A Comparison of the 2011 Trends in International Mathematics and Science Study (TIMSS) Assessment Items and the 2011 National Assessment of Educational Progress (NAEP) Frameworks

October 2013

Stephen Provasnik
Project Officer
National Center for Education Statistics

Chien-Yu Lin
David Darling
James Dodson
AVAR

Abstract

This paper reports the results of two expert panels that examined the similarities and differences between the Trends in International Mathematics and Science Study (TIMSS) and the National Assessment of Educational Progress (NAEP) by classifying TIMSS 2011 grade 4 and 8 mathematics and science items to the NAEP 2011 Mathematics and Science Frameworks. They found that the TIMSS and NAEP mathematics and science assessments assess similar content areas and that almost all TIMSS mathematics items fit the NAEP mathematics framework, but not all TIMSS science items fit the NAEP science framework. Both TIMSS mathematics and science items cover only a subset of the NAEP target grade's framework. More than 20 percent of TIMSS science items fall outside the content specified by the NAEP framework. In addition, some mathematics and science items that do cover specific NAEP content, do so at different grade levels in NAEP than in TIMSS.

Introduction

One of the objectives of the National Center for Education Statistics (NCES) is to provide a comprehensive picture of how U.S. students perform in key subject areas. In the United States, national data on fourth- and eighth-grade students' mathematics and science achievement come primarily from two sources: the Trends in International Mathematics and Science Study (TIMSS) and the National Assessment of Educational Progress (NAEP), also known as the "Nation's Report Card," These two assessments attempt to measure the level of fourth- and eighth-graders' mathematics and science achievement (i.e., what students know and can do) by asking students a broad range of questions in both subject areas. The two assessments, however, were created for different purposes. TIMSS is meant to provide internationally comparable trend data on student mathematics and science achievement at the national level, while NAEP is meant to provide data on student mathematics and science achievement for the nation, national population subgroups, states, and selected large urban school districts. To understand the similarities and differences between TIMSS and NAEP, NCES has conducted a series of comparison studies to compare the

TIMSS assessments and the "main" NAEP assessments across years. [1] Previous comparisons of TIMSS and NAEP have found distinctive differences between the two in terms of subject matter coverage and level of difficulty (NCES 2007; Neidorf et al. 2006; Neidorf, Binkley, and Stephens 2006; NCES 2001). This paper presents the findings from the item comparison study of these two assessments that were conducted in 2011.

Adopting the method used in the previous NCES comparison study (NCES 2007), this study compared TIMSS 2011 items with the NAEP framework[2] in two expert panel meetings convened in Washington, D.C. in October and December 2011: the first meeting focused on the mathematics assessments and the second meeting focused on the science assessments. Both Expert Panels consisted of nationally recognized content experts familiar with the TIMSS and NAEP assessments (see Appendix A for a list of panel members). Each Expert Panel was tasked to analyze all the TIMSS items in their subject expertise and to "map" those items to the NAEP assessment framework, by item content and item grade-level. In addition, item formats were compared.

The Mathematics Expert Panel completed a "content match" analysis by classifying all the TIMSS 2011 fourth- and eighth-grade mathematics items into the following three categories specified in the NAEP 2011 framework for mathematics: (a) content area, (b) content area's subtopic, and (c) subtopic's objective by grade level. The Science Expert Panel completed its

[1] The name "main" NAEP refers to the cyclical NAEP state-level assessments. The "main" NAEP assessment is distinguished from both (a) the original NAEP national assessment of U.S. students (now called the "long-term trend" NAEP), which began in the 1970s based on different frameworks and is administered every four years under different testing conditions than the "main" NAEP, and (b) the Trial Urban District Assessment (TUDA) NAEP, a multiyear study of the feasibility of district-level NAEP in selected urban districts.

[2] An assessment's "framework" lays out a logical organization for the assessment's content and serves as the blueprint to guide the development of the assessment instrument by identifying the content to be assessed. For a brief introduction to the NAEP framework, see http://nces.ed.gov/nationsreportcard/frameworks.asp.

"content match" analysis by classifying all the TIMSS 2011 fourth- and eighth-grade science items into the following four categories specified in the NAEP 2009 or 2011 framework[3] for science: (a) content area, (b) content area's topic, (c) topic's subtopic, and (d) subtopic's content statement by grade level. Content areas, topics, and subtopics are the same across grade levels while objectives (for mathematics) and content statements (for science) are specific to the grade level.

The purpose of such an item-by-item review is to assess how well the TIMSS 2011 fourth- and eighth-grade item coverage matches the NAEP 2011 framework. With an "ideal" content match, all TIMSS items would map to the NAEP framework's finest level of detail (*objectives* in the case of mathematics and *content statement*s in the case of science) and all NAEP framework's subtopics and objectives or content statement would be addressed by TIMSS items. An ideal content match would indicate that the TIMSS' content coverage is virtually identical with NAEP's content coverage. Because these assessments were created for different purposes, an ideal content match is unlikely. At the same time, though, because these assessments attempt to measure mathematics and science achievement at the same levels, it is expected that the TIMSS 2011 fourth- and eighth-grade item coverage matches the NAEP 2011 framework to some degree. This "degree" of content match is measured by two ratios:

1. The number of **subtopics** in the NAEP 2011 framework that are addressed by TIMSS 2011 items over the total number of the framework's **subtopics**.

[3] As explained more fully later in the section Population Assessed, for this study, the NAEP 2009 framework is used for fourth-grade science and NAEP 2011 for eighth-grade science.

2. The number of **objectives** (for mathematics) or **content statements** (for science) in the NAEP 2011 framework that are addressed by TIMSS 2011 items over the total number of the framework's **objectives** or **content statements**.[4]

Besides the content match analysis, both Expert Panels also completed a "grade-level fit" analysis by classifying all TIMSS items at the most appropriate of the NAEP framework's three grade levels. In this item-by-item analysis, the Expert Panels reviewed all TIMSS 2011 items to determine (a) what percentage of the fourth-grade TIMSS 2011 items were consistent with the general content and targeted skills specified in the NAEP 2011 mathematics or NAEP 2009 science framework for fourth grade, and (b) what percentage of TIMSS items were more consistent with the content and targeted skills specified for the eighth grade or twelfth grade in the respective NAEP framework. The same type of analysis was completed for all eighth-grade TIMSS 2011 items: the Expert Panels reviewed each TIMSS 2011 eighth-grade item to determine (a) what percentage of the eighth-grade TIMSS 2011 items were consistent with the general content and targeted skills specified in the NAEP 2011 mathematics or science framework for eighth grade, and (b) what percentage were more consistent with the content and targeted skills specified for the fourth or twelfth grade in the NAEP 2011 frameworks. The purpose of this "grade-level fit" analysis is to assess the comparability of TIMSS items in terms of NAEP's classification of the key content to be measured at grade level and of content difficulty.

This paper reports the results from both Expert Panels' analyses, classifications, and comparisons of TIMSS 2011 and NAEP 2011. In addition, it provides additional comparative

[4] Note that this analysis does not compare whether TIMSS items cover the NAEP framework's objectives or content statements in similar percentages as they are actually covered by NAEP items. Because the NAEP item pool does not cover all objectives or content statements with the same number of items, some objectives or content statements may be addressed by multiple items, some by only one item, and, at times, it is possible that an assessment may have no items addressing a particular objective or content statement. Likewise, this analysis does not compare TIMSS items to the NAEP framework's cognitive dimensions which were not examined by the Expert Panels.

information about TIMSS and NAEP to understand these findings. Comparisons of trends in NAEP and TIMSS mathematics and science assessments are beyond the scope of this paper.

This paper is organized into four sections. The first section provides background information about TIMSS and NAEP and describes differences in their purpose, assessed population, and sample size. The next two sections compare the content coverage in mathematics and science according to the assessment frameworks of TIMSS and NAEP, and then present the findings of content match and grade-level fit analyses, as well as the item format comparisons. The last section of this paper summarizes the main findings from this comparison study. This paper does not include examples of TIMSS or NAEP items; however, a portion of each assessment's items have been publicly released and are available on the NCES website. For examples of TIMSS items, go to http://nces.ed.gov/timss/educators.asp; for examples of NAEP items, go to http://nces.ed.gov/nationsreportcard/about/booklets.asp.

1. A general comparison between TIMSS and NAEP assessments

TIMSS is a project of the International Association for the Evaluation of Educational Achievement (IEA), an independent, international cooperative of national research institutions and governmental research agencies. Since first administered in 1995, TIMSS has reported every four years on the mathematics and science achievement of students in fourth and eighth grades in participating countries and education systems around the world.[5] TIMSS 2011 is the fifth in a series of TIMSS assessments (Mullis et al. 2009). The TIMSS 2011 assessment framework

[5] TIMSS 1999 tested eighth-grade only; there was no assessment of fourth-grade students that year.

differs from the previous TIMSS 2007 framework in two ways: (a) the target percentages for the cognitive domains were adjusted, and (b) the substantive points under its objectives were consolidated to reduce the overall number of objectives.[6]

NAEP is a national survey conducted by the U.S. Department of Education's National Center for Education Statistics (NCES).[7] Since first administered in 1990, the main NAEP assessment has reported on the mathematics achievement of fourth- and eighth-grade students nationally and in participating U.S. states every two years (except between 1992 and 2003 when it reported only twice).[8] Since 1996, the main NAEP assessment has reported on the science achievement of fourth- and eighth-grade students nationally and in participating U.S. states roughly every four years, plus it reported on eighth-grade students in 2011.[9] The NAEP mathematics and science frameworks used to create the 2011 NAEP mathematics and science assessments (hereafter referred to as the "NAEP 2011 frameworks," although they were originally adopted for the 2009 NAEP assessment) differ in various ways from NAEP's previous (pre-2009) versions of these frameworks.[10] The NAEP 2011 framework for mathematics (*Mathematics Framework for the 2011 National Assessment of Educational Progress*, National Assessment Governing Board [NAGB] 2010a) differs from NAEP's pre-2009 mathematics framework in that it includes new objectives for twelfth grade and a new subtopic of "mathematical reasoning" for fourth, eighth,

[6] The TIMSS target percentages for the cognitive domains changed between 2007 and 2011 in science as follows: At grade 4, *applying* changed from 35 to 40 percent, and *reasoning* from 25 to 20 percent; at grade 8, *knowing* changed from 30 to 35 percent, and *reasoning* from 35 to 30 percent. The consolidation of objectives between the 2007 and 2011 framework reduced the total number of objectives (for which items needed to be written to) in mathematics from 38 to 28 for fourth grade and from 55 to 41 for eighth grade; and in science from 38 to 29 for fourth grade and from 66 to 50 for eighth grade. See Mullis, I.V.S. et al. 2009 and Mullis, I.V.S. et al. 2005.

[7] The National Assessment Governing Board (NAGB), appointed by the Secretary of Education but independent of the Department, sets policy for NAEP and is responsible for developing the framework and test specifications that serve as the blueprint for the assessments.

[8] The NAEP mathematics assessment at grade 4 and 8 was administered in 1990, 1992, 1996, 2000, 2003, 2005, 2007, 2009, and 2011.

[9] The NAEP science assessment at grade 4 and 8 was administered in 1996, 2000, 2005, and 2009, and at grade 8 in 2011.

[10] The NAEP 2011 frameworks are identical to the NAEP 2009 frameworks, which were used to create the NAEP 2009 mathematics and science assessments and were released in 2009. The NAEP 2009 frameworks were revisions of NAEP's existing mathematics and science frameworks.

and twelfth grades (NAGB 2010a).[11] The NAEP 2011 framework for science (*Science Framework for the 2011 National Assessment of Educational Progress*, NAGB 2010b) differs substantially from NAEP's pre-2009 science framework due to a major revision in 2009. The revision included changes in the content to be assessed across grades, changes to the detailed content statements at each grade level, and the redefinition and reorganization of the cognitive dimensions from the original three categories (i.e., conceptual understanding, scientific investigation, and practical reasoning) to four science practices (i.e., identifying science principles, using science principles, using scientific inquiry, and using technological design) (NAGB 2010b).[12]

.

Differences in Assessment Purpose

TIMSS is the only source for internationally comparable data on the mathematics and science achievement of students in fourth and eighth grade, and on other related aspects, such as curricula and classroom practices across participating countries and other education systems. The TIMSS frameworks are developed through collaboration with international mathematics and science experts, and national research coordinators from each participating education system. Therefore, TIMSS assesses students' knowledge, skills, and competencies viewed as essential for science and mathematics curricula among the participating education systems and reports results

[11] While there were no new objectives for grade 4 and 8, the objectives were reorganized under the new reasoning subtopic as appropriate.
[12] The extensive changes to the NAEP science framework were made by NAEP's steering and planning committees based on a variety of existing standards and assessment frameworks, including *National Science Education Standards* (National Research Council [NRC] 1996), *Benchmarks for Science Literacy* (American Association for the Advancement of Science [AAAS] 1993), TIMSS, the Program for International Student Assessment (PISA), and state standards. As a result of these changes, the "main" NAEP science trend line breaks in 2005 and a new trend line begins in 2009.

as average scores and with benchmarks of performance using four achievement levels: Low, Intermediate, High, and Advanced.

NAEP is a key source of data on mathematics and science achievement in the United States at specific stages of schooling (i.e., fourth, eighth, and twelfth grades), using benchmarks of performance with three achievement levels: *Basic*, *Proficient*, and *Advanced*.[13] NAGB establishes the frameworks and benchmarks, which are based on the collaborative input from a wide range of experts and participants from government, education, business, and public sectors in the United States. These frameworks are intended to reflect the best thinking from these experts about knowledge, skills, and competencies needed by U.S. students in these subjects at various grades.

Both TIMSS and NAEP aim to measure school-based curricular attainment. However, NAEP tailors its content to what U.S. experts and educators deem that students in the United States should know and be able to do. By comparison, TIMSS is constructed by international experts and educators to broadly assess mathematics and science as taught internationally in participating education systems, and thus covers curricula commonly taught around the world but not necessarily in the United States.

Differences in Assessed Population

TIMSS and NAEP are both sample-based assessments; each program administers the assessment to a nationally representative sample of U.S. students so that results can be

[13] The achievement levels were developed by a representative panel of teachers, education specialists, and members of the general public in the 1990s (for more information, see http://nces.ed.gov/nationsreportcard/set-achievement-lvls.asp). These achievement levels are still considered "trial" and, therefore, should be used and interpreted with caution.

generalized to the larger population. TIMSS and NAEP both use grade-based samples targeted at the same grade levels, so the sample populations by definition appear identical. However, in practice, there are differences in how the assessment is administered that introduce differences in the population assessed.

The TIMSS target population is students enrolled in the grades that represent 4 or 8 years of formal schooling, counting from the first year of the International Standard Classification of Education (ISCED)[14] Level 1 in the participating education systems, providing that the mean age at the time of testing is at least 9.5 years and 13.5 years, respectively for the two grades. In the United States, ISCED Level 1 begins at first grade, so TIMSS results reflect the performance of U.S. students in fourth and eighth grades. The NAEP target population is students in fourth, eighth, and twelfth grades, and thus reflects the performance of U.S. students in the same elementary and middle school grades as TIMSS. However, NAEP does not always assess students at all three grade levels or in all subjects with each administration.

In 2011, the NAEP mathematics assessment tested students in fourth and eighth grades, whereas the science assessment was administered to students in eighth grade only. In 2009, the NAEP science assessment was administered to students in fourth, eighth, and twelfth grades. Thus, to compare the TIMSS assessment at fourth-grade with the NAEP science assessment at fourth grade, this study compares the TIMSS 2011 science assessment's fourth-grade sample size, framework, and item contents with the NAEP 2009 science assessment's fourth-grade sample size, framework, and item contents.

[14] The ISCED was developed by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) to facilitate the comparability of educational levels across countries. ISCED Level 1 begins with the first year of required, formal, academic learning (UNESCO 1999). In the United States, Kindergarten is at ISCED Level 0 because it is not universally required.

Differences in how the assessments are administered introduce differences in the population assessed in at least two ways—through accommodations provided and in the timing of the administration.

TIMSS and NAEP differ in the accommodations they provide to students with special needs, for example students with disabilities and English language learners (i.e., students who are learning English and who are not yet proficient in the English language). TIMSS does not provide accommodations to such students, and students requiring accommodations are excluded from the assessment. However, schools, in some cases, may provide a minor accommodation for students taking TIMSS, such as supplying a Spanish-English Dictionary, a magnification device, or seating near a light. In contrast, NAEP offers accommodations to students with disabilities and English language learners (e.g., bilingual booklets, read-aloud questions, and small group sessions) so that they can demonstrate their content knowledge and skills on NAEP.[15] As a result, NAEP's exclusion rates for student participants are lower than those for TIMSS: the overall exclusion rates for the NAEP 2011 mathematics and science assessments were 2.5 percent and 1.6 percent, respectively, compared with 7.2 percent for both the TIMSS 2011 mathematics and science assessments.[16] This difference is known to have some influence on the characteristics of the populations being assessed in these two assessments; however, how much influence is unknown.

[15] NAEP began offering student accommodations in 1996. NAEP offers a comprehensive set of accommodations to increase access to testing for students with disabilities and English language learners. To ensure that appropriate accommodations are determined for individual students, NAEP relies on school staff to make inclusion and accommodations decisions for those students selected for the assessment. The accommodations allowed on NAEP and those allowed in states are often similar, but there may be some differences. For more information, see http://nces.ed.gov/nationsreportcard/about/inclusion.asp.

[16] NAEP exclusion rates for mathematics come from National Center for Education Statistics (2011). *The Nation's Report Card: Mathematics 2011* (NCES 2012-458). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C., Table A-6, and for science from http://nationsreportcard.gov/science_2011/exclusion.asp. The TIMSS exclusion rates come from table A-1 in (NCES 2013-009).

TIMSS and NAEP also differ in the timing of their assessment windows. TIMSS administers the science and mathematics assessment in the Northern Hemisphere in April and May. NAEP administers its assessments between January and the beginning of March. This difference means that the assessed populations from the same grade will not be exactly comparable because the TIMSS administration measures students with (at least one and as much as four) more months of schooling than the NAEP administration.

Differences in Sample Size

TIMSS and NAEP both provide accurate and reliable measures of students' achievement and provide trend information on students' performance over time; however, the precision of these measures differs because of their respective sample sizes (table 1).[17] The NAEP samples include many more U.S. students than does TIMSS, as is apparent from table 1, and, thus, is able to (a) measure U.S. students' performance at a finer level of precision and (b) detect smaller variations in U.S. students' performance.

Table 1. Number of students and schools participating in TIMSS 2011 and NAEP 2009/2011, by grade

Subject (grade)	Number of students	Number of schools
TIMSS[1]		
2011 Mathematics and Science (fourth grade)	12,600	370
2011 Mathematics and Science (eighth grade)	10,400	500
NAEP[2]		
2011 Mathematics (fourth grade)	209,000	8,500
2011 Mathematics (eighth grade)	175,200	7,610
2009 Science (fourth grade)[3]	156,500	9,330
2011 Science (eighth grade)	122,000	7,290

[17] The precision of the measures can be described by the size of the standard errors associated with scale scores. The standard errors for the U.S. national sample in TIMSS 2011 for mathematics and science at grades 4 and 8 ranged from 1.8 to 2.6. In NAEP 2011, the comparable range in standard errors was 0.2 to 0.3.

[1] The sample sizes in TIMSS 2011 reported in this table exclude students and schools in the TIMSS 2011 benchmarking state samples.
[2] The NAEP sample sizes reported in this table are for main NAEP 2011 and main NAEP 2009, respectively, and exclude students and schools only in NAEP TUDA samples.
[3] The NAEP 2009 science assessment at fourth grade was not administered in Alaska, Kansas, Nebraska, Vermont, and the District of Columbia.
NOTE: The TIMSS assessment includes both the mathematics assessment and the science assessment. Numbers have been rounded to the nearest hundred for students and the nearest ten for schools.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Trends in International Mathematics and Science Study (TIMSS) 2011; and National Assessment of Educational Progress (NAEP) 2011 mathematics and science assessments, NAEP 2009 science assessment.

2. Comparison between TIMSS 2011 and NAEP 2011 Mathematics Assessments

The TIMSS and NAEP mathematics assessments assess similar content areas, and almost all TIMSS items (100 percent of grade 4 items and 99 percent of grade 8 items) fit the NAEP mathematics framework. However, TIMSS items cover only a subset of the NAEP target grade's framework; parts of the NAEP mathematics framework are not covered by any TIMSS items at the target grade level. On the other hand, some TIMSS items do cover NAEP content specified by the NAEP framework at higher (and sometimes at lower) grade levels.

Mathematics Framework Comparisons

The content areas for the TIMSS 2011 mathematics framework and NAEP 2011 mathematics framework are organized somewhat differently. In the TIMSS 2011 mathematics framework, the following content areas (called "cognitive domains" in TIMSS) are assessed in the following proportions:

Table 2. TIMSS content domains and the percentage of the fourth- and eighth-grade TIMSS 2011 mathematics assessment devoted to each domain

Grade 4		Grade 8	
Content domain	Percent of assessment	Content domain	Percent of assessment
Number	50	Number	30
Geometric Shapes and Measures	35	Algebra	30
Data Display	15	Geometry	20
		Data and Chance	20

NOTE: The percentages in this table are based on the number of target score points (also referred to as "testing time") and not the number of items in each content domain.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS) 2011.

Within these content areas, TIMSS further delineates topics at each grade level.

In the NAEP 2011 mathematics framework, the content areas assessed are:

Table 3. NAEP content areas and the percentage of fourth- and eighth-grade NAEP 2011 mathematics assessment devoted to each area

	Percent of items	
Content area	Grade 4	Grade 8
Number Properties and Operations	40	20
Measurement	20	15
Geometry	15	20
Data Analysis Statistics and Probability	10	15
Algebra	15	30

NOTE: NAEP content area targets are set as a distribution of items, not score points or test time. Thus, the target percentages in this table are based on the number of items and not the target score points in each content area.
SOURCE: National Assessment of Educational Progress (NAEP) 2011 mathematics assessments.

Within these content areas, NAEP also further delineates detailed subtopics and objectives by grade level.

Both frameworks also specify a cognitive dimension by which students are to be assessed (in addition to the content dimension). The cognitive dimension and the intended or "target" distribution of items classified into it are defined differently in the two assessments' frameworks.

For TIMSS, the cognitive dimension has three cognitive domains: *knowing*, *applying*, and *reasoning*. The *knowing* domain covers facts and concepts, the *applying* domain focuses on applying knowledge and conceptual understanding to solve problems, and the *reasoning* domain encompasses solving unfamiliar, complex or multistep problems (Mullis et al. 2009). Each grade in TIMSS has a different target distribution of item across these three cognitive domains: the fourth-grade assessment has most items in *knowing* and *applying* (40 percent and 40 percent, respectively), while the plurality of the items in the eighth-grade assessment are in the domain of *applying* (40 percent, as compared with 35 percent in *knowing* and 25 percent in *reasoning*).

In NAEP mathematics, the cognitive dimension has three levels of cognitive complexity: *low*, *moderate*, and *high*. These three levels of cognitive complexity form an ordered description of the demands an item or a task make on a student. For example, a low-level item might ask students to recall a property, a moderate-level item might ask students to make a connection between two properties, and a high-level item might ask a student to analyze the assumption made in a mathematical model (NAGB 2010a). Both grades in NAEP have the same target item distribution across these three levels of cognitive complexity based on the amount of time the items are expected to take: 25 percent of testing time is expected to be spent on low-complexity items, 50 percent of testing time on moderate-complexity items, and 25 percent of testing time on high-complexity items.

For a more detailed comparison of the grade 8 TIMSS and NAEP frameworks for mathematics, see *A Comparison of the 2011 Grade 8 NAEP and TIMSS Mathematics and Science Frameworks* (NCES 2013-462); available at http://nces.ed.gov/nationsreportcard/pubs/studies/2013462.asp.

Mathematics Item Comparisons

The contents of all 179 fourth-grade and 217 eighth-grade TIMSS 2011 mathematics items were reviewed by a panel of mathematics curriculum experts (see Appendix A for a list of the Expert Panel members). To assess the level of correspondence between the TIMSS 2011 and NAEP 2011 assessments, the Mathematics Expert Panel classified the TIMSS 2011 mathematics assessment items to the *Mathematics Framework for the 2011 National Assessment of Educational Progress* (NAGB 2010a). NAEP's test specification documents were also made available to the Expert Panel and used when the Expert Panel determined that more clarification was needed to complete the classification process. If disagreements arose during the classification process, the Expert Panelists discussed their differing viewpoints to reach consensus on a classification. The Expert Panel's final classifications were recorded and analyzed to determine how well the TIMSS items mapped to the content areas defined in the NAEP framework.

The rest of this section describes the content match, as well as other differences and similarities that the Expert Panel found in item content, grade-level fit, and item format, between the TIMSS 2011 and NAEP 2011 mathematics assessments.

Item content

As explained at the start, the Mathematics Expert Panel completed a "content match" analysis by classifying all the TIMSS 2011 fourth-grade and eighth-grade mathematics items into the following three categories specified in the NAEP 2011 framework for mathematics: (a)

content area, (b) content area's subtopic, and (c) subtopic's objective by grade level. This was an item-by-item review to see how many TIMSS items fit in the NAEP framework and how much of the NAEP framework was covered by TIMSS items.

The Expert Panel's content match analysis found that all of the TIMSS 2011 mathematics items could be classified to the NAEP 2011 mathematics framework's content categories down to the subtopic level. At the finest level of detail (i.e., grade level objectives), 1 percent of the fourth-grade items and 3 percent of the eighth-grade items could not be matched to a specific objective within the NAEP 2011 mathematics framework. These results are consistent with the findings from the previous study[18] and indicate that nearly all (but not all) TIMSS items map to the NAEP framework's objectives, which suggests a strong content match.

To assess how much of the NAEP framework's subtopics and objectives are addressed by TIMSS items, the Mathematics Expert Panel's classification of TIMSS items to the NAEP framework was examined. Table 4 summarizes how many of NAEP framework's subtopics were addressed by TIMSS fourth-grade items.[19] In total, 18 of the NAEP mathematics framework's 21 fourth-grade *subtopics* were addressed by TIMSS fourth-grade items, and 7 of its 14 eighth-grade subtopics were addressed by TIMSS fourth-grade items as well. Table 5 summarizes how many of NAEP's fourth- and eighth-grade framework's *objectives* were addressed by one or more TIMSS fourth-grade items. In total, 41 out of the NAEP mathematics framework's 65 objectives for fourth grade were addressed by one or more TIMSS fourth-grade items. In

[18] The previous comparison study (NCES 2007) found that 2 percent of TIMSS 2007 fourth-grade mathematics items and 3 percent of TIMSS 2007 eighth-grade mathematics items could not be classified to the finest level of detail within the NAEP 2005/2007 mathematics frameworks.

[19] Because some of the TIMSS fourth-grade items addressed fourth-grade objectives under these subtopics and some addressed eighth-grade objectives under them, separate ratios and percentages are presented for the framework's two grade levels. Coverage of subtopics and objectives at grades different than the target grade are provided for a complete picture of TIMSS' items' coverage of the NAEP framework.

addition, 14 of the NAEP mathematics framework's 100 objectives for *eighth* grade were addressed by TIMSS fourth-grade items. Most of these eighth-grade objectives elaborated or expanded on fourth-grade objectives addressed by the TIMSS items; however, one eighth-grade objective (under *Ratios and proportional reasoning*) represented an area that TIMSS fourth-grade items addressed only as an eighth-grade objective and not as a fourth-grade objective.

According to the Expert Panel's classifications of TIMSS fourth-grade items, the fourth-grade NAEP content area with the most coverage in TIMSS at fourth grade was *Measurement*: all its subtopics were addressed by TIMSS items, and 7 out of 10 of its objectives for fourth grade plus 3 out of 12 of its objectives for eighth grade were addressed by TIMSS fourth-grade items. The NAEP content area with the least coverage in TIMSS was *Data Analysis, Statistics, and Probability*: only 2 of its 3 fourth-grade subtopics were addressed by TIMSS items; with only 4 out of 9 of its objectives for fourth grade and 1 out of 22 of its objectives for eighth grade addressed by TIMSS fourth-grade items.

Table 4. The ratio and percentage of NAEP 2011 mathematics content area's subtopics for grades 4 and 8 addressed by TIMSS 2011 grade 4 mathematics items

	TIMSS grade 4 item coverage of NAEP grade 4 subtopics		TIMSS grade 4 item coverage of NAEP grade 8 subtopics	
NAEP Framework content areas	Ratio	Percentage	Ratio	Percentage
A: Number Properties and Operations	5/6	83	3/6	50
B: Measurement	2/2	100	1/3	33
C: Geometry	5/5	100	2/5	40
D: Data Analysis, Statistics, and Probability	2/3	67	1/4	25
E: Algebra	4/5	80	0/5	0

NOTE: Shading indicates areas of coverage. The ratios in this table have as the numerator the number of NAEP subtopics within the named content area that are addressed by one or more TIMSS fourth-grade items, with the denominator being the total number of subtopics within the named content area. The ratios and percentages do not indicate the depth of TIMSS item's coverage of any subtopic as a subtopic may be addressed by a single item or by multiple items. See Appendix B for the data underlying this table.
SOURCE: TIMSS-NAEP Comparison Mathematics Expert Panel, 2011.

Table 5. The ratio and percentage of NAEP 2011 mathematics content area's objectives for grades 4 and 8 addressed by TIMSS 2011 grade 4 mathematics items

	TIMSS grade 4 item coverage of NAEP grade 4 objectives, by subtopic		TIMSS grade 4 item coverage of NAEP grade 8 objectives, by subtopic	
NAEP Framework content area and subtopic	Ratio	Percentage	Ratio	Percentage
A: Number Properties and Operations				
1. Number sense	6/6	100	1/8	13
2. Estimation	1/3	33	0/4	0
3. Number operations	4/6	67	1/4	25
4. Ratios and proportional reasoning	0/1	0	1/4	25
5. Properties of number and operations	1/3	33	0/5	0
6. Mathematical reasoning using numbers	1/1	100	0/2	0
B: Measurement				
1. Measuring physical attributes	5/6	83	3/6	50
2. System of measurement	2/4	50	0/5	0
3. Measurement in triangles	NA	NA	0/1	0
C: Geometry				
1. Dimension and shape	3/4	75	4/6	67
2. Transformations of shapes and preservation of properties	2/4	50	3/5	60
3. Relationships between geometric figures	3/4	75	0/5	0
4. Position, direction, and coordinate geometry	1/2	50	0/4	0
5. Mathematical reasoning in geometry	1/1	100	0/1	0
D: Data Analysis, Statistics, and Probability				
1. Data representation	3/3	100	1/5	20
2. Characteristics of data sets	1/2	50	0/5	0
3. Experiments and samples	NA	NA	0/3	0
4. Probability	0/4	0	0/9	0
5. Mathematical reasoning with data	NA	NA	NA	NA
E: Algebra				
1. Patterns, relations, and functions	3/5	60	0/5	0
2. Algebraic representation	2/2	100	0/5	0
3. Variables, expressions, and operations	1/2	50	0/2	0
4. Equations and inequalities	1/1	100	0/5	0
5. Mathematical reasoning in algebra	0/1	0	0/1	0

NOTE: Shading indicates areas of coverage. The ratios in this table have as the numerator the number of NAEP objectives within the named subtopic that are addressed by one or more TIMSS fourth-grade items, with the denominator being the total number of objectives within the named subtopic. The ratios and percentages do not indicate the depth of TIMSS item's coverage of any objective as an objective may be addressed by a single item or by multiple items. It is also important to bear in mind that the pool of NAEP fourth-grade items does not cover all the NAEP framework's objectives in the same depth (i.e., with the same number of items per objective). Some objectives may be addressed by multiple items, some by only one item, and, at times, it is possible that an assessment may have no NAEP items addressing a particular objective. See Appendix B for the data underlying this table.
SOURCE: TIMSS-NAEP Comparison Mathematics Expert Panel, 2011.

Turning to the analysis of the TIMSS eighth-grade items, table 6 summarizes how many of the NAEP framework's subtopics were addressed by TIMSS eighth-grade items. Because some of the TIMSS eighth-grade items addressed eighth-grade objectives under these subtopics and some addressed fourth- or twelfth-grade objectives under them, separate ratios and percentages are presented for the framework's different grade levels. In total, 20 of the NAEP mathematics framework's 23 eighth-grade *subtopics* were addressed by TIMSS eighth-grade items, and 6 of its 23 twelfth-grade subtopics along with 3 of its 21 fourth-grade subtopics were addressed as well. Table 7 summarizes how many of NAEP's fourth-, eighth-, and twelfth-grade framework's o*bjectives* that were addressed by one or more TIMSS eighth-grade items. In total, 57 out of the NAEP mathematics framework's 100 objectives for eighth grade were addressed by one or more TIMSS eighth-grade items. In addition, 8 of the NAEP mathematics framework's 117 objectives for *twelfth* grade and 4 of the NAEP mathematics framework's 65 objectives for *fourth* grade were addressed by TIMSS eighth-grade items.

Table 6. The ratio and percentage of NAEP 2011 mathematics content area's subtopics for grades 4, 8, and 12 addressed by TIMSS 2011 grade 8 mathematics items

	TIMSS grade 8 item coverage of NAEP grade 4 subtopics		TIMSS grade 8 item coverage of NAEP grade 8 subtopics		TIMSS grade 8 item coverage of NAEP grade 12 subtopics	
NAEP Framework content areas	Ratio	Percentag	Ratio	Percentag	Ratio	Percentage
A: Number Properties and Operations	0/6	0	6/6	100	1/6	17
B: Measurement	1/2	50	2/3	67	1/3	33
C: Geometry	1/5	20	4/5	80	1/5	20
D: Data Analysis, Statistics, and Probability	1/3	33	3/4	75	0/4	0
E: Algebra	0/5	0	5/5	100	3/5	60

NOTE: Shading indicates areas of coverage. The ratios in this table have as the numerator the number of NAEP subtopics within the named content area that are addressed by one or more TIMSS eighth-grade items, with the denominator being the total number of subtopics within the named content area. The ratios and percentages do not indicate the depth of TIMSS item's coverage of any subtopic as a subtopic may be addressed by a single item or by multiple items. See Appendix B for the data underlying this table.
SOURCE: TIMSS-NAEP Comparison Mathematics Expert Panel, 2011

Table 7. The ratio and percentage of NAEP 2011 mathematics content area's objectives for grades 4, 8, and 12 addressed by TIMSS 2011 grade 8 mathematics items

NAEP Framework content area and subtopic	TIMSS grade 8 item coverage of NAEP grade 4 objectives, by subtopic		TIMSS grade 8 item coverage of NAEP grade 8 objectives, by subtopic		TIMSS grade 8 item coverage of NAEP grade 12 objectives, by subtopic	
	Ratio	Percentage	Ratio	Percentage	Ratio	Percentage
A: Number Properties and Operations						
1. Number sense	0/6	0	6/8	75	0/4	0
2. Estimation	0/3	0	1/4	25	0/3	0
3. Number operations	0/6	0	3/4	75	0/5	0
4. Ratios and proportional reasoning	0/1	0	3/4	75	0/2	0
5. Properties of number and operations	0/3	0	4/5	80	1/4	25
6. Mathematical reasoning using numbers	0/1	0	1/2	50	0/2	0
B: Measurement						
1. Measuring physical Attributes	1/6	17	4/6	67	1/6	17
2. System of measurement	0/4	0	0/5	0	0/5	0
3. Measurement in triangles	NA	NA	1/1	100	0/7	0
C: Geometry						
1. Dimension and shape	0/4	0	2/6	33	0/4	0
2. Transformations of shapes and preservation of properties	2/4	50	5/5	100	0/6	0
3. Relationships between geometric figures	0/4	0	3/5	60	1/7	14
4. Position, direction, and coordinate geometry	0/2	0	1/4	25	0/8	0
5. Mathematical reasoning in geometry	0/1	0	0/1	0	0/5	0
D: Data Analysis, Statistics, and Probability						
1. Data representation	1/3	33	3/5	60	0/6	0
2. Characteristics of data sets	0/2	0	3/5	60	0/7	0
3. Experiments and samples	NA	NA	0/3	0	0/5	0
4. Probability	0/4	0	3/9	33	0/9	0
E: Algebra						
1. Patterns, relations, and functions	0/5	0	2/5	40	0/7	0
2. Algebraic representation	0/2	0	4/5	80	1/7	14
3. Variables, expressions, and operations	0/2	0	2/2	100	2/7	29
4. Equations and inequalities	0/1	0	5/5	100	2/6	33
5. Mathematical reasoning in algebra	0/1	0	1/1	100	0/3	0

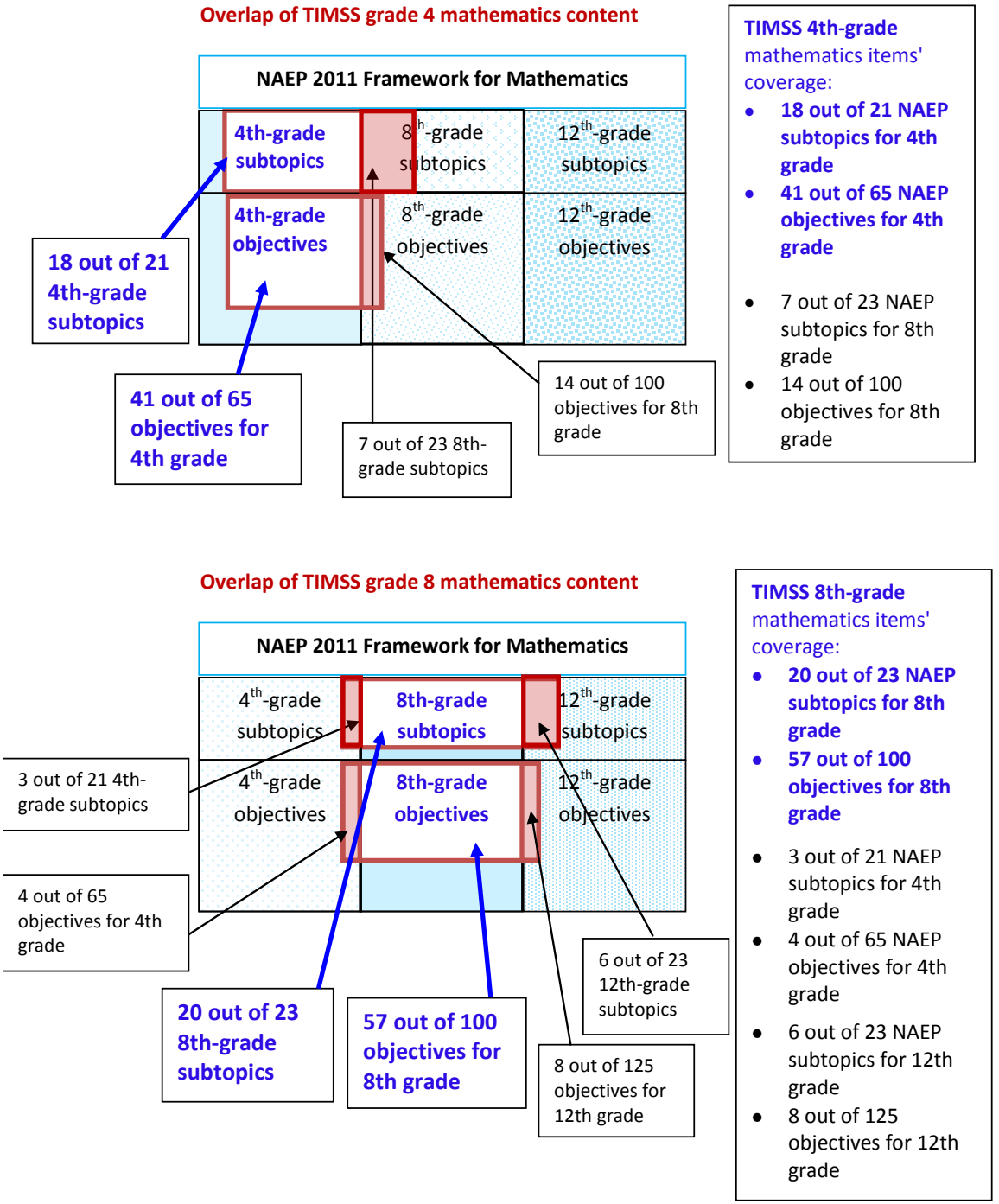
NOTE: Shading indicates areas of coverage. The percentages in this table indicate the number of NAEP's fourth-, eighth-, and twelfth-grade framework's objectives that were addressed by one or more TIMSS eighth-grade items; they do not indicate the depth of TIMSS item's coverage of any objectives. Moreover, it is important to bear in mind that the pool of NAEP eighth-grade items does not cover all the NAEP framework's objectives in the same depth (i.e., with the same number of items per objective). Some objectives may be addressed by multiple items, some by only one item, and, at times, it is possible that an assessment may have no items addressing a particular objective. At twelfth grade, Geometry and Measurement are combined into a single content area in the NAEP framework. See Appendix B for the data underlying this table.
SOURCE: TIMSS-NAEP Comparison Mathematics Expert Panel, 2011.

According to the Expert Panel's classifications of TIMSS eighth-grade items, the NAEP content area with the most coverage in TIMSS was *Algebra*: all its subtopics were addressed by TIMSS items, and 14 out of 18 of its objectives for eighth grade plus 5 out of 30 of its objectives

for twelfth grade were addressed by TIMSS eighth-grade items. The eighth-grade NAEP content area with the least coverage in TIMSS was *Measurement*: only 2 of its 3 eighth-grade subtopics were addressed by TIMSS items, with only 5 out of 12 of its objectives for eighth grade and 1 out of 18 of its objectives for twelfth grade addressed by TIMSS eighth-grade items.

This analysis of TIMSS items indicates that 18 out of 21 of NAEP framework's fourth-grade subtopics and 20 out of 23 of its eighth-grade subtopics are addressed by TIMSS fourth- and eighth-grade items, respectively. About two-thirds of its fourth-grade objectives (41/65) are addressed by TIMSS fourth-grade items, with some addressed at the eighth-grade level. About three-fifths (57/100) of its eighth-grade objectives are addressed by TIMSS eighth-grade items, with some addressed at the twelfth-grade level. Taken together with the finding that nearly all TIMSS items map to the NAEP framework at the most specific level of detail, one can conclude that TIMSS 2011 mathematics items neatly fit into the NAEP framework and cover most of the framework's objectives, but not all. This is to say, the NAEP framework defines a broader array of assessment content than TIMSS covers, but TIMSS items cover a significant proportion or subset of the whole NAEP framework (see exhibit 1). This level of congruence at both fourth- and eighth-grade can be described as a strong content match.

Exhibit 1: Overlap of TIMSS 2011 mathematics content with NAEP 2011 Framework for Mathematics



Grade-level fit

In addition to classifying all the TIMSS 2011 mathematics items to the NAEP 2011 mathematics framework's three content categories (i.e., content area, subtopic, and objective), the Expert Panel also completed a separate grade-level fit analysis. The Expert Panel used the NAEP 2011 mathematics framework's objectives' grade level as their reference for completing this classification. However, there were occasions when the Expert Panel could not match a particular TIMSS item to a *specific* NAEP framework objective; however, they thought that that item was implied in the NAEP 2011 mathematics framework at a particular grade based on their knowledge of the NAEP mathematics framework and their content expertise in mathematics.[20] Items that were neither specified nor implied by the NAEP 2011 mathematics framework at any particular grade were classified as "no fit."

For TIMSS 2011 fourth-grade mathematics items, the Expert Panel determined that 89 percent of the items aligned with the fourth grade NAEP 2011 mathematics framework, while 11 percent aligned better with the eighth grade NAEP 2011 mathematics framework (table 8). Most of the TIMSS 2011 fourth-grade items that fit better with the eighth grade NAEP 2011 mathematics framework were geometry items. The Expert Panel also noted that the TIMSS 2011 fourth-grade mathematics assessment included some items that are not commonly familiar to U.S. fourth-graders, such as drawing symmetrical figures, visualizing paper folding, and drawing a specific type of angle (e.g., a right angle) or recognizing an angle (e.g., identifying a 45- or 90-degree angle).

[20] For example, the NAEP fourth-grade framework talks about constructing graphs but does not specify constructing data tables. Thus, item M041182 in TIMSS, which asks fourth-grade students to construct a data table, members of the Expert Panel considered to be "implied" even though it is not spelled out in the NAEP framework.

For TIMSS 2011 eighth-grade mathematics items, the Expert Panel determined that 85 percent of the items aligned with the eighth-grade NAEP 2011 mathematics framework, while 3 percent of the items aligned better with the fourth-grade NAEP 2011 mathematics framework and 11 percent of the items aligned better with the twelfth-grade NAEP 2011 mathematics framework. About half of the TIMSS 2011 mathematics items that aligned better with the twelfth-grade NAEP 2011 mathematics framework were geometry items, and about half were algebra problems that required multiple steps to solve an equation. The Expert Panel identified 1 percent of the TIMSS 2011 eighth-grade mathematics items to be "no fit" (table 8). That is to say, they determined that these items assessed specific content or targeted skills (e.g., plotting angles on a polar graph, and measuring an angle defined with polar coordinates) that were not specified, either explicitly or implicitly, in the NAEP framework at any of the three grade levels.

Table 8. Percentage distribution of TIMSS 2011 mathematics items across the NAEP 2011 grade 4, grade 8, and grade 12 mathematics frameworks

Grade level "fit" of TIMSS items mapped to the NAEP 2011 mathematics framework's objectives	TIMSS 2011 mathematics items' grade level	
	Fourth grade (n=179)	Eighth grade (n=217)
Fourth grade	89	3
Eighth grade	11	85
Twelfth grade	0	11
No fit	0	1

NOTE: The letter "n" is an abbreviation for the total number of assessment items. "No fit" identifies the percentage of items the Expert Panelists were unable to classify as consistent with a specific objective in the NAEP framework or consistent with the content or targeted skills within a grade level as described in the NAEP 2011 framework.
SOURCE: TIMSS-NAEP Comparison Mathematics Expert Panel, 2011.

Item format

Both TIMSS and NAEP assess students using multiple-choice and constructed-response item formats. Multiple-choice items provide, typically, four or five possible options where only one is correct. Constructed-response items are open-ended items that require students to construct written responses.

TIMSS 2011 has a relatively equal proportion of items in the two formats at both the fourth- and eighth-grade levels, whereas NAEP 2011 has a greater number of multiple-choice than constructed-response items at both grade levels (table 9). At fourth grade, 70 percent of the NAEP mathematics items are multiple-choice and 30 percent are constructed-response. At eighth grade, 74 percent of the NAEP mathematics items are multiple-choice and 26 percent are constructed-response. Also TIMSS relies on multiple-choice items with four response options, while NAEP relies on multiple-choice items with four response options at fourth grade and five response options and eighth grade.

Table 9. Percentage distribution of TIMSS 2011 and NAEP 2011 mathematics items, by item format and grade level

Item format	Fourth grade		Eighth grade	
	TIMSS 2011 (n=179)	NAEP 2011 (n=158)	TIMSS 2011 (n=217)	NAEP 2011 (n=155)
Multiple choice	53	70	54	74
Constructed response	47	30	46	26

NOTE: The letter "n" is an abbreviation for the total number of assessment items in a category.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS) 2011; and National Assessment of Educational Progress (NAEP) 2011 mathematics.

3. Comparison between TIMSS 2011 and NAEP 2011 Science Assessments[21]

Both the TIMSS and NAEP science assessments assess similar content areas; however, not all TIMSS items fit the NAEP science framework. As with mathematics, TIMSS items cover a subset of the NAEP target grade's framework; parts of the NAEP science framework are not covered by any TIMSS items at the target grade level. However, more than 20 percent of TIMSS science items fall outside the content specified by the NAEP framework and some items that do cover specific NAEP content, do so at different grade levels in NAEP than in TIMSS.

Science Framework Comparisons

The TIMSS 2011 and the NAEP 2011 science frameworks define similar content areas in science. In the TIMSS 2011 science framework, the following content areas (called "cognitive domains" in TIMSS) are assessed in the following proportions.

Table 10. TIMSS content domains and the percentage of the fourth- and eighth-grade TIMSS 2011 science assessment devoted to each domain

Grade 4		Grade 8	
Content domain	Percent of assessment	Content domain	Percent of assessment
Physical science	35	Chemistry	20
Life science	45	Physics	25
Earth science	20	Biology	35
		Earth science	20

NOTE: The percentages in this table are based on the number of target score points (also referred to as "testing time") and not the number of items in each content domain.

[21] As noted previously, in 2011 the NAEP science assessment was administered to students in eighth grade only. In 2009, the NAEP science assessment was administered to students in fourth, eighth, and twelfth grades. Thus, to provide a comparison between the TIMSS and the NAEP science assessments at fourth grade for this comparative study, the NAEP 2009 science assessment at fourth grade was included. The NAEP 2011 science assessment is based on the same NAEP framework used for the NAEP 2009 science assessment. Since the NAEP science assessment was not administered to fourth-grade students in 2011, this substitution permits a comparison between the TIMSS and the NAEP science assessments at fourth grade. This comparison applies to the populations assessed, the frameworks, and the items only; it does not include a comparison of assessment results.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS) 2011.TIMSS further delineates topics at each grade level.

In the NAEP 2011 science framework, the content areas assessed are.

Table 11. NAEP content areas and the percentage of fourth- and eighth-grade NAEP science assessment devoted to each area

	Percent of items	
Content area	Grade 4	Grade 8
Physical science	33.3	30
Life science	33.3	30
Earth and space sciences	33.3	40

NOTE: The percentages in this table are based on the percentage of student response time and not the number of items in each content area.
SOURCE: National Assessment of Educational Progress (NAEP) 2009 and 2011 science assessments.

Within these content areas, NAEP also further delineates detailed topics, subtopics, and content statements by grade level.

Both frameworks specify cognitive dimensions by which students are to be assessed (in addition to the content dimension). The dimension(s) and the intended or "target" distribution of items classified into this/these dimensions, however, are defined differently in the two assessments' frameworks. For TIMSS, the cognitive dimension assessed within science has the same three domains as within mathematics: *knowing*, *applying*, and *reasoning*. Each TIMSS item is developed to measure one of these cognitive categories (in addition to the content knowledge). The *knowing* domain covers science facts, procedures and concepts; the *applying* domain focuses on applying knowledge and conceptual understanding to solve a science problem; and the *reasoning* domain encompasses solving unfamiliar, complex or multi-step science problems (Mullis et al. 2009). Each grade in TIMSS has a different target distribution of items across these three cognitive domains: the fourth-grade assessment has most items in

knowing and *applying* (40 percent and 40 percent, respectively), while in the eighth-grade assessment items are more evenly distributed: 35 percent in the domains of *applying* and *knowing* and 30 percent in *reasoning*.

NAEP, on the other hand, defines four *science practices* in its framework: *identifying science principles*, *using science principles*, *using scientific inquiry*, and *using technological design*. The science practices are associated with four sets of cognitive demands that the items place on students: (1) "knowing that" (declarative knowledge), (2) "knowing how" (procedural knowledge) (3) "knowing why" (schematic knowledge), and (4) "knowing when and where to apply knowledge"(strategic knowledge) (NAGB 2010b). In 2011, the proportion of assessment time devoted to each of these science practices was 25 percent for identifying science principles, 35 percent using science principles, 30 percent using scientific inquiry, and 10 percent using technological design. NAEP science items are developed to measure specific content knowledge and to require students to apply science practice skills (i.e., one of the four science practices) when answering a question/responding to an item. Thus, students must possess knowledge, skill, and competence in both a specific *content area* and *science practice* to answer each item.

For a more detailed comparison of the TIMSS and NAEP frameworks for science, see *A Comparison of the 2011 Grade 8 NAEP and TIMSS Mathematics and Science Frameworks* (NCES 2013-462); available at http://nces.ed.gov/nationsreportcard/pubs/studies/2013462.asp.

Science Item Comparisons

The content of all 196 fourth-grade and 239 eighth-grade TIMSS 2011 science items were reviewed by a panel of science curriculum experts (see Appendix A for a list of the Expert Panel members). To assess the level of correspondence between the TIMSS 2011 and NAEP 2009/2011 science assessments, the Science Expert Panel classified the TIMSS 2011 science assessment items to the *Science Framework for the 2011 National Assessment of Educational Progress* (NAGB 2010b). NAEP's test specification documents were available to the Expert Panel and used only when the Expert Panel determined that more clarification was needed to complete the classification process. If disagreements arose during the classification process, the Expert Panelists discussed the differing viewpoints to reach consensus on a classification. The Expert Panel's final classifications were recorded and analyzed to determine how well the TIMSS items mapped to the content areas defined in the NAEP framework.

The rest of this section describes the content match, as well as other differences and similarities that the Expert Panel found in item content, grade-level fit, and item format, between the TIMSS 2011 and NAEP 2009/2011 science assessments.

Item content

As explained earlier, the Science Expert Panel completed a "content match" analysis by classifying all the TIMSS 2011 fourth-grade and eighth-grade science items into the following four categories specified in the NAEP 2009 or 2011 framework[22] for science: (a) content area, (b) content area's topic, (c) topic's subtopic, and (d) subtopic's content statement by grade level.

[22] As explained in the section Population Assessed, the NAEP 2009 framework is used for fourth-grade science and NAEP 2011 for eighth-grade science in this study.

Content areas, topics, and subtopics are the same across grade levels while content statements are specific to the grade level. This was an item-by-item review to see how many TIMSS items fit in the NAEP framework and how much of the NAEP framework was covered by TIMSS items. In addition, a "grade-level fit" analysis was conducted.

For fourth grade, the Expert Panel's content match analysis found that all of the TIMSS 2011 science items could be classified to content areas in the NAEP 2009 science framework, but that not all the fourth-grade items could be classified to a lower level of the NAEP framework (i.e., to topics, subtopics, or content statements). Table 12 summarizes the Expert Panel's classification by level and indicates that 10 percent of the fourth-grade TIMSS 2011 science items did not address any topic in the NAEP science framework; 18 percent did not address any subtopic; and 31 percent did not address any content statement at any of the three grade levels.

For eighth grade, the Expert Panel's content match analysis of the TIMSS science items found that 98 percent of TIMSS items could be classified to a content area within the NAEP 2011 framework and that 2 percent could not be classified to the NAEP framework at the most general level.[23] Again, table 12 summarizes the Expert Panel's classification by level and indicates that, at the lower levels of classification, 11 percent of the eighth-grade TIMSS 2011 science items did not address any topic in the NAEP science framework; 13 percent did not address any subtopic; and 23 percent did not address any content statement at any of the three grade levels.

[23] TIMSS 2011 science items that could not be classified at the most general level were items that asked about conducting laboratory experiments. In the opinion of the Expert Panel, these items were about experimental practices in the abstract and not about any particular scientific content in the NAEP framework.

Table 12. Percentage distribution of TIMSS 2011 science items across the four levels of the NAEP 2009/2011 science framework based on the analysis of the Expert Panel

NAEP 2009/2011 science framework	TIMSS 2011 fourth grade (n=196)	TIMSS 2011 eighth grade (n=239)
Content Area	100	98
Topic	90	89
Subtopic	82	87
Content Statement	69	77

NOTE: The letter "n" is an abbreviation for "the total number of assessment items in a category."
SOURCE: TIMSS-NAEP Comparison Science Expert Panel, 2011.

In the previous comparison study (NCES 2007), it was found that 12 percent of TIMSS 2007 fourth-grade science items and 20 percent of TIMSS 2007 eighth-grade science items could not be classified to the finest level of detail within the NAEP 2005 science frameworks. Given that the current study finds that 31 percent of TIMSS 2011 fourth-grade and 23 percent of TIMSS 2011 eighth-grade science items could not be classified to the finest level of detail in the NAEP 2009 or 2011 science framework, it would seem that the revisions to the 2009 NAEP science framework have made NAEP less like TIMSS at fourth grade than it was before the revisions. Furthermore, it would seem that TIMSS science items cover content that is outside the NAEP framework's specified contents.

In their deliberations, the Science Expert Panel noted several factors that account in part for the number of TIMSS items that do not map to the NAEP 2009 or 2011 framework. First, at both grade levels, a number of topics assessed in TIMSS were not explicitly specified in the NAEP framework in the judgment of the Expert Panel, including dissolving, mixture and its separations, parallel circuits, specific properties of light and sound, body systems, and health-related questions. Second, the Expert Panel noted that the TIMSS assessment has several items that assessed skills for conducting scientific inquiries (such as controlling variables or interpreting

findings in a graph), but that did not require science content knowledge per se. To answer these "non-content-specific" items, students need only to apply specific scientific practices. In NAEP assessments, by comparison, all items combine one of the four *science practices* (explained earlier) with some specific science content.[24] Therefore, the kinds of "non-content-specific" items or items assessing generic scientific inquiry skills found in the TIMSS 2011 science assessment are not consistent with the NAEP framework. Third, the Expert Panel noted that the TIMSS assessment, especially at fourth grade, has some items that assess experiential knowledge rather than school-learned scientific knowledge. This is to say, the Expert Panel believed that students could answer these items based on physical experiences in daily life (e.g., an item about shadows or submerging an object in liquid) rather than on the basis of specific content knowledge or skill in the framework. Consequently, the Expert Panel found it difficult to find a specific place in the NAEP framework to classify these items.

To assess the second criterion for the content match—how much of the NAEP framework's subtopics and content statements are addressed by TIMSS items—the Science Expert Panel's classification of TIMSS items to the NAEP framework was examined. Table 13 summarizes how many of the NAEP framework's subtopics were addressed by TIMSS fourth-grade items.[25] In total, 16 of the NAEP science framework's 17 fourth-grade *subtopics* were addressed by TIMSS fourth-grade items, and 12 of its 18 eighth-grade subtopics were addressed as well. Table 14

Table 13. The ratio and percentage of NAEP 2011 science content area's subtopics for grades 4 and 8 addressed by TIMSS 2011 grade 4 science items

[24] NAEP assessment science items are created to address a specific content statement and require students to apply specific *science practice* skills to answer each item correctly.

[25] Because some of the TIMSS items addressed fourth-grade content statements under these subtopics and some addressed eighth-grade content statements under them, separate ratios and percentages are presented for the two grade levels.

NAEP Framework content areas and topics	TIMSS grade 4 item coverage of NAEP grade 4 subtopics		TIMSS grade 4 item coverage of NAEP grade 8 subtopics	
	Ratio	Percentage	Ratio	Percentage
Life Science				
A: Structures and Functions of Living Systems	3/3	100	3/3	100
B: Changes in Living Systems	2/2	100	2/2	100
Physical Science				
A: Matter	2/2	100	2/2	100
B: Energy	2/2	100	1/2	50
C: Motion	1/2	50	0/2	0
Earth and Space Sciences				
A: Earth in Space and Time	2/2	100	2/2	100
B: Earth Structures	1/1	100	1/2	50
C: Earth Systems	3/3	100	1/3	33

NOTE: Shading indicates areas of coverage. The ratios in this table have as the numerator the number of NAEP subtopics within the named topic that are addressed by one or more TIMSS fourth-grade items, with the denominator being the total number of subtopics within the named topic. The ratios and percentages do not indicate the depth of TIMSS item's coverage of any subtopic as a subtopic may be addressed by a single item or by multiple items. See Appendix C for the data underlying this table.
SOURCE: TIMSS-NAEP Comparison Science Expert Panel, 2011.

Table 14. The ratio and percentage of NAEP 2011 science content area's content statements for grades 4 and 8 addressed by TIMSS 2011 grade 4 science items

NAEP Framework content area, topic, and subtopic	TIMSS grade 4 item coverage of NAEP grade 4 content statements, by subtopic		TIMSS grade 4 item coverage of NAEP grade 8 content statements, by subtopic	
	Ratio	Percentage	Ratio	Percentage
Life Science				
A: Structures and Functions of Living Systems				
A1: Organization and Development	1/1	100	1/2	50
A2: Matters of Energy and Transformations	1/1	100	1/3	33
A3: Interdependence	2/2	100	2/3	67
B: Changes in Living Systems				
B1: Heredity and Reproduction	2/2	100	2/2	100
B2: Evolution and Diversity	1/1	100	2/2	100
Physical Science				
A: Matter				
A1: Properties of Matter	4/5	80	2/5	40
A2: Changes in Matter	1/1	100	1/2	50
B: Energy				
B1: Forms of Energy	1/4	25	3/4	75
B2: Energy Transfer and Conservation	1/1	100	0/2	0
C: Motion				
C1: Motion at the Macroscopic Level	0/2	0	0/1	0
C2: Forces Affecting Motion	2/2	100	0/2	0
Earth and Space Sciences				
A: Earth in Space and Time				
A1: Objects in the Universe	2/2	100	1/2	50
A2: History of Earth	1/1	100	1/2	50
B: Earth Structures				
B1: Properties of Earth Materials	2/3	67	0/3	0
B2: Tectonics	NA	NA	1/3	33
C: Earth Systems				
C1: Energy in Earth Systems	1/1	100	0/2	0
C2: Climate and Weather	1/2	50	0/1	0
C3: Biogeochemical Cycle	2/2	100	1/2	50

NOTE: Shading indicates areas of coverage. The ratios in this table have as the numerator the number of NAEP content statements within the named subtopic that are addressed by one or more TIMSS fourth-grade items, with the denominator being the total number of content statements within the named subtopic. The ratios and percentages do not indicate the depth of TIMSS item's coverage of any content statement as a content statement may be addressed by a single item or by multiple items. It is also important to bear in mind that the pool of NAEP fourth-grade items does not cover all the NAEP framework's content statements in the same depth (i.e., with the same number of items per content statement). Some content statements may be addressed by multiple items, some by only one item, and, at times, it is possible that an assessment may have no NAEP items addressing a particular content statement. See Appendix C for the data underlying this table.
SOURCE: TIMSS-NAEP Comparison Science Expert Panel, 2011.

summarizes how many of NAEP's fourth- and eighth-grade framework's *content statements* were addressed by one or more TIMSS fourth-grade items. In total, 25 out of the NAEP science framework's 33 content statements for fourth grade were addressed by one or more TIMSS fourth-grade items. In addition, 18 of the NAEP science framework's 43 content statements for *eighth* grade were addressed by TIMSS fourth-grade items.

According to the Expert Panel's classifications of TIMSS fourth-grade science items, the fourth-grade NAEP content area with the most coverage in TIMSS at fourth-grade was *Life Science*: all its subtopics and content statements were addressed by TIMSS items, plus 8 out of 12 of its content statements for eighth grade were addressed by TIMSS fourth-grade items. The NAEP content area with the least coverage in TIMSS at fourth grade was *Physical Science*; though, 5 of its 6 fourth-grade subtopics were addressed by TIMSS items, and 9 out of 15 of its content statements for fourth grade and 6 out of 16 of its content statements for eighth grade were addressed by TIMSS fourth-grade items.

Turning to the analysis of the TIMSS eighth-grade items, table 15 summarizes how many of the NAEP framework's *subtopics* were addressed by TIMSS eighth-grade items. Because some of the TIMSS eighth-grade items addressed eighth-grade content statements under these subtopics and some addressed fourth- or twelfth-grade content statements under them, separate ratios and percentages are presented for the framework's different grade levels. In total, 17 out of the NAEP science framework's 18 eighth-grade subtopics were addressed by TIMSS eighth-grade items, and 4 of its 17 twelfth-grade subtopics and 11 of its 17 fourth-grade subtopics were addressed as well. Table 16 summarizes how many of NAEP's fourth-, eighth-, and twelfth-grade framework's *content statements* were addressed by one or more TIMSS eighth-grade

items. In total, 33 out of the NAEP science framework's 43 content statements for eighth grade were addressed by one or more TIMSS eighth-grade item. In addition, 4 of the NAEP science framework's 48 content

Table 15. The ratio and percentage of NAEP 2011 science content area's subtopics for grades 4 and 8 addressed by TIMSS 2011 grade 8 science items

	TIMSS grade 8 item coverage of NAEP grade 4 subtopics		TIMSS grade 8 item coverage of NAEP grade 8 subtopics		TIMSS grade 8 item coverage of NAEP grade 12 subtopics	
NAEP Framework content areas and topics	Ratio	Percentage	Ratio	Percentage	Ratio	Percentage
Life Science						
A: Structures and Functions of Living Systems	2/3	67	3/3	100	0/3	0
B: Changes in Living Systems	1/2	50	2/2	100	0/2	0
Physical Science						
A: Matter	2/2	100	2/2	100	1/2	50
B: Energy	1/2	50	2/2	100	1/2	50
C: Motion	2/2	100	1/2	50	0/2	0
Earth and Space Sciences						
A: Earth in Space and Time	2/2	100	2/2	100	1/2	50
B: Earth Structures	0/1	0	2/2	100	0/1	0
C: Earth Systems	1/3	33	3/3	100	1/3	33

NOTE: Shading indicates areas of coverage. The ratios in this table have as the numerator the number of NAEP subtopics within the named topic that are addressed by one or more TIMSS eighth-grade items, with the denominator being the total number of subtopics within the named topic. The ratios and percentages do not indicate the depth of TIMSS item's coverage of any subtopic as a subtopic may be addressed by a single item or by multiple items. See Appendix C for the data underlying this table.
SOURCE: TIMSS-NAEP Comparison Science Expert Panel

Table 16. The ratio and percentage of NAEP 2011 science content area's content statements for grades 4, 8, and 12 addressed by TIMSS 2011 grade 8 science items

	TIMSS grade 8 item coverage of NAEP grade 4 content statements, by subtopic		TIMSS grade 8 item coverage of NAEP grade 8 content statements, by subtopic		TIMSS grade 8 item coverage of NAEP grade 12 content statements, by subtopic	
NAEP Framework content area, topic, and subtopic	Ratio	Percentage	Ratio	Percentage	Ratio	Percentage
Life Science						
A: Structures and Functions of Living Systems						
A1: Organization and Development	0/1	0	1/2	50	0/3	0
A2: Matters of Energy and Transformations	1/1	100	3/3	100	0/3	0
A3: Interdependence	1/2	50	3/3	100	0/1	0
B: Changes in Living Systems						
B1: Heredity and Reproduction	2/2	100	2/2	100	0/2	0
B2: Evolution and Diversity	0/1	0	2/2	100	0/3	0
Physical Science						
A: Matter						
A1: Properties of Matter	3/5	60	4/5	80	1/4	25
A2: Changes in Matter	1/1	100	2/2	100	0/3	0
B: Energy						
B1: Forms of Energy	2/4	50	3/4	75	0/4	0
B2: Energy Transfer and Conservation	0/1	0	1/2	50	1/5	20
C: Motion						
C1: Motion at the Macroscopic Level	1/2	50	0/1	0	0/2	0
C2: Forces Affecting Motion	1/2	50	1/2	50	0/5	0
Earth and Space Sciences						
A: Earth in Space and Time						
A1: Objects in the Universe	1/2	50	2/2	100	0/3	0
A2: History of Earth	1/1	100	1/2	50	1/4	25
B: Earth Structures						
B1: Properties of Earth Materials	0/3	0	2/3	67	NA	NA
B2: Tectonics	NA	NA	2/3	67	0/1	0
C: Earth Systems						
C1: Energy in Earth Systems	0/1	0	1/2	50	0/1	0
C2: Climate and Weather	0/2	0	1/1	100	1/1	100
C3: Biogeochemical Cycle	2/2	100	2/2	100	0/3	0

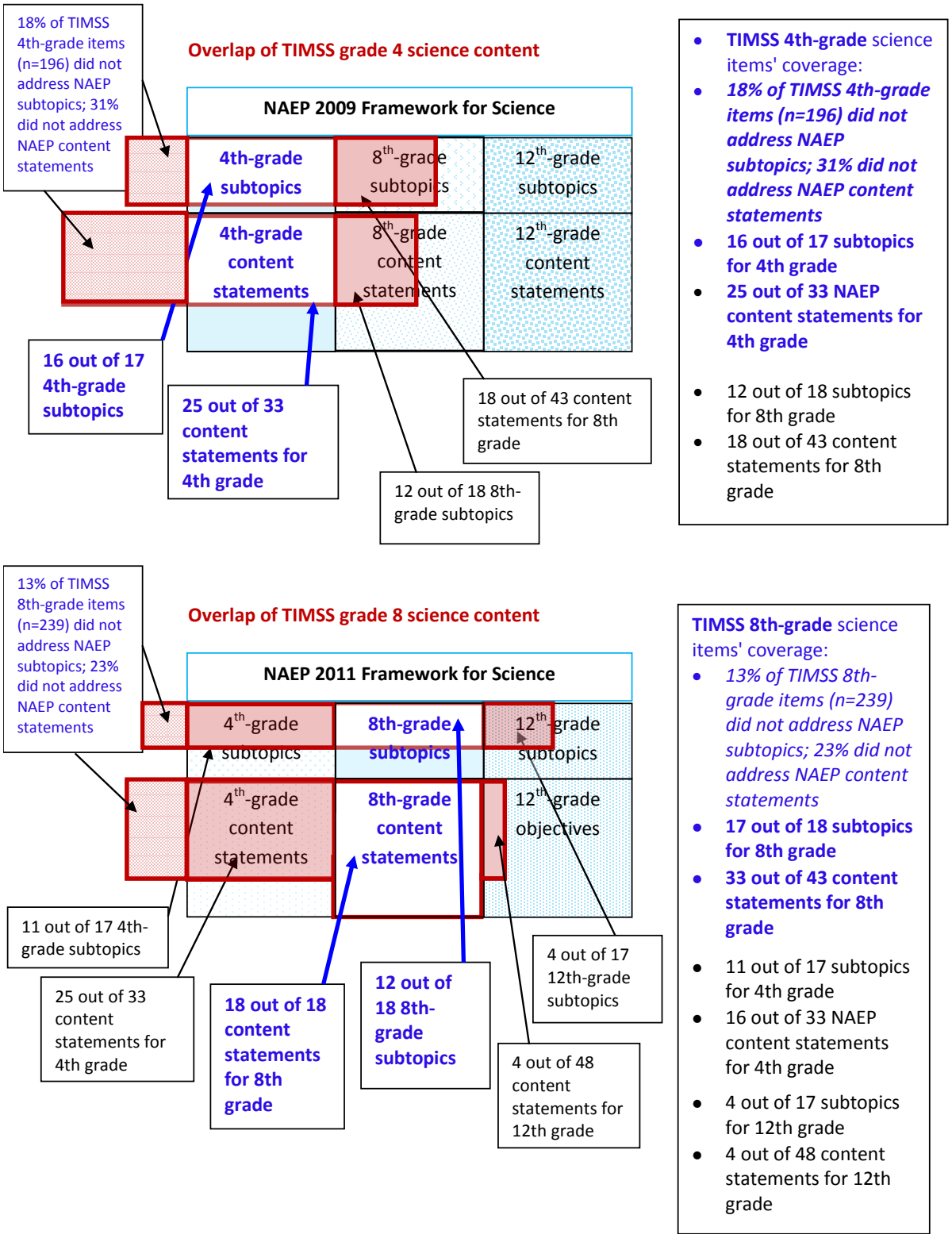
NOTE: Shading indicates areas of coverage. The percentages in this table indicate the number of NAEP's fourth-, eighth-, and twelfth-grade framework's content statements that were addressed by one or more TIMSS eighth-grade items; they do not indicate the depth of TIMSS item's coverage of any content statements. Moreover, it is important to bear in mind that the pool of NAEP eighth-grade items does not cover all the NAEP framework's content statements in the same depth (i.e., with the same number of items per content statement). Some content statements may be addressed by multiple items, some by only one item, and, at times, it is possible that an assessment may have no items addressing a particular content statement. See Appendix C for the data underlying this table.
SOURCE: TIMSS-NAEP Comparison Science Expert Panel, 2011.

statements for twelfth grade and 16 of the NAEP science framework's 33 content statements for fourth grade were addressed by TIMSS eighth-grade items.

According to the Expert Panel's classifications of TIMSS eighth-grade science items, the NAEP content area with the most coverage in TIMSS at eighth grade was *Life Science*: all its subtopics and 11 out of 12 of its eighth-grade content statements were addressed by TIMSS items. The NAEP eighth-grade content area with the least coverage in TIMSS at eighth grade was *Physical Science;* though 5 of its 6 eighth-grade subtopics were addressed by TIMSS items, and 11 out of 16 of its content statements for eighth grade, 2 out of 23 of its content statements for twelfth grade, and 8 out of 15 of its content statements for fourth grade were addressed by TIMSS eighth-grade items.

This analysis of TIMSS items indicates that 16 out of 17 of NAEP framework's fourth-grade subtopics and 17 out of 18 of its eighth-grade subtopics are addressed by TIMSS fourth- and eighth-grade items, respectively. About three-quarters (25/33) of both its fourth-grade content statements and three-quarters (33/43) of its eighth-grade content statements are addressed by TIMSS eighth-grade items. Thus while there are many TIMSS items that do not match the NAEP framework, the items that do map to the NAEP framework cover it fairly thoroughly. This level of congruence suggests that TIMSS does not have a strong content match at fourth- and eighth-grade (even though it does cover most of the NAEP framework at the finest level of detail) because 31 percent of TIMSS 2011 fourth-grade and 23 percent of TIMSS 2011 eighth-grade science items could not be classified to the finest level of detail in the NAEP 2009 or 2011 science framework (see exhibit 2).

Exhibit 2: Overlap of TIMSS 2011 science content with NAEP 2009 and 2011 Frameworks for Science



Grade-level fit

In addition to classifying all the TIMSS 2011 science items to the NAEP 2011 science framework's four content categories (i.e., content area, topic, subtopic, and content statement), the Expert Panel also completed a separate grade-level fit analysis. For most items, the Expert Panel used the NAEP 2011 content statements by grade-level as their reference for completing this classification. However, there were occasions when the Expert Panel could not find a *specific* content statement for an item, but thought that that item was implied in the NAEP 2011 science framework at a particular grade based on their knowledge of NAEP and the content expertise.[26] Items that were neither specified nor implied by the NAEP 2011 science framework at any particular grade were classified as "no fit."

For TIMSS 2011 fourth-grade science items, the Expert Panel determined that 61 percent of the items aligned with the fourth grade NAEP 2011 science framework, while 22 percent aligned better with the eighth grade NAEP 2011 framework (table 17). In addition, the Expert Panel identified 18 percent of the TIMSS 2011 fourth-grade science items to be "no fit." That is to say, they determined that these items assessed content or target skills that could not be found in the NAEP 2011 science framework at any of the three grade levels.

For TIMSS 2011 eighth-grade science items, the Expert Panel determined that 69 percent of the items aligned with the eighth grade NAEP 2011 science framework, while 11 percent of the items aligned better with the fourth grade NAEP 2011 science framework and 3 percent aligned"

[26] For example, the NAEP fourth-grade framework talks about "states of matter" and "physical properties of matter" but does not specify understanding of a mixture. Thus, item S031410 in TIMSS, which asks fourth-grade students to identify "salt water" as a mixture, members of the Expert Panel considered to be "implied" even though it is not spelled out in the NAEP framework.

Table 17. Percentage distribution of the TIMSS 2011 science items across NAEP 2011 grade 4, grade 8, and grade 12 science frameworks

	TIMSS 2011 Items	
	Fourth grade (n=194)[1]	Eighth grade (n=238)[1]
NAEP fourth grade	61	11
NAEP eighth grade	22	69
NAEP twelfth grade	0	3
No fit	18	18

[1]Two items in fourth grade and one item in eighth grade could not be mapped to any grade level of the NAEP framework because the Expert Panel felt that the information about the items that they had was not sufficient enough to make grade-level placements. Therefore, these three items were taken out of the total number of the items in this analysis.
NOTE: The letter "n" is an abbreviation for the total number of assessment items. "No fit" identifies the percentage of items the Expert Panelists were unable to classify as consistent with a specific content statement in the NAEP framework or consistent with the content or targeted skills within a grade level as described in the NAEP 2011 framework.
SOURCE: TIMSS-NAEP Comparison Science Expert Panel, 2011.

better with the twelfth grade in the NAEP 2011 science framework (table 17). In addition, the Expert Panel identified 18 percent of the TIMSS 2011 eighth-grade science items to be "no fit.

The Expert Panel also noted that because TIMSS is based on common curricular elements among the participating education systems while NAEP mainly addresses common U.S. curricular elements, some TIMSS items contain content that might be less familiar to U.S. students in a science class and/or at the corresponding grade level. For example, at eighth grade, the TIMSS items that focused on simple machines, forces and pressure, the immune system, and the origin and history of life on earth might be unfamiliar among U.S. eighth-graders in the judgment of the Expert Panel. These themes could be taught in other subject areas or at different grade levels. In addition, the Expert Panel pointed out that contextual information (e.g., region-specific references) associated with certain items may influence how students interpret and respond to the items. For example, some items had to do with specific contextual information

that some U.S. students might be unfamiliar with (depending on their region), such as artesian wells and hedgehogs. In addition, some topics, such as geologic time and atomic structure, were (in the context of certain TIMSS items) considered to be outside the curriculum covered by many U.S. students by the respective grade-level of the assessment.

Item format

The TIMSS 2011 and NAEP 2009/2011 science assessments use multiple-choice and constructed-response formats for test items. The proportions of TIMSS and NAEP science items in these formats are shown in table 18. At both fourth grade and eighth grade, the TIMSS and NAEP assessments have more multiple-choice items than constructed-response items (table 18).

Table 18. Percentage distribution of TIMSS 2011 and NAEP 2009/2011 science items by item format and grade level

	Fourth grade		Eighth grade	
Item format	TIMSS 2011 (n=196)	NAEP 2009 (n= 143)	TIMSS 2011 (n=239)	NAEP 2011 (n=149)
Multiple choice	60	68	53	66
Constructed response	40	32	47	34

NOTE: The letter "n" is an abbreviation for the total number of assessment items in a category.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS) 2011; and National Assessment of Educational Progress (NAEP) 2011 science.

At fourth grade, TIMSS 2011 has 60 percent multiple-choice items and 40 percent constructed-response items. NAEP 2009 has 68 percent multiple-choice items and 32 percent constructed-response items. At eighth grade, the proportions of multiple-choice items and constructed-response items are relatively equal (53 percent and 47 percent, respectively) for TIMSS 2011. By comparison, NAEP 2011 has larger proportion of multiple-choice items (66 percent) than constructed-response items (34 percent). Unlike in mathematics at grade 8, in

science TIMSS and NAEP both rely on multiple-choice items with four response options at grade 4 and 8.

5. Summary

In reporting results on how U.S. students perform, the National Center for Education Statistics (NCES) draws on multiple sources of national data in order to capitalize on the information presented in national and international assessments. In the United States, data on fourth-grade and eighth-grade students' mathematics and science achievement come primarily from two sources: the National Assessment of Educational Progress (NAEP) and the Trends in International Mathematics and Science Study (TIMSS). TIMSS provides internationally comparable data on student performance, while NAEP tracks performance nationally as well as in state and national population subgroups. This comparative study of TIMSS 2011 and NAEP 2009/2011 revealed important similarities and differences between the two assessments.

In the mathematics portion of the comparative study, the TIMSS 2011 and NAEP 2011 mathematics frameworks both specify five similar mathematical content areas to be assessed (*number*, *measurement*, *geometry*, *data*, and *algebra*). However, there are key differences between the two assessments. While both frameworks also specify a cognitive dimension by which students are to be assessed (in addition to the content dimension, the dimension is defined differently. For TIMSS, it has the three domains (*knowing*, *applying*, and *reasoning*), and each grade in TIMSS has a different distribution across these three cognitive domains. For NAEP, it has three levels of cognitive complexity (*low*, *moderate*, and *high*), and the distribution of items across grade levels is set at fourth grade and eighth grade (with targets of 25 percent, 50 percent,

and 25 percent, respectively). In terms of item content, both TIMSS and NAEP emphasize number at fourth grade and shift the focus to algebra at eighth grade.

Item-by-item content match analyses of the mathematics assessments show a strong content correspondence between TIMSS 2011 and NAEP 2011. Only 1 percent of the fourth-grade items and 3 percent of the TIMSS 2011 eighth-grade items could not be fit to a *specific* objective within the NAEP 2011 mathematics framework. Grade-level fit analyses revealed some mismatch between the two assessments. For TIMSS fourth-grade mathematics items, 11 percent were found to align better with the eighth grade in the NAEP mathematics framework. For TIMSS eighth-grade mathematics items, 14 percent were found to align better with the fourth grade or the twelfth grade in the NAEP mathematics framework (3 percent and 11 percent, respectively), while 1 percent of the TIMSS eighth-grade items were found to be "no fit" at any grade level in the NAEP 2011 mathematics framework. Finally, both TIMSS and NAEP assess students using multiple-choice and constructed-response item formats; however, TIMSS has a relatively equal proportion of items in the two formats at both grade levels, whereas NAEP has a greater proportion of multiple-choice items at both grade levels.

Turning to science, the TIMSS 2011 and NAEP 2009/2011 science frameworks assess similar content areas. However, there are key differences between the two assessments. In particular, TIMSS covers a broader range of topics than NAEP, with 31 percent of TIMSS 2011 fourth-grade and 23 percent of TIMSS 2011 eighth-grade science items not matching any of the NAEP 2009 or 2011 science framework's content statements. For this reason, the item-by-item content match analyses of the science assessments did not show the sort of strong content correspondence between TIMSS 2011 and NAEP 2011, even though the remaining TIMSS 2011

science items do cover most of the NAEP framework at the finest level of detail. Grade-level fit analyses, likewise, revealed a degree of mismatch between the two assessments. For TIMSS fourth-grade science items, 22 percent were found to align best with the eighth grade in the NAEP science framework, while an additional 18 percent were considered "no fit" for either of the three grade levels (4th, 8th, or 12th) in the NAEP science framework. For TIMSS eighth-grade science items, 14 percent were found to align best with the fourth grade or twelfth grade in the NAEP science framework (11 percent and 3 percent, respectively), while an additional 18 percent were considered "no fit" for either of the three grade levels (4th, 8th, or 12th) in the NAEP science framework.

Finally, both assessments use similar proportions of multiple-choice and constructed-response item formats, though the proportion of multiple-choice items is slightly greater in NAEP than in TIMSS at both the fourth- and eighth-grades.

In short, the purpose of the assessments, the content coverage, and the grade-level correspondence of the assessment items distinguish TIMSS 2011 from NAEP 2009/2011. The item differences are more noteworthy in the science assessments than in the mathematics assessments. Thus, it is important to bear in mind these differences when interpreting U.S. students' achievement, nationally and internationally, on NAEP and TIMSS.

4. References

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

Mullis, I.V.S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., and Preuschoff. C. (2009). *TIMSS 2011 Assessment Frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Mullis, I.V.S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Aroroa, A., and Erberber, E. (2005). *TIMSS 2007 Assessment Frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

National Assessment Governing Board (NAGB), U.S. Department of Education. (2010a). *Mathematics Framework for the 2011 National Assessment of Educational Progress*. Washington, D.C.: U. S. Government Printing Office.

National Assessment Governing Board (NAGB), U.S. Department of Education. (2010b). *Science Framework for the 2011 National Assessment of Educational Progress*. Washington, D.C.: U.S. Government Printing Office.

National Center for Education Statistics (NCES), U.S. Department of Education. (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). Retrieved June 4, 2012, from http://nces.ed.gov/pubs2001/200107.pdf.

National Center for Education Statistics (NCES), U.S. Department of Education. (2007). *Comparing TIMSS with NAEP and PISA in Mathematics and Science*. Retrieved June 4, 2012, from http://nces.ed.gov/timss/pdf/Comparing_TIMSS_NAEP_%20PISA.pdf.

National Center for Education Statistics (NCES), U.S. Department of Education. (2011). *Table of countries participating in TIMSS from 1995 – 2011*. Retrieved June 4, 2012, from http://nces.ed.gov/timss/countries.asp.

National Center for Education Statistics (NCES), U.S. Department of Education. (2013a). *NAEP-TIMSS Linking Study: Technical Report* (NCES 2013-461). Available at: http://nces.ed.gov/nationsreportcard/pubs/studies/2013461.asp.

National Center for Education Statistics (NCES), U.S. Department of Education. (2013b). *A Comparison of the 2011 Grade 8 NAEP and TIMSS Mathematics and Science Frameworks* (NCES 2013-462). Available at: http://nces.ed.gov/nationsreportcard/pubs/studies/2013462.asp.

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

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 June 4, 2012, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006029.

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. Retrieved June 4, 2012, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006026.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (1999). Classifying Educational Programmes Manual for ISCED-97 Implementation in OECD Countries (1999 Edition). Paris: Author. Retrieved April 9, 2008, from http://www.oecd.org/dataoecd/7/2/1962350.pdf.

Appendix A

Math Expert Panel Members

Carl Cowen
Indiana University-Purdue University of Indiana

Bill Hopkins
Charles A. Dana Center

Mary Ann Huntley
Cornell University

Jeane M. Joyner
Meredith College, Raleigh, North Carolina

Jeremy Kilpatrick
University of Georgia

Mary Lindquist
Columbus State University in Columbus, Georgia

James Milgram
Stanford University

Linda Dager Wilson
Consultant

Science Expert Panel Members

Alicia Cristina Alonzo
Michigan State University

Mary Thandi Buthelezi
Wheaton College

Audrey Champagne
University at Albany, State University of New York

Alex DeCaria
Millersville University of Pennsylvania

Tony Heiting
Iowa Department of Education

Jacqueline Miller
Center for Science Education

Ellen Mingione
Delaware Department of Education

Harold Pratt
Educational Consultants, Inc.

Patrick Sean Smith
Horizon Research, Inc.

Appendix B

The following table presents the results of the Mathematics Expert Panel's work to classify each of the TIMSS 2011 grade 4 and 8 mathematics items to the NAEP 2011 Mathematics Framework.

This table reproduces the NAEP Mathematics Framework's organization on the left. The numbers and letters listed under the content areas and subtopics specify the objectives, by grade level and order in the NAEP Mathematics Framework. Thus "4a" identifies the first objective under the subtopic "Number sense" as being a fourth-grade objective with the label "a" (to distinguish it from the other fourth-grade objectives). Note that the objectives are all labeled alphabetically but that some letters are missing. For example, there is no "8c" under the subtopic "Number sense." This is because there is alignment between objectives with the same letter at grade 4, 8, and 12.

The columns on the right identify the number of TIMSS 2011 items, at each grade, that were matched by the Expert Panel to the objective in the corresponding row. Thus, the first "3" under TIMSS Grade 4 Items indicates that 3 different TIMSS grade 4 items were mapped by the Expert Panel to objective 4a. The zero under TIMSS Grade 8 Items means that no grade 8 items were mapped to objective 4a under the subtopic "Number sense."

Table B-1: Mathematics Expert Panel Classification of TIMSS 2011 Mathematics Items to the NAEP 2011 Mathematics Framework's Objectives

NAEP Framework Content Area and Subtopic	TIMSS Grade 4 Items	TIMSS Grade 8 Items
A: Number Properties and Operations		
1. Number sense		
4a	3	0
4b	1	0
4c	1	0
4d	1	0
4e	9	0
4i	9	0
8a	0	1
8b	0	2
8d	1	6
8e	0	1
8f	0	0
8g	0	0
8h	0	1
8i	0	4
12d	0	0
12f	0	0
12g	0	0
12i	0	0

2. Estimation		
4a	0	0
4b	3	0
4c	0	0
8a	0	0
8b	0	2
8c	0	0
8d	0	0
12b	0	0
12c	0	0
12d	0	0
3. Number operations		
4a	9	0
4b	3	0
4c	0	0
4d	0	0
4e	1	0
4f	17	0
8a	0	9
8d	0	1
8e	0	0
8f	1	11
12a	**0**	**0**
12b	0	0
12c	0	0
12d	0	0
12f	0	0
4. Ratios and proportional reasoning		
4a	0	0
8a	0	2
8b	0	0
8c	1	4
8d	0	6
12c	0	0
12d	0	0
5. Properties of number and operations		
4a	0	0
4b	4	0
4e	0	0
8a	0	0
8b	0	2
8c	0	1
8d	0	1
8e	0	1
12c	0	0
12d	0	0
12e	0	1

12f	0	0
6. Mathematical reasoning using numbers		
4a	2	0
8a	0	0
8b	0	2
12a	0	0
12b	0	0
B: Measurement		
1. Measuring physical attributes		
4a	0	0
4b	6	1
4c	5	0
4e	2	0
4f	5	0
4g	1	0
8b	1	1
8c	0	0
8e	0	1
8f	1	4
8h	1	2
8i	0	0
12b	0	0
12c	0	0
12d	0	8
12f	0	0
12h	0	0
12i	0	0
2. System of measurement		
4a	0	0
4b	3	0
4d	3	0
4e	0	0
8a	0	0
8b	0	0
8c	0	0
8d	0	0
8e	0	0
12a	0	0
12b	0	0
12d	0	0
12e	0	0
12f	0	0
3. Measurement in triangles		
8a	0	1
12a	0	0
12b	0	0

12c	0	0
12d	0	0
12e	0	0
12f	0	0
12g	0	0
C: Geometry		
1. Dimension and shape		
4a	0	0
4b	1	0
4c	1	0
4f	4	0
8a	0	0
8b	3	0
8c	0	1
8d	1	1
8e	1	0
8f	1	0
12c	0	0
12d	0	0
12e	0	0
12f	0	0
2. Transformations of shapes and preservation of properties		
4a	3	1
4c	4	1
4d	0	0
4e	0	0
8a	3	1
8c	1	3
8d	2	5
8e	0	1
8f	0	2
12a	0	0
12b	0	0
12c	0	0
12d	0	0
12e	0	0
12g	0	0
3. Relationships between geometric figures		
4a	0	0
4b	1	0
4c	2	0
4f	1	0
8b	0	0
8c	0	0
8d	0	3
8f	0	1
8g	0	1

12b	0	0
12c	0	0
12d	0	0
12e	0	0
12f	0	0
12g	0	2
12h	0	0
4. Position, direction, and coordinate geometry		
4a	4	0
4d	0	0
8a	0	1
8b	0	0
8c	0	0
8d	0	0
12a	0	0
12b	0	0
12c	0	0
12d	0	0
12e	0	0
12f	0	0
12g	0	0
12h	0	0
5. Mathematical reasoning in geometry		
4a	1	0
8a	0	0
12a	0	0
12b	0	0
12c	0	0
12d	0	0
12e	0	0
D: Data Analysis, Statistics, and Probability		
1. Data representation		
4a	9	0
4b	3	3
4c	10	0
8a	0	11
8b	0	0
8c	0	7
8d	1	2
8e	0	0
12a	0	0
12b	0	0
12c	0	0
12d	0	0
12e	0	0
12f	0	0

2. Characteristics of data sets		
4b	0	0
4d	1	0
8a	0	5
8b	0	0
8c	0	1
8d	0	2
8e	0	0
12a	0	0
12b	0	0
12c	0	0
12d	0	0
12e	0	0
12f	0	0
12g	0	0
3. Experiments and samples		
8a	0	0
8b	0	0
8d	0	0
12a	0	0
12b	0	0
12c	0	0
12d	0	0
12e	0	0
4. Probability		
4a	0	0
4b	0	0
4e	0	0
4g	0	0
8a	0	1
8b	0	3
8c	0	0
8d	0	5
8e	0	0
8f	0	0
8g	0	0
8h	0	0
8j	0	0
12a	0	0
12b	0	0
12c	0	0
12d	0	0
12e	0	0
12h	0	0
12i	0	0
12j	0	0
12k	0	0

5. Mathematical reasoning with data		
12a	0	0
12b	0	0
12c	0	0
12d	0	0
12e	0	0
E: Algebra		
1. Patterns, relations, and functions		
4a	1	0
4b	4	0
4c	7	0
4d	0	0
4e	0	0
8a	0	12
8b	0	7
8c	0	0
8e	0	0
8f	0	0
12a	0	0
12b	0	0
12e	0	0
12g	0	0
12h	0	0
12i	0	0
12j	0	0
2. Algebraic representation		
4a	2	0
4c	6	0
8a	0	4
8b	0	1
8c	0	0
8d	0	1
8f	0	1
12a	0	1
12b	0	0
12d	0	0
12e	0	0
12f	0	0
12g	0	0
12h	0	0
3. Variables, expressions, and operations		
4a	0	0
4b	1	0
8b	0	11
8c	0	8
12b	0	0

12c	0	1
12d	0	0
12e	0	8
12f	0	0
12g	0	0
12h	0	0
4. Equations and inequalities		
4a	4	0
8a	0	3
8b	0	4
8c	0	4
8d	0	1
8e	0	2
12a	0	1
12c	0	0
12d	0	2
12e	0	0
12f	0	0
12g	0	0
5. Mathematical reasoning in algebra		
4a	0	0
8a	0	1
12a	0	0
12b	0	0
12c	0	0

NOTE: The published NAEP Mathematics Framework (NAGB 2010a) that was used by the Expert Panel listed 100 objectives at grade 8; however, other sources list 101 objectives.

Appendix C

The following table presents the results of the Science Expert Panel's work to classify each of the TIMSS 2011 grade 4 and 8 science items to the NAEP 2011 Science Framework.

This table reproduces the NAEP Science Framework's organization on the left. The letter and numbers listed under the content areas, topics, and subtopics specify the content statements, by content area, grade level, and order in the NAEP Science Framework. Thus "L4.1" identifies the first Life Science ("L") content statement under the subtopic "A1: Organization and Development" as being a fourth-grade content statement (4) with the label ".1" (to distinguish it from the other L4 content statements). Content statements with the letter "P" cover Physical science, and "E" indicates Earth science. Note that the content statements are all labeled sequentially but that the next content statement in a sequence may appear under a different subtopic. For example, "L4.3" appears under the subtopic "A3: Interdependence."

The columns on the right identify the number of TIMSS 2011 items, at each grade, that were matched by the Expert Panel to the objective in the corresponding row. Thus, the first "2" under TIMSS Grade 4 Items indicates that 2 different TIMSS grade 4 items were mapped by the Expert Panel to content statement L4.1. The zero under TIMSS Grade 8 Items means that no grade 8 items were mapped to content statement L4.1.

Table C-1: Science Expert Panel Classification of TIMSS 2011 Science Items to the NAEP 2011 Science Framework's Content Statements[27]

NAEP Framework Content Area, Topic, Subtopic, and Content Statement	TIMSS Grade 4 Items	TIMSS Grade 8 Items
Life Science		
A: Structures and Functions of Living Systems		
A1: Organization and Development		
L4.1	2	0
L8.1	2	5
L8.2	0	0
L12.1	0	0
L12.2	0	0
L12.3	0	0
A2: Matters of Energy and Transformations		
L4.2	6	3
L8.3	0	3
L8.4	2	5
L8.5	0	1
L12.4	0	0
L12.5	0	0
L12.6	0	0

[27] Note that NAEP 2009 and NAEP 2011 Science Frameworks are the same.

	A3: Interdependence		
	L4.3	4	1
	L4.4	2	0
	L8.6	5	10
	L8.7	0	5
	L8.8	4	1
	L12.7	0	0
B: Changes in Living Systems			
	B1: Heredity and Reproduction		
	L4.5	6	2
	L4.6	2	1
	L8.9	2	3
	L8.10	1	4
	L12.8	0	0
	L12.9	0	0
	L12.10	0	0
	B2: Evolution and Diversity		
	L4.7	13	0
	L8.11	2	7
	L8.12	1	5
	L12.11	0	0
	L12.12	0	0
	L12.13	0	0
Physical Science			
A: Matter			
	A1: Properties of Matter		
	P4.1	4	1
	P4.2	6	1
	P4.3	7	0
	P4.4	0	0
	P4.5	4	2
	P8.1	0	7
	P8.2	0	1
	P8.3	0	0
	P8.4	1	22
	P8.5	8	13
	P12.1	0	0
	P12.2	0	1
	P12.3	0	0
	P12.4	0	0
	A2: Changes in Matter		
	P4.6	3	1
	P8.6	0	14
	P8.7	1	9

	P12.5	0	0
	P12.6	0	0
	P12.7	0	0
B: Energy			
	B1: Forms of Energy		
	P4.7	0	1
	P4.8	0	0
	P4.9	5	3
	P4.10	0	0
	P8.8	1	1
	P8.9	0	2
	P8.10	1	2
	P8.11	1	0
	P12.8	0	0
	P12.9	0	0
	P12.10	0	0
	P12.11	0	0
	B2: Energy Transfer and Conservation		
	P4.11	4	0
	P8.12	0	1
	P8.13	0	0
	P12.12	0	0
	P12.13	0	0
	P12.14	0	3
	P12.15	0	0
	P12.16	0	0
C: Motion			
	C1: Motion at the Macroscopic Level		
	P4.12	0	1
	P4.13	0	0
	P8.14	0	0
	P12.17	0	0
	P12.18	0	0
	C2: Forces Affecting Motion		
	P4.14	1	0
	P4.15	3	2
	P8.15	0	0
	P8.16	0	1
	P12.19	0	0
	P12.20	0	0
	P12.21	0	0
	P12.22	0	0
	P12.23	0	0
Earth and Space Sciences			
A: Earth in Space and Time			

	A1: Objects in the Universe		
	E4.1	8	2
	E4.2	1	0
	E8.1	1	7
	E8.2	0	3
	E12.1	0	0
	E12.2	0	0
	E12.3	0	0
	A2: History of Earth		
	E4.3	2	2
	E8.3	4	1
	E8.4	0	0
	E12.4	0	1
	E12.5	0	0
	E12.6	0	0
	E12.7	0	0
B: Earth Structures			
	B1: Properties of Earth Materials		
	E4.4	0	0
	E4.5	1	0
	E4.6	2	0
	E8.5	0	1
	E8.6	0	0
	E8.7	0	3
	B2: Tectonics		
	E8.8	1	1
	E8.9	0	3
	E8.10	0	0
	E12.8	0	0
C: Earth Systems			
	C1: Energy in Earth Systems		
	E4.7	1	0
	E8.11	0	1
	E8.12	0	0
	E12.9	0	0
	C2: Climate and Weather		
	E4.8	1	0
	E4.9	0	0
	E8.13	0	1
	E12.10	0	1
	C3: Biogeochemical Cycle		
	E4.10	5	1
	E4.11	1	2
	E8.14	0	4
	E8.15	1	2
	E12.11	0	0

E12.12	0	0
E12.13	0	0