NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series *June 2001*

The Working Paper Series was initiated to promote the sharing of the valuable work experience and knowledge reflected in these preliminary reports. These reports are viewed as works in progress, and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series.

U.S. Department of Education
Office of Educational Research and Improvement

NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

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)

Working Paper No. 2001-07 **June 2001**

David Nohara

Arnold A. Goldstein
Project Officer
National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement

U.S. Department of Education
Rod Paige
Secretary

National Center for Education Statistics
Gary W. Phillips
Acting Commissioner

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

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

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

> National Center for Education Statistics
> Office of Educational Research and Improvement
> U.S. Department of Education
> 1990 K Street, NW
> Washington, DC 20006

June 2001

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

Suggested Citation

U.S. Department of Education, National Center for Education Statistics. *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, by David Nohara. Arnold A. Goldstein, project officer. Washington, DC: 2001.

Contact:

Arnold A. Goldstein
arnold.goldstein@ed.gov
202-502-7344

Working Paper Foreword

In addition to official NCES Publications, NCES staff and individuals commissioned by NCES produce preliminary research reports that include analyses of survey results, and presentations of technical, methodological, and statistical evaluation issues.

The ***Working Paper Series*** was initiated to promote the sharing of the valuable work experiences and knowledge reflected in these preliminary reports. These reports are viewed as works in progress, and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series.

Copies of Working Papers can be downloaded as pdf files from the NCES Electronic Catalog (http://nces.ed.gov/pubsearch/), or contact Sheilah Jupiter by phone at (202) 502-7444, or by e-mail at sheilah.Jupiter@ed.gov, or by mail at U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 1990 K Street NW, Room 9048, Washington, DC 20006.

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)

Prepared by
David Nohara

Prepared for:

U.S. Department of Education
Office of Educational Research and Improvement
National Center for Education Statistics

June 2001

Executive summary

This report compares the eighth-grade science and mathematics portions of NAEP 2000 with TIMSS-R (the repeat of the Third International Mathematics and Science Study) and the scientific literacy and mathematics literacy portions of PISA (the OECD's Programme for International Student Assessment). It is based on the work of expert panels in mathematics and science education who examined items on each of the three assessments in terms of content, response type, context, requirements for multi-step reasoning, and other characteristics. For all of the characteristics except content, the panels used sets of descriptors developed specifically for this comparison. In the area of curriculum content, panel members compared the three assessments to the NAEP "Fields of Science" and mathematics "Content Strands." The assessments were thus compared using a set of common criteria, which, in almost all cases, were different from the criteria used to develop each assessment. This system of classification was intended to facilitate a comparison of the three assessments and *not* to make judgments regarding their quality. Each assessment was developed based on a different underlying philosophy and set of frameworks. As a result, while sharing many common characteristics, the assessments each have different emphases on content and item type.

In both science and mathematics, there are significant differences between the assessments in most areas examined, many of which can be traced to differences in the purpose of each assessment. Both NAEP and TIMSS-R seek to assess students' mastery of basic knowledge, concepts, and subject-specific thinking skills tied to extensive frameworks of curriculum topics. As a result, both assessments have large numbers of items covering a broad range of topics, with items generally focused on a single, identifiable piece of knowledge, concept, or skill. Some items draw on a combination of topic areas or are more focused on students' scientific or mathematical thinking abilities than on content topic, but these items were in the minority. In contrast, the purpose of PISA is to assess students' abilities to handle everyday situations that require scientific and mathematical skills. As a result, PISA items fit less well on frameworks of curriculum topics and are more often set in real-world contexts. More specific findings for the two different subjects are as follows:

Science

Whereas NAEP items addressed each of the three NAEP Fields of Science in roughly equal proportions, TIMSS-R contained relatively more items emphasizing physical science and PISA contained relatively more items emphasizing Earth science.

Percentage of items that address the NAEP Fields of Science

	NAEP	TIMSS-R	PISA
Earth science	32	22	43
Physical science	33	50	37
Life science	35	30	34

Note: Percentages for TIMSS-R and PISA do not add to 100 since some items were given more than one category designation.

Multiple-choice was the most common response type on all three assessments (73 percent on TIMSS-R, 60 percent on PISA, and 50 percent on NAEP). NAEP had the highest proportion of items requiring extended responses, 43 percent, compared to 21 percent on TIMSS-R and 23 percent on PISA.

Sixty-six percent of PISA items were judged to build connections to relevant practical situations or problems, compared to 23 percent of NAEP items and 16 percent of TIMSS-R items.

PISA had the highest proportion of items requiring multi-step reasoning, 77 percent, compared to 44 percent for NAEP and 31 percent for TIMSS-R.

Based on the factors examined, PISA was judged to be the most difficult of the three assessments. Not only did it rank highest on three of four factors associated with difficulty (response type, context, multi-step reasoning, and mathematical skill), but it contained the largest proportion of items with combinations of two or more of those factors (71 percent, compared to 37 percent for NAEP and 17 percent for TIMSS-R).

Mathematics

The most commonly addressed NAEP mathematics Content Strand on both NAEP and TIMSS-R was number sense, properties, and operations, addressed by 32 percent of NAEP items and 46 percent of TIMSS-R items, compared to only 9 percent of PISA items. The most commonly addressed topic on PISA was data analysis, addressed by 31 percent of items, compared to 14 percent on NAEP and 11 percent on TIMSS-R.

Percentage of items that address the NAEP mathematics Content Strands

	NAEP	TIMSS-R	PISA
Number sense, properties, and operations	32	46	9
Measurement	15	15	25
Geometry and spatial sense	20	12	22
Data analysis, statistics, and probability	14	11	31
Algebra and functions	20	19	19

Note: Percentages for TIMSS-R and PISA do not add to 100 since some items were given more than one category designation.

Extended response items comprised a relatively small proportion of items on all three assessments, 10 percent on NAEP, 3 percent on TIMSS-R, and 12 percent on PISA. The most common response type on NAEP and TIMSS-R was multiple-choice (60 percent of NAEP items and 77 percent of TIMSS-R items, compared to 34 percent of PISA items). The most common response type on PISA was short answer (50 percent of items).

All but one PISA item (97 percent) were judged to present students with real-life situations or scenarios as settings for problems, compared to 48 percent of NAEP items and 44 percent of TIMSS-R items.

TIMSS-R had the highest proportion of items requiring computation (beyond simple computation), 34 percent, compared to 27 percent on NAEP and 25 percent on PISA. Some of these items focus primarily on students' computational abilities, which the panel members placed in the "number sense, properties, and operations" Content Strand. Other items, however, were placed in other Content Strands. In these cases, computation can be seen as an additional element of difficulty. PISA had the highest proportion of items requiring computation but that were not classified in the “number sense, properties, and operations” Content Strand, 19 percent, compared to 12 percent on NAEP and 10 percent on TIMSS-R.

NAEP and PISA contained similar proportions of items requiring multi-step reasoning, 41 and 44 percent respectively. On TIMSS-R, the proportion was somewhat lower, 31 percent.

Almost all PISA items (91 percent) required the interpretation of figures or other graphical data. On NAEP and TIMSS-R, the proportions were closer to half, 56 and 45 percent, respectively.

Based on four factors associated with item difficulty (response type, context, multi-step reasoning, and computation (excluding items classified as “number sense, properties, and operations”)), PISA was judged to be the most difficult of the three assessments, ranking highest on all four factors. It also included the highest percentage of items with two or more of the four factors, 59 percent, compared to 39 percent on NAEP and 24 percent on TIMSS-R.

Project Purpose

For the past 31 years, the National Assessment of Educational Progress (NAEP) has provided educators, policy makers, and the general public with indicators of U.S. student achievement in mathematics, science, reading, writing, geography, U.S. history, and other subjects. In addition to providing overall indicators of student proficiency, the results have been used to gauge progress toward state and national achievement goals, compare achievement levels across states, and to track changes over time. As states have undertaken substantial efforts to raise their students' academic performance, NAEP results have taken on increased significance since they provide external benchmarks and indicators of progress. They are not the only indicators, however. Most notably, the international assessments in mathematics, science, and reading conducted by the International Association for the Evaluation of Educational Achievement (IEA) and the mathematics and science assessments contained in the International Assessment of Educational Progress (IAEP) have assessed similar subject areas and grade levels, but allow comparisons between U.S. students and their counterparts in many other countries throughout the world. In addition, the Organisation for Economic Cooperation and Development (OECD) recently launched the Programme for International Student Assessment (PISA), an assessment of reading literacy, mathematical literacy, and scientific literacy for 28 OECD member countries (of which the United States is one) and several additional non-OECD countries.

With two of these international assessments, the inaugural administration of PISA and the repeat of the IEA's Third International Mathematics and Science Study (TIMSS-R), roughly coinciding with the year 2000 administration of NAEP, there will soon be an unprecedented amount of data regarding U.S. students' achievement in mathematics and science. If all three assessments addressed the same body of knowledge, required the same type of cognitive skills, were administered to students of the same ages and grades, and reported results in the same manner, one would expect performance of U.S. students on the three assessments to be quite similar. The assessments are not the same, however. The three assessments are targeted toward slightly different student populations, place differing emphases on content areas within science and mathematics, include questions requiring different types of responses and thinking skills, and report results in different ways. Consequently, it may not be easy for someone unfamiliar with the details of the three assessments to grasp fully what each says about U.S. students' knowledge and abilities and to reconcile apparent differences in performance across the three.

This publication is intended to help those interested in learning more about the assessments, including their purposes, their similarities and differences, and the relative emphasis each one places on the various content areas and types of knowledge. It is based on the work of expert panels in science and math education and testing who analyzed each assessment item in various categories. It is not intended to facilitate the translation of performance on one of the three into a projected performance on one of the others, nor is it intended as an evaluation of the quality of any of the assessments. But this report should help those wishing to understand the differences between the three assessments and how they might influence performance.

Background on the three assessments

NAEP

The National Assessment of Educational Progress (NAEP) serves as the primary source of information on U.S. students' knowledge and skills in the various subject areas it assesses. Since 1969, assessments have been conducted on a periodic basis, providing educators and policy makers both snapshots of current levels of achievement and trend data based on changes from previous assessments. It addresses knowledge and skills commonly found in school curricula and national curriculum documents, including both specific content topics and broader thinking skills. Assessments are given to fourth-, eighth-, and twelfth-grade students. At the fourth- and eighth-grade levels in reading, writing, mathematics, and science, representative samples are also constructed for each participating state, allowing them to compare their students' achievement with state goals and with average achievement of students in other states and the nation. The most recently administered NAEP assessments were the 2000 assessments in mathematics, science, and reading. In 2001, assessments will be administered in U.S. history and geography. The next assessments in science and mathematics will take place in 2004.

A total of 195 items were developed for the 2000 eighth-grade science assessment and 165 for the 2000 eighth-grade mathematics assessment.[1] However, each individual student was given only a portion of the items in either subject. Both science and mathematics are primarily paper-and-pencil assessments, but the science assessment also includes several sets of items that require students to perform experiments and the mathematics assessment includes items that allow students to use calculators and ones that involve the use of manipulatives, such as cardboard shapes, rulers, and protractors.

Because the other two assessments included in this study were given to students of only one age group, only the eighth-grade NAEP assessments are considered here. Unless otherwise stated, hereafter, "NAEP" refers to the eighth-grade assessment.

TIMSS-R

TIMSS-R is a repeat of the Third International Mathematics and Science Study (TIMSS). The original TIMSS was administered in 1995 in a total of 41 countries at three different grade levels: fourth, eighth, and the final grade of secondary school. As the name indicates, TIMSS was the third international comparative study of both science and mathematics achievement conducted by the International Association for the Evaluation of Educational Achievement (IEA), although it was the first time assessments in the two subjects were conducted together. The original TIMSS had three student populations and three assessments: Population I, students in the two grades enrolling the largest number of 9-year-old students (third and fourth grade in most countries); Population II, students in the two grades enrolling the largest number of 13-year-olds (seventh and eighth grade in most countries); and Population III, students in the final

[1] Several items had two or more parts. The totals mentioned in this report are based on counting each part of an item as a separate item.

year of secondary education.[2] TIMSS-R, administered in 1999 to students in 38 countries, was essentially a repeat of the Population II assessment. It is based on the same framework as TIMSS, and approximately one third of the assessment items are identical to those on the TIMSS Population II assessment.

A total of 144 items were included in the TIMSS-R science assessment and 164 in the mathematics assessment.[1] As in the case of NAEP, each student was given only a subset of the items, but whereas in NAEP, separate assessments exist for each subject, on TIMSS-R, science and mathematics items were placed together in students' assessment booklets.

PISA

The first PISA (Programme for International Student Assessment) assessments were administered in 2000 to 15-year-old students in 32 countries. The stated goal of the PISA program is to measure the "cumulative yield" of education systems, that is, students' knowledge and abilities near the end of their primary-secondary educational careers. It focuses on students' ability to function in situations common in adult life in a mathematically literate society, as opposed to their mastery of detailed sets of curriculum topics.

PISA features separate assessments in the domains of reading literacy, mathematical literacy, and scientific literacy. In each administration cycle of PISA, one of the three domains is to be designated the "major" domain, with approximately two thirds of assessment time devoted to it. In the first cycle of PISA, reading literacy was designated the major domain. In the second cycle, in 2003, mathematical literacy will be the major domain and in 2006, the major domain will be science. In cases where a domain is not the major domain, since less time is available for it, the assessments do not attempt to cover the full range of all aspects identified in the assessment frameworks. For example, although the mathematics framework includes a set of six "big ideas," only two of them were addressed in the first cycle of assessments, "space and shape" and "change and growth." The fact that mathematical literacy and scientific literacy were minor domains in the first PISA cycle meant that far fewer items were developed for PISA in these areas than for either NAEP or TIMSS-R (35 in scientific literacy and 32 in mathematics literacy). PISA also differs from NAEP and TIMSS-R in that most items are grouped together, in groups of two to four, around a common situation described partly by text, graph, or chart, with the sequence of questions increasing in complexity or difficulty.

[2] There were two additional assessments at the Population III level, advanced mathematics and physics, involving two additional groups of students, those students taking or who had taken those courses.

Assessment Frameworks

All three assessments are based on multi-dimensional frameworks that outline the important facts, concepts, and competencies to be covered on the assessments and other desirable characteristics for items. These frameworks are summarized in Figures 1, 2, and 3. In all three frameworks, there is one dimension consisting of content topics and sub-topics (e.g., “algebra” or “life science”) and at least one describing non-topic-based cognitive processes (e.g., “reasoning”). Although these various dimensions may make each framework as a whole appear somewhat complex, they reflect the idea that the importance of any subject comes not just from its body of facts and concepts, but also from processes and skills related to it, not associated with any one topic or sub-topic. In other words, while it is important, for example, for students to have a grasp of scientific facts and concepts, it is also important that they be able to construct a logical chain of reasoning using their science knowledge, regardless of whether they are examining rocks, cells, or circuits.

It is possible to make several general statements about how the different dimensions of the frameworks guide the development of each assessment. First, the different topics and categories within each dimension serve to ensure balance within that dimension. Before the assessment items are written, recommendations are made regarding the proportion of items that should address each topic or fall in each category. For example, the group responsible for designing the NAEP mathematics framework recommended that 15 percent of items on the eighth-grade assessment address “measurement” and that items be evenly distributed across the three categories of Mathematical Abilities. Another common feature of framework categories and topics is that they are not mutually exclusive: all three frameworks recognize that a single item may address more than one content topic or involve more than one type of cognitive skill.

Beyond these general similarities, however, there are significant differences in the purpose of each assessment that affect the dimensions included in the frameworks and their relative influence on item development. One important purpose of both NAEP and TIMSS-R is to measure students’ mastery of knowledge, skills, and concepts. As a result, the content-related dimensions of the NAEP and TIMSS-R frameworks are highly detailed and serve as primary considerations in item development. (Only the major headings are presented in Figures 1 and 2.) In contrast, PISA’s focus is on science and mathematics as they are encountered outside of school, thus the content-related dimensions of PISA are less elaborate and, in the case of mathematics, a secondary consideration for item development. Instead, the dimensions developed in the greatest detail and that serve as primary considerations for item development deal with skills and competencies associated with the subjects but which are not necessarily tied to specific curriculum topics. Although roughly analogous dimensions to the NAEP and TIMSS frameworks exist in PISA, they are not elaborated in as much detail and are given less prominence.

There are other differences as well. For example, while each framework has several dimensions, with the possible exception of the content-related dimensions, they do not correspond well across the assessments. One could argue that the Performance Expectations of the TIMSS-R mathematics framework encompasses both Mathematical Abilities and Mathematical Power of NAEP, but there is nothing on the NAEP or TIMSS-R frameworks comparable to the Situations dimension of the PISA framework. Even in the content-related dimensions, not all topics from one framework can be located easily on another. That such differences exist between frameworks covering the same disciplines demonstrates the idea that there can be different, yet equally valid, ways of conceptualizing and describing these subjects. To the extent that these differences in frameworks will likely influence item development, it will be useful to reflect back on them after the three assessments have been compared.

Figure 1: NAEP Frameworks

Science
Fields of Science (with subtopics) Earth science Solid earth Water Air Earth in space Physical science Matter and its transformations Energy and its transformations Motion Life science Change and evolution Cells and their functions Organisms Ecology
Knowing and Doing Science Conceptual understanding Scientific investigation Practical reasoning
Themes Models Systems Patterns of change
The Nature of Science

Mathematics
Content Strands Number sense, properties, and operations Measurement Geometry and spatial sense Data analysis, statistics, and probability Algebra and Functions
Mathematical Abilities Conceptual understanding Procedural knowledge Problem solving
Mathematical Power Reasoning Connections Communication

Figure 2: TIMSS Frameworks

Science	Mathematics
Content Earth sciences Life sciences Physical sciences Science, technology, and mathematics History of science and technology Environmental and resource issues Nature of science Science and other disciplines	**Content** Numbers Measurement Geometry: position, visualization, and shape Geometry: symmetry, congruence, and similarity Proportionality Functions, relations, and equations Data representation, probability, and statistics Elementary analysis Validation and structure Other content
Performance expectations Understanding Theorizing, analyzing, and solving problems Using tools, routine procedures, and science processes Investigating the natural world Communicating	**Performance expectations** Knowing Using routine procedures Investigating and problem solving Mathematical reasoning Proportionality Communicating
Perspectives Attitudes towards science, mathematics, and technology Careers in science, mathematics, and technology Participation in science and mathematics by underrepresented groups Science, mathematics, and technology to Increase interest Safety in science performance Scientific habits of mind	**Perspectives** Attitudes towards science, mathematics, and technology Careers in science, mathematics, and technology Participation in science and mathematics by underrepresented groups Science, mathematics, and technology to increase interest Scientific and mathematical habits of mind

Figure 3: PISA Frameworks

Science

Scientific Processes

- Recognising scientifically investigable questions
- Identifying evidence needed in a scientific investigation
- Drawing or evaluating conclusions
- Communicating valid conclusions
- Demonstrating understanding of scientific concepts

Scientific Concepts

Scientific themes

- Structure and properties of matter
- Atmospheric change
- Chemical and physical changes
- Energy transformations
- Forces and movement
- Form and function
- Human biology
- Physiological change
- Biodiversity
- Genetic control
- Ecosystems
- Earth and its place in the universe
- Geological change

Areas of Application

- Science in life and health
- Science in Earth and environment
- Science in technology

Situations

- Personal
- Community
- Global
- Historical

Mathematics

MAJOR ASPECTS

Mathematical Competency Classes[3]

- Class 1: reproduction, definitions, and computations
- Class 2: connections and integration for problem solving
- Class 3: mathematical thinking, generalisation, and insight

Mathematical "big ideas"

- Chance
- Change and growth
- Space and shape
- Quantitative reasoning
- Uncertainty
- Dependency and relationships

MINOR ASPECTS

Mathematical Curricular Strands

- Number
- Measurement
- Estimation
- Algebra
- Functions
- Geometry
- Probability
- Statistics
- Discrete mathematics

Situations

- Personal
- Educational
- Occupational
- Public
- Scientific

[3] There is another framework of mathematical competencies, including mathematical thinking; argumentation; modelling; problem posing and solving; representation; symbolic, formal and technical skills; communication; and aids and tools skills. However, the system of competency classes is used instead for the purposes of item development.

Comparing the three assessments

In the preceding background discussion on the assessments and their frameworks, clear differences can be seen in the purposes and philosophical underpinnings of each assessment. Most significant is the fact that while both NAEP and TIMSS-R seek to find out how well students have mastered curriculum-based scientific and mathematical knowledge and skills, the purpose of PISA is to assess students' scientific and mathematical "literacy," that is, their ability to apply scientific and mathematical concepts and thinking skills to everyday, non-school situations. At the same time, it is not always clear how the stated intentions of each assessment will influence what students are asked to do on them. The frameworks differ in structure, content, and nomenclature, making direct comparisons between them difficult, but they also suggest considerable overlap. While one assessment's unique way of conceiving and describing science or mathematics may lead to particular types of items, it is possible that those same items could also fit within the framework of one of the other assessments. Therefore, if the goal is to identify similarities and differences in what students are asked to do on each assessment, it is useful to (1) examine each item, and (2) use a common set of categories and descriptive terms for items across all three assessments.

The methodology for this study is based on a 1997 report to NCES comparing the 1996 NAEP science and mathematics assessments and the original TIMSS.[4] That study and this one relied on panels of experts in science and mathematics to develop criteria for comparison and to review individual items. The 1997 panels identified several important characteristics of items and categories to describe them, most of which were retained for use in this study, with slight modification in some cases. Because differences in the natures of science and mathematics can be reflected in assessment items and because the science and mathematics panels worked separately, the specific questions asked by two groups differ somewhat. In general, however, these characteristics address three questions:

1) Do the assessments cover the same topics?

2) Do the assessments ask the same type of questions?

3) Do the assessments ask the students to use similar types of thinking skills?

Based on how the panels rated items on each characteristic, it is then possible to develop profiles of each assessment, both in terms of individual characteristics and as a whole.

It is important to recognize that placing items in several of the categories below requires judgment on the part of panel members. The panel ratings discussed in this report are those agreed upon by the panels after discussion; their initial individual ratings may have been different. While the consensus process is appropriate for discussing the characteristics of one assessment in relation to those of another, caution should be taken in using the same judgments as absolute statements regarding an individual item or assessment.

---

[4] Don McLaughlin, Senta Raizen, and Fran Stancavage, *Validation Studies of the Linkage Between NAEP and TIMSS Eighth Grade Science Assessments* (Educational Statistical Services Institute, 1997); and Don McLaughlin, John Dossey, and Fran Stancavage, *Validation Studies of the Linkage Between NAEP and TIMSS Fourth and Eighth Grade Mathematics Assessments* (Educational Statistical Services Institute, 1997).

Do the assessments cover the same topics?

Content categories: Although all three assessments are based on multi-dimensional frameworks, with content topic being just one dimension, since U.S. curricula are still, for the most part, structured according to topics within subject areas, the topics addressed remains one of the most important characteristic of any science or mathematics assessment. For the purpose of comparability, panelists were asked to place each item into a category and subcategory of the NAEP "Fields of Science" and the mathematics "Content Strands." (See Figure 1.) While the content frameworks from either TIMSS-R or PISA could also have been used to compare the three assessments, because the purpose of this project was to compare these two assessments to NAEP, the NAEP content frameworks were chosen. As will be seen in the section on the results of the science assessment comparison, NAEP science items are distributed almost equally across the three Fields of Science. It is important not to attach too much significance to NAEP's appearance of balance, since it would probably appear otherwise if analyzed on one of the other two frameworks, both of which organize science topics in different ways.

Using the framework of one assessment to describe items from another assessment inevitably results in several challenges. First, because the frameworks do not cover the exact same set of content topics, there are likely to be items on both TIMSS-R and PISA that do not fit, or do not fit well, within a single NAEP category. They may address several different topics, or none at all. To address this problem, both the science and mathematics panels listed more than one content category or subcategory for items that addressed more than one category or subcategory.

One problem to which the solution is somewhat more elusive is the fact that not all items were developed to address a particular content topic or set of topics. As noted earlier, the differences between the three frameworks are not simply ones of how the same set of curriculum topics is arranged, but rather of how science and mathematics are approached. In NAEP and TIMSS-R, the approaches are similar; both are centered on curriculum frameworks. PISA, on the other hand, places the primary emphasis on students' ability to use science and mathematics in real-life situations. Addressing curriculum topics was only a secondary consideration. In fact, while the PISA framework does include a list of curriculum topics, unlike NAEP or TIMSS-R, the assessments are not designed to cover the full range of topics, at least not in a single year or when the domains are minor, as was the case for mathematical and scientific literacy in the first cycle. Therefore, while a PISA item might address an identifiable science or mathematics topic, its significance within the PISA framework may come instead from its relation to a different objective, such as assessing a non-topic-bound cognitive skill or either of the "big ideas." The fact that a large number of items can be placed in an externally developed content category does not necessarily mean that assessing that category was the primary purpose of the assessment. The same is also true to a lesser extent for TIMSS-R, and even NAEP, since the frameworks for both of those assessments also include dimensions addressing non-topic-specific scientific and mathematical thinking skills. Although panelists noted cases where items did not fit particularly well on the framework or contained no identifiable science or mathematics curriculum topic, describing the three assessments solely in terms of curriculum topics can not adequately represent the nature of any of them.

Scientific vocabulary (science only): The science panel also examined items to see if they required knowledge of a specialized scientific word. In reviewing the items, they adopted the following three criteria for this question: *1) that knowledge of the term be required to answer the question, 2) that the item not contain a definition of the term, and 3) that the term be one encountered primarily in science class or textbooks, and not have moved into general use.*

Panelists encountered several items which included advanced scientific terms but which either defined them or did not require knowledge of them in order to answer the question. Panelists also found numerous items that included scientific terms that have, over the past several years, moved from the domain of science into more general usage. While there were cases of words that fit clearly into one category or the other, whether a word is part of the general parlance or whether it remains in the domain of science is admittedly a subjective judgment. In spite of the potential for subjectivity, the panel felt that drawing such a distinction was useful nevertheless since a student's knowledge of more general scientific terms and facts may be more the result of influences outside the school than of science instruction.

Do the assessments ask the same type of questions?

Response type: Written assessments can utilize a number of response types, including multiple-choice, short answer, extended response, and drawing or other non-verbal response. Response types are selected based on the information on students' knowledge being sought and on practical considerations of assessment administration. The significance of response type for comparing the three assessments comes from the fact that some response types are associated with higher order thinking skills. While it is certainly true that a multiple-choice item can require advanced reasoning and that an extended response item can be easy for most students, in general, items that require students to explain or justify their answer involve an additional level of reasoning and communication skill not found in multiple-choice or short answer items. On these items, it is not enough to know, infer, or guess the correct answer; students must also be able to explain why they think it is correct. Figure 4 presents the response type classifications used by the science and mathematics panels. It should be noted that items were given only one designation. Items that allowed alternative answers and also required extended free response were generally classified as FRA (free response allowing alternative answers).

Figure 4: Response type classifications

Science	Mathematics
MC—Multiple choice	MC—Multiple choice
FRS—free response with a single short answer	FRS—free response with a single short answer
FRJ—free response involving an explanation or justification	FRJ—free response involving an explanation or justification
FRA—free response allowing alternative answers	FRA—free response allowing alternative answers
	FRD—free response requiring drawing

Context: The context of an item refers to whether it is presented in a manner seen only in the study of mathematics or science, or whether it uses situations, language, or visual information relevant to the world outside of school. The context of an item is important for two reasons. First, it can affect the difficulty of an item. If the context requires students to translate the item into scientific or mathematical terms or concepts, then it requires more thinking than if the item were stated more directly. Students taught primarily in the context of the subject itself may have difficulty with problems presented in a real-world context. In some cases, however, if the real-world context makes the problem more familiar to students or makes it less abstract, they may perform better on it. The context of an item is also important because being able to use scientific and mathematical knowledge in real-world settings is a prominent goal of many curricula and education reform efforts.

Because of the natures of the two subjects, the science and mathematics panels viewed the issue of context somewhat differently. In mathematics, problems that deal solely in the language of mathematics are common and are clearly distinguishable from those incorporating non-mathematical references. Thus the mathematics panel used a simple "yes/no" rating for this category. But, since science is based on observations and explorations of the world around us, science problems devoid of any references to the world outside school are far less common. A more useful distinction is between items that use real-world contexts but focus primarily on the underlying scientific concepts and theories and items that focus on the practical implications of a given situation. In both cases, students must possess knowledge of science, but in the latter, they must consider the practical implications of the situation described. Some items also present situation where students are performing particular actions, presumably outside of school, but where the actions more closely resemble scientific investigations than something students would do in the course of their everyday lives. The panel desired to distinguish items with practical implications from those concerned solely with underlying scientific theories and concepts or those that are essentially scientific experiments. To accomplish this, the science panel rated items according to whether or not they *"build connections to relevant practical situations or problems (either personal or societal), likely to occur outside a science class, lab, or scientific investigation."*

Do the assessments ask the students to use similar types of thinking skills?

Multi-step reasoning: Educators and researchers often draw a distinction between basic skills, such as recalling facts or using routine procedures, and thinking skills, such as developing a solution strategy for an unfamiliar type of problem. There are many systems of describing such skills, but no one method prevails, nor is using them ever free of subjectivity. In this project, panel members focused on reasoning, specifically, whether the item required multi-step solutions. Their definition of "multi-step" was as follows:

> "*requires the transformation of information involving an intermediate image, construct, or sub-problem in order to frame the question in a manner that can then be answered*"

Classifying an item as multi-step requires assumptions about the way students think and solve problems, assumptions that cannot be correct in all cases. Asked about the potential impact of an environmental change, some students may be able to create a mental image of the processes involved and work through the different cause-and-effect relationships while others may simply recall the answer as a fact or theory they had learned in class. Students unable to recall a particular mathematical formula or solution strategy—either because they forgot it or because they never learned it in the first place—might still be able to solve the problem through reasoning or trial-and-error. For some students the problem is a simple one of recalling material previously learned but for others it is far more complex. Whether students use recall or reasoning depends primarily on what they have been taught and what they have learned, both of which will differ from student to student. In examining the multi-step reasoning requirements of items, panel members based their judgments on the knowledge and skills commonly taught in science and mathematics by and in the eighth grade.

Mathematical skills (in science items): Because mathematical thinking is another type of skill that can be found in some science items and that can add to the difficulty of an item, reviewers identified items requiring mathematical skill, excluding extremely basic skills, such as addition or subtraction of whole numbers less than ten.

Computation (in mathematics items): Computation, although a separate curriculum topic itself, often is required in all the other areas of school mathematics and may introduce an added challenge for students in these areas. By the eighth grade, however, most students will have had enough exposure in and practice with some basic computation skills such that they should not add any difficulty to an item. Examples of such computation skills include computation with whole numbers, fractions with common denominators, decimals, elementary percents, and familiar direct proportions. Thus, mathematics panel members identified items requiring computation, making a distinction between two types of items:

> Items requiring no computation or extremely basic computation—These may include some computation, but mastery of these skills is assumed by eighth grade. Computation should not be an obstacle to most students in responding to such items.
>
> Items requiring computation—The computational skill requirements may not necessarily be new, but they will be an obstacle for some students. They will result in variations in performance between students.

Interpretation or use of figures and graphs (in mathematics items)—Mathematics panel members identified items that involved the use and interpretation of figures or visual data, including drawings, charts, figures, or graphs, or the use of manipulatives, such as cardboard shapes. Although processing graphical information is generally considered to require skills different from those involved in the processing of words or mathematical symbols, it does not always add to the difficulty of an item. Some types of charts and figures may be fairly complex and require more effort to comprehend, but others may be quite familiar to students and may actually facilitate students' understanding of the problem.

A larger question: Are the assessments of comparable levels of difficulty?

The level of difficulty of an assessment is one of its most important characteristics, especially when comparing with other assessments and when examining student performance. Perhaps the most direct measure of difficulty is student performance, but since the students taking the assessments were of different ages and grade levels, and since data on student performance were not available for all three assessments at the time this report was written, it is not examined here.

Instead of using actual student performance, difficulty is discussed here in terms of the characteristics that are thought to make items more difficult, several of which have been discussed above. The ***content*** of an item will increase difficulty if students have had little or no exposure to it or if it is particularly complex. Items with certain ***response types*** will be more difficult than others, particularly if they require students to explain or justify their answers. Placing the item in a real-world ***context*** may make it more difficult if it requires the student to translate between the concrete and the abstract and between unfamiliar situations and their existing knowledge. And items will also be more difficult if they require ***multi-step reasoning*** or ***computation***. The influence of these factors, of course, is not uniform, and several of them involve subjective judgments. In general, though, they provide several possible reasons why students might find one item, or an entire assessment, more difficult than another.

Results of the comparison: science

Content

Reviewers placed each item from the three science assessments in the NAEP categories and subcategories of Fields of Science. Figure 6 presents the percent and number of items that address each of the three NAEP Fields of Science and their 11 subcategories. In terms of areas of emphasis, NAEP includes roughly equivalent proportions of items across the three fields of science while TIMSS-R places greater emphasis on physical science than on Earth science or life science. On NAEP, 32 percent of items address Earth science, 33 percent address physical science, and 35 percent address life science, whereas 50 percent of TIMSS-R items address physical science, compared to 30 percent in life science and 22 percent in Earth science. On PISA, the emphasis is more equally distributed than on TIMSS-R but less so than on NAEP: 43 percent of items address Earth science compared to 37 percent for physical science and 34 percent for life science. The fact that NAEP appears more "balanced" than both TIMSS-R and PISA is not an indication of quality, but rather reflects the different emphases of the assessments. Furthermore, had the content frameworks of one of the other two assessments been used, it is unlikely that NAEP would appear as balanced.

Figure 5: Percent and number of items that address NAEP Fields of Science categories and subcategories

		NAEP (195 items)		TIMSS-R (144 items)		PISA (35 items)	
		Percent	Number of items	Percent	Number of items	Percent	Number of items
Earth Science	Solid Earth	18	35	9	13	3	1
	Water	3	6	3	5	9	3
	Air	6	11	7	10	29	10
	Earth in Space	5	10	3	5	11	4
	Earth Science Total	32	62	22	32	43	15
Physical Science	Matter and its Transformations	14	27	23	33	17	6
	Energy and its Transformations	7	13	11	16	9	3
	Motion	12	24	16	23	14	5
	Physical Science Total	33	64	50	72	37	13
Life Science	Change and Evolution	10	20	6	9	3	1
	Cells and Their Functions	4	7	1	1	9	3
	Organisms	10	20	18	26	17	6
	Ecology	12	24	6	8	6	2
	Life Science Total	35	69	30	43	34	12

Notes: Percentages and number of items may not add to totals and category totals due to the fact that, in a small number of cases in NAEP and TIMSS and a significant number of instances in PISA, items were assigned more than one category or subcategory designation, or none at all. For example, an item may have been given two different subcategory classifications within the same field. In this case, the item is counted twice at the subcategory level but only once at the category level.

Looking at subcategories, all three assessments included a relatively large number of items dealing with Matter and Its Transformations. This was the most common subcategory in TIMSS-R, with 23 percent of items addressing it, and the second most common subcategory in both NAEP (14 percent) and PISA (17 percent, the same as Organisms). Motion was another subcategory that was relatively common on all three assessments: it was addressed by 12 percent of items on NAEP, 16 percent of items on TIMSS-R, and 14 percent of items on PISA. However, these were the only two subcategories addressed by a relatively large share of items on all three assessments. As Figure 5 illustrates, there were several cases where a topic emphasized on one assessment received little attention on the others. For example, Organisms was a common topic on both TIMSS-R and PISA, addressed by 18 and 17 percent of items respectively, but it was addressed by only 4 percent of items on NAEP. The most commonly addressed subcategory on NAEP was Solid Earth (18 percent) but only 9 percent of TIMSS-R items and 3 percent of PISA items addressed this topic. The most commonly addressed subcategory on PISA was Air (29 percent), a topic addressed by 7 percent of TIMSS-R items and 6 percent of NAEP items. These differences in topic emphasis indicate that if a single group of students were to take all three assessments, their relative performance on each could be significantly affected by the content of their science instruction.

Although panel members gave each item a category and subcategory designation, they did encounter several cases where the NAEP framework could not easily accommodate the content topics of TIMSS-R or PISA items. Examples of such topics include nutrition, health, chemistry, biochemistry, and levels of organization (e.g., cells, tissue, etc.). They also encountered a number of items that appeared more closely connected to framework dimensions other than content topic. For example, an item may have asked a student to design or draw conclusions from an experiment. In this case, while the field of science in which the experiment was conducted may have been clear, a successful response would depend more on students' ability to reason or think scientifically than on their content knowledge. This finding is not surprising since by design, most items on all three assessments addressed more than one dimension of their frameworks. Panel members found items on all three assessments whose primary emphasis appeared to be scientific thinking, other cognitive processes, or knowledge about the nature of science—notably more on PISA than on NAEP or TIMSS-R.

It is also important to note that while virtually all items could be placed somewhere on the framework, some items addressed more than one category or subcategory. This was much more common on PISA than on the other two assessments, perhaps a reflection of the fact that PISA was designed less as an assessment of curriculum-based knowledge and skills than as an assessment of the ability to use scientific knowledge in real-world situations. Although the NAEP Fields of Science serve as useful means of comparing the three assessments, the significance of each individual item is best understood by examining the complete frameworks of each assessment, including the non-content-based frameworks.

Science-specific vocabulary

Relatively few items on any of the assessments required knowledge of science-specific vocabulary, that is, facts or words one would only encounter in science classes or textbooks. (See Figure 6.) Panel members did, however, find items that included such vocabulary but either did not require knowledge of them to answer the question or that provided a definition of the term, either explicitly or implicitly.

Figure 6: Percent and number of items that require knowledge of science-specific vocabulary

	Percent	Number of items
NAEP	7	14
TIMSS-R	6	9
PISA	3	1

Response type

Multiple-choice was the dominant response type for items on all three assessments, but the extent of that dominance varied between assessments. As illustrated by Figure 7, almost three fourths of TIMSS-R items were multiple-choice (73 percent), compared to half of NAEP items and 60 percent of PISA items. NAEP included the greatest number of questions that required extended responses, 43 percent, compared to 21 percent of TIMSS-R items and 23 percent of PISA items.

Figure 7: Percent and number of items requiring different response types

	Multiple-choice		Free Response: short answer		Extended Free Response: Requires justification		Extended Free Response: allows alternative answers	
	Percent	Number of items	Percent	Number of items	Percent	Number of items	Percent	Number of items
NAEP	50	98	7	13	22	43	21	41
TIMSS-R	73	105	6	9	12	17	9	13
PISA	60	21	17	6	6	2	17	6

Context

As an indicator of the extent to which the assessments are based in real-world situations, science panel members identified items that "built connections to relevant practical situations or problems (either personal or societal), likely to occur outside a science class, lab, or scientific investigation." As would be expected based on its stated purpose, PISA had the highest proportion of such items, 66 percent, compared to 23 percent of items on NAEP and 16 percent on TIMSS-R.

Figure 8: Percent and number of items that build connections to relevant practical situations or problems

	Percent	Number of items
NAEP	23	44
TIMSS-R	16	23
PISA	66	23

Mathematical skills

A relatively small proportion of items on all three science assessments involved mathematical skills. PISA had the highest proportion, 20 percent, followed by NAEP, 13 percent, and TIMSS-R, 8 percent. On the items that did require mathematical skills, the most common skill required was interpreting charts and graphs. Other skills included basic computation and calculating proportions.

Figure 9: Percent and number of items that involve mathematical skills

	Percent	Number of items
NAEP	12	24
TIMSS-R	8	11
PISA	20	7

Multi-step Reasoning

PISA had the highest proportion of items requiring multi-step reasoning, 77 percent, compared to 44 percent for NAEP and 31 percent for TIMSS-R. In this case, multi-step reasoning is defined as “the transformation of information involving an intermediate image, construct, or sub-problem in order to frame the question in a manner that can then be answered.” Because whether or not students use reasoning or simply recall information learned in science class may depend on the content of their science instruction, panelists had to make certain assumptions about students’ base of knowledge. Since they were examining the 8^{th}-grade NAEP assessment and since the target student population for TIMSS-R, 13-year-olds, corresponds roughly to the 8^{th} grade, they did so based on the content of typical U.S. science curricula through the 8^{th} grade. (It should be noted, however, that the target population for PISA is somewhat older, 15 years old.)

Figure 10: Percent and number of items that require multi-step reasoning

	Percent	Number of items
NAEP	44	85
TIMSS-R	31	44
PISA	77	27

Initially, panel members were concerned that the definition used would be too broad and would suppress important distinctions between levels of reasoning. They therefore looked specifically for items within those identified as requiring reasoning that stood out as clearly more challenging than the others. In fact, such items were rare; reviewers found only a few on each assessment, making an additional category unnecessary.

Reading

Reviewers also noted that PISA science items involved more reading than those on either NAEP or TIMSS. All but one of the PISA items were parts of item groups, two or more items based on a passage of text, a chart or figure, or a combination of the two. The performance-based items on NAEP also required students to follow sets of written instructions, but they comprised a much smaller proportion of items, 21 of 195 items, or 11 percent. In general, a substantial amount of reading will add to the difficulty of items, and will present more of a challenge to some students than to others. Although no indicator was developed to describe the amount of reading associated with items, panel members felt that it was significantly more of a factor in the overall difficulty of PISA, with its extensive use of long passages of text, than on NAEP or TIMSS.

Overall difficulty

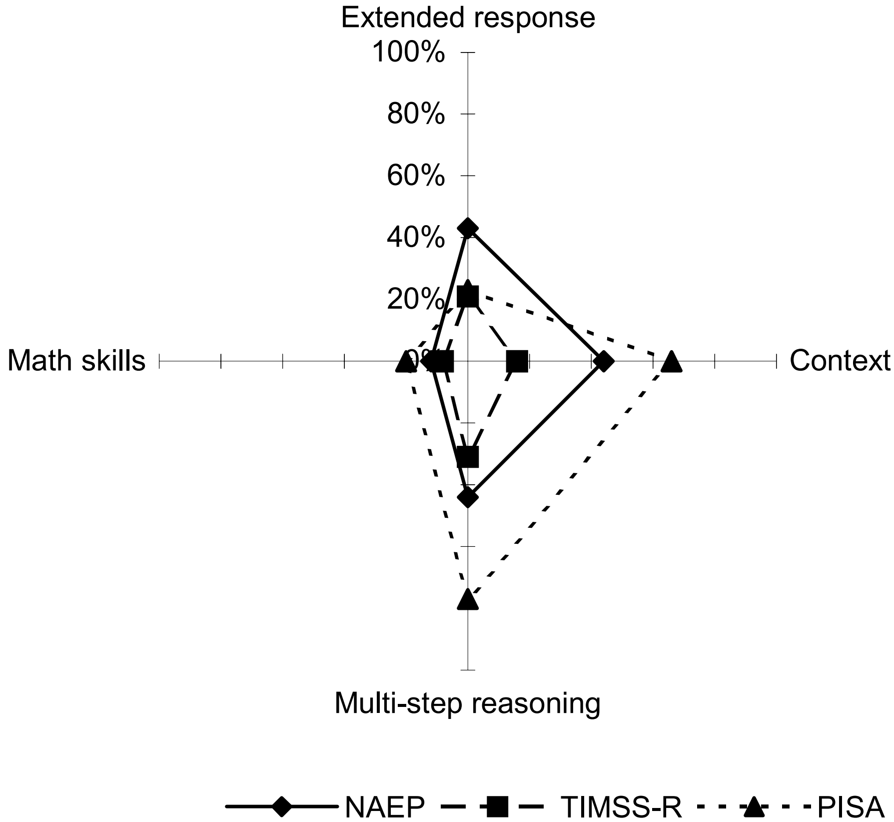
No single indicator was used to describe item difficulty, in part due to the fact that there are many factors that contribute to it, several of which were examined separately by panel members. Although all of the factors discussed above could influence difficulty to some degree, as they were analyzed here, some are less useful indicators than are others. The curricular content of an item will play an important role, since students who have been exposed to the topic in science class or elsewhere will have a clear advantage over those for whom the topic is new. With the differences in topic emphases across the three assessments, it is possible that some students' science education may make them better prepared for one assessment than for another. But, since the inclusion of a topic will affect different students in different ways, it is not a useful indicator of overall difficulty. The presence of science-specific vocabulary could also play an important role, particularly if it is at an advanced level, but it was rare on all three assessments, and thus not a useful comparative indicator.

Examining the remaining factors—response type, context, multi-step reasoning, and mathematical skill—it is possible to develop limited profiles of overall difficulty. Figure 11 presents these four factors on a multi-dimensional plot, with one line representing each of the factors:

> <u>Extended response</u>—the percent of items requiring extended responses (either with justification or with alternative answers),
> <u>Context</u>—the percent of items set in relevant non-school contexts,
> <u>Multi-step reasoning</u>—the percent of items requiring the transformation of information involving an intermediate image, construct, or sub-problem in order to frame the question in a manner that can then be answered, and
> <u>Mathematical skill</u>—the percent of items requiring mathematical skill, excluding extremely basic skills, such as addition or subtraction of whole numbers less than ten..

Looking at all four factors, PISA ranks higher than the other two on three of the four factors and NAEP ranks higher than TIMSS-R on all four.

Figure 11: Science difficulty factors



Another way to use these factors to examine difficulty is to calculate the percentage of items that include combinations of them, based on the reasoning that if these factors do indeed contribute to item difficulty, the more of them present on a single item, the more difficult that item will be. Figure 12 presents the percent and number of items on each assessment that were judged to contain 0, 1, 2, 3, or 4 of the factors associated with difficulty. In this analysis as well, PISA appears to be the most difficult of the three, followed by NAEP. Seventy-one percent of PISA items included 2 or more difficulty factors, compared to 37 percent for NAEP and 17 percent for TIMSS-R.

Figure 12: Percent and number of items with different numbers of difficulty factors

	0 factors		1 factor		2 factors		3 factors		4 factors	
	percent	Number	percent	number	percent	Number	Percent	number	percent	number
NAEP	36	70	27	52	19	38	18	35	0	0
TIMSS-R	56	81	26	38	8	12	8	12	1	1
PISA	14	5	14	5	51	18	11	4	9	3

It is important to recognize that neither of these analyses provides a complete or conclusive prediction of the difficulty of the assessments. Other factors will exert a significant influence, most importantly the content and methods of students' science education in relation to the knowledge and skills addressed on the assessments. Students' science backgrounds will cause them to find items of a given topic relatively simple but those of another topic difficult. Similarly, based on how they have learned and practiced science, they may, for example, find items set in a real-world context easier to understand than those based in the context of scientific theory. Therefore, these analyses should be understood as characterizations of the assessments based on judgments on a limited number of factors thought to be associated with item difficulty.

Summary

There are clear differences between the assessments on a number of factors, differences that in many ways reflect differences in purpose. Both NAEP and TIMSS-R seek to assess the science knowledge of eighth-grade students in relation to extensive frameworks of content topics and subtopics. Not surprisingly, both assessments contain large numbers of items, most of which focus on students' knowledge of basic scientific concepts. While many items address scientific thinking and knowledge of scientific processes—NAEP contains several items requiring students to perform actual experiments— the vast majority of items address a single, identifiable curriculum topic. In contrast, PISA is designed to assess the abilities of older students—15 years old—to function in situations requiring scientific knowledge and skills they are likely to encounter as adults. As a result, PISA contains a large number of items that integrate more than one curriculum topic, focus on students' ability to reason and think scientifically, and require students to read and interpret extended passages of text or charts and figures similar to ones found in newspapers or other common media.

Results of the comparisons: mathematics

Content

When assessment items were placed in the NAEP mathematics Content Strands, there were clear differences in the content emphases of the three assessments. (See Figure 13.) While approximately one fifth of the items on all three assessments dealt with Algebra and Functions, the degrees of emphases on the other four categories differed considerably. On NAEP, the most commonly addressed category was Number Sense, Properties, and Operations. This was true to a greater extent on TIMSS-R: 32 percent of NAEP items addressed this topic, compared to 46 percent of TIMSS-R items. In contrast, only 9 percent of PISA items addressed this category. On PISA, the most commonly addressed topic was Data Analysis, Statistics, and Probability (31 percent of items), whereas on both NAEP and TIMSS-R, it was the least commonly addressed topic (14 percent of NAEP items and 11 percent of TIMSS-R items). These differences in distribution across content categories should not be viewed as indicators of quality, but rather as partial reflections of the different purposes of the assessments.

Figure 13: Percent and number of items that address NAEP mathematics Content Strands

	NAEP (165 items)		TIMSS-R (164 items)		PISA (32 items)	
	Percent	Number of items	Percent	Number of items	Percent	Number of items
Number sense, properties, and operations	32	52	46	76	9	3
Measurement	15	24	15	24	25	8
Geometry and spatial sense	20	33	12	20	22	7
Data analysis, statistics, and probability	14	23	11	18	31	10
Algebra and functions	20	33	19	31	19	6

Notes: Percentages may not add to 100 and number of items in each content strand may not add to item totals due to the fact that, in a small number of cases, items were assigned more than one category designation, or none at all.

If topic subcategories are examined, differences between the assessments become even clearer. As stated, 31 percent of PISA items were classified as data analysis items. Of those, 8 of 10 items related to a common subcategory, "read, interpret, and make predictions using tables and graphs." (See Appendix A.) This means that 25 percent of PISA items related to this one subcategory, compared to only 4 percent of NAEP items and 7 percent of TIMSS-R items. The most commonly addressed subcategory on both NAEP and TIMSS-R was "use computation and estimation in application," a subcategory of Number sense, Properties, and Operations. Thirteen percent of all NAEP items addressed this subcategory, as did 20 percent of TIMSS-R items. On PISA, there was only one item that addressed it.

In general, NAEP and TIMSS-R addressed similar sets of subcategories within each of the five Content Strands, albeit with different distributions among those subcategories. PISA, with a much smaller number of items than either NAEP or TIMSS-R, 32 compared to 165 and 164, did not have near the coverage across subcategories that NAEP and TIMSS-R did. This is a direct result of the intentions of the assessment designers. Whereas the focus of PISA was on students' abilities to use mathematical skills and reasoning in everyday situations, with content being only a secondary consideration, NAEP and TIMSS-R were far more focused on assessing a large and varied range of mathematical skills. Although they also addressed mathematical thinking skills, most items had a clearly identifiable content component.

Response type

Over 75 percent of items on all three assessments were either multiple-choice or short answer. (See Figure 14.) On TIMSS-R, these types of items accounted for all but four percent of items, with 77 percent of all items being multiple-choice and 20 percent being short-answer.[5] On NAEP, 60 percent of items were multiple-choice and 16 percent were short answer. PISA differed from the other two assessments in that there were more short answer items, 50 percent of all items, than multiple-choice, 34 percent. Only NAEP included a significant number of items that required students to draw, 13 percent. While some of these items clearly required spatial reasoning and thereby added a different element of difficulty, other items appeared more basic, for example, requiring students to add a bar or data point to a graph. The only response types that were judged to consistently add difficulty to the items were the extended free responses, which required a justification, allowed for alternative correct answers, or both. On none of the assessments were these items particularly common, 10 percent on NAEP, 3 percent on TIMSS-R, and 9 percent on PISA.

Figure 14: Percent and number of items requiring different response types

	Multiple Choice		Free Response: short answer		Free Response: Drawing		Extended Free Response: requires justification		Extended Free Response: allows alternative answers	
	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N
NAEP	60	99	16	27	13	22	8	14	2	3
TIMSS-R	77	126	20	32	1	2	2	3	1	1
PISA	34	11	50	16	3	1	3	1	9	3

Context

Panel members looked for items that presented students with real-life situations, defined as items not presented strictly in the language of mathematics. This characteristic is significant because connecting mathematics to the world outside of school is a major goal of many mathematics education reform initiatives. It is also significant because it means that students have to choose for themselves the operations and solutions most appropriate for the problem and figure out how

[5] Both figures are rounded up, such that the percentage for both of these two response types combined is 96 percent rather than 97 percent.

they relate to the information provided, thereby adding to the difficulty of an item. All three assessments contained many items situated in real-world contexts, 48 percent of items on NAEP, 44 percent of items on TIMSS-R, and all but one item on PISA, 97 percent.

Figure 15: Percent and number of items that present students with real-life situations or scenarios as settings for the problem

	Percent	Number of items
NAEP	48	79
TIMSS-R	44	72
PISA	97	31

In reviewing PISA items, panel members noted that several items set in real-life situations presented students with significantly more challenging contexts than others. These contexts either were highly unique, that is, not typically encountered in mathematics instruction or textbooks, or required significantly more thought regarding how the nature of the context affects the mathematics involved in the problem. This type of item can be contrasted with standard word problems typically used in mathematics classes, which can be described as "proxies for reality." Panel members looked for this type of item on subsets of NAEP and TIMSS items, but found only a few.[6]

Computation

Panel members looked for items requiring computation, restricting their search only to those items whose computational tasks, although included in most school curricula by the eighth grade, would nevertheless result in variation in student performance. This definition excludes items that include computation judged to be basic enough that it should not be a factor in student success with the item, such as computation involving whole numbers, simple money and measurement problems, and simple fractions. Panel members found a roughly similar percentage of items that required computation on all three assessments: 27 percent on NAEP, 34 percent on TIMSS-R, and 25 percent on PISA. (See Figure 16.)

Figure 16: Percent and number of items that require computation

	All items		Excluding items classified as "number sense, properties, and operations"	
	Percent	Number of items	As a percentage of all items	Number of items
NAEP	27	44	12	19
TIMSS-R	34	55	10	17
PISA	25	8	19	6

[6] The subsets examined were items not appearing in the 1996 NAEP or the original TIMSS, plus an additional block of repeated NAEP items. Of the 51 NAEP items examined, 2 were judged to be in a more challenging context than other items set in real-world contexts. None of the 116 TIMSS-R items were.

When computation is required on an item whose primary content topic is *not* computation, it can add another element of difficulty to the item. Since the NAEP Content Strand of "Number Sense, Properties, and Operations" is the strand most closely associated with computation, looking at the number of items in other content strands that also include computation should provide another indicator of difficulty. Although a large proportion of items requiring computation did fall into the category of "Number Sense, Properties, and Operations," excluding items from that category still leaves a significant number of items that require computation. (See Figure 16.) When the numbers of these items are compared to the numbers of all the items on the assessments, PISA has the highest proportion of items with this additional degree of difficulty, 19 percent, compared to 12 percent on NAEP and 10 percent on TIMSS-R.[7]

Initially, the mathematics panel created an additional level of computational difficulty to describe computation that is either highly complex or is advanced for the eighth-grade level. Items in this category might involve, for example, negative integer exponents, computing with symbolic expressions, or the Pythagorean Theorem. However, panel members found no items on any of the three assessments that fell in this category.

It should be noted that two of the three assessments, NAEP and PISA, allowed students to use calculators. On NAEP, students were allowed to use calculators on designated item blocks (3 blocks consisting of 36 items, or 22 percent of all items). On PISA, the policy was to allow students to have access to calculators, but also to design the items so that the need for calculators was minimal.

Multi-step reasoning

Although virtually all items require some degree of reasoning, panel members attempted to distinguish those items that required students to take more than one step to solve, that is, items that require students to generate an intermediate image, construct, or sub-problem before solving the original problem. Examples of this type of item are ones that require the student to read and interpret a scenario stated in words, a chart, or a diagram or to identify the information needed to solve a problem and derive that information from data given in the item. PISA had the highest proportion of such items, 44 percent, followed by NAEP, 41 percent, and TIMSS-R, 31 percent.

Figure 17: Percent and number of items that require multi-step reasoning

	Percent	Number of items
NAEP	41	68
TIMSS-R	31	51
PISA	44	14

In examining the multi-step reasoning requirements of items, panel members noted one difference between PISA and both NAEP and TIMSS-R related to multi-step thinking. On PISA, items were often clustered together in groups of two to four, centered around a single situation which may involve a figure or chart, with questions increasing in complexity and difficulty. Whereas a

[7] Since the purpose is to assess the extent to which this type of added difficulty affects the assessments as wholes, the denominators used to calculate these percentages are the numbers of all the items on the assessments, rather than the total number of items not classified as "number sense, properties, and operations".

single item on NAEP or TIMSS-R might require students to go through several sub-steps in order to answer the question, some PISA clusters were in essence multi-step tasks, but with each component item representing a single step of that task. In these cases, while an individual item may not have required students to engage in multi-step reasoning, by answering each of the items, students were being led on a multi-step path.

Interpret figures and charts

All three assessments included a large proportion of items that required the use or interpretation of figures or visual data, including drawings, charts, figures, or graphs or the manipulation of physical objects, such as cardboard shapes. PISA had the highest proportion of such items, 91 percent, followed by NAEP, 56 percent, and TIMSS-R, 45 percent. These items were distributed across the five content strands, with the proportions for geometry on NAEP and TIMSS-R higher than the overall proportions of geometry items on the assessments. Subcategories in which this type of item commonly fell included:

- "read, interpret, and make predictions using tables and graphs" (from the Data Analysis, Statistics, and Probability Content Strand),
- "represent numbers and operations in a variety of equivalent forms using models, diagrams, and symbols" (Number Sense, Properties, and Operations),
- "describe, extend, interpolate, transform, and create a wide variety of patterns and functional relationships" (Algebra and Functions),
- "estimate the size of an object or compare objects with respect to a given attribute" (Measurement), and
- "identify the relationship (congruence, similarity) between a figure and its image under a transformation (Geometry).

Figure 18: Percent and number of items that require interpretation of figures

	Percent	Number of items
NAEP	56	92
TIMSS-R	45	73
PISA	91	29

Figures or other graphical data will not have a uniform effect on item difficulty. To the extent that interpreting figures involves a unique set of cognitive skills and often introduces additional steps to the solution process, they can make items more difficult. At the same time, however, a figure or chart can provide additional information in a format other than words, possibly aiding the student's comprehension and development of a solution strategy. Panel members did find several items—all but one on PISA—whose figures they judged to be significantly more complex than the others. In contrast to the standard types of figures and charts used widely in mathematics instruction and familiar to many students, these figures presented information in a novel fashion, requiring more interpretation and analysis on the part of students.

Overall difficulty

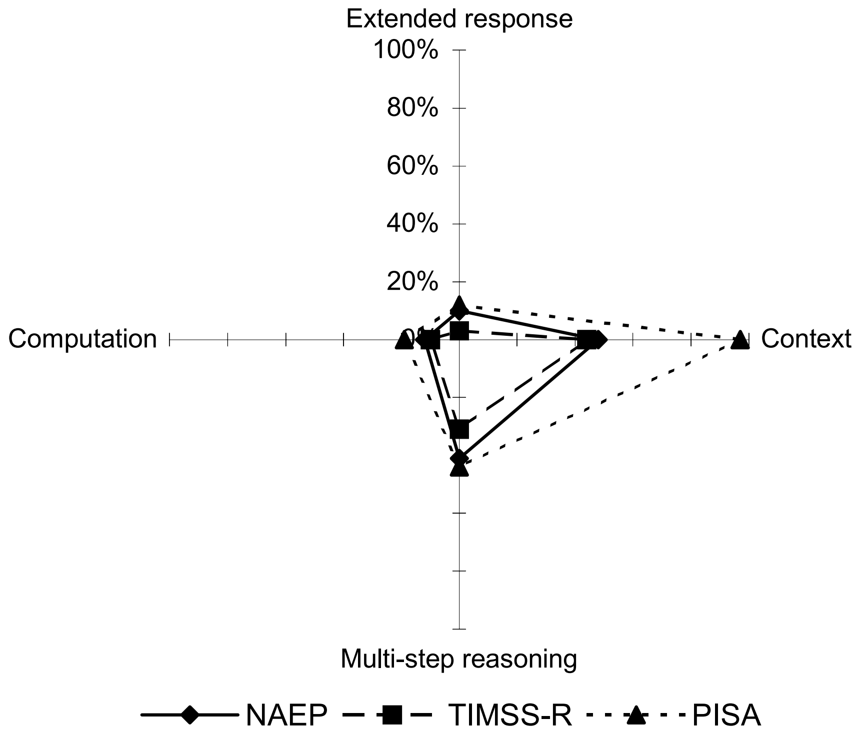
Panel members identified several factors that could contribute to the relative difficulty of the assessments. Key among them are the topics to which students have been exposed and the manner in which they learned mathematics. While many, if not most, students will have had exposure to a broad range of topics and contexts, because different assessments have different emphases is content areas and question types, students' mathematics education may cause them to be better prepared for one assessment than for the others. For example, almost half of TIMSS-R items focused on the content strand of Number sense, Properties, and Operations, more than on NAEP and much more than on PISA, where nearly one third of items instead focused on Data Analysis, Statistics, and Probability, specifically, on reading and interpreting tables and graphs. Almost all PISA items were set in real-life contexts, several of which were judged to be considerably different from the typical word problems used in mathematics instruction.

Of the factors examined, four are likely to make items more difficult for most students in most cases. These include the response type, the context of the item, requirements for multi-step reasoning, and the amount of computation. Figure 19 presents each of these factors together for each assessment on four-line graphs, where:

> Extended response represents the percentage of extended response items, including free-response items that require students to justify their answer, that allow for more than one correct answer, or both,
>
> Context represents the percent of items that presented students with real-life situations, ones not presented strictly in the language of mathematics,
>
> Multi-step reasoning represents the percent of items requiring students to generate an intermediate image, construct, or sub-problem before solving the original problem, and
>
> Computation represents the number of items requiring computation outside the "Number Sense, Properties, and Operations" content strand as a percentage of all items. This is not to say that number sense items are not difficult, but rather that the presence of a computation requirement does not present an *additional* degree of difficulty as it would in an item classified in another content strand.

Looking only at these four factors, PISA appears to be the most difficult: it has the highest percentages in all four categories. It stands out in particular for the high degree of contextualization of items. NAEP and TIMSS-R have similar profiles, with NAEP having more extended response items, more items set in real-world contexts, and more items requiring multi-step reasoning, while TIMSS-R has a slightly greater computational requirement.

Figure 19: Mathematics difficulty factors



PISA also has the highest proportion of items with multiple difficulty factors. On 59 percent of PISA items, panel members found two or more difficulty factors, compared to 39 percent on NAEP and 24 percent on TIMSS-R. Although items exhibiting only one or none of the four characteristics can be more difficult than items exhibiting several of them, especially if the content is unfamiliar to the students, in general since each characteristic represents a different source of variation in student performance, items with a greater number of difficulty factors will present a greater degree of challenge for students.

Figure 20: Percent and number of mathematics items with 0, 1, 2, 3, and 4 difficulty factors

	0 factors		1 factor		2 factors		3 factors		4 factors	
	Percent	number	Percent	Number	percent	Number	percent	number	Percent	number
NAEP	27	45	35	57	27	44	10	16	2	3
TIMSS-R	37	61	39	64	21	34	3	5	0	0
PISA	0	0	41	15	47	13	9	3	3	1

Summary

The three mathematics assessments differ significantly in terms of purpose, target age groups, content emphasis, the type of questions that were asked, and overall degree of difficulty. PISA is intended to be an assessment of mathematical literacy, that is, students' ability to deal with situations they are likely to encounter as adults that require posing and solving mathematical problems. This intention is reflected in the items, which are typically presented in real-life contexts, require the interpretation of charts and graphs, and require a combination of skills and knowledge from different topic areas. PISA includes a much larger proportion of items that involve the interpretation of charts and graphs. It is meant to measure the cumulative effects of a nation's school system, thus the target age for students is 15, an age when most students are still in the school system, but close to the point of entry into the adult world. NAEP and TIMSS-R, on the other hand, are designed for younger students and focus more on knowledge and skills as they relate to a broad range of clearly defined curriculum topics. Comparing NAEP and TIMSS-R, although both contain a large proportion of items dealing with Number sense, Properties, and Operations, the proportion on TIMSS-R is greater than on NAEP (47 percent compared to 32 percent) and TIMSS-R contains a slightly larger percentage of items that require computation. NAEP also contains a larger proportion of geometry items than TIMSS-R, 20 percent compared to 12 percent. In terms of overall difficulty, while the factors examined here cannot provide a definitive indicator of difficulty for each item, PISA items typically have more of the characteristics associated with increased difficulty.

Appendix A: Percent of all mathematics items classified by NAEP mathematics Content Strands

Appendix A.1: Number sense, properties, and operations

		NAEP	TIMSS-R	PISA
1	Relate counting, grouping, and place value	2	2	0
2	Represent numbers and operations in a variety of equivalent forms using models, diagrams, and symbols	6	9	3
3	Compute with numbers (that is, add, subtract, multiply, divide)	3	6	0
4	Use computation and estimation in applications	13	20	3
5	Apply ratios and proportional thinking in a variety of situations	4	8	3
6	Use elementary number theory	2	.5	0
	Total	32	46	9

Appendix A.2: Measurement

		NAEP	TIMSS-R	PISA
1	Estimate the size of an object or compare objects with respect to a given attribute	2	2	0
2	Select and use appropriate measurement instruments	3	2	0
3	Select and use appropriate units of measurement according to type of unit and size of unit	0	.5	0
4	Estimate, calculate, or compare perimeter, area, volume, and surface area in meaningful contexts to solve mathematical and real-world problems	5	5	19
5	Apply given measurement formulas for perimeter, area, volume, and surface area in problem settings	.5	1	3
6	Convert from one measurement to another within the same system	2	1	0
7	Determine precision, accuracy, and error	.5	.5	0
8	Make and read scale drawings	1	.5	6
9	Select appropriate methods of measurement	0	0	0
10	Apply the concept of rate to measurement situations	0	0	0
	Total	15	15	25*

*Note: The total listed for PISA is less than the sum of the percentages of the subcategories since one PISA item classified as measurement was given two different subcategories designations.

Appendix A.3: Geometry and spatial sense

		NAEP	TIMSS-R	PISA
1	Describe, visualize, draw, and construct geometric figures	4	.5	6
2	Investigate and predict results of combining, subdividing, and changing shapes	4	.5	9
3	Identify the relationship between a figure and its image under a transformation	4	2	0
4	Describe the intersection of two or more geometric figures	.5	0	0
5	Classify figures in terms of congruence and similarity, and informally apply these relationships using proportional reasoning where appropriate	2	2	0
6	Apply geometric properties and relationships in solving problems	1	4	0
7	Establish and explain relationships involving geometric concepts	0	2	0
8	Represent problem situations with geometric models and apply properties of figures in meaningful contexts to solve mathematical and real-world problems	4	0	6
9	Represent geometric figures and properties algebraically using coordinates and vectors	0	0	0
	Total	20	12	22

Appendix A.4: Data analysis, statistics, and probability

		NAEP	TIMSS-R	PISA
1	Read, interpret, and make predictions using tables and graphs	4	7	25
2	Organize and display data and make inferences	2	.5	0
3	Understand and apply sampling, randomness, and bias in data collection	1	.5	0
4	Describe measures of central tendency and dispersion in real-world situations	3	.5	0
5	Use measures of central tendency, correlation, dispersion, and shapes of distribution to describe statistical relationships (intended for 12th grade assessment only)	0	0	0
6	Understand and reason about the use and misuse of statistics in our society	.5	0	6
7	Fit a line or curve to a set of data and use this line or curve to male predictions about the data, using frequency distributions where appropriate (intended for 12th grade assessment only)	0	0	0
8	Design a statistical experiment to study a problem and communicate the outcomes	0	0	0
9	Use basic concepts, trees, and formulas for combinations, permutations, and other counting techniques to determine the number of ways an event can occur	0	0	0
10	Determine the probability of a simple event	3	.5	0
11	Apply the basic concept of probability to real-world situations	0	1	0
	Total	14	11	31

Appendix A.5: Algebra and functions

		NAEP	TIMSS-R	PISA
1	Describe, extend, interpolate, transform, and create a wide variety of patterns and functional relationships	5	6	3
2	Use multiple representations for situations to translate among diagrams, models, and symbolic expressions	2	4	0
3	Use number lines and rectangular coordinate systems as representational tools	4	1	0
4	Represent and describe solutions to linear equations and inequalities to solve mathematical and real-world problems	4	5	0
5	Interpret contextual situations and perform algebraic operations on real numbers and algebraic expressions to solve mathematical and real-world problems	2	2	.9
6	Solve systems of equations and inequalities	0	0	0
7	Use mathematical reasoning	2	.5	3
8	Represent problem situations with discrete structures (simple level at 8th grade)	0	0	0
9	Solve polynomial equations with real and complex roots using a variety of algebraic and graphical methods and using appropriate tools (intended for 12th grade assessment only)	0	0	0
10	Approximate solutions of equations (bisection, sign changes, and successive approximations) (simple level at 8th grade)	0	0	0
11	Use appropriate notation and terminology to describe functions and their properties (intended for 12th grade assessment only)	0	0	0
12	Compare and apply the numerical, symbolic, and graphical properties of a variety of functions and families of functions, examining general parameters and their effect on curve shape (simple level at 8th grade)	0	0	0
13	Apply function concepts to model and deal with real-world situations (simple level at 8th grade)	.5	0	.5
14	Use trigonometry (intended for 12th grade assessment only)	0	0	0
	Total	20	19	19

Appendix B: Note on Methodology

The method of comparing the three assessments used in this report is largely based on a study conducted in 1997 to compare the 1996 NAEP mathematics and science assessments with the original TIMSS.[8] In that study, categories of item characteristics were developed for science and mathematics and panels of reviewers gave each item a set of ratings in each category. Most of these categories were retained for this study. Since a large number of the items on the 1996 NAEP assessments and the original TIMSS were repeated on the 2000 NAEP assessments and TIMSS-R, doing so allowed the possibility of using the original item ratings for these repeated items.

This current study also involved two panels, including one person on each panel who had participated in the original study. Panel members were provided with the categories and criteria used in the 1997 study, examples of how items were rated in each category, and item sets for the three assessments. Item sets consisted of newly introduced items on NAEP and TIMSS-R and the complete set of PISA items. In NAEP 2000, 60 of the 195 science items were new and 30 of the 165 mathematics items were. For TIMSS-R, the numbers of new items were 96 for science and 116 for mathematics. In the first step of the review process, reviewers worked independently to rate items in the different categories. Each panel then came together for a two-day meeting to discuss their ratings. Before addressing the items, they first discussed the rating categories. Both groups chose to make slight modifications in the rating system, converting some yes/no categories in ones using a three-point scale. They then reviewed the items, one by one, discussed any differences in how they had rated them, and gave a final consensus rating to each item. After reviewing all the new items, they then looked at how their ratings fit with how items were rated in the original study. Since there were a few categories—some intentional and others not—where they had used a different set of criteria than the original panels, they then rated all the items in these categories that were repeated from the 1996 NAEP assessments and the original TIMSS in the same way they had the new items.

The table below presents the rating categories and data sources for those categories for the items from the 1996 NAEP assessments and the original TIMSS repeated in the 2000 NAEP assessments and TIMSS-R. In science, data on these repeated items were taken from the 1997 study for three categories: content, response type, and mathematical skills. New ratings were developed in the categories of science vocabulary, context, and multi-step reasoning. In mathematics, ratings for the repeated items were taken from the 1997 in all categories except computation.

[8] Don McLaughlin, Senta Raizen, and Fran Stancavage, *Validation Studies of the Linkage Between NAEP and TIMSS Eighth Grade Science Assessments* (Educational Statistical Services Institute, 1997); and Don McLaughlin, John Dossey, and Fran Stancavage, *Validation Studies of the Linkage Between NAEP and TIMSS Fourth and Eighth Grade Mathematics Assessments* (Educational Statistical Services Institute, 1997).

Use of ratings from 1997 study for items repeated in NAEP and TIMSS-R, by category

	Science		Mathematics	
	1997 ratings	2000 ratings	1997 ratings	2000 ratings
Content	✔		✔	
Science vocabulary		✔	(NA)	(NA)
Response type	✔		✔	
Context		✔	✔	
Multi-step reasoning		✔	✔	
Mathematical skills	✔		(NA)	(NA)
Computation	(NA)	(NA)		✔
Interpretation of figures and charts	(NA)	(NA)	✔	

For purposes of comparing the balance of 1997 and 2000 ratings used in this report, the total number of ratings can be calculated by multiplying the number of items in the assessments by the number of categories. For science, there was a total of 374 items across all three assessments (195 on NAEP, 144 on TIMSS-R, and 35 on PISA). Multiplying this number by the number of categories, six, results in 2,244 ratings. The number of 1997 ratings retained for this study is 549, which is equal to the number of repeated items, 183 (135 on NAEP and 48 on TIMSS-R), multiplied by the number of categories in which 1997 ratings were used, three. Thus the percentage of ratings taken from the 1997 study is 24 percent (549 divided by 2244 multiplied by 100). Calculated in this manner, in mathematics, 42 percent of all item ratings came from the 1997 study: 183 (repeated items) multiplied by 5 (categories in which 1997 data were retained), divided by 361 (total items across the three assessments) multiplied by 6 (rating categories), multiplied by 100.

Appendix C: Project Participants

Science Panel

Angelo Collins
Knowles Foundation for Science Teaching

Kathleen Hogan
Institute of Ecosystem Studies

Senta Raizen
National Center for Improving Science Education

Mathematics Panel

John Dossey
Illinois State University

Mary Lindquist
Columbus State University

Thomas Romberg
University of Wisconsin, Madison

Arnold Goldstein
National Center for Education Statistics

David Nohara
Project Consultant

Authors of and participants in 1997 study:

Don McLaughlin
Educational Statistical Services Institute

John Dossey (mathematics)
Illinois State University

Senta Raizen (science)
National Center for Improving Science Education

Fran Stancavage
Educational Statistical Services Institute

Appendix D: Item samples exemplifying classification categories

Sources of samples: Since publicly released items were not available at the time this report was prepared, samples are taken from publicly released items from the 1996 administration of NAEP, the original TIMSS (1996), and a special OECD report prepared to introduce the PISA program, *Measuring Student Knowledge and Skills* (2000). Although none of these items appeared on the assessments reviewed by panel members, they are used to illustrate the type of items that were placed in the various rating categories.

Science

Item requiring knowledge of science-specific vocabulary

Sample S1: What does a mitochondrion do in a cell? It (a) controls the transport of substances leaving and entering the cell, (b) contains the information to control the cell, (c) produces a form of energy that the cell can use, (d) breaks down waste products in the cell. (NAEP)

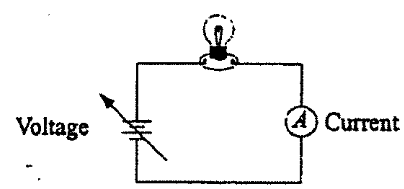
Item not requiring knowledge of science-specific vocabulary

Sample S2: Which of the following energy sources is the best example of a nonrenewable resource? (a) coal, (b) wind, (c) water, (d) sunlight. (NAEP)

Items requiring free response, short answer (FRS)

Sample S3:

Some students used an ammeter *A* to measure the current in the circuit for different voltages.



The table shows some of the results.

Complete the table.

Voltage (volts)	Current (milliamperes)
1.5	10
3.0	20
6.0	

(TIMSS)

Items requiring extended free responses with justification (FRJ)

Sample S4: When operating, ordinary incandescent light bulbs produce a lot of heat in addition to light. Fluorescent light bulbs produce much less heat when operating. If you wanted to conserve electricity, which type of bulb should you use? Explain your answer. (NAEP)

Sample S5: Juanita did several experiments to germinate corn. She summed up her results as follows:

1. Moist grains of corn germinate in the light.
2. Moist grains of corn germinate in the dark.

What can you conclude from her results?

(TIMSS)

Items requiring extended free responses with alternative correct answers (FRA)

Sample S6: Maria's house is near a stream. She wants to put her vegetable garden close to the edge of the stream. Discuss one advantage and one disadvantage of putting the garden there. (NAEP)

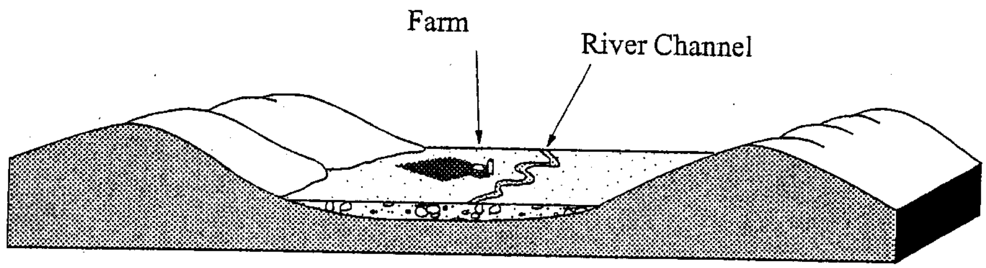
Sample S7: José caught influenza. Write down one way he could have caught it. (TIMSS)

Sample S8: As early as the 11th century, Chinese doctors were manipulating the immune system. By blowing pulverised scabs from a smallpox victim into their patients' nostrils, they could often induce a mild case of the disease that prevented a more severe onslaught later on. In the 1700s, people rubbed their skins with dried scabs to protect themselves from the disease. These primitive practices were introduced into England and the American colonies. In 1771 and 1772, during a smallpox epidemic, a Boston doctor named Zabdiel Boylston scratched the skin on his six-year-old son and 285 other people and rubbed pus from smallpox scabs into the wounds. All but six of his patients survived. Give two other pieces of information that you would need to dcecide how successful Boylston's approach was. (PISA—*one of two items based on the text*)

Items building connections to relevant practical situations or problems

Sample S9: Mrs. Sanchez grows crops on her farm in a hilly region where soil erosion is a big problem. Which of the following would normally help most to protect the soil on her farm from eroding? (a) rotating her crops on a yearly basis, (b) using contour plowing, (c) irrigating her crops more frequently, (d) using more chemical pesticides. (NAEP)

Sample S10: The diagram shows a river flowing through a wide plain. The plain is covered with several layers of soil and sediment.



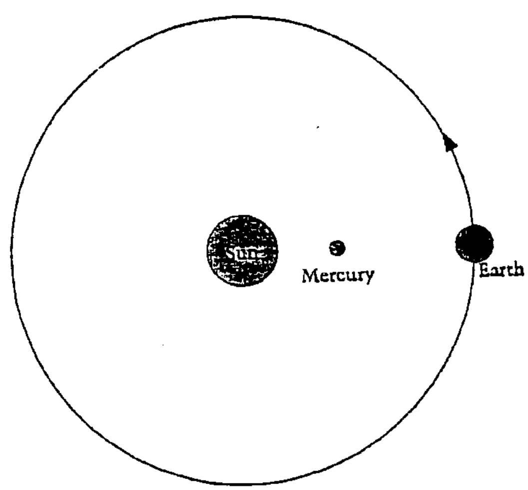
Write down one reason why this plain is a good place for farming.

Write down one reason why this plain is NOT a good place for farming. (TIMSS)

Sample S11: Drivers are advised to leave more space between their vehicle and the one in front when they are travelling more quickly than when they are travelling more slowly because faster cars take longer to stop. Explain why a faster car takes longer to stop than a slower one. (PISA)

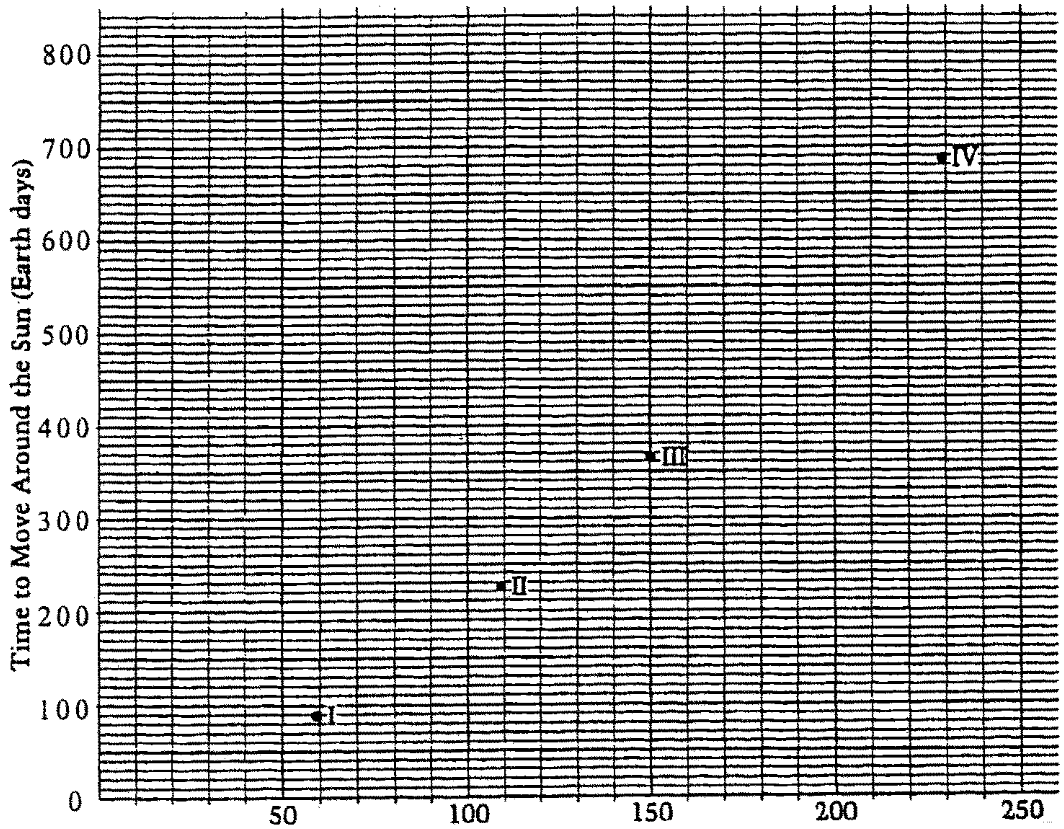
Items requiring mathematical skills

Sample S12: In the picture below, the Earth and Mercury are lined up with the Sun. After Mercury has circled the Sun once, the Earth would be in a different location. Using the information from the graph on [the following page], put an X on the Earth's orbit in the picture to show the new location of the Earth. *(Part of a theme block on planetary orbits in the solar system)*



[Additional graph]

The planets move at different speeds and require different amounts of time to circle the Sun. The following graph shows the number of Earth days it takes for each of the four planets to move around the Sun once.



(NAEP)

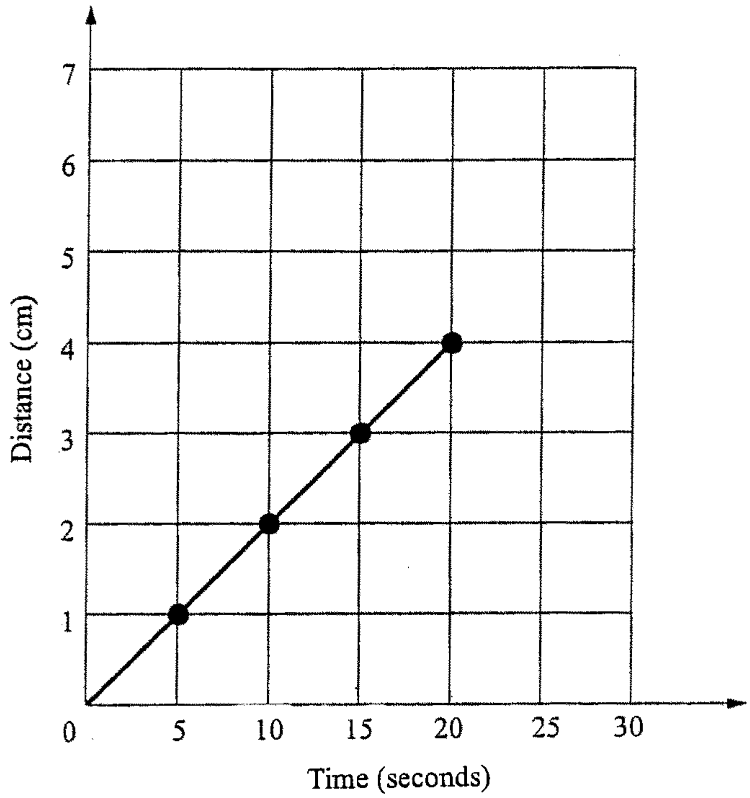
Average Distance from Sun (million kilometers)

I Mercury III Earth

II Venus IV Mars

(NAEP)

Sample S13: The graph shows the progress made by an ant moving along a straight line.



If the ant keeps moving at the same speed, how far will it have traveled at the end of 30 seconds?

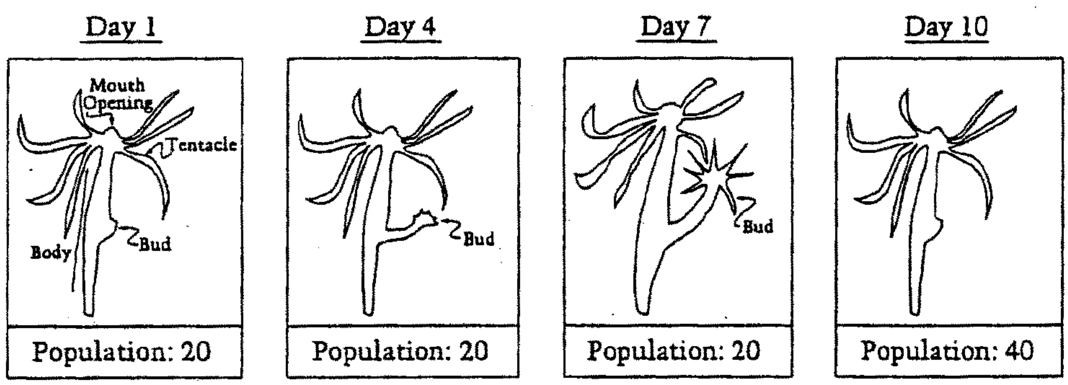
(a) 5 cm, (b) 6 cm, (c) 20 cm, (d) 30 cm.
(TIMSS)

Items requiring multi-step reasoning

Sample S14: Hydras are tiny (l-centimeter long) animals that live in streams and ponds. The picture below shows an adult hydra drawn larger than actual size.

Evita and Michael used 20 hydras for a class science project. They kept the hydras in a glass dish about 5 centimeters high, fed them regularly, and bubbled air into the water to make sure the hydras had enough oxygen.

Evita and Michael observed the hydras every day for 10 days. Each day they drew in their lab notebook the appearance of a typical hydra and recorded the total number of hydras. Their records for day 1, day 4, day 7, and day 10 are shown below.



Evita and Michael predicted that if they fed the hydras twice as much food, the population would double their number in five days. Describe an experiment with appropriate controls that Evita and Michael could do to test this hypothesis. (NAEP—one of three questions based on text and diagrams)

Sample S15: Suppose you want to investigate how the human heart rate changes with changes in activity. What materials would you use and what procedures would you follow? (TIMSS)

Sample S16: Suppose that on one stretch of narrow road Peter *[presented in larger section of text as someone researching traffic behavior]* finds that after the lane lines are painted the traffic changes as in this table:

Speed	Traffic moves more quickly
Position	Traffic keeps nearer edges of road
Distance apart	No change

On the basis of these results it was decided that lane lines should be painted on all narrow roads. Do you think this was the best decision? Give your reasons for agreeing or disagreeing. (PISA—one of four questions based on common theme)

Mathematics

Items requiring a free response, short answer (FRS)

Sample M1: A cereal company packs its oatmeal into cylindrical containers. The height of each container is 10 inches and the radius of the bottom is 3 inches. What is the volume of the box to the nearest cubic inch? (The formula for the volume of a cylinder is $V=\pi r^2 h$.)

Answer: ________________ cubic inches
(NAEP)

Sample M2: The numbers in the sequence 2, 7, 12, 17, 22, … increase by five. The numbers in the sequence 3, 10, 17, 24, 31, … increase by sevens. The number 17 occurs in both sequences. If the two sequences are continued, what is the next number that will be seen in both sequences?

Answer: ____________________
(TIMSS)

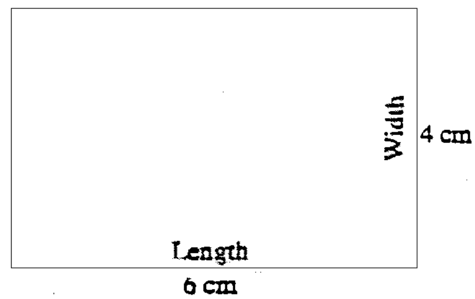
Items requiring a free response and drawing (FRD)

Sample M3: (Part of a block in which students worked with cardboard manipulatives shaped as squares and triangles. Shapes labeled Q are isosceles right triangles.)

You will need the 2 pieces labeled Q. Please find those 2 pieces now.

Use the 2 pieces labeled Q to make a square. Trace the square and draw the line to show where the two pieces meet. (NAEP)

Sample M4:



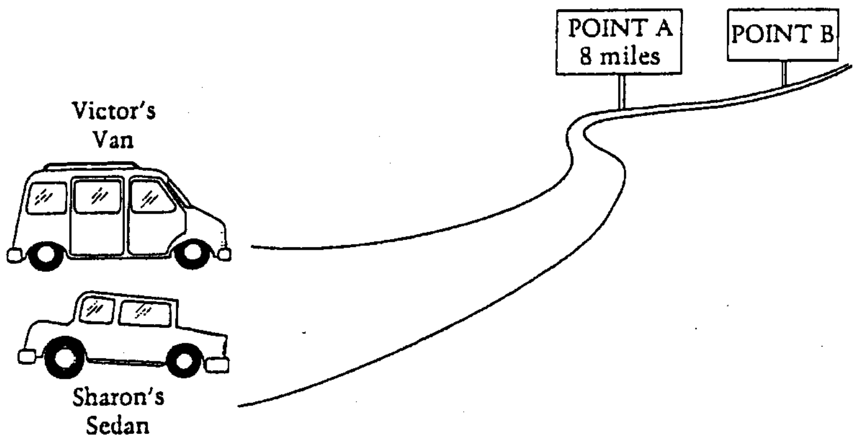
In the space below, draw a new rectangle whose length is one and one half times the length of the rectangle above, and whose width is half the width of the rectangle above. Show the length and width of the new rectangle in centimeters on the figure.

(Students were provided with a grid of 1 x 1 centimeter squares on which to draw a response.)

(TIMSS)

Items requiring extended responses and a justification

Sample M5



Victor's van travels at a rate of 8 miles every 10 minutes. Sharon's sedan travels at a rate of 20 miles every 25 minutes.

If both cars start at the same time, will Sharon's sedan reach point A, 8 miles away, before, at the same time, or after Victor's van?

Explain your reasoning. ____

____

If both cars start at the same time, will Sharon's sedan reach point B (at a distance further down the road) before, at the same time, or after Victor's van?

Explain your reasoning. ____

____

NOTE: Students had access to calculators for this question.

(NAEP)

Sample M6

Theresa wants to record 5 songs on tape. The length of time each song plays for is shown in the table.

Song	Amount of Time
1	2 minutes 41 seconds
2	3 minutes 10 seconds
3	2 minutes 51 seconds
4	3 minutes
5	3 minutes 32 seconds

ESTIMATE to the nearest minute the total time taken for all five songs to play and explain how this estimate was made.

Estimate: ____________________

Explain:

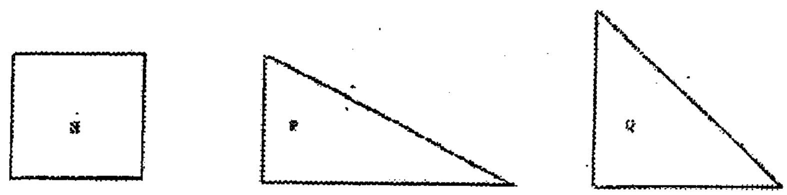
NOTE: The estimate and the explanation were scored as two separate components and therefore counted as two separate items in the current analysis. It is the "explanation" item which is an example of free response with justification.

(TIMSS)

Sample M7: *(Part of a 4-question series on symmetry)* John said: "I know a rule for being able to tell when a 4-sided figure has a folding line of symmetry *[defined earlier]*. If the triangles on each side of the line have the same size and shape, it has a folding line of symmetry." Explain why you either agree or disagree with John. (PISA)

Items requiring extended free responses with alternative correct answers (FRA)

Sample M8: Laura was asked to choose 1 of the 3 shapes N, P, and Q that is different from the other 2. Laura chose shape N. Explain how N is different from shapes P and Q.



(NAEP)

Items presenting real-life situations or scenarios as settings for the problem

Sample M9: Anita is making bags of treats for her sister's birthday party. She divides 65 pieces of candy equally among 15 bags so that each bag contains as many pieces as possible. How many pieces will she have left? (a) 33 (b) 5 (c) 4 (d) 3 (e) 0.33 (NAEP)

Sample M10: Jan had a bag of marbles. She gave half of them to James and then a third of the marbles still in the bag to Pat. She then had 6 marbles left. How many marbles were in the bag to start with? (a) 18 (b) 24 (c) 30 (d) 36 (TIMSS)

Sample M11: You have driven two thirds of the distance in your car. You started with a full tank and your tank is now one quarter full. Do you have a problem? (PISA)

Items requiring computation

Sample M12: A car odometer registered 41,256.9 miles when a highway sign warned of a detour 1,200 feet ahead. What will the odometer read when the car reaches the detour? (5,280 feet = 1 mile)

A) 42,456.9
B) 41,279.9
C) 41,261.3
D) 41,259.2
E) 41,257.1

(NAEP)

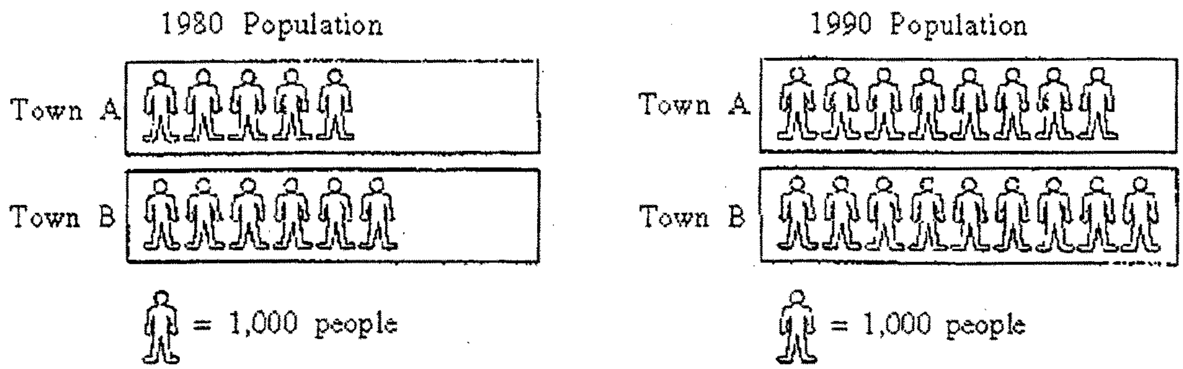
Sample M13: Multiply: 0.203 x 0.56 =

Answer: ________________

(TIMSS)

Items requiring multi-step reasoning

Sample M14:



5. In 1980, the populations of Town A and Town B were 5,000 and 6,000, respectively. The 1990 populations of Town A and Town B were 8,000 and 9,000, respectively.

Brian claims that from 1980 to 1990 the populations of the two towns grew by the same amount. Use mathematics to explain how Brian might have justified his claim.

Darlene claims that from 1980 to 1990 the population of Town A had grown more. Use mathematics to explain how Darlene might have justified her claim.

(NAEP)

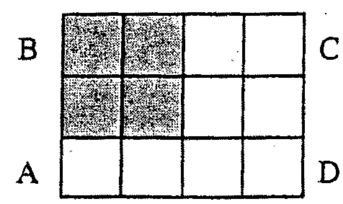
Sample M15: A car has a fuel tank that holds 35 liters of fuel. The car consumes 7.5 liters of fuel for each 100 kilometers driven. A trip of 250 kilometers was started with a full tank of fuel. How much fuel remained in the tank at the end of the trip?

(a) 16.25 L (b) 17.65 L (c) 18.75 L (d) 23.75 L
(TIMSS)

Sample M16: Would it be possible to establish a coinage system (or a stamp system) based on only the denominations 3 and 5? More specifically, what amounts could be reached on that basis? If possible, would such a system be desirable? (PISA)

Items requiring use or interpretation of figures or other visual data

Sample M17



In the figure above, what fraction of rectangle ABCD is shaded?

(a) $\frac{1}{6}$ (b) $\frac{1}{5}$ (c) $\frac{1}{4}$ (d) $\frac{1}{3}$ (e) $\frac{1}{2}$

(NAEP)

Sample M18

The figure consists of 5 squares of equal size. The area of the whole figure is 405 cm^2.

Find the area of one square. Answer ________ square centimeters.

Find the length of the side of one square. Answer_________ centimeters

Find the perimeter of the whole figure in centimeters. Answer ________ centimeters.

(TIMSS)

Sample M19:

Tepla Loroupe won the 1998 Rotterdam marathon. "It was easy", she said, "the course was quite flat".

Here you see a graph of the differences in elevation of the Rotterdam marathon course:

[*Differences in level of the course — in metres relative to the starting point*]

(PISA)

What was the difference between the highest and the lowest points of the course?

Listing of NCES Working Papers to Date

Working papers can be downloaded as pdf files from the NCES Electronic Catalog (http://nces.ed.gov/pubsearch/). You can also contact Sheilah Jupiter at (202) 502–7444 (sheilah_jupiter@ed.gov) if you are interested in any of the following papers.

Listing of NCES Working Papers by Program Area

No.	Title	NCES contact
Baccalaureate and Beyond (B&B)		
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
Beginning Postsecondary Students (BPS) Longitudinal Study		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
2001-04	Beginning Postsecondary Students Longitudinal Study: 1996-2001 (BPS:1996/2001) Field Test Methodology Report	Paula Knepper
Common Core of Data (CCD)		
95–12	Rural Education Data User's Guide	Samuel Peng
96–19	Assessment and Analysis of School-Level Expenditures	William J. Fowler, Jr.
97–15	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
2000–12	Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary School Universe Survey	Beth Young
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
Data Development		
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Decennial Census School District Project		
95–12	Rural Education Data User's Guide	Samuel Peng
96–04	Census Mapping Project/School District Data Book	Tai Phan
98–07	Decennial Census School District Project Planning Report	Tai Phan
Early Childhood Longitudinal Study (ECLS)		
96–08	How Accurate are Teacher Judgments of Students' Academic Performance?	Jerry West
96–18	Assessment of Social Competence, Adaptive Behaviors, and Approaches to Learning with Young Children	Jerry West
97–24	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97–36	Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research	Jerry West
1999–01	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2001–02	Measuring Father Involvement in Young Children's Lives: Recommendations for a Fatherhood Module for the ECLS-B	Jerry West
2001–03	Measures of Socio-Emotional Development in Middle Childhood	Elvira Hausken
2001–06	Papers from the Early Childhood Longitudinal Studies Program: Presented at the 2001 AERA and SRCD Meetings	Jerry West

No.	Title	NCES contact
Education Finance Statistics Center (EDFIN)		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
96–19	Assessment and Analysis of School-Level Expenditures	William J. Fowler, Jr.
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
1999–16	Measuring Resources in Education: From Accounting to the Resource Cost Model Approach	William J. Fowler, Jr.
High School and Beyond (HS&B)		
95–12	Rural Education Data User's Guide	Samuel Peng
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
HS Transcript Studies		
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
International Adult Literacy Survey (IALS)		
97–33	Adult Literacy: An International Perspective	Marilyn Binkley
Integrated Postsecondary Education Data System (IPEDS)		
97–27	Pilot Test of IPEDS Finance Survey	Peter Stowe
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2000–14	IPEDS Finance Data Comparisons Under the 1997 Financial Accounting Standards for Private, Not-for-Profit Institutes: A Concept Paper	Peter Stowe
National Assessment of Adult Literacy (NAAL)		
98–17	Developing the National Assessment of Adult Literacy: Recommendations from Stakeholders	Sheida White
1999–09a	1992 National Adult Literacy Survey: An Overview	Alex Sedlacek
1999–09b	1992 National Adult Literacy Survey: Sample Design	Alex Sedlacek
1999–09c	1992 National Adult Literacy Survey: Weighting and Population Estimates	Alex Sedlacek
1999–09d	1992 National Adult Literacy Survey: Development of the Survey Instruments	Alex Sedlacek
1999–09e	1992 National Adult Literacy Survey: Scaling and Proficiency Estimates	Alex Sedlacek
1999–09f	1992 National Adult Literacy Survey: Interpreting the Adult Literacy Scales and Literacy Levels	Alex Sedlacek
1999–09g	1992 National Adult Literacy Survey: Literacy Levels and the Response Probability Convention	Alex Sedlacek
2000–05	Secondary Statistical Modeling With the National Assessment of Adult Literacy: Implications for the Design of the Background Questionnaire	Sheida White
2000–06	Using Telephone and Mail Surveys as a Supplement or Alternative to Door-to-Door Surveys in the Assessment of Adult Literacy	Sheida White
2000–07	"How Much Literacy is Enough?" Issues in Defining and Reporting Performance Standards for the National Assessment of Adult Literacy	Sheida White
2000–08	Evaluation of the 1992 NALS Background Survey Questionnaire: An Analysis of Uses with Recommendations for Revisions	Sheida White
2000–09	Demographic Changes and Literacy Development in a Decade	Sheida White
National Assessment of Educational Progress (NAEP)		
95–12	Rural Education Data User's Guide	Samuel Peng
97–29	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Steven Gorman
97–30	ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results	Steven Gorman
97–31	NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress	Steven Gorman
97–32	Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questionnaires)	Steven Gorman
97–37	Optimal Rating Procedures and Methodology for NAEP Open-ended Items	Steven Gorman

No.	Title	NCES contact
97–44	Development of a SASS 1993–94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
2001-07	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)	Arnold Goldstein
National Education Longitudinal Study of 1988 (NELS:88)		
95–04	National Education Longitudinal Study of 1988: Second Follow-up Questionnaire Content Areas and Research Issues	Jeffrey Owings
95–05	National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors	Jeffrey Owings
95–06	National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data	Jeffrey Owings
95–07	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
95–12	Rural Education Data User's Guide	Samuel Peng
95–14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
98–06	National Education Longitudinal Study of 1988 (NELS:88) Base Year through Second Follow-Up: Final Methodology Report	Ralph Lee
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
National Household Education Survey (NHES)		
95–12	Rural Education Data User's Guide	Samuel Peng
96–13	Estimation of Response Bias in the NHES:95 Adult Education Survey	Steven Kaufman
96–14	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman
96–20	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96–21	1993 National Household Education Survey (NHES:93) Questionnaires: Screener, School Readiness, and School Safety and Discipline	Kathryn Chandler
96–22	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
96–29	Undercoverage Bias in Estimates of Characteristics of Adults and 0- to 2-Year-Olds in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
96–30	Comparison of Estimates from the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97–02	Telephone Coverage Bias and Recorded Interviews in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97–03	1991 and 1995 National Household Education Survey Questionnaires: NHES:91 Screener, NHES:91 Adult Education, NHES:95 Basic Screener, and NHES:95 Adult Education	Kathryn Chandler
97–04	Design, Data Collection, Monitoring, Interview Administration Time, and Data Editing in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97–05	Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97–06	Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97–08	Design, Data Collection, Interview Timing, and Data Editing in the 1995 National Household Education Survey	Kathryn Chandler
97–19	National Household Education Survey of 1995: Adult Education Course Coding Manual	Peter Stowe

No.	Title	NCES contact
97–20	National Household Education Survey of 1995: Adult Education Course Code Merge Files User's Guide	Peter Stowe
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
97–28	Comparison of Estimates in the 1996 National Household Education Survey	Kathryn Chandler
97–34	Comparison of Estimates from the 1993 National Household Education Survey	Kathryn Chandler
97–35	Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey	Kathryn Chandler
97–38	Reinterview Results for the Parent and Youth Components of the 1996 National Household Education Survey	Kathryn Chandler
97–39	Undercoverage Bias in Estimates of Characteristics of Households and Adults in the 1996 National Household Education Survey	Kathryn Chandler
97–40	Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey	Kathryn Chandler
98–03	Adult Education in the 1990s: A Report on the 1991 National Household Education Survey	Peter Stowe
98–10	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
National Longitudinal Study of the High School Class of 1972 (NLS-72)		
95–12	Rural Education Data User's Guide	Samuel Peng
National Postsecondary Student Aid Study (NPSAS)		
96–17	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
2000–17	National Postsecondary Student Aid Study:2000 Field Test Methodology Report	Andrew G. Malizio
National Study of Postsecondary Faculty (NSOPF)		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
Postsecondary Education Descriptive Analysis Reports (PEDAR)		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
Private School Universe Survey (PSS)		
95–16	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95–17	Estimates of Expenditures for Private K–12 Schools	Stephen Broughman
96–16	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
96–26	Improving the Coverage of Private Elementary-Secondary Schools	Steven Kaufman
96–27	Intersurvey Consistency in NCES Private School Surveys for 1993–94	Steven Kaufman
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2000–15	Feasibility Report: School-Level Finance Pretest, Private School Questionnaire	Stephen Broughman
Recent College Graduates (RCG)		
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
Schools and Staffing Survey (SASS)		
94–01	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94–02	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94–03	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94–04	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk

No.	Title	NCES contact
94–06	Six Papers on Teachers from the 1990–91 Schools and Staffing Survey and Other Related Surveys	Dan Kasprzyk
95–01	Schools and Staffing Survey: 1994 Papers Presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk
95–02	QED Estimates of the 1990–91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates	Dan Kasprzyk
95–03	Schools and Staffing Survey: 1990–91 SASS Cross-Questionnaire Analysis	Dan Kasprzyk
95–08	CCD Adjustment to the 1990–91 SASS: A Comparison of Estimates	Dan Kasprzyk
95–09	The Results of the 1993 Teacher List Validation Study (TLVS)	Dan Kasprzyk
95–10	The Results of the 1991–92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation	Dan Kasprzyk
95–11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
95–12	Rural Education Data User's Guide	Samuel Peng
95–14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
95–15	Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey	Sharon Bobbitt
95–16	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95–18	An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey	Dan Kasprzyk
96–01	Methodological Issues in the Study of Teachers' Careers: Critical Features of a Truly Longitudinal Study	Dan Kasprzyk
96–02	Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association	Dan Kasprzyk
96–05	Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey	Dan Kasprzyk
96–06	The Schools and Staffing Survey (SASS) for 1998–99: Design Recommendations to Inform Broad Education Policy	Dan Kasprzyk
96–07	Should SASS Measure Instructional Processes and Teacher Effectiveness?	Dan Kasprzyk
96–09	Making Data Relevant for Policy Discussions: Redesigning the School Administrator Questionnaire for the 1998–99 SASS	Dan Kasprzyk
96–10	1998–99 Schools and Staffing Survey: Issues Related to Survey Depth	Dan Kasprzyk
96–11	Towards an Organizational Database on America's Schools: A Proposal for the Future of SASS, with comments on School Reform, Governance, and Finance	Dan Kasprzyk
96–12	Predictors of Retention, Transfer, and Attrition of Special and General Education Teachers: Data from the 1989 Teacher Followup Survey	Dan Kasprzyk
96–15	Nested Structures: District-Level Data in the Schools and Staffing Survey	Dan Kasprzyk
96–23	Linking Student Data to SASS: Why, When, How	Dan Kasprzyk
96–24	National Assessments of Teacher Quality	Dan Kasprzyk
96–25	Measures of Inservice Professional Development: Suggested Items for the 1998–1999 Schools and Staffing Survey	Dan Kasprzyk
96–28	Student Learning, Teaching Quality, and Professional Development: Theoretical Linkages, Current Measurement, and Recommendations for Future Data Collection	Mary Rollefson
97–01	Selected Papers on Education Surveys: Papers Presented at the 1996 Meeting of the American Statistical Association	Dan Kasprzyk
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
97–10	Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993–94 School Year	Dan Kasprzyk
97–11	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97–12	Measuring School Reform: Recommendations for Future SASS Data Collection	Mary Rollefson
97–14	Optimal Choice of Periodicities for the Schools and Staffing Survey: Modeling and Analysis	Steven Kaufman
97–18	Improving the Mail Return Rates of SASS Surveys: A Review of the Literature	Steven Kaufman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
97–23	Further Cognitive Research on the Schools and Staffing Survey (SASS) Teacher Listing Form	Dan Kasprzyk
97–41	Selected Papers on the Schools and Staffing Survey: Papers Presented at the 1997 Meeting of the American Statistical Association	Steve Kaufman

No.	Title	NCES contact
97–42	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
97–44	Development of a SASS 1993–94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98–01	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
98–02	Response Variance in the 1993–94 Schools and Staffing Survey: A Reinterview Report	Steven Kaufman
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
98–05	SASS Documentation: 1993–94 SASS Student Sampling Problems; Solutions for Determining the Numerators for the SASS Private School (3B) Second-Stage Factors	Steven Kaufman
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
98–12	A Bootstrap Variance Estimator for Systematic PPS Sampling	Steven Kaufman
98–13	Response Variance in the 1994–95 Teacher Follow-up Survey	Steven Kaufman
98–14	Variance Estimation of Imputed Survey Data	Steven Kaufman
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
98–16	A Feasibility Study of Longitudinal Design for Schools and Staffing Survey	Stephen Broughman
1999–02	Tracking Secondary Use of the Schools and Staffing Survey Data: Preliminary Results	Dan Kasprzyk
1999–04	Measuring Teacher Qualifications	Dan Kasprzyk
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
1999–08	Measuring Classroom Instructional Processes: Using Survey and Case Study Fieldtest Results to Improve Item Construction	Dan Kasprzyk
1999–10	What Users Say About Schools and Staffing Survey Publications	Dan Kasprzyk
1999–12	1993–94 Schools and Staffing Survey: Data File User's Manual, Volume III: Public-Use Codebook	Kerry Gruber
1999–13	1993–94 Schools and Staffing Survey: Data File User's Manual, Volume IV: Bureau of Indian Affairs (BIA) Restricted-Use Codebook	Kerry Gruber
1999–14	1994–95 Teacher Followup Survey: Data File User's Manual, Restricted-Use Codebook	Kerry Gruber
1999–17	Secondary Use of the Schools and Staffing Survey Data	Susan Wiley
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2000–18	Feasibility Report: School-Level Finance Pretest, Public School District Questionnaire	Stephen Broughman
Third International Mathematics and Science Study (TIMSS)		
2001–01	Cross-National Variation in Educational Preparation for Adulthood: From Early Adolescence to Young Adulthood	Elvira Hausken
2001-07	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)	Arnold Goldstein

Listing of NCES Working Papers by Subject

No.	Title	NCES contact
Adult education		
96–14	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman
96–20	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96–22	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
98–03	Adult Education in the 1990s: A Report on the 1991 National Household Education Survey	Peter Stowe
98–10	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
1999–11	Data Sources on Lifelong Learning Available from the National Center for Education Statistics	Lisa Hudson
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Adult literacy—see Literacy of adults		
American Indian – education		
1999–13	1993–94 Schools and Staffing Survey: Data File User's Manual, Volume IV: Bureau of Indian Affairs (BIA) Restricted-Use Codebook	Kerry Gruber
Assessment/achievement		
95–12	Rural Education Data User's Guide	Samuel Peng
95–13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
97–29	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Larry Ogle
97–30	ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results	Larry Ogle
97–31	NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress	Larry Ogle
97–32	Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questions)	Larry Ogle
97–37	Optimal Rating Procedures and Methodology for NAEP Open-ended Items	Larry Ogle
97–44	Development of a SASS 1993–94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
2001-07	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)	Arnold Goldstein
Beginning students in postsecondary education		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
2001-04	Beginning Postsecondary Students Longitudinal Study: 1996-2001 (BPS:1996/2001) Field Test Methodology Report	Paula Knepper
Civic participation		
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
Climate of schools		
95–14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng

No.	Title	NCES contact
Cost of education indices		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
Course-taking		
95–12	Rural Education Data User's Guide	Samuel Peng
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
Crime		
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
Curriculum		
95–11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
Customer service		
1999–10	What Users Say About Schools and Staffing Survey Publications	Dan Kasprzyk
2000–02	Coordinating NCES Surveys: Options, Issues, Challenges, and Next Steps	Valena Plisko
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
Data quality		
97–13	Improving Data Quality in NCES: Database-to-Report Process	Susan Ahmed
Data warehouse		
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
Design effects		
2000–03	Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets	Ralph Lee
Dropout rates, high school		
95–07	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
Early childhood education		
96–20	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96–22	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
97–24	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97–36	Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research	Jerry West
1999–01	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West
2001–02	Measuring Father Involvement in Young Children's Lives: Recommendations for a Fatherhood Module for the ECLS-B	Jerry West
2001–03	Measures of Socio-Emotional Development in Middle School	Elvira Hausken
2001–06	Papers from the Early Childhood Longitudinal Studies Program: Presented at the 2001 AERA and SRCD Meetings	Jerry West

No.	Title	NCES contact
Educational attainment		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
Educational research		
2000–02	Coordinating NCES Surveys: Options, Issues, Challenges, and Next Steps	Valena Plisko
Employment		
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
2001–01	Cross-National Variation in Educational Preparation for Adulthood: From Early Adolescence to Young Adulthood	Elvira Hausken
Engineering		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
Faculty – higher education		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
Fathers – role in education		
2001–02	Measuring Father Involvement in Young Children's Lives: Recommendations for a Fatherhood Module for the ECLS-B	Jerry West
Finance – elementary and secondary schools		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
96–19	Assessment and Analysis of School-Level Expenditures	William J. Fowler, Jr.
98–01	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
1999–16	Measuring Resources in Education: From Accounting to the Resource Cost Model Approach	William J. Fowler, Jr.
2000–18	Feasibility Report: School-Level Finance Pretest, Public School District Questionnaire	Stephen Broughman
Finance – postsecondary		
97–27	Pilot Test of IPEDS Finance Survey	Peter Stowe
2000–14	IPEDS Finance Data Comparisons Under the 1997 Financial Accounting Standards for Private, Not-for-Profit Institutes: A Concept Paper	Peter Stowe
Finance – private schools		
95–17	Estimates of Expenditures for Private K–12 Schools	Stephen Broughman
96–16	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
2000–15	Feasibility Report: School-Level Finance Pretest, Private School Questionnaire	Stephen Broughman
Geography		
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
Graduate students		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico

No.	Title	NCES contact
Imputation		
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
Inflation		
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
Institution data		
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
Instructional resources and practices		
95–11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
1999–08	Measuring Classroom Instructional Processes: Using Survey and Case Study Field Test Results to Improve Item Construction	Dan Kasprzyk
International comparisons		
97–11	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97–16	International Education Expenditure Comparability Study: Final Report, Volume I	Shelley Burns
97–17	International Education Expenditure Comparability Study: Final Report, Volume II, Quantitative Analysis of Expenditure Comparability	Shelley Burns
2001–01	Cross-National Variation in Educational Preparation for Adulthood: From Early Adolescence to Young Adulthood	Elvira Hausken
2001-07	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)	Arnold Goldstein
Libraries		
94–07	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrol Kindel
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
Limited English Proficiency		
95–13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
Literacy of adults		
98–17	Developing the National Assessment of Adult Literacy: Recommendations from Stakeholders	Sheida White
1999–09a	1992 National Adult Literacy Survey: An Overview	Alex Sedlacek
1999–09b	1992 National Adult Literacy Survey: Sample Design	Alex Sedlacek
1999–09c	1992 National Adult Literacy Survey: Weighting and Population Estimates	Alex Sedlacek
1999–09d	1992 National Adult Literacy Survey: Development of the Survey Instruments	Alex Sedlacek
1999–09e	1992 National Adult Literacy Survey: Scaling and Proficiency Estimates	Alex Sedlacek
1999–09f	1992 National Adult Literacy Survey: Interpreting the Adult Literacy Scales and Literacy Levels	Alex Sedlacek
1999–09g	1992 National Adult Literacy Survey: Literacy Levels and the Response Probability Convention	Alex Sedlacek
1999–11	Data Sources on Lifelong Learning Available from the National Center for Education Statistics	Lisa Hudson
2000–05	Secondary Statistical Modeling With the National Assessment of Adult Literacy: Implications for the Design of the Background Questionnaire	Sheida White
2000–06	Using Telephone and Mail Surveys as a Supplement or Alternative to Door-to-Door Surveys in the Assessment of Adult Literacy	Sheida White
2000–07	“How Much Literacy is Enough?” Issues in Defining and Reporting Performance Standards for the National Assessment of Adult Literacy	Sheida White
2000–08	Evaluation of the 1992 NALS Background Survey Questionnaire: An Analysis of Uses with Recommendations for Revisions	Sheida White
2000–09	Demographic Changes and Literacy Development in a Decade	Sheida White

No.	Title	NCES contact
Literacy of adults – international		
97–33	Adult Literacy: An International Perspective	Marilyn Binkley
Mathematics		
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
1999–08	Measuring Classroom Instructional Processes: Using Survey and Case Study Field Test Results to Improve Item Construction	Dan Kasprzyk
2001-07	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)	Arnold Goldstein
Parental involvement in education		
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
1999–01	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West
2001–06	Papers from the Early Childhood Longitudinal Studies Program: Presented at the 2001 AERA and SRCD Meetings	Jerry West
Participation rates		
98–10	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
Postsecondary education		
1999–11	Data Sources on Lifelong Learning Available from the National Center for Education Statistics	Lisa Hudson
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Postsecondary education – persistence and attainment		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D’Amico
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D’Amico
Postsecondary education – staff		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
Principals		
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
Private schools		
96–16	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2000–15	Feasibility Report: School-Level Finance Pretest, Private School Questionnaire	Stephen Broughman
Projections of education statistics		
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D’Amico
Public school finance		
1999–16	Measuring Resources in Education: From Accounting to the Resource Cost Model Approach	William J. Fowler, Jr.

No.	Title	NCES contact
2000–18	Feasibility Report: School-Level Finance Pretest, Public School District Questionnaire	Stephen Broughman
Public schools		
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
98–01	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
1999–02	Tracking Secondary Use of the Schools and Staffing Survey Data: Preliminary Results	Dan Kasprzyk
2000–12	Coverage Evaluation of the 1994–95 Public Elementary/Secondary School Universe Survey	Beth Young
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
Public schools – secondary		
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
Reform, educational		
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
Response rates		
98–02	Response Variance in the 1993–94 Schools and Staffing Survey: A Reinterview Report	Steven Kaufman
School districts		
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
School districts, public		
98–07	Decennial Census School District Project Planning Report	Tai Phan
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
School districts, public – demographics of		
96–04	Census Mapping Project/School District Data Book	Tai Phan
Schools		
97–42	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
Schools – safety and discipline		
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
Science		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
2001-07	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)	Arnold Goldstein
Software evaluation		
2000–03	Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets	Ralph Lee

No.	Title	NCES contact
Staff		
97–42	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
Staff – higher education institutions		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
Staff – nonprofessional		
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
State		
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
Statistical methodology		
97–21	Statistics for Policymakers or Everything You Wanted to Know About Statistics But Thought You Could Never Understand	Susan Ahmed
Students with disabilities		
95–13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
Survey methodology		
96–17	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
97–15	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97–35	Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey	Kathryn Chandler
98–06	National Education Longitudinal Study of 1988 (NELS:88) Base Year through Second Follow-Up: Final Methodology Report	Ralph Lee
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
98–16	A Feasibility Study of Longitudinal Design for Schools and Staffing Survey	Stephen Broughman
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
1999–17	Secondary Use of the Schools and Staffing Survey Data	Susan Wiley
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
2000–02	Coordinating NCES Surveys: Options, Issues, Challenges, and Next Steps	Valena Plisko
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2000–12	Coverage Evaluation of the 1994–95 Public Elementary/Secondary School Universe Survey	Beth Young
2000–17	National Postsecondary Student Aid Study:2000 Field Test Methodology Report	Andrew G. Malizio
2001-04	Beginning Postsecondary Students Longitudinal Study: 1996-2001 (BPS:1996/2001) Field Test Methodology Report	Paula Knepper
2001-07	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)	Arnold Goldstein
Teachers		
98–13	Response Variance in the 1994–95 Teacher Follow-up Survey	Steven Kaufman
1999–14	1994–95 Teacher Followup Survey: Data File User's Manual, Restricted-Use Codebook	Kerry Gruber
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
Teachers – instructional practices of		
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
Teachers – opinions regarding safety		
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk

No.	Title	NCES contact
Teachers – performance evaluations		
1999–04	Measuring Teacher Qualifications	Dan Kasprzyk
Teachers – qualifications of		
1999–04	Measuring Teacher Qualifications	Dan Kasprzyk
Teachers – salaries of		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
Training		
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Variance estimation		
2000–03	Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets	Ralph Lee
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
Violence		
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
Vocational education		
95–12	Rural Education Data User's Guide	Samuel Peng
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson