple modular arithmetic)	•	(8)	12

See notes at end of exhibit.

Exhibit A-1. NAEP mathematics framework and specifications summary: 2003—Continued

B	**MEASUREMENT**	**Grade(s)**		
B1	**Estimate the size of an object or compare objects with respect to a given attribute (e.g., length, area, capacity, volume, weight/mass)**			
		4	**8**	**12**
B2	**Select and use appropriate measurement instruments (e.g., manipulatives such as a ruler, meter stick, protractor, thermometer, scales for weight or mass, and gauges)**			
		4	**8**	**12**
B3	**Select and use appropriate units of measurement, according to**			
3a:	Type of unit	4	8	12
3b:	Size of unit	4	8	12
B4	**Estimate, calculate (using basic principles or formulas), or compare perimeter, area, volume, and surface area in meaningful contexts to solve mathematical and real-world problems.**			
4a:	Solve problems involving perimeter and area (e.g., triangles, quadrilaterals, other polygons, circles, and combined forms)	(4)	8	12
4b:	Solve problems involving volume and surface area (e.g., rectangular solids, cylinders, cones, pyramids, prisms, and combined forms)	(4)	(8)	12
B5	**Apply given measurement formulas for perimeter, area, volume, and surface area in problem settings**			
		•	**8**	**12**
B6	**Convert from one measurement to another within the same system (customary or metric)**			
		•	**8**	**12**
B7	**Determine precision, accuracy, and error**			
7a:	Apply significant digits in meaningful contexts	•	8	12
7b:	Determine appropriate size of unit of measurement in problem situations	•	8	12
7c:	Apply concepts of accuracy of measurement in problem situations	•	8	12
7d:	Apply absolute and relative error in problem situations	•	•	12
B8	**Make and read scale drawings**			
		•	**8**	**12**
B9	**Select appropriate methods of measurement (e.g., direct or indirect)**			
		4	**8**	**12**
B10	**Apply the concept of rate to measurement situations**			
		•	**8**	**12**

See notes at end of exhibit.

Exhibit A-1. NAEP mathematics framework and specifications summary: 2003—Continued

C	GEOMETRY AND SPATIAL SENSE	Grade(s)		
C1	**Describe, visualize, draw and construct geometric figures**			
1a:	Draw or sketch a figure given a verbal description	4	8	12
1b:	Given a figure, write a verbal description of its geometric qualities	•	8	12
C2	**Investigate and predict results of combining, subdividing, and changing shapes (e.g., paper folding, dissecting, tiling, and rearranging pieces of solids)**			
		4	**8**	**12**
C3	**Identify the relationship (congruence, similarity) between a figure and its image under a transformation**			
3a:	Use motion geometry (informal: lines of symmetry, flips, turns, and slides)	4	8	12
3b:	Use transformations (translations, rotations, reflections, dilations, symmetry)			
	i. synthetic	•	(8)	12
	ii. algebraic	•	•	12
C4	**Describe the intersection of two or more geometric figures**			
4a:	Two-dimensional	•	8	12
4b:	Planar cross-section of a solid	•	8	12
C5	**Classify figures in terms of congruence and similarity, and informally apply these relationships using proportional reasoning where appropriate**			
		•	**8**	**12**
C6	**Apply geometric properties and relationships in solving problems**			
6a:	Use concepts of 'between', 'inside', 'on' and 'outside'	4	8	•
6b:	Use the Pythagorean relationship to solve problems	•	8	12
6c:	Apply properties of ratio and proportion with respect to similarity	•	(8)	12
6d:	Solve problems involving right triangle trigonometric applications	•	•	12
C7	**Establish and explain relationships involving geometric concepts**			
7a:	Make conjectures	4	8	12
7b:	Validate and justify conclusions and generalizations	4	8	12
7c:	Use informal induction and deduction	(4)	8	12
C8	**Represent problem situations with geometric models and apply properties of figures in meaningful contexts to solve mathematical and real-world problems**			
		4	**8**	**12**
C9	**Represent geometric figures and properties algebraically using coordinates and vectors**			
9a:	Use properties of lines (including distance, midpoint, slope, parallelism, and perpendicularity) to describe figures algebraically	•	(8)	12
9b:	Algebraically describe conic sections and their properties	•	•	12
9c:	Use vectors in problem situations (addition, subtraction, scalar multiplication, dot product)	•	•	12

See notes at end of exhibit.

Exhibit A-1. NAEP mathematics framework and specifications summary: 2003—Continued

D	**DATA ANALYSIS, STATISTICS AND PROBABILITY**	**Grade(s)**		
D1	**Read, interpret, and make predictions using tables and graphs**			
1a:	Read and interpret data	4	8	12
1b:	Solve problems by estimating and computing with data	4	8	12
1c:	Interpolate and extrapolate from data	•	8	12
D2	**Organize and display data and make inferences**			
2a:	Use tables, histograms (bar graphs), pictograms, and line graphs	4	8	12
2b:	Use circle graphs and scattergrams	•	8	12
2c:	Use stem-and-leaf plots and box-and-whisker plots	•	8	12
2d:	Make decisions about outliers	•	8	12
D3	**Understand and apply sampling, randomness, and bias in data collection**			
3a:	Given a situation, identify sources of sampling error	•	8	12
3b:	Describe a procedure for selecting an unbiased sample	•	8	12
3c:	Make generalizations based on sample results	•	8	12
D4	**Describe measures of central tendency and dispersion in real-world situations**			
		(4)	**8**	**12**
D5	**Use measures of central tendency, correlation, dispersion, and shapes of distributions to describe statistical relationships**			
5a:	Use standard deviation and variance	•	•	12
5b:	Use the standard normal distribution	•	•	12
5c:	Make predictions and decisions involving correlation	•	•	12
D6	**Understand and reason about the use and misuse of statistics in our society**			
6a:	Given certain situations and reported results, identify faulty arguments or misleading presentations of the data	(4)	8	12
6b:	Appropriately apply statistics to real-world situations	(4)	8	12
D7	**Fit a line or curve to a set of data and use this line or curve to make predictions about the data, using frequency distributions where appropriate**			
		•	**•**	**12**
D8	**Design a statistical experiment to study a problem and communicate the outcomes**			
		•	**8**	**12**
D9	**Use basic concepts, trees, and formulas for combinations, permutations, and other counting techniques to determine the number of ways an event can occur**			
		•	**8**	**12**

See notes at end of exhibit.

Exhibit A-1. NAEP mathematics framework and specifications summary: 2003—Continued

D	DATA ANALYSIS, STATISTICS AND PROBABILITY	Grade(s)		
D10	**Determine the probability of a simple event**			
10a:	Estimate probabilities by use of simulation	•	8	12
10b:	Use sample spaces and the definition of probability to describe events	4	8	12
10c:	Describe and make predictions about expected outcomes	•	8	12
D11	**Apply the basic concept of probability to real-world situations**			
11a:	Informal use of probabilistic thinking	4	8	12
11b:	Use probability related to independent and dependent events	•	8	12
11c:	Use probability related to simple and compound events	•	•	12
11d:	Use conditional probability	•	•	12

See notes at end of exhibit.

Exhibit A-1. NAEP mathematics framework and specifications summary: 2003—Continued

E	ALGEBRA AND FUNCTIONS	Grade(s)		
E1	**Describe, extend, interpolate, transform and create a wide variety of patterns and functional relationships**			
1a:	Recognize patterns and sequences	4	8	12
1b:	Extend a pattern or functional relationship	4	8	12
1c:	Given a verbal description, extend or interpolate with a pattern (complete a missing term)	•	8	12
1d:	Translate patterns from one context to another	(4)	8	12
1e:	Create an example of a pattern or functional relationship	4	8	12
1f:	Understand and apply the concept of a variable	(4)	8	12
E2	**Use multiple representations for situations to translate among diagrams, models, and symbolic expressions**			
		4	**8**	**12**
E3	**Use number lines and rectangular coordinate systems as representational tools**			
3a:	Identify or graph sets of points on a number line or in a rectangular coordinate system	4	8	12
3b:	Identify or graph sets of points in a polar coordinate system	•	8	12
3c:	Work with applications using coordinates	•	8	12
3d:	Transform the graph of a function	•	(8)	12
E4	**Represent and describe solutions to linear equations and inequalities to solve mathematical and real-world problems**			
4a:	Solution sets of whole numbers	4	8	12
4b:	Solution sets of real numbers	(4)	8	12
E5	**Interpret contextual situations and perform algebraic operations on real numbers and algebraic expressions to solve mathematical and real-world problems**			
5a:	Perform basic operations, using appropriate tools, on real numbers in meaningful contexts (including grouping and order of multiple operations involving basic operations, exponents and roots)	•	8	12
5b:	Solve problems involving substitution in expressions and formulas	•	8	12
5c:	Solve meaningful problems involving a formula with one variable	•	8	12
5d:	Use equivalent forms to solve problems	•	8	12
E6	**Solve systems of equations and inequalities using appropriate methods**			
6a:	Solve systems graphically	•	8	12
6b:	Solve systems algebraically	•	•	12
6c:	Solve systems using matrices	•	•	12

See notes at end of exhibit.

Exhibit A-1. NAEP mathematics framework and specifications summary: 2003—Continued

E	**ALGEBRA AND FUNCTIONS**	**Grade(s)**		
E7	**Use mathematical reasoning**			
7a:	Make conjectures	4	8	12
7b:	Validate and justify conclusions and generalizations	4	8	12
7c:	Use informal induction and deduction	(4)	8	12
E8	**Represent problem situations with discrete structures**			
8a:	Use finite graphs and matrices	•	(8)	12
8b:	Use sequences and series	•	•	12
8c:	Use recursive relations (including numerical and graphical iteration and finite differences)	•	•	12
E9	**Solve polynomial equations with real and complex roots using a variety of algebraic and graphical methods and using appropriate tools**			
		•	**•**	**12**
E10	**Approximate solutions of equations (bisection, sign changes, successive approximations)**			
		•	**(8)**	**12**
E11	**Use appropriate notation and terminology to describe functions and their properties (includes domain, range, function composition, inverses)**			
		•	**•**	**12**
E12	**Compare and apply the numerical, symbolic, and graphical properties of a variety of functions and families of functions, examining general parameters and their effect on curve shape**			
		•	**(8)**	**12**
E13	**Apply function concepts to model and deal with real-world situations**			
		•	**(8)**	**12**
E14	**Use trigonometry**			
14a:	Use triangle trigonometry to model problem situations	•	•	12
14b:	Use trigonometric and circular functions to model real-world phenomena	•	•	12
14c:	Apply concepts of trigonometry to model solve real-world problems	•	•	12

NOTE: *Content strands* are identified by capital letters (A, B, C,…), *topics* are identified by numbers (1, 2, 3,…), and *subtopics* are identified by lowercase letters (a, b, c, …). Topics and subtopics can be assessed at those grade levels indicated by 4, 8, and 12 on the right side of the exhibit. Parentheses around a grade level indicate that a topic may be introduced at a simple level at that grade. If a topic or subtopic should not be addressed at a specific grade level, it is indicated by a dot (•).
SOURCE: U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003

A Number		Grade(s)	
A1	Whole Numbers	4	8
A2	Fractions and Decimals	4	8
A3	Integers	•	8
A4	Ratio, Proportion, and Percent	4	8

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

A NUMBER	
A1 Whole Numbers	
Grade 4	**Grade 8**
1a: Represent whole numbers using words, diagrams, or symbols, including recognizing and writing numbers in expanded form. 1b: Demonstrate knowledge of place value. 1c: Compare and order whole numbers. 1d: Identify sets of numbers according to common properties such as odd and even, multiples, or factors. 1e: Compute with whole numbers. 1f: Estimate computations by approximating the numbers involved. 1g: Solve routine and non-routine problems, including real-life problems.	1a: Demonstrate knowledge of place value and of the four operations. 1b: Find and use factors or multiples of numbers, and identify prime numbers. 1c: Express in general terms and use the principles of commutativity, associativity, and distributivity. 1d: Evaluate powers of numbers, and square roots of perfect squares to 144. 1e: Solve problems by computing, estimating, or approximating.
A2 Fractions and Decimals	
Grade 4	**Grade 8**
2a: Recognize fractions as parts of unit wholes, parts of a collection, locations on number lines, divisions of whole numbers. 2b: Identify equivalent fractions. 2c: Compare and order fractions. 2d: Show understanding of decimals. 2e: Represent fractions or decimals using words, numbers, or models. 2f: Add and subtract fractions with the same denominator. 2g: Add and subtract with decimals. Notes: Grade 4 fractions items will involve denominators of 2, 3, 4, 5, 6, 8, 10, or 12. Grade 4 decimals items will involve decimals to tenths and/or hundredths.	2a: Compare and order fractions. 2b: Compare and order decimals. 2c: Demonstrate knowledge of place value for decimals. 2d: Represent decimals and fractions using words, numbers, or models (including number lines). 2e: Recognize and write equivalent fractions. 2f: Convert fractions to decimals and vice versa. 2g: Relate operations with fractions or decimals to situations and models. 2h: Compute with fractions and decimals, including use of commutativity, associativity, and distributivity. 2i: Approximate decimals to estimate computations. 2j: Solve problems involving fractions. 2k: Solve problems involving decimals.

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

A NUMBER	
A3 Integers	
Grade 4	**Grade 8**
Not assessed at this level.	3a: Represent integers using words, numbers, or models (including number lines). 3b: Compare and order integers. 3c: Show an understanding of addition, subtraction, multiplication, and division with integers. 3d: Compute with integers. 3e: Solve problems using integers.
A4 Ratio, Proportion, and Percent	
Grade 4	**Grade 8**
4a: Solve problems involving simple proportional reasoning.	4a: Identify and find equivalent ratios. 4b: Divide a quantity in a given ratio. 4c: Convert percents to fractions or decimals, and vice versa. 4d: Solve problems involving percents. 4e: Solve problems involving proportions.

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

B Measurement		Grade(s)	
B1	Attributes and Units	4	8
B2	Tools, Techniques and Formulas	4	8

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

B MEASUREMENT	
B1 Attributes and Units	
Grade 4	**Grade 8**
1a: Use given non-standard units to measure length, area, volume, and time (e.g., paper clips for length, tiles for area, sugar cubes for volume). 1b: Select appropriate standard units to measure length, area, mass/weight,* angle, and time (e.g., kilometers for car trips, centimeters for human height). 1c: Use conversion factors between standard units (e.g., hours to minutes, grams to kilograms). 1d: Recognize that total measures of length, area, volume, angle, and time do not change with position, decomposition into parts, or division. *More properly mass, but weight expressed in grams or kilograms is the common usage at these levels. Countries in which mass is the common usage for grades 4 and/or 8 will frame items accordingly.	1a: Select and use appropriate standard units to find measures of length, area, volume, perimeter, circumference, time, speed, density, angle, mass/weight.* 1b: Use relationships among units for conversions within systems of units, and for rates. * More properly mass, but weight expressed in grams or kilograms is the common usage at these levels. Countries in which mass is the common usage for grades 4 and/or 8 will frame items accordingly.

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

B MEASUREMENT	
B2 Tools, Techniques, and Formulas	
Grade 4	**Grade 8**
2a: Use instruments with linear or circular scales to measure length, weight, time, and temperature in problem situations (e.g., dimensions of a window, weight of a parcel). 2b: Estimate length, area, volume, weight, and time in problem situations (e.g., height of a building, volume of a block of material). 2c: Calculate areas and perimeters of squares and rectangles of given dimensions. 2d: Compute measurements in simple problem situations (e.g., elapsed time, change in temperature, difference in height or weight).	2a: Use standard tools to measure length, weight, time, speed, angle, and temperature in problem situations and to draw line segments, angles, and circles of a given size. 2b: Estimate length, circumference, area, volume, weight, time, angle, and speed in problem situations (e.g., circumference of a wheel, speed of a runner). 2c: Compute with measurements in problem situations (e.g., add measures, find average speed on a trip, find population density). 2d: Select and use appropriate measurement formulas for perimeter of a rectangle, circumference of a circle, areas of plane figures (including circles), surface area and volume of rectangular solids, and rates. 2e: Find measures of irregular or compound areas by covering with grids or dissecting and rearranging pieces. 2f: Give and interpret information about the precision of measurements (e.g., upper and lower bounds of a length reported as 8 centimeters to the nearest centimeter).

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

C Geometry		Grade(s)	
C1	**Lines and Angles**	**4**	**8**
C2	**Two- and Three-Dimensional Shapes**	**4**	**8**
C3	**Congruence and Similarity**	**4**	**8**
C4	**Locations and Spatial Relationships**	**4**	**8**
C5	**Symmetry and Transformations**	**4**	**8**

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

C GEOMETRY	
C1 Lines and Angles	
Grade 4	**Grade 8**
1a: Classify angles as greater than, equal to, or less than a right angle (or 90°). 1b: Identify and describe parallel and perpendicular lines. 1c: Compare given angles and place them in order of size.	1a: Classify angles as acute, right, straight, obtuse, reflex, complementary, and supplementary. 1b: Recall the relationships for angles at a point, angles on a line, vertically opposite angles, angles associated with a transversal cutting parallel lines, and perpendicularity. 1c: Know and use the properties of angle bisectors and perpendicular bisectors of lines.
C2 Two- and Three-Dimensional Shapes	
Grade 4	**Grade 8**
2a: Know and use vocabulary associated with familiar two- and three-dimensional shapes. 2b: Identify common geometric shapes in the environment. 2c: Classify two- and three-dimensional shapes according to their properties. 2d: Know properties of geometric figures and use them to solve routine problems. 2e: Decompose shapes and rearrange the parts to form simpler shapes.	2a: Recall properties of geometric shapes: triangles (scalene, isosceles, equilateral, right) and quadrilaterals (scalene, trapezoid, parallelogram rectangle, rhombus, square). 2b: Use properties of familiar geometric shapes in a compound figure to make conjectures about properties of the compound figure. 2c: Recall properties of other polygons (regular pentagon, hexagon, octagon, decagon). 2d: Construct or draw triangles and rectangles of given dimensions. 2e: Apply geometric properties to solve routine and non-routine problems 2f: Use Pythagorean theorem (not proof) to solve problems (e.g., find the length of a side of a right-angled triangle given the lengths of the other two sides; or, given the lengths of three sides of a triangle, determine whether the triangle is right-angled).

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

C GEOMETRY	
C3 Congruence and Similarity	
Grade 4	**Grade 8**
3a: Identify triangles that have the same size and shape (congruent). 3b: Identify triangles that have the same shape but different sizes (similar).	3a: Identify congruent triangles and their corresponding measures. 3b: Identify congruent quadrilaterals and their corresponding measures. 3c: Consider the conditions of congruence to determine whether triangles with given corresponding measures (at least three) are congruent. 3d: Identify similar triangles and recall their properties. 3e: Use properties of congruence in mathematical and practical problem situations. 3f: Use properties of similarity in mathematical and practical problem situations.
C4 Locations and Spatial Relationships	
Grade 4	**Grade 8**
4a: Use informal coordinate systems to locate points in a plane. 4b: Relate a net to the shape it will make. 4c: Recognize relationships between two-dimensional and three-dimensional shapes when shown nets and different two-dimensional views of three-dimensional objects.	1a: Locate points using number lines, coordinate grids, and maps. 1b: Use ordered pairs, equations, intercepts, intersections, and gradients to locate points and lines in the Cartesian plane. 1c: Recognize relationships between two-dimensional and three-dimensional shapes when shown nets and different two-dimensional views of three-dimensional objects.
C5 Symmetry and Transformations	
Grade 4	**Grade 8**
5a: Recognize line symmetry. 5b: Draw two-dimensional symmetrical figures. 5c: Recognize translation, reflection, and rotation.	5a: Recognize line and rotational symmetry for two-dimensional shapes. 5b: Draw two-dimensional symmetrical figures. 5c: Recognize, or demonstrate by sketching, translation, reflection, rotation, and enlargement. 5d: Use transformations to explain or establish geometric properties.

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

D Data		Grade(s)	
D1	Data Collection and Organization	4	8
D2	Data Representation	4	8
D3	Data Interpretation	4	8
D4	Uncertainty and Probability	•	8

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

D DATA	
D1 Data Collection and Organization	
Grade 4	**Grade 8**
1a: Match a set of data with appropriate characteristics of situations or contexts (e.g., outcomes from rolling a die). 1b: Organize a set of data by one characteristic (e.g., height, color, age, shape).	1a: Match a set of data, or a data display, with appropriate characteristics of situations or contexts (e.g., monthly sales of a product for a year). 1b: Organize a set of data by one or more characteristics using a tally chart, table, or graph. 1c: Recognize and describe possible sources of error in collecting and organizing data (e.g., bias, inappropriate grouping). 1d: Select the most appropriate data collection method (e.g., survey, experiment, questionnaire) to answer a given question, and justify the choice.
D2 Data Representation	
Grade 4	**Grade 8**
2a: Read data directly from tables, pictographs, bar graphs, and pie charts. 2b: Display data using tables, pictographs, and bar graphs. 2c: Compare and match different representations of the same data.	2a: Read data from charts, tables, pictographs, bar graphs, pie charts, and line graphs. 2b: Display data using charts, tables, pictographs, bar graphs, pie charts, and line graphs. 2c: Compare and match different representations of the same data.

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

D DATA	
D3 Data Interpretation	
Grade 4	**Grade 8**
3a: Compare characteristics of related data sets (e.g., given data or representations of data on student heights in two classes, identify the class with the shortest/tallest person). 3b: Draw conclusions from data displays.	3a: Compare characteristics of data sets, using mean, median, range, and shape of distribution (in general terms). 3b: Interpret data sets (e.g., draw conclusions, make predictions, and estimate values between and beyond given data points). 3c: Evaluate interpretations of data with respect to correctness and completeness of interpretation. 3d: Use and interpret data sets to answer questions.
D4 Uncertainty and Probability	
Grade 4	**Grade 8**
Not assessed at this grade.	4a: Judge the likelihood of an event as certain, more likely, equally likely, less likely, or impossible. 4b: Use data from experiments to estimate probabilities for favorable outcomes. 4c: Use problem conditions to calculate theoretical probabilities for possible outcomes.

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

E Algebra		Grade(s)	
E1	**Patterns**	**4**	**8**
E2	**Algebraic Expressions**	**•**	**8**
E3	**Equations and Formulas**	**4**	**8**
E4	**Relationships**	**4**	**8**

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

E ALGEBRA	
E1 Patterns	
Grade 4	**Grade 8**
1a: Extend and find missing terms of numeric and geometric patterns. 1b: Match numeric and geometric patterns with descriptions. 1c: Describe relationships between adjacent terms in a sequence or between the number of the term and the term.	1a: Extend numeric, algebraic, and geometric patterns or sequences using words, symbols, or diagrams; find missing terms. 1b: Generalize pattern relationships in a sequence, or between adjacent terms, or between the number of the term and the term, using words or symbols.
E2 Algebraic Expressions	
Grade 4	**Grade 8**
Not assessed at this grade.	2a: Find sums, products, and powers of expressions containing variables. 2b: Evaluate expressions for given numeric values of the variable(s). 2c: Simplify or compare algebraic expressions to determine equivalence. 2d: Model situations using expressions.
E3 Equations and Formulas	
Grade 4	**Grade 8**
3a: Show understanding of equality using equations, areas, volumes, masses/weights. 3b: Find the missing number in an equation (e.g., if 17 + __ = 29, what number would go in the blank to make the equation true?). 3c: Model simple situations involving unknowns with an equation. 3d: Solve problems involving unknowns.	3a: Evaluate formulas given the values of the variables. 3b: Use formulas to answer questions about given situations. 3c: Indicate whether a value (or values) satisfies a given equation. 3d: Solve simple linear equations and inequalities, and simultaneous (two variables) equations. 3e: Write linear equations, inequalities, or simultaneous equations that model given situations. 3f: Solve problems using equations or formulas.

See notes at end of exhibit.

Exhibit A-2. TIMSS mathematics framework and specifications summary: 2003—Continued

E ALGEBRA	
E4 Relationships	
Grade 4	**Grade 8**
4a: Generate pairs of numbers following a given rule (e.g., multiply the first number by 3 and add 2 to get the second number). 4b: Write, or select, a rule for a relationship given some pairs of numbers satisfying the relationship. 4c: Graph pairs of numbers following a given rule. 4d: Show why a pair of numbers follows a given rule. (E.g., a rule for a relation between two numbers is "multiply the first number by 5 and subtract 4 to get the second number." Show that when the first number is 2 and the second number is 6 the rule is followed.)	4a: Recognize equivalent representations of functions as ordered pairs, tables, graphs, words, or equations. 4b: Given a function in one representation, generate a different but equivalent representation. 4c: Recognize and interpret proportional, linear, and nonlinear relationships (travel graphs and simple piecewise functions included). 4d: Write or select a function to model a given situation. 4e: Given a graph of a function, identify attributes such as intercepts on axes and intervals where the function increases, decreases, or is constant.

NOTE: *Content domains* are identified by capital letters (A, B, C,…), *topic areas* are identified by numbers (1, 2, 3,…), and *objectives* are identified by lowercase letters (a, b, c, …). Topic areas can be assessed at those grade levels indicated by 4 and 8 on the right side of the table. If a topic area should not be addressed at a specific grade level, it is indicated by a dot (•).
SOURCE: International Study Center, Lynch School of Education, Boston College, *TIMSS Assessment Frameworks and Specifications 2003, 2nd ed.,* 2003.

Appendix B: Levels of Mathematical Complexity

Exhibit B-1 is the document that was used by the expert panel for the *mathematical complexity* classifications. The summary document is based on the prepublication version of the NAEP 2005 framework that was available at the time of the expert panel meeting but has been reformatted and adapted slightly to facilitate the classification process.

Exhibit B-1. Levels of mathematical complexity adapted from the NAEP 2005 mathematics framework

Low Complexity	Moderate Complexity	High Complexity
This category relies heavily on the recall and recognition of previously learned concepts and principles. Items typically specify what the student is to do, which is often to carry out some procedure that can be performed mechanically. It is not left to the student to come up with an original method or solution.	Items in the moderate-complexity category involve more flexibility of thinking and choice among alternatives than do those in the low-complexity category. They require a response that goes beyond the habitual, is not specified, and ordinarily has more than a single step. The student is expected to decide what to do, using informal methods of reasoning and problem-solving strategies, and to bring together skill and knowledge from various domains.	High-complexity items make heavy demands on students, who must engage in more abstract reasoning, planning, analysis, judgment, and creative thought. A satisfactory response to the item requires that the student think in an abstract and sophisticated way.
The following are some, but not all, of the demands that items in the low-complexity category might make: • Recall or recognize a fact, term, or property; • Recognize and example of a concept; • Compute a sum, difference, product, or quotient; • Recognize an equivalent representation; • Perform a specified procedure; • Evaluate an expression in an equation or formula for a given variable; • Solve a one-step word problem; • Draw or measure simple geometric figures; or • Retrieve information from a graph, table, or figure.	The following illustrate some of the demands that items of moderate complexity might make: • Represent a situation mathematically in more than one way; • Select and use different representations, depending on situation and purpose; • Solve a word problem requiring multiple steps; • Compare figures or statements; • Provide a justification for steps in a solution process; • Interpret a visual representation; • Extend a pattern; • Retrieve information from a graph, table, or figure and use it to solve a problem requiring multiple steps; • Formulate a routine problem, given data and conditions; or • Interpret a simple argument.	Items at the level of high complexity may ask the student to do any of the following: • Describe how different representations can be used for different purposes; • Perform a procedure having multiple steps and multiple decision points; • Analyze similarities and differences between procedures and concepts; • Generalize a pattern; • Formulate an original problem, given a situation; • Solve a novel problem; • Solve a problem in more than one way; • Explain and justify a solution to a problem; • Describe, compare, and contrast solution methods; • Formulate a mathematical model for a complex situation; • Analyze the assumptions made in a mathematical model; • Analyze or produce a deductive argument; or • Provide a mathematical justification.

SOURCE: U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2005 National Assessment of Educational Progress*, 2004.

Appendix C: Expert Panel

Members and Staff

Expert Panel Members

Jim Braswell
Educational Testing Service

Mary Anne Dorofee
Educational Testing Service

Canda Mueller Engheta
Kansas State Department of Education

Ramesh Gangolli
University of Washington

Jeane Joyner
Meredith College

Jeremy Kilpatrick
University of Georgia

Patricia Klag
Educational Testing Service

Mary Lindquist
Columbus State University

Marge Petit
National Center for the Improvement of Educational Assessment

Norman Webb
Wisconsin Center for Education Research

Linda Wilson
Consultant

Staff

National Center for Education Statistics

Marilyn Binkley

Patrick Gonzales

Andrew Malizio

Eugene Owen

Elois Scott

Education Statistics Services Institute

Kristy David

Janine Emerson

Kim Gattis

Dana Kelly

Teresa Smith Neidorf

Margaret Woodworth

David Nohara
Independent Project Consultant

Appendix D: Methodological Notes and Supplementary Data

Considerations in Selecting Classification Methods

The cross-classification approach (classification of items in one assessment to the other assessment framework) was selected for the examination of content and grade match so that there would be multiple content profiles for each assessment. This method also prevents each assessment from being evaluated through only one perspective, which may or may not be reflective of its purposes. This approach takes advantage of the multiple ways of describing content found in the frameworks and enables direct comparisons across the assessments.

For the classifications based on cognitive processes and skills, a common classification system was chosen—levels of *mathematical complexity* from the NAEP 2005 framework. There were several reasons for deciding, first, to use only a single classification system and, second, for selecting this NAEP 2005 dimension for the common system. With regard to the first decision, the study organizers recognized that classifications in these dimensions likely would require more discretionary judgment than those in content areas, and thought a single rubric would be the most realistic to implement under the time constraints. With regard to the second decision, although all three 2003 mathematics assessment frameworks compared in this study include a dimension related to cognitive skills and processes, the classification system from the NAEP 2005 framework focuses on the characteristics of items rather than on inferred cognitive abilities of students, which may vary widely from student to student. Furthermore, it is expected to be in use in NAEP for several years to come, making it a potentially valuable link to similar comparison studies in the future.

Reliability Analyses

For the classification of items, the expert panel was divided into three groups to review items by content area (as described in section 3). To measure the extent to which the different content area groups were interpreting the common rubric in similar ways, a common set of items was classified by all three groups with respect to *mathematical complexity* level. The degree to which the three groups classified these items in the same categories on this dimension serves as a measure of the reliability of these classifications. The set of 60 items (30 from NAEP and 30 from TIMSS), which reflects approximately 9 percent of the total item classifications across both assessments, was taken from across the mathematics content areas and grade levels. This was not a random sample, but a representative set chosen to cover the main categories addressed in the study (content area and grade level). Some effort was also made to ensure that there were at least some items from each of the cognitive categories based on the original assessment developers' classifications (*cognitive domains* in TIMSS and *mathematical abilities* in NAEP). Reliability items were classified at regular intervals throughout the classification process. Given the limited time available for the expert panel meeting, the 30 items from each assessment was the maximum number of items that could be included in the reliability set.

The reliability set of 30 items from each assessment (NAEP and TIMSS) was composed of 10 items from the *number* category and 5 items from each of *measurement, geometry, data*, and *algebra*. This reflects 10 items from each of the three primary content groups into which the expert panel was divided (*number, measurement/geometry*, and *data/algebra*). For the reliability results, each group contributed 10 classifications from their primary content areas and another 20 secondary classifications for reliability items from the other content areas. Due to time constraints, one of the three groups only provided secondary classifications for 5 of the 20 items outside their primary content area. This group contributed primary classifications for items in *data* and *algebra*, and 5 secondary classifications from the *number* category. Therefore, the set of 15 items that have classifications from all three groups do not reflect any items from *measurement* and *geometry*; for those items, reliability data are based on only two classifications. With respect to other item characteristics (e.g., grade level, item format, and NAEP *mathematical ability* category), the set is still balanced and representative of the full reliability set.

The multiple classification data for the reliability set were analyzed based on the percentage of classifications where there was agreement. Classification reliability statistics were computed in two ways, as follows:

- The percentage of total comparisons: based on the number of comparisons where there was agreement between any two groups (i.e., groups 1 and 2, groups 2 and 3 and groups 1 and 3) across ALL items; and

- The percentage of items: based on the number of items where there was agreement across ALL three groups.

The results from these two types of analyses are shown in tables D-1 and D-2.

The results of the reliability analyses shown in table D-1 (based on number of comparisons across any two groups) were checked to evaluate any impact of removing the 5 *number* and 10 *measurement/geometry* NAEP items with data from only two groups. The results showed no change in the reported percentage agreement (78 percent). Thus, the full set of reliability data were used for the analyses shown.

There was reasonably high agreement across groups on classification to the three levels of *mathematical complexity* (low, moderate, and high). The results indicate 79 percent agreement for all comparisons between any two groups across all items (table D-1). These results reflect agreement across all groups for 69 percent of all items (table D-2). When broken down into the NAEP and TIMSS items, the results are similar. For items where there was not total agreement across groups, disagreement was always to “adjacent” categories (i.e., low/moderate and moderate/high). There were no instances of disagreement between low and high complexity.

Table D-1. Reliability of mathematical complexity level classifications for mathematics items in NAEP 2003 and TIMSS 2003, by number of comparisons and percentage agreement

Number of comparisons and percentage agreement	NAEP 2003	TIMSS 2003	Overall
Total number of comparisons across items	60	90	150
Number of comparisons with agreement between groups	47	72	119
Percentage agreement	78	80	79

NOTE: Data are based on 30 NAEP items and 30 TIMSS items that were classified by three expert panel groups and reflect all comparisons between any two groups (i.e., groups 1 and 2; groups 2 and 3; and groups 1 and 3). One group classified only 15 of 30 NAEP items in the reliability set, meaning that agreement on the classification for 15 NAEP items was based on the classifications of two groups instead of three. This results in a total number of comparisons of 60 for NAEP items (instead of 90) and 150 overall.
SOURCE: Expert panel classifications of selected fourth- and eighth-grade mathematics items from the National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment and the Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment.

Table D-2. Reliability of mathematical complexity level classifications for mathematics items in NAEP 2003 and TIMSS 2003, by number of items and percentage agreement

Number of items and percentage agreement	NAEP 2003	TIMSS 2003	Overall
Total number of items	15	30	45
Number of items with agreement across all groups	10	21	31
Percentage agreement	67	70	69

NOTE: Data are based on 30 NAEP items and 30 TIMSS items that were classified by three expert panel groups and reflect comparisons across all three groups for each item. One group classified only 15 of 30 NAEP items in the reliability set; the percentage agreement for NAEP items includes only the items that were classified by all three groups.
SOURCE: Expert panel classifications of selected fourth- and eighth-grade mathematics items from the National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment and the Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment.

In sum, the main focus of the present study is a content comparison—classification of items to the content framework of the other assessment—which was done by the separate content-area subpanels. The reliability tables are included only to provide some indication of the extent to which the expert panelists agreed on the other metric that used a common rubric (*mathematical complexity* level). Expert panelists typically spent less time reviewing and classifying the items in the reliability set that were outside of their primary content areas, and the results from these secondary classifications should not be viewed as a complete replication of the process used by the primary group which was most familiar with the items in the respective content areas. Therefore, only the primary group classifications were used in the reporting of results for *mathematical complexity* level.

Data Processing

After the expert panel meeting, the facilitators of each group met to review the methods used and the data collected to ensure consistency. In some cases, methods or reporting conventions were slightly different between groups. For these cases, the facilitators reviewed their notes and the notes of individual panel members to standardize the data. Datasets were produced that included the standardized expert panel classifications for all items from each assessment (including multiple classifications on the reliability set) as well as original classification information for each item provided by the assessment developers. The raw data containing all original panelist classifications and comments from each subgroup were also available for analysts and were consulted in the writing of this report.

Supplementary Data on NAEP/PISA Comparisons

Some supplementary data to inform the discussion of the NAEP/PISA content comparisons in Section 6.1 are shown in Table D-3. This table shows the distribution of PISA items classified across topics in the NAEP *content strands*, indicating the degree of overlap between PISA items from each *overarching idea* and NAEP topics.

Table D-3. Distribution of PISA 2003 mathematics items across topics within the NAEP 2003 mathematics framework content strands, by PISA overarching idea category

NAEP content strand and topic	PISA overarching idea: Change and Relationships	Quantity	Space and shape	Uncertainty
Number sense, properties, and operations				
Represent numbers and operations using models, diagrams, and symbols	0	2	0	0
Compute with numbers	1	0	1	0
Use computations and estimation in applications	1	9	1	0
Apply ratios and proportional thinking	1	0	0	1
Use elementary number theory	0	2	0	0
Measurement				
Estimate the size of an object or compare objects with respect to a given attribute	0	0	1	0
Estimate, calculate or compare perimeter, area, volume, and surface area	0	0	7	0
Convert from one measurement to another within the same system	0	2	0	0
Make and read scale drawings	0	0	1	0
Apply the concept of rate to measurement situations	1	1	0	0
Geometry and spatial sense				
Describe, visualize, draw and construct geometric figures	0	0	1	0
Investigate and predict results of combining, subdividing, and changing shapes	0	0	2	0
Identify the relationship (congruence, similarity) between a figure and its image under a transformation	0	0	1	0
Apply geometric properties and relationships in solving problems	0	0	1	0
Represent situations with geometric models and apply properties of figures	0	0	3	0
Data analysis, statistics, and probability				
Read, interpret, and make predictions using tables and graphs	9	3	0	4
Organize and display data and make inferences	1	0	0	0
Understand and apply sampling, randomness, and bias in data collection	0	0	0	1
Describe measures of central tendency and dispersion	0	0	0	5
Understand and reason about the use and misuse of statistics in our society	0	0	0	4
Use basic concepts, trees, and formulas for combinations, permutations, and other counting techniques	0	2	0	0
Determine the probability of a simple event	0	0	0	2
Apply the basic concept of probability	0	0	0	3
Algebra and functions				
Describe, extend, interpolate, transform and create patterns and functional relationships	0	2	0	0
Use multiple representations to translate among diagrams, models, and symbolic expressions	1	0	0	0
Interpret situations and perform algebraic operations	3	0	0	0
Compare and apply the numerical, symbolic, and graphical properties of a variety of functions	1	0	0	0
Apply function concepts to model real-world situations	1	0	0	0

NOTE: Data reflect the number of PISA items classified by the panel to a topic or subtopic at any grade level in the NAEP 2003 framework. Items classified to multiple topics were counted in all relevant topics. Four PISA items that were classified to a NAEP strand but not to a topic are not reflected in this table: three items from *change and relationships* classified as *measurement*, *algebra* or *geometry* and one item from *space and shape* classified as *geometry*.
SOURCE: Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*, 2003; and U.S. Department of Education, National Assessment Governing Board, *Mathematics Framework for the 2003 National Assessment of Educational Progress*, 2002.

Appendix E: Example Items

Exhibit E-1. Index of example items from NAEP 2003, TIMSS 2003, and PISA 2003

Example Number	Description of characteristics illustrated in text
1	NAEP *conceptual understanding* item (low *mathematical complexity*)
2	NAEP *problem solving* item (low *mathematical complexity*)
3	TIMSS *reasoning* item (low *mathematical complexity*)
4	PISA *reproduction* item (moderate *mathematical complexity*)
5	NAEP short constructed-response item (high *mathematical complexity*)
6	NAEP extended constructed-response item (high *mathematical complexity*)
7	TIMSS extended constructed-response item (moderate *mathematical complexity*)
8	PISA complex multiple-choice item (moderate *mathematical complexity*)
9	PISA open constructed-response item (low *mathematical complexity*)
10	TIMSS item described as "basic computation" by the expert panel (low *mathematical complexity*)
11	TIMSS item described as "select an operation or procedure to solve a problem" by the expert panel (low *mathematical complexity*)
12	NAEP item described as "basic computation" by the expert panel (low *mathematical complexity*)
13	NAEP fourth-grade *number* item classified at eighth-grade level on TIMSS mathematics framework (moderate *mathematical complexity*)
14	TIMSS eighth-grade *measurement* item classified at fourth-grade level on NAEP mathematics framework (moderate *mathematical complexity*)
15	TIMSS fourth-grade *geometry* item not classified at the topic level on the NAEP mathematics framework (low *mathematical complexity*)
16	NAEP fourth-grade *geometry* item with a general match to the TIMSS mathematics framework—classified at the topic level but not to a specific subtopic (moderate *mathematical complexity*)
17	NAEP cross-grade *geometry* item (grades 4, 8, and 12) classified at the fourth-grade level on TIMSS mathematics framework (low *mathematical complexity*)
18	TIMSS eighth-grade *geometry* item classified at the twelfth-grade level on the NAEP mathematics framework (low *mathematical complexity*)
19	TIMSS eighth-grade *geometry* item classified to NAEP topic of *apply geometric properties and relationships in solving problems* (low *mathematical complexity*)
20	TIMSS fourth-grade *data* item classified to NAEP topic of *read, interpret, and make predictions using tables and graphs* (low *mathematical complexity*)
21	NAEP eighth-grade *algebra* item classified to TIMSS *number* content domain (low *mathematical complexity*)
22	NAEP *problem solving* item classified to PISA *reproduction competency cluster* (moderate *mathematical complexity*)
23	NAEP *problem solving* item classified to PISA *connections competency cluster* and judged as appropriate for PISA (moderate *mathematical complexity*)
24	NAEP *problem solving* item judged as not likely to appear on the PISA mathematics assessment (moderate *mathematical complexity*)
25	NAEP *problem solving* item judged as requiring revision to appear on the PISA mathematics assessment (moderate *mathematical complexity*)
26	PISA task requiring graphical interpretation (low-moderate *mathematical complexity*)

EXAMPLE 1

NAEP multiple-choice item – grade 4

Which of the following has only 3 angles?

A) A triangle
B) A square
C) A rectangle
D) A cube

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

Answer Key: A

Framework classifications

NAEP 2003 framework[1]	TIMSS 2003 framework[2]
Geometry and spatial sense	Geometry
Describe, visualize, draw, and construct geometric figures	Two- and three-dimensional shapes
Grade 4	Grade 4
Conceptual understanding	
Mathematical complexity level:[3] low	

[1] Classified by NAEP assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 2

NAEP multiple-choice item – grades 4, 8, and 12

The perimeter of a square is 36 inches. What is the length of one side of the square?

A) 4 inches
B) 6 inches
C) 9 inches
D) 18 inches

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

Answer Key: C

Framework classifications

NAEP 2003 framework[1]	TIMSS 2003 framework[2]
Measurement	Measurement
Estimate, calculate (using basic principles or formulas), or compare perimeter, area, volume, and surface area in meaningful contexts to solve mathematical and real-world problems	Tools, techniques, and formulas
Grade 4	Grade 4
Problem solving	
Mathematical complexity level:[3] low	

[1] Classified by NAEP assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 3

TIMSS multiple-choice item – grade 8

(3, 6), (6, 15), (8, 21)

Which of these describes how to get the second number from the first number in every ordered pair above?

A) Add 3
B) Subtract 3
C) Multiply by 2
D) Multiply by 2 and then add 3
E) Multiply by 3 and then subtract 3

SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment.

Answer Key: E

Framework classifications

TIMSS 2003 framework[1]	NAEP 2003 framework[2]
Algebra	Algebra and functions
Relationships	Describe, extend, interpolate, transform, and create a wide variety of patterns and functional relationships
Grade 8	Grade 8
Reasoning	
Mathematical complexity level:[3] low	

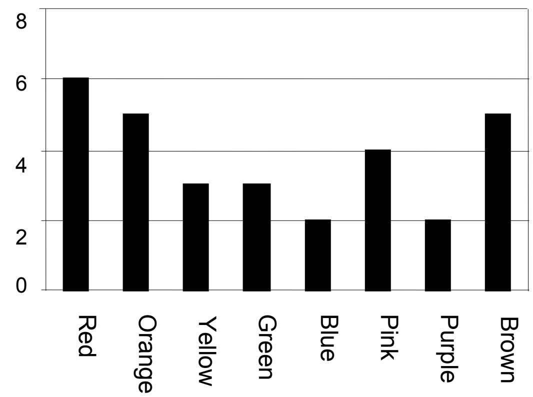
[1] Classified by TIMSS assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 4

PISA multiple-choice item

Colored Candies

Robert's mother lets him pick one candy from a bag. He can't see the candies. The number of candies of each color in the bag is shown in the following graph.



What is the probability that Robert will pick a red candy?

A) 10%
B) 20%
C) 25%
D) 50%

SOURCE: Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment.

Answer Key: B

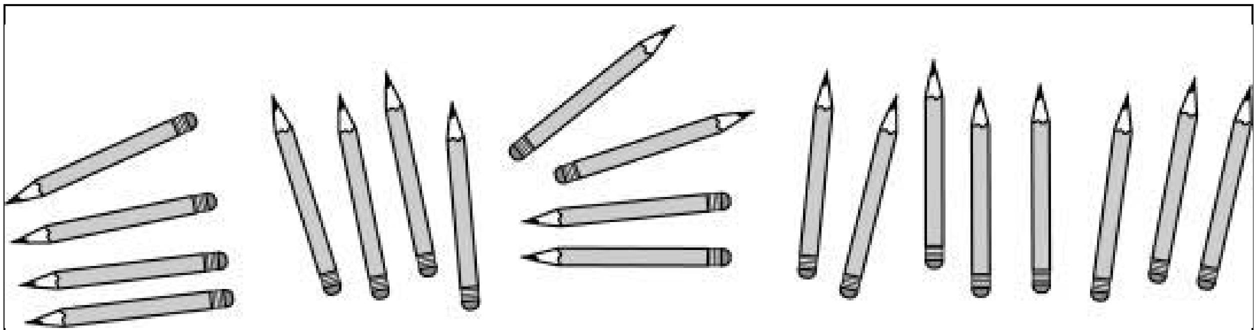
Framework classifications

PISA 2003 framework[1]		NAEP 2003 framework[2]	
Overarching idea	Uncertainty	Content strand	Data analysis, statistics, and probability
Competency cluster	Reproduction	Topic	Determine the probability of a simple event
Situation or context	Personal	Grade level	8
Mathematical complexity level:[3] moderate			

[1] Classified by PISA assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 5

NAEP short constructed-response item – grade 4



Together, Sara and Brendan have 20 pencils. Sara says 1/4 of the pencils are hers. Brendan says 15 of the pencils belong to him. Explain how they both could be right. Use words or drawings.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

Scoring guide
In this question the student was given information in two different ways—a fractional part and a number of items—and the student needed to justify that these two interpretations of the same situation were consistent. To answer the question, the student needed to observe that the fractional part has meaning in terms of the number of items, or that the number of items can be represented as a fractional part of the whole amount. Students were permitted to use a calculator.

Correct
They can both be right because ¼ of 20 = 5 and 20 – 5 = 15
OR
¼ is 5 and ¾ is 15
OR

Sara ¼ or 5	Brendan ¾ or 15

Partial
¼ is 5
OR
¾ is 15
OR
"Sara has 5" (5 must be connected to Sara; if states Sara has 5 because 20 – 15 = 5, item is scored as incorrect).

Incorrect
Incorrect response (includes 20 – 15 = 5, 5 + 15 = 20, and switching names)

EXAMPLE 5—continued

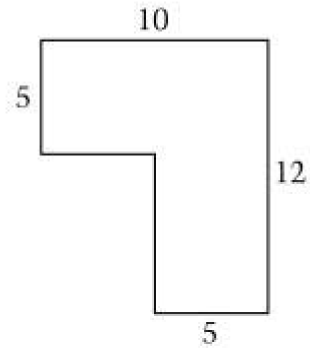
Framework classifications

NAEP 2003 framework[1]	TIMSS 2003 framework[2]
Number sense, properties, and operations	Number
Apply ratios and proportional thinking in a variety of situations	Fractions and decimals
Grade 4	Grade 4
Conceptual understanding	
Mathematical complexity level:[3] high	

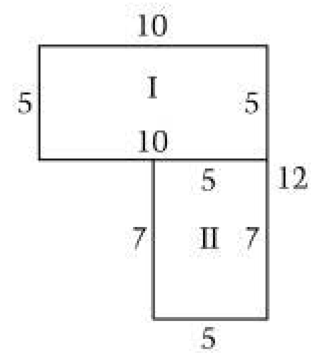
[1] Classified by NAEP assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 6

NAEP extended constructed-response item – grade 8



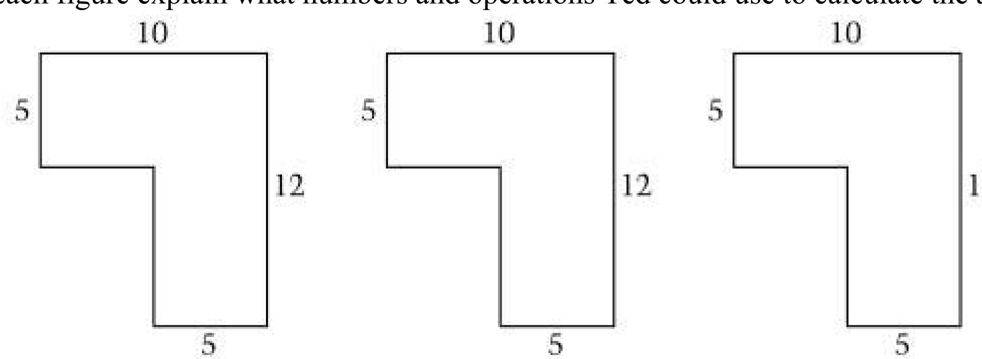
Ted wants to purchase floor covering for the hallway shown above. He knows there are many ways to find the area of the hallway. One way is to divide the hallway into the sections shown below and then add together the area of each section.



Area of Hallway = Area of Region I + Area of Region II

Area = (5 x 10) + (7 x 5)

Use the figures below to show 3 other ways that Ted can divide the hallway to find its area. Below each figure explain what numbers and operations Ted could use to calculate the area.

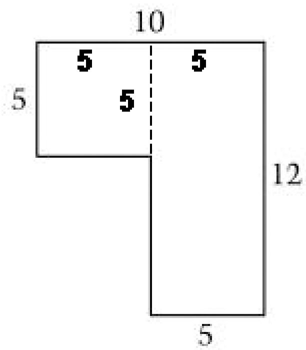


SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

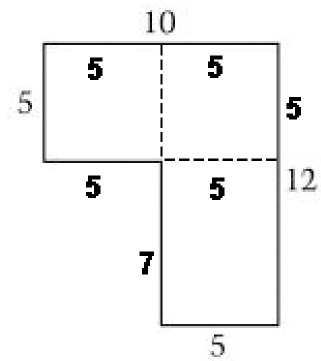
EXAMPLE 6—continued

Solution:

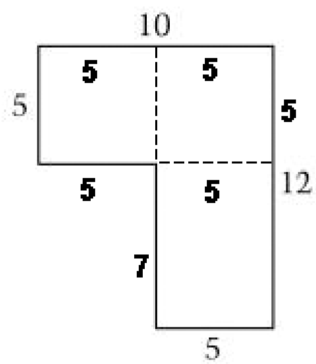
Possible answers include:



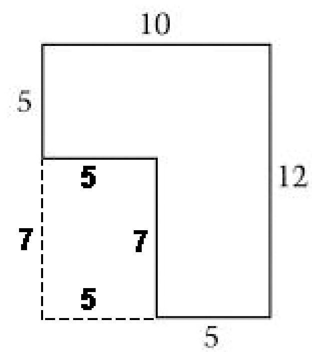
Area = (5 × 5) + (5 × 12)



Area = (5 × 5) + (5 × 5) + (7 × 5)



Area = $5\left(\frac{7+12}{2}\right)+5\left(\frac{5+10}{2}\right)$



Area = (10 × 12) + (7 × 5)
For this method the missing piece must be indicated

EXAMPLE 6—continued

Scoring guide

In this question the student was asked to draw and explain three different ways to divide an L-shaped region to determine the area. The student was also required to give an expression representing the area for each of the different divisions of the region (however, the student was not asked to calculate the area of the region). There are many possible ways to do this but to earn full credit the student needed to show three different divisions of the region, label the lengths in each figure correctly, and write an expression for the area consistent with each figure. This question requires visualization and knowledge of one or more formulas for finding area. Any division of the figures into rectangles, triangles, trapezoids, or parallelograms is acceptable. Student either needs to show (5 x 5) + (5 x 12), for example OR label all appropriate dimensions in the figure to give credit for 25 + 60. However, "25 + 60" is not acceptable if dimensions are not labeled.
Extended Three figures divided correctly with no incorrect labels and three correct expressions for area.
Satisfactory Three figures divided correctly with no incorrect labels and two correct expressions for area.
Partial Two figures divided correctly with no incorrect labels and one or two correct expressions for area of those figures. OR Three figures divided correctly with no incorrect labels and one correct expression for area.
Minimal One figure divided correctly with no incorrect labels and correct expression for area of that figure. OR Two or three figures divided correctly with no incorrect labels and no correct (or missing) expressions for area of figures.
Incorrect Incorrect response.

Framework classifications

NAEP 2003 framework[1] Measurement Estimate, calculate (using basic principles or formulas), or compare perimeter, area, volume, and surface area in meaningful contexts or solve mathematical and real-world problems Grade 8 Problem solving	TIMSS 2003 framework[2] Measurement Tools, techniques, and formulas Grade 8
PISA 2003 framework[2] Overarching idea — Space and shape Competency cluster — Reproduction Situation or context — Educational/occupational	
Mathematical complexity level:[3] high	

[1] Classified by NAEP assessment developers
[2] Classified by expert panel. NAEP grade 8 *problem solving* items were classified to both the TIMSS and PISA frameworks.
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 7

TIMSS extended constructed-response item – grade 8

Betty, Frank, and Darlene have just moved to Zedland. They each need to get phone service. They received the following information from the telephone company about the two different phone plans it offers.

They must pay a set fee each month and there are different rates for each minute they talk. These rates depend on the time of the day or night they use the phone, and on which payment plan they choose. Both plans include time for which phone calls are free. Details of the two plans are shown in the table below.

Plan	Monthly Fee	Rate per minute		Free minutes per month
		Day (8 am – 6 pm)	Night (6 pm – 8 am)	
Plan A	20 zeds	3 zeds	1 zed	180
Plan B	15 zeds	2 zeds	2 zeds	120

Betty talks for less than 2 hours per month. Which plan would be less expensive for her?

Less expensive plan ___________

Explain your answer in terms of both the monthly fee and free minutes.

SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment.

Scoring guide

Correct response Plan B with explanation that includes free minutes and explicit reference to lower monthly fee for Plan B.
Partial response Plan B with explicit reference to lower monthly fee and no reference to free minutes.
Incorrect response Plan B with inadequate (only free minutes) or no explanation. OR Plan A with or without explanation.

EXAMPLE 7—continued

Framework classifications

TIMSS 2003 framework[1]	NAEP 2003 framework[2]
Data	Data analysis, statistics, and probability
Data interpretation	Read, interpret, and make predictions using tables and graphs
Grade 8	Grade 8
Reasoning	
Mathematical complexity level:[3] moderate	

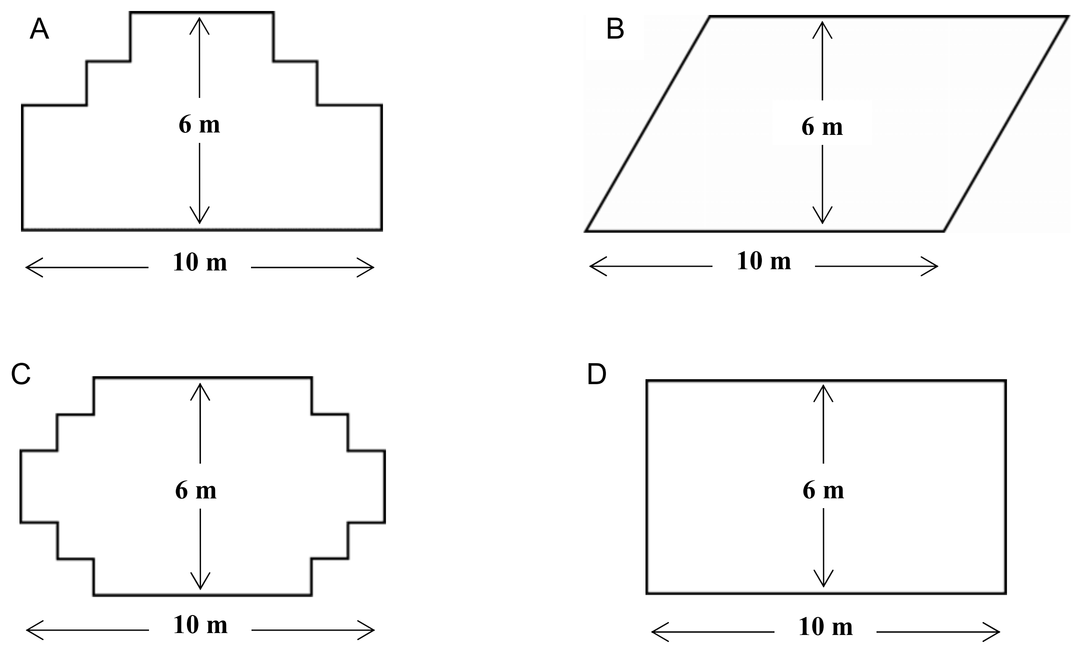
[1] Classified by TIMSS assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 8

PISA complex multiple-choice item

<u>Carpenter</u>

A carpenter has 32 meters of timber and wants to make a border around a garden bed. He is considering the following designs for the garden bed.



Circle either "Yes" or "No" for each design to indicate whether the garden bed can be made with 32 meters of timber.

Garden bed design	Using this design, can the garden bed be made with 32 meters of timber?
Design A	Yes / No
Design B	Yes / No
Design C	Yes / No
Design D	Yes / No

SOURCE: Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment.

EXAMPLE 8—continued

Scoring guide

Full credit Exactly four correct. Design A Yes Design B No Design C Yes Design D Yes
Partial credit Exactly three correct.
No credit Two or fewer correct.

Framework classifications

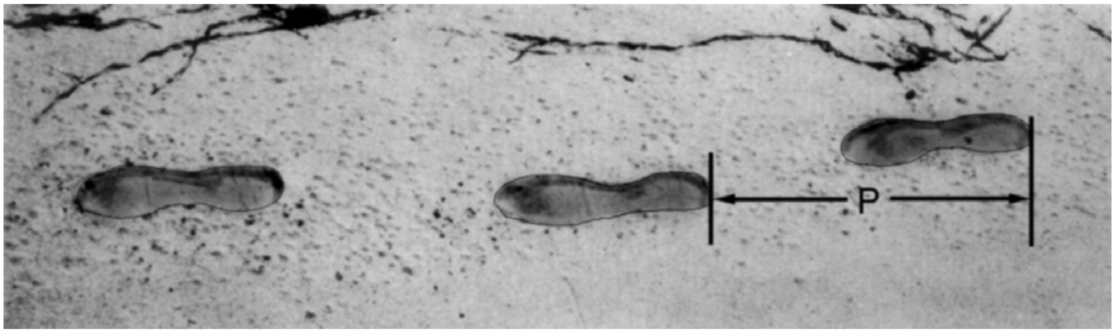
PISA 2003 framework[1]		NAEP 2003 framework[2]	
Overarching idea	Space and shape	Content strand	Measurement
Competency cluster	Connections	Topic	Estimate, calculate (using basic principles or formulas), or compare perimeter, area, volume, and surface area in meaningful contexts to solve mathematical and real-world problems.
Situation or context	Educational/occupational		
		Grade level	8
Mathematical complexity level:[3] moderate			

[1] Classified by PISA assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 9

PISA open constructed-response item

<u>Walking</u>



The picture shows the footprints of a man walking. The pacelength P is the distance between the rear of two consecutive footprints.

For men, the formula, $\frac{n}{P} = 140$, gives an approximate relationship between n and P where,

n = number of steps per minute, and

P = pacelength in meters.

Question 1: **WALKING**

If the formula applies to Heiko's walking and Heiko takes 70 steps per minute, what is Heiko's pacelength? Show your work.

SOURCE: Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2003 Mathematical Literacy Assessment.

EXAMPLE 9—continued

Scoring guide

Full credit 0.5 m OR 50 cm OR ½ (unit not required) • 70/ P = 140 [substitute numbers in the formula only]. 70 = 140 P P = 0.5 • 70/140
Partial credit Correct substitution of numbers in the formula, but incorrect answer, or no answer. • 70/ P = 140 [substitute numbers in formula only]. • 70/ P = 140 70 = 140 P P = 2 [correct substitution, but working out is incorrect]. OR Correctly manipulated the formula into $P = n/140$, but no further correct working.
No credit Other responses

Framework classifications

PISA 2003 framework[1]		NAEP 2003 framework[2]	
Overarching idea	Change and relationships	Content strand	Algebra and functions
Competency cluster	Reproduction	Topic	Interpret contextual situations and perform algebraic operations on real numbers and algebraic expressions to solve mathematical and real-world problems
Situation or context	Personal		
		Grade level	8
Mathematical complexity level:[3] low			

[1] Classified by PISA assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 10

TIMSS short constructed-response item – grade 4

$204 \div 4 =$ Answer: _____________ SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment.

Scoring guide

Correct response 51
Incorrect response Incorrect

Framework classifications

TIMSS 2003 framework[1]	NAEP 2003 framework[2]
Number	Number sense, properties, and operations
Whole numbers	Compute with numbers (i.e., add, subtract, multiply, divide)
Grade 4	Grade 4
Knowing facts and procedures	
Mathematical complexity level:[3] low	

[1] Classified by TIMSS assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 11

TIMSS multiple choice item – grade 4

It takes Chris 4 minutes to wash a window. He wants to know how many minutes it will take him to wash 8 windows at this rate. He should

A) multiply 4 x 8
B) divide 8 by 4
C) subtract 4 from 8
D) add 8 and 4

SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment.

Answer Key: A

Framework classifications

TIMSS 2003 framework[1]	NAEP 2003 framework[2]
Number Whole numbers Grade 4 Solving routine problems	Number sense, properties, and operations Compute with numbers (i.e., add, subtract, multiply, divide) Grade 4
Mathematical complexity level:[3] low	

[1] Classified by TIMSS assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 12

NAEP multiple-choice item – grades 4, 8, and 12

Add:

$$\begin{array}{r} 238 \\ +\ \ 462 \\ \hline \end{array}$$

A) 600
B) 690
C) 700
D) 790

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

Key: C

Framework classifications

NAEP 2003 framework[1]	TIMSS 2003 framework[2]
Number sense, properties, and operations	Number
Compute with numbers (i.e., add, subtract, multiply, divide)	Whole numbers
Grade 4	Grade 4
Procedural knowledge	
Mathematical complexity level:[3] low	

[1] Classified by NAEP assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 13

NAEP multiple-choice item – grade 4

Estela wants to buy 2 notebooks that cost \$2.79 each, including tax. If she has one-dollar bills and no coins, how many one-dollar bills does she need? A) 3 B) 4 C) 5 D) 6 SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

Answer Key: D

Framework classifications

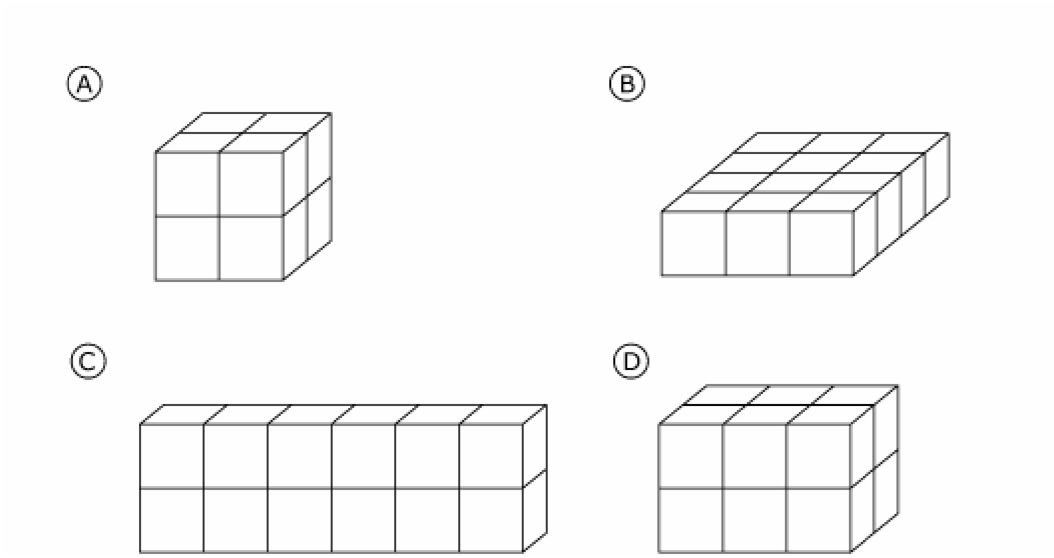
NAEP 2003 framework[1]	TIMSS 2003 framework[2]
Number sense, properties, and operations	Number
Use computations and estimation in applications	Fractions and decimals
Grade 4	Grade 8
Problem solving	
Mathematical complexity level:[3] moderate	

[1] Classified by NAEP assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 14

TIMSS multiple-choice item – grade 8

All the small blocks are the same size. Which stack of blocks has a different volume from the others?



SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment.

Answer Key: A

Framework classifications

TIMSS 2003 framework[1]	NAEP 2003 framework[2]
Measurement Tools, techniques, and formulas Grade 8 Using concepts	Measurement Estimate, calculate (using basic principles or formulas), or compare perimeter, area, volume, and surface area in meaningful contexts to solve mathematical and real-world problems. Grade 4
Mathematical complexity level:[3] moderate	

[1] Classified by TIMSS assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 15

TIMSS multiple-choice item – grade 4

All of the pupils in a class cut out paper shapes. The teacher picked one out and said, "This shape is a triangle." Which of these statements MUST be correct?

A) The shape has three sides.
B) The shape has a right angle.
C) The shape has equal sides.
D) The shape has equal angles.

SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment.

Answer Key: A

Framework classifications

TIMSS 2003 framework[1]	NAEP 2003 framework[2]
Geometry Two- and three-dimensional shapes Grade 4 Knowing facts and procedures	Geometry and spatial sense No match to topic Grade 4
Mathematical complexity level:[3] low	

[1] Classified by TIMSS assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 16

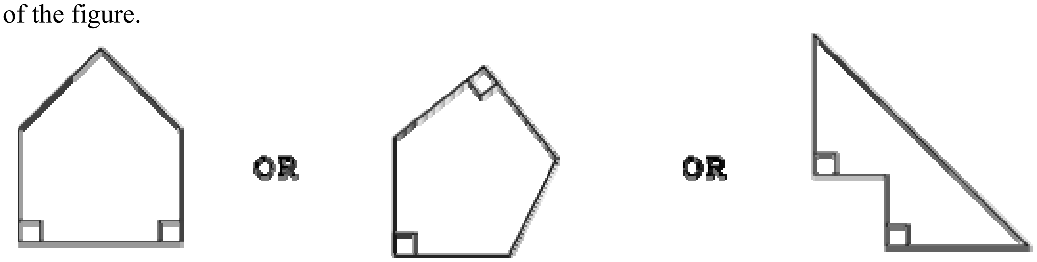
NAEP short constructed-response item – grade 4

In the space below, draw a closed figure with 5 sides. Make 2 of the angles right angles. SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

Scoring guide

Solution:
Figure must be closed and have 5 sides and 2 or more right angles.

Right angles do not have to be marked, but should appear to be right angles. Two right angles must be on the inside of the figure.



In this question the student needed to show geometric understanding by drawing a closed figure with 5 sides and at least 2 right angles. Students did not have a ruler or protractor.
Correct Correct response.
Incorrect Figure drawn is a five-pointed star with a pentagon shown in the interior. The pentagon may or may not have right angles. OR Figure has 5 sides and only 1 right angle. OR No right angles in the figure drawn. OR Figure is not 5-sided or is not closed.

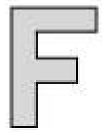
Framework classifications

NAEP 2003 framework[1]	TIMSS 2003 framework[2]
Geometry and spatial sense	Geometry
Describe, visualize, draw, and construct geometric figures	Two- and three-dimensional shapes
Grade 4	Grade 4
Conceptual understanding	
Mathematical complexity level:[3] moderate	

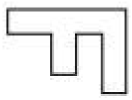
[1] Classified by NAEP assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

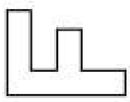
EXAMPLE 17

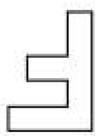
NAEP multiple-choice item – grades 4, 8, and 12

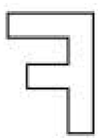


The figure above is shaded on the top side and white on the under side. If the figure were flipped over, its white side could look like which of the following figures?

A) 

B) 

C) 

D) 

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

Answer Key: D

Framework classifications

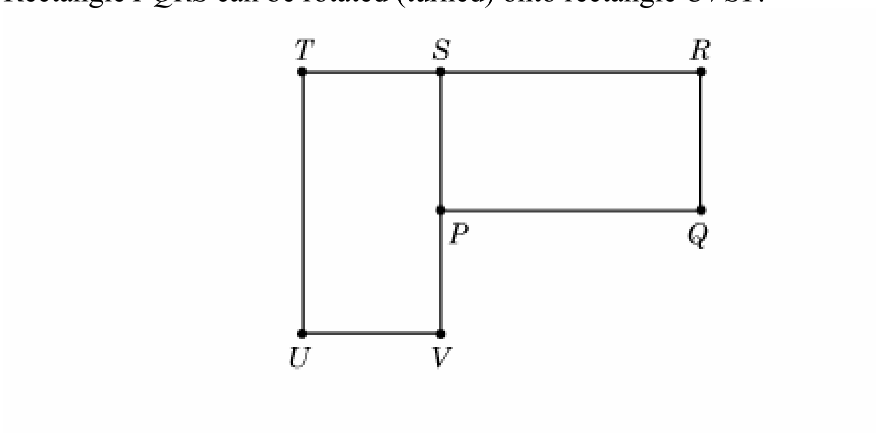
NAEP 2003 framework[1]	TIMSS 2003 framework[2]
Geometry and spatial sense	Geometry
Identify the relationship (congruence, similarity) between a figure and its image under a transformation	Symmetry and transformations
Grade 4	Grade 4
Conceptual understanding	
Mathematical complexity level:[3] low	

[1] Classified by NAEP assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 18

TIMSS multiple-choice item – grade 8

Rectangle *PQRS* can be rotated (turned) onto rectangle *UVST*.



What point is the center of rotation?

A) *P*
B) *R*
C) *S*
D) *T*
E) *V*

SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment.

Answer Key: C

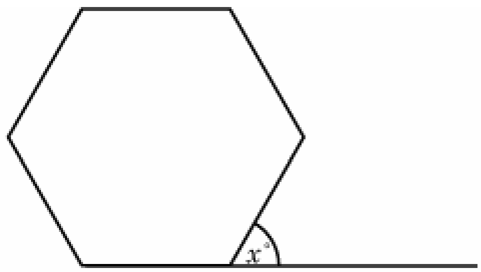
Framework classifications

TIMSS 2003 framework[1]	NAEP 2003 framework[2]
Geometry	Geometry and spatial sense
Symmetry and transformations	Identify the relationship (congruence, similarity) between a figure and its image under a transformation
Grade 8	Grade 12
Reasoning	
Mathematical complexity level:[3] low	

[1] Classified by TIMSS assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 19

TIMSS short constructed-response item – grade 8



The figure above is a regular hexagon. What is the value of x?

Answer: ___________

SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment.

Scoring guide

Correct response 60 degrees
Incorrect response 120 degrees OR Other incorrect

Framework classifications

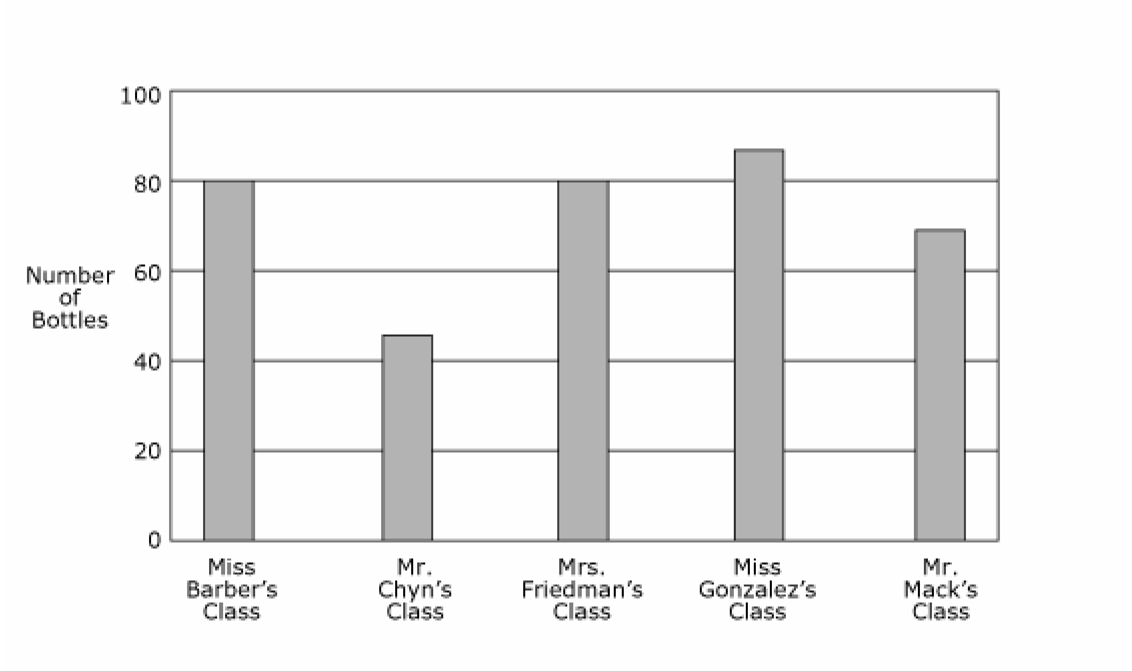
TIMSS 2003 framework[1]	NAEP 2003 framework[2]
Geometry	Geometry and spatial sense
Two- and three-dimensional shapes	Apply geometric properties and relationships in solving problems
Grade 8	Grade 8
Solving routine problems	
Mathematical complexity level:[3] low	

[1] Classified by TIMSS assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 20

TIMSS multiple-choice item – grade 4

Central School had a bottle collection. Children in each class brought empty bottles to school. The principal made a bar graph of the number of bottles from five classes.



Which two classes collected exactly 80 bottles?

A) Miss Barber's and Mrs. Friedman's classes
B) Miss Barber's and Mr. Mack's classes
C) Mrs. Friedman's and Miss Gonzalez's classes
D) Miss Gonzalez's and Mr. Mack's classes

SOURCE: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS) 2003 Assessment.

Key: A

Framework classifications

TIMSS 2003 framework[1]	NAEP 2003 framework[2]
Data	Data analysis, statistics, and probability
Data representation	Read, interpret, and make predictions using tables and graphs
Grade 4	Grade 4
Solving routine problems	
Mathematical complexity level:[3] low	

[1] Classified by TIMSS assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 21

NAEP multiple-choice item – grade 8

$3 + 15 \div 3 - 4 \times 2 =$

A) -9
B) -2
C) 0
D) 4
E) 5

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

Key: C

Framework classifications

NAEP 2003 framework[1]	TIMSS 2003 framework[2]
Algebra and functions	Number
Interpret contextual situations and perform algebraic operations on real numbers and algebraic expressions to solve mathematical and real-world problems	Whole numbers
Grade 8	Grade 8
Procedural knowledge	
Mathematical complexity level:[3] low	

[1] Classified by NAEP assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 22

NAEP multiple choice item – grades 8 and 12

Fifteen boxes each containing 8 radios can be repacked in 10 larger boxes each containing how many radios?

A) 8
B) 10
C) 12
D) 80
E) 120

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

Answer key: C

Framework classifications

NAEP 2003 framework[1]	TIMSS 2003 framework[2]
Number sense, properties, and operations	Number
Use computations and estimation in applications	Whole numbers
Grade 8	Grade 8
Problem solving	

PISA 2003 framework[2]	
Overarching idea	Quantity
Competency cluster	Reproduction
Situation or context	Not classified
Mathematical complexity level:[3] moderate	

[1] Classified by NAEP assessment developers
[2] Classified by expert panel. NAEP grade 8 *problem solving* items were classified to both the TIMSS and PISA frameworks.
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 23

NAEP short constructed-response item – grades 8 and 12

One store, Price Pleasers, reduces the price each week of a \$100 stereo by 10 percent of the original price. Another store, Bargains Plus, reduces the price each week of the same \$100 stereo by 10 percent of the previous week's price. After 2 weeks, how will the prices at the two stores compare? (A) The price will be cheaper at Price Pleasers. (B) The price will be the same at both stores. (C) The price will be cheaper at Bargain Plus. Explain your reasoning. SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

Scoring guide

In this question the student was asked to compare the sale price of a stereo, after 3 weeks, based on two different ways for reducing the price. In one store, the price was reduced each week by a fixed amount (10% of \$100, or \$10). In the other store the price was reduced each week by a varying amount (10% of the current price, which is less each week). To earn full credit, the student needed to indicate that the price would be less at the first store after 3 weeks and explain how the solution was obtained. Students were permitted to use a calculator.
Solution: A. Cheaper at Price Pleasers At Price Pleasers the stereo would be \$80 after 2 weeks. At Bargain Plus, it would cost \$81. OR Successive 10% reductions of the original price will yield greater savings than successive reductions of 10% of the reduced price.
Correct Correct response—cheaper at Price Pleasers with an explanation that compares price at each store after 2 weeks (\$80 vs. \$81). OR Cheaper at Price Pleasers with an explanation that generalizes as described in solution above. NOTE: Score CORRECT if incorrect answer is B or C with a clear statement that Price Pleasers is cheaper and explanation is correct and complete.
Partial Cheaper at Price Pleasers with anything less than a complete explanation. OR Computes the correct amount for at least 2 weeks for either Price Pleasers or Bargain Plus, but conclusion is missing, incomplete, or incorrect (if the store is not identified the score is still a 2)
Incorrect Incorrect response.

EXAMPLE 23—continued

Framework classifications

NAEP 2003 framework[1]	TIMSS 2003 framework[2]
Number sense, properties, and operations Apply ratios and proportional thinking in a variety of situations Grade 8 Problem solving	Number Ratio, proportion, and percent Grade 8

PISA 2003 framework[2]	
Overarching idea	Change and relationships
Competency cluster	Connections
Situation or context	Public/personal
Mathematical complexity level:[3] moderate	

[1] Classified by NAEP assessment developers
[2] Classified by expert panel. NAEP grade 8 *problem solving* items were classified to both the TIMSS and PISA frameworks.
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 24

NAEP extended constructed-response item – grade 8

While she was on vacation, Tara sent 14 friends either a letter or a postcard. She spent $3.84 on postage. If it costs $0.20 to mail a postcard and $0.33 to mail a letter, how many letters did Tara send?
Show what you did to get your answer.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

Scoring guide

This question was a word problem that asked students to consider two values—the number of letters and the number of postcards—even though the student was only asked for the number of letters. This question could be solved in several ways. A student could reason numerically to find the number of letters and the number of postcards, possible by using a guess-and-check strategy or by creating a table. Another possibility was to set up and solve a system of two linear equations in two unknowns. To earn full credit, students needed to show how they obtained the answer. Students were permitted to use a calculator.

Solution:
8 letters
.20 (6) + .33(8) = $3.84
Students may use a variety of strategies to solve this, including guess and check, formal algebra, or others. For example,

# postcards	# letters	Total cost
1	13	4.49
2	12	4.36
3	11	4.23
4	10	4.10
5	9	3.97
6	8	**3.84**
7	7	3.71
8	6	3.58

OR
$x + y = 14$
$.20x + .33y = 3.84$
Therefore $.20x + .33(14-x) = 3.84$
So $x = 6$ and y = 8

Extended
Correct response.

Satisfactory
Correct, complete process is indicated, but answer is not 8 and has a minor computational error.
OR
Shows correct, complete process but does not indicate answer.

Partial
Correct, complete process in indicated, but answer is not 8 and there are several computational errors. (Process must clearly illustrate a correct strategy, such as a table or equations.)
OR
Correct response of 8 but shows no work or incomplete work.

Minimal
Process is incorrect because it ignores one or more pieces of given information.
OR
Process is correct but incomplete (process may be guess and check or another process which may lead to the correct answer, i.e., a chart but no equation, but goal is not clearly defined) and answer is not 8

Incorrect
Incorrect response.

EXAMPLE 24—continued

Framework classifications

<table>
<tr><td colspan="2"><u>NAEP 2003 framework[1]</u>
Algebra and functions

Solve systems of equations and inequalities using appropriate methods

Grade 8

Problem solving</td><td><u>TIMSS 2003 framework[2]</u>
Algebra

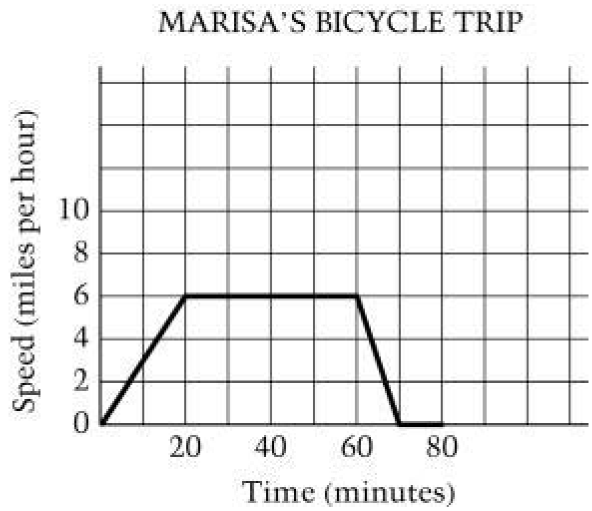
Equations and formulas

Grade 8</td></tr>
<tr><td colspan="3"><u>PISA 2003 framework[2]</u></td></tr>
<tr><td>Overarching idea</td><td colspan="2">Change and relationships</td></tr>
<tr><td>Competency cluster</td><td colspan="2">Connections</td></tr>
<tr><td>Situation or context</td><td colspan="2">Public/personal</td></tr>
<tr><td colspan="3">Mathematical complexity level:[3] moderate</td></tr>
</table>

[1] Classified by NAEP assessment developers
[2] Classified by expert panel. NAEP grade 8 *problem solving* items were classified to both the TIMSS and PISA frameworks.
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

EXAMPLE 25

NAEP extended constructed-response item – grade 8



The graph above represents Marisa's riding speed throughout her 80-minute bicycle trip. Use the information in the graph to describe what could have happened on the trip, including her speed throughout the trip.

During the first 20 minutes, Marisa

____

____

From 20 minutes to 60 minutes, she

____

____

From 60 minutes to 80 minutes, she

____

____

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2003 Mathematics Assessment.

EXAMPLE 25—continued

Scoring guide

In this question the student was asked to translate across representations by interpreting information presented graphically and giving a verbal description (story) of Marisa's bicycle trip. The given graph presented Marisa's speed on the trip as a function of time. To earn full credit, the student needed to give both quantitative and qualitative information about the situation, including the observation that Marisa was at a stop during the last time interval.

Solution:

During the first 20 minutes, Marisa increased her speed from 0 to 6 mph.

From 20 to 60 minutes, she remained at 6 mph.

From 60 to 80 minutes, she decreased her speed from 6 mph to 0 mph and stopped. (Must have both decrease and stop from this interval).

Responses may be presented in the following ways.

Category A	**Category B**	**Category C**
Speed from 0 to 6	Increase	Downhill
Speed at 6	Remained the same	Float road (level)
Speed from 6 to 0; stop	Decrease	Uphill

- Responses may mix parts of more than one category
- Speed from 0 to 6 may be expressed as 0-6 (likewise for speed from 6 to 0 as 6-0)
- Responses may include extraneous correct information.
- "Stop" after 70 minutes may be expressed as "had no speed" or "maintained speed of 0," "stayed at that speed" having stated 0 mph.

Extended

All of Category A and all of Category B or Category C.

Satisfactory

All of Category A and incomplete Category B or Category C.
OR
Category A without stop and all of Category B or Category C.

Partial

All of Category A and no Category B or Category C.
OR
All of Category B and Category C, with stop.

Minimal

Category A without stop.
OR
Category B only.
OR
Category C only.
OR
Response that accounts for each of three parts of trip.

Incorrect

Incorrect response.

EXAMPLE 25—continued

Framework classifications

NAEP 2003 framework[1]	TIMSS 2003 framework[2]
Data analysis, statistics, and probability	Data
Read, interpret, and make predictions using tables and graphs	Data interpretation
Grade 8	Grade 8
Problem solving	

PISA framework[2]	
Overarching idea	Change and relationships
Competency cluster	Connections
Situation or context	Educational/occupational

Mathematical complexity level:[3] moderate

[1] Classified by NAEP assessment developers
[2] Classified by expert panel. NAEP grade 8 *problem solving* items were classified to both the TIMSS and PISA frameworks.
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.

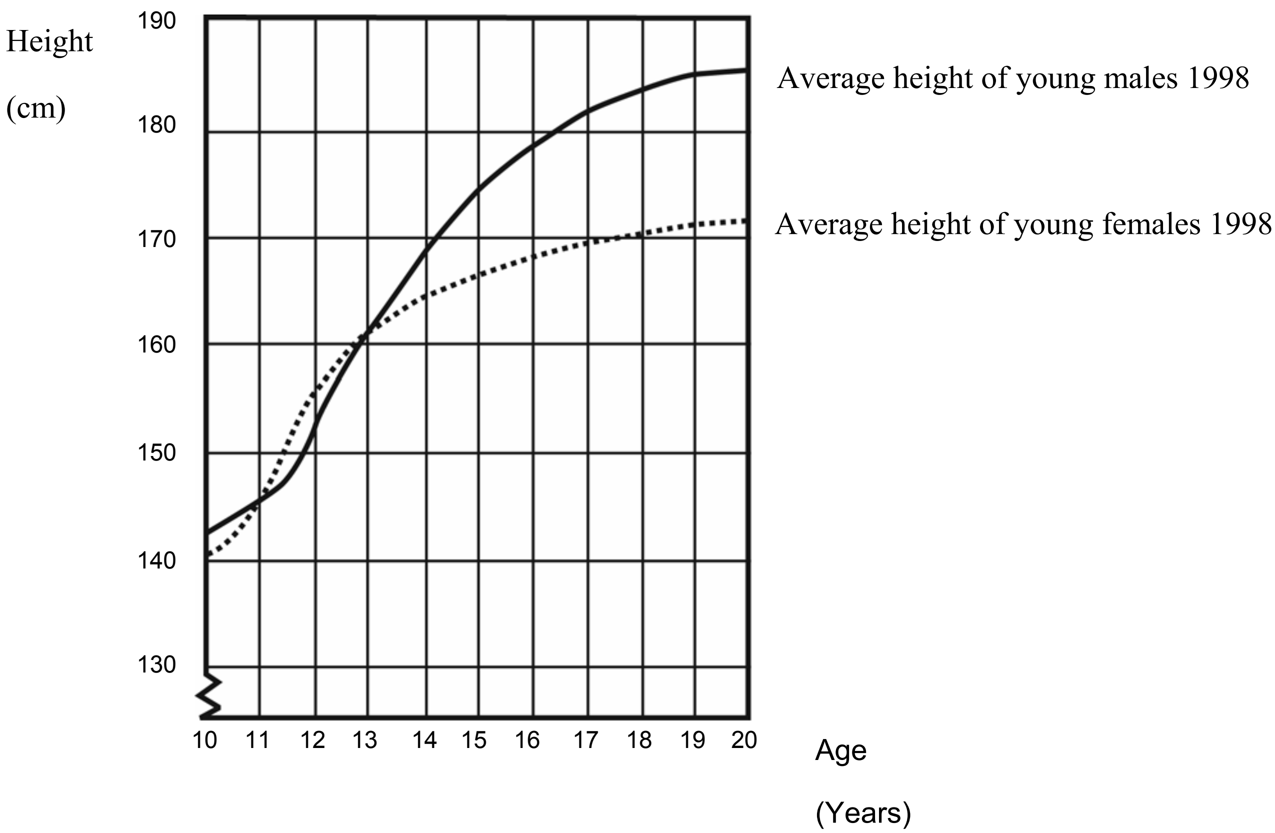
EXAMPLE 26

PISA task
Question 1: closed constructed-response item
Question 2: closed constructed-response item
Question 3: open constructed-response item

Growing Up

YOUTH GROWS TALLER

In 1998 the average height of both young males and young females in the Netherlands is represented in this graph.



EXAMPLE 26—continued

Question 1

Since 1980 the average height of 20-year-old females has increased by 2.3 cm, to 170.6 cm. What was the average height of a 20-year-old female in 1980?

Answer: ______________________cm

Question 2

According to this graph, on average, during which period in their life are females taller than males of the same age?

____

____

Question 3

Explain how the graph shows that on average the growth rate for girls slows down after 12 years of age.

____

____

____

SOURCE: Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA), 2003 Mathematical Literacy Assessment.

EXAMPLE 26—continued

Scoring guide

Question 1

Full credit 168.3 cm (unit already given).
No credit Other responses.

Question 2

Full credit Gives the correct interval, from 11-13 years. OR States that girls are taller than boys when they are 11 and 12 years old. (This answer is correct in daily-life language because it means the interval from 11-13).
Partial credit Other subsets of (11, 12, 13), not included in the full credit section.
No credit Other responses.

Question 3

Full credit The key here is that the response should refer to the "change" of the gradient of the graph for female. This can be done explicitly or implicitly. Full credit is for explicitly mentioning about the steepness of the curve of the graph or for implicit comparison using the actual amount of growth before 12 years and after 12 years of age. Refers to the reduced steepness of the curve from 12 years onwards, using daily-life language, not mathematical language. OR Refers to the reduced steepness of the curve from 12 years onwards, using mathematical language. OR Comparing actual growth (comparison can be implicit).
No credit Student indicates that female height drops below male height, but does NOT mention the steepness of the female graph or a comparison of the female growth rate before and after 12 years. OR Other incorrect responses. For example, the response does not refer to the characteristics of the graph, as the question clearly asks about how the graph shows the answer.

EXAMPLE 26—continued

Framework classifications

Question 1

PISA 2003 framework[1]		NAEP 2003 framework[2]	
Overarching idea	Change and relationships	Content strand	Number sense, properties, and operations
Competency cluster	Reproduction	Topic	Use computations and estimation in applications
Situation or context	Scientific	Grade level	8
Mathematical complexity level:[3] low			

Question 2

PISA 2003 framework[1]		NAEP 2003 framework[2]	
Overarching idea	Change and relationships	Content strand	Data analysis, statistics, and probability
Competency cluster	Reproduction	Topic	Read, interpret, and make predictions using tables and graphs
Situation or context	Scientific	Grade level	8
Mathematical complexity level:[3] moderate			

Question 3

PISA 2003 framework[1]		NAEP 2003 framework[2]	
Overarching idea	Change and relationships	Content strand	Data analysis, statistics, and probability
Competency cluster	Connections	Topic	Read, interpret, and make predictions using tables and graphs
Situation or context	Scientific	Grade level	8
Mathematical complexity level:[3] low			

[1] Classified by PISA assessment developers
[2] Classified by expert panel
[3] Mathematical complexity level classifications were made by the expert panel based on the definitions in the NAEP 2005 framework.