NATIONAL CENTER FOR EDUCATION STATISTICS

Institute of Education Sciences

U.S. Department of Education
NCES 2007-049

Problem Solving in the PISA and TIMSS 2003 Assessments

Technical Report

December 2006

John A. Dossey
Sharon S. McCrone
Illinois State University

Christine O'Sullivan
K-12 Consulting

Patrick Gonzales
Project Officer
National Center for Education Statistics

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Mark Schneider
Commissioner

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

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

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

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

December 2006

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.

This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES World Wide Web Electronic Catalog address shown above.

Suggested Citation

Dossey, J.A., McCrone, S.A., and O'Sullivan, C. (2006). *Problem Solving in the PISA and TIMSS 2003 Assessments* (NCES 2007-049). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from http://nces.ed.gov/pubsearch.

Content Contact
Patrick Gonzales
(415) 920-9229
patrick.gonzales@ed.gov

Acknowledgments

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

Thanks to Patrick Gonzales of the National Center for Education Statistics (NCES) and Mariann Lemke, formerly with NCES, now with Chicago Public Schools, for their input on the design of, and role in reviewing, the report. The authors are also grateful to the following reviewers at NCES: Jack Buckley, Shelley Burns, Valena Plisko, Taslima Rahman, Marilyn Seastrom, and Elois Scott; and at the Education Statistics Services Institute: Rachel Dinkes, Todd Thomas, and Jed Tank. In addition, the authors wish to acknowledge the insightful contributions of Ina Mullis (Boston College) and Barry McCrae (Australian Council for Education Research).

Several individuals with the Education Statistics Services Institute (ESSI) provided support in the form of research assistance and/or review: Gillian Hampden-Thompson, Anindita Sen, and Jamie Johnston.

For permission to use secure items and publish released items, the authors would like to thank the International Association for the Evaluation of Educational Achievement (IEA) and the Organization for Economic Cooperation and Development (OECD).

Executive Summary

In 2003, two major international assessments of student learning were conducted in the United States. The Trends in International Mathematics and Science Study (TIMSS) is administered under the auspices of the International Association for the Evaluation of Educational Achievement (IEA) to measure trends in the performance of grade 4 and 8 students in school mathematics and science in participating countries. The Program for International Student Assessment (PISA) is an international assessment administered by the Organization for Economic Cooperation and Development (OECD) to 15-year-old students around the world that emphasizes mastery of processes, understanding of concepts, and the ability to function in various situations within the domains of reading literacy, mathematical literacy, and scientific literacy. Both studies were carried out in the United States by the National Center for Education Statistics (NCES), part of the Institute of Education Sciences in the U.S. Department of Education.

In addition to measuring student capabilities in mathematics and science, both assessments included a special focus on problem solving in 2003. TIMSS 2003 added special sets of questions (problem-solving and inquiry, or PSI, questions) within its mathematics and science assessments to probe student problem solving in greater depth. PISA 2003 created a separate special study on cross-disciplinary (C-D) problem solving as part of its 2003 assessment. PISA's items focused on students' problem-solving capabilities in settings that transcended normal curricular content boundaries. TIMSS and PISA both contained aspects of problem solving in the assessment frameworks that guided the development of their general assessments of mathematics and science. The PISA assessment had a separate framework focusing entirely on problem solving to guide its added special assessment of problem solving in cross-disciplinary settings. In this report, the mathematics tasks containing problem solving from the general TIMSS and PISA assessments are analyzed and compared. In a like manner, the science tasks containing problem solving from the two assessments are analyzed and compared. Finally, an analysis and comparison are made of the problem-solving items from the special PISA C-D problem-solving assessment with subsets of the TIMSS mathematics and science items that were identified as PSI items.

While both assessments measured student performance, their goals in doing so differed. TIMSS 2003 focused on what students achieved as a result of what they had studied in school. PISA 2003 focused on how well students could use their knowledge and skills when faced with a problem in a real-life context. These differing approaches are reflected in the way problem solving was incorporated into the general mathematics and science sections of each assessment and into the special C-D problem-solving section of the PISA assessment.

Given that problem solving played an important role in each assessment, NCES commissioned this review of the problem-solving aspects of each study in order to compare and contrast the nature of problem solving in each assessment. Based on an expert review of the assessments, their respective assessment frameworks, and their items, this report analyzes

- how problem solving was actualized in grade 8 TIMSS 2003 and PISA 2003,[1]

[1] Grade 4 TIMSS items were not included in the analyses. It was considered more appropriate to compare the items from PISA and grade 8 TIMSS because of the relative closeness in the target age/grade.

- the ways in which these assessments' items measured students' capabilities to solve problems, and
- how these similarities and differences in approach may relate to the interpretation of results of these two assessments.

The report's authors develop and use a definition for problem solving to identify items in the two assessments that address students' problem-solving capabilities. For this study, an item was considered to measure problem solving if a student was not likely to have a known strategy to immediately apply that would lead to a correct answer. The authors' broad range of experiences with curricula, assessments, and classrooms for the age and grade levels assessed contributed to the identification of the problem-solving items.

Based on this definition, a significantly greater amount (53 percent) of PISA 2003 items (including mathematics literacy items, science literacy items, and special cross-disciplinary problem-solving items) were found to measure problem solving than the (32 percent) of TIMSS 2003 items (including mathematics and science items) so coded.[2]

Analyses were conducted by content area (mathematics and science) to compare the characteristics of problem-solving items within each content area. For example, what topics do problem-solving items in mathematics cover in TIMSS compared to PISA? Are problem-solving items in science more likely to require interpretations of figures in TIMSS or in PISA? This analysis provides in-depth information about the distribution and nature of the problem-solving items included in each content area of the assessment. In addition, the PSI items that were added to the mathematics and science assessments in TIMSS to examine students' problem-solving and inquiry skills were isolated as a set of TIMSS items and compared with the items contained in the PISA special assessment on C-D problem solving. While the PSI items from the TIMSS assessment were included in the foregoing comparisons of the mathematics and science assessments, the PISA C-D problem-solving items were not analyzed as part of the foregoing mathematics and science comparisons.

Items that were identified as problem-solving items in the TIMSS and PISA mathematics, science, and C-D problem-solving assessments were analyzed in terms of six types of item characteristics: (1) content coverage; (2) cognitive processes; (3) problem-solving attributes; (4) item formats; (5) computational aspects; and (6) translation of representations.

Findings of Mathematics Assessment Comparisons

Based on the definition of problem solving used in this report, 74 of the 194 items (38 percent) in the TIMSS 2003 mathematics assessment and 41 of the 85 items (48 percent) in the PISA 2003 mathematical literacy assessment were identified as problem-solving items.

---

[2] While these percentages deal with the assessments as a whole, readers should be aware that many comparisons between subsets of items drawn from the respective assessments in this report involve disproportionate numbers of items, which sometimes are quite small. Thus, a higher percentage of items does not always correspond to a greater number of items. It is for this reason that numbers, percentages, and their statistical significance are presented in the data tables and text supporting the analyses.

The two assessments were not found to differ statistically in the percentages of items distributed across the mathematics content areas. A mapping the PISA items onto the TIMSS framework, combined with a comparison of the relative weighting of item content within the PISA and TIMSS assessments was completed. Though TIMSS appeared to have a larger absolute percentage of items dealing with *geometry* and *measurement*, and the PISA assessment appeared to have a larger absolute percentage of items focusing on problem solving in *algebra* and *data*, these differences were not found to be significant. Though not significant, the distribution of the problem-solving items among the content areas in the mathematics portions of the two assessments appear to mirror the overall differences in emphases found in comparison to the mathematics content in the National Assessment of Educational Progress (NAEP) (Neidorf et al. 2006).[3]

A comparison of the clusters of cognitive competencies associated with the items in mathematics appeared to indicate that the TIMSS items identified as problem solving focused more on students *using concepts* and *solving routine problems* (combined) and on *reasoning* than the PISA problem-solving items. However, there were no significant differences found in the proportion of items assigned to the varied competency clusters. Moreover, a comparison of the attributes related to the PISA and TIMSS mathematics items identified as problem solving indicated that a larger percentage of the TIMSS items required students to *identify variables or relationships* than the PISA items.

TIMSS 2003 mathematics items identified as problem solving were also found to require more *drawing or sketching* by students than the PISA 2003 mathematics items, while a larger percentage of the PISA items than the TIMSS items were found to require students to *interpret statistical representations*. Finally, it was also found that, whereas there was a larger percentage of TIMSS mathematics items identified as problem solving designed in a multiple choice format, there was a larger percentage of PISA mathematics that were designed as closed short constructed response items.

Findings of Science Assessment Comparisons

In science, 49 of the 189 items (26 percent) in the TIMSS 2003 science assessment and 17 of the 35 items (49 percent) in the PISA 2003 scientific literacy assessment were identified as problem-solving items. The difference in the total numbers of items in TIMSS science (189) and PISA science (35) reflects the fact that scientific literacy was a minor domain in the PISA assessment in 2003. Hence, comparisons between the science assessments should be made carefully. While this difference in proportions significantly favored the PISA assessment from the problem solving perspective, there were no differences detected in the proportions of items allocated to the subject matter content areas across the two assessments (except in the number of items in sets; table 8).

In both the TIMSS and PISA 2003 science assessments, a little over a quarter of the items identified as problem-solving were classified in the content category of *life science* (27 and 29

---

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

percent, respectively). Nearly half of the PISA science problem-solving items focused on *earth and environmental science*. TIMSS science problem-solving items were generally spread across the categories of *life science*, *chemistry*, *physics*, and *environmental science*, with the smallest percentage focused on *earth science*. Though the results appeared to indicate that problem-solving items that covered *chemistry* and *physics* were more prevalent in TIMSS than in PISA (40 vs. 24 percent), this was not found to be a significant difference.

When the cognitive processes required by the science problem-solving items were examined, it was found that the percentage of items addressing the corresponding cognitive domains and competency clusters in TIMSS and PISA was relatively similar across the two assessments and that none of the items considered as problem solving were classified in the TIMSS *factual knowledge* domain.[4] In both TIMSS and PISA, at least 80 percent of the science problem-solving items required students to *identify variables or relationships*. Moreover, there was a significantly larger percentage of PISA science items identified as problem solving than TIMSS items that required students to *critically evaluate information*, while there was a significantly larger percentage of TIMSS science items identified as problem solving than PISA items that *required science knowledge*. Comparisons of the science problem-solving items found no significant differences detected between TIMSS and PISA in the distribution of skills required of students to complete the items, or item formats.

Findings of the Comparison of PISA Special Study Items With TIMSS PSI Items

Twenty-three (70 percent) of the 33 TIMSS PSI items that were embedded in the TIMSS 2003 mathematics and science assessments, and 15 (79 percent) of the 19 items in the special PISA 2003 C-D problem-solving assessment, were identified as problem-solving items, based on the definition of problem solving used in this report.[5]

A higher percentage of the TIMSS PSI items than PISA items identified as problem solving required students to *identify variables or relationships*. The PISA assessment placed significantly greater emphasis than TIMSS on items requiring the *interpretation of information from a reading passage* (87 vs. 35 percent), while a larger percentage of the TIMSS PSI items than the PSI C-D items were designed with open short constructed response formats.

Summary

While it is possible to compare and contrast the problem-solving items contained in each assessment, the findings reflect the goals of the assessment programs as defined by the frameworks developed by the sponsoring organizations for each assessment. That is, the TIMSS 2003 assessment items identified as problem solving tend to focus on students' knowledge and ability to perform particular skills or procedures taught in school curricula, while the PISA 2003 assessment

---

[4] PISA categorizes cognitive processes into competency clusters. These clusters are composed of varied combinations of competency levels or skills necessary for mathematics, science, and problem-solving tasks.
[5] The other four PISA C-D items measured whether students *understand* a problem, which is the first step toward problem solving. These four items are the first items in the stimulus materials for a C-D set of items. The other 10 TIMSS PSI items, like the PISA items, deal either with assessing the understanding of a problem or performing a calculation to provide a basis for attacking or addressing the problem in a related follow-up item.

items identified as problem solving tend to focus on broader interpretive and application outcomes associated with literacy objectives. The specific analyses of problem-solving items by content coverage, cognitive processes, item formats employed, and problem-solving attributes provide evidence of these tendencies. Where the purposes of the PISA and TIMSS assessments were somewhat more similar—in the problem-solving items in the special studies areas (TIMSS PSI and PISA C-D)—some differences existed in terms of item format and the role of the problem-solving attributes.

The analysis of the problem-solving items in these two assessments, especially those in the areas of TIMSS PSI and PISA C-D, indicates a need for further research on the role of students' reading skills in the measurement of problem-solving performance. In both the TIMSS and PISA assessments, items were presented with opening passages presenting a context for items that follow. Although such an investigation is beyond the scope of this report, it is important to note that students' reading capabilities may be an important factor in the measurement of problem-solving abilities.

Contents

Contents (cont.)

Page

List of Tables

Table **Page**

List of Exhibits

Exhibit **Page**

Section I: Introduction

When examining the outcomes of education at local, state, national, or international levels, one of the major concerns of educators is whether students are able to employ the knowledge and skills they have acquired in formal schooling and through daily living experiences to solve problems. Students' capabilities to solve problems are necessary not only for the demands of everyday life—personal, social, and public decisionmaking—but also for their future careers and their ability to continue learning in formal education settings.

The purpose of this report is to compare and contrast features of the problem-solving tasks found in the 2003 Trends in International Mathematics and Science Study (TIMSS) and the 2003 Program for International Student Assessment (PISA). The portions of the TIMSS assessment analyzed for this study were the general mathematics and science assessments for grade 8 students. The PISA assessment involved a single population: 15-year-olds. Thus, the portions of the PISA assessment analyzed were the general mathematics and science assessment, plus the items in a separate study of on cross-disciplinary (C-D) problem solving as part of its 2003 assessment. Section I contains the definition of problem solving and an outline of how problem solving is assessed. Section II contains an overview of the TIMSS and PISA assessment frameworks with a special focus on aspects that emphasize or are related to problem solving. This section also contains a summary of the analysis process used to compare problem solving in TIMSS and PISA. More details on the methods used in this study can be found in appendix F. Sections III and IV contain the comparisons of problem solving in the TIMSS and PISA general mathematics and science assessments, while section V contains a comparison of the items developed for the separate PISA study on cross-disciplinary (C-D) problem solving and subsets of items from the TIMSS mathematics and science assessments that focus on problem-solving and inquiry (PSI) skills. Section VI contains a summary of the findings related to these comparisons and a brief discussion of additional research that would further an understanding of student problem solving and its assessment.

What Is Problem Solving?

If one were to pose the question of what constitutes problem solving to the general population, the answers would probably include some variation of solving word problems in school mathematics, completing puzzles commonly found in popular magazines, or resolving real-life situations (such as how to clear a blocked drain). These answers are all valid. However, in order to more accurately define the nature of problem solving and to describe what steps a person must take in order to solve a problem, a finer and more comprehensive definition is required of what constitutes a *problem* together with an understanding of what activities constitute *problem-solving behaviors.*

Considerable thought, writing, and research exist in an attempt to define and better understand the act of problem solving in mathematics and science (Henderson and Pingry 1953; Polya 1945, 1962, 1965). Education researchers and curriculum specialists have also worked to understand the learner's capabilities to solve problems and to promote problem solving in mathematics and science classrooms (Charles and Silver 1988; Kilpatrick, Swafford, and Findell 2001; Lester 1980; Schoenfeld 1992; Silver 1985).

A set of common factors defining problem solving appears through the literature on the subject (see Mayer 1985; Case 1985; Resnick and Ford 1981; Bransford, Brown, and Cocking 1999; and English 2002 for reviews of the literature). These commonly cited factors are the role of the knowledge base; the existence of strategies and the learner's ability to apply strategic knowledge; the role of monitoring and control employed by the individual while engaged with the problem situation; the role of individuals' beliefs and attitudes and their regulation of the willingness to engage problems; and the individual's use of cognitive practices during problem solving. Although not all of these factors were taken into consideration for this analysis of problem solving in international assessments, criteria to identify problem-solving items were developed based upon this literature.

The key to determining when a situation is a *problem* is the development of a working definition. This, in turn, allows for items to be coded to compare and contrast the two assessments of interest. When faced with an item calling for some sort of resolution, one might ask, "Is it possible to…?" or "How could one…?" Question posing of this sort is the first sign that the situation is a problem, although the presence of a question is not sufficient alone to indicate that a problem exists. Indeed, what may be a problem for one person may not be a problem for another person. The key to determining if a given situation is a problem is to see how an individual reacts to the situation. If the individual assumes responsibility for trying to address the questions evoked by the situation, and then sets the goal of trying to resolve the situation, the possibility of the situation being a problem exists. Next, the individual must search for or develop a strategy to resolve the situation. If no strategy is easily found, the situation constitutes a problem. However, if the strategy the individual would normally employ in similar settings works in the given setting, the situation is not considered a problem, but merely an *exercise*. That is to say, an *exercise* is a situation in which the individual is familiar with the knowledge or tools needed for resolution of the situation and is able to apply that knowledge. While some exercises may be more difficult than others to resolve, the individual's attempts to resolve them are not significantly impeded.

Thus, *problem-solving* strategies are required in situations where an individual's known attempts or ideas for resolving a situation do not work. In these cases, the individual must consider new vantage points or simplify the problem to a workable one. The behavior of the individual and the nature of the approaches used by that individual provide evidence that he or she is working on a *problem*. *Exercises* are tasks that call for students to exhibit knowledge and skills in a manner and setting that students have practiced. When given an *exercise*, students know exactly which procedures apply and how to proceed. When given a *problem*, the knowledge and skills needed are not immediately clear to students. One can ask how skillfully students answer such items, as skill is a measure of their ability to remember and reapply a previously learned method of deriving answers to exercises. Illustrative examples of items containing exercises and problems are shown in appendix A.

For the review of the TIMSS and PISA assessments, a *problem* exists when

- the context allows students to be engaged,
- students do not have a known strategy to immediately apply, and
- the situation calls for a solution.

Not all assessment items can be uniformly classified as problems or nonproblems, as noted earlier. Items must be considered in terms of intended grade/age levels and the content or experiences related to the problem students have likely experienced in school or life settings.

How Can Problem Solving Be Assessed?

Earlier examinations of work on assessing students' problem-solving capabilities identify various important components of assessments, such as ways of addressing inductive and analogical reasoning (Csapó 1997; Vosniadou and Ortony 1989) or methods of addressing varying levels of cognitive complexity (Bloom, Hasting, and Madaus 1971; Collis, Romberg, and Jurdak 1986). Richard Mayer (1992) argued that problem-solving assessments must require the problem solver to engage in higher order thinking (or cognitive) processes by presenting tasks that require the invention of novel solution strategies. These recommendations suggest that authentic tests of problem solving must include a significant number of items that require student-produced responses. However, just because an item is an open response item does not guarantee that it measures problem solving. An item asking for the sum of a set of numbers could be a constructed response item, but still just an exercise. Items that measure problem solving should acquire samples of students' thinking and actions or provide evidence of students' prior knowledge and their ability to integrate concepts, representations, and processes. However, many multiple-choice questions still contain problems and elicit problem-solving skills from students. Most individuals involved in the study of problem solving in research or practice-based settings agree that the major focus in assessing and describing student problem solving is that of describing the cognitive acts students make in addressing, solving, and reporting solutions to problems (Pellegrino, Chudowsky, and Glaser 2001).

In order to carry out an evaluation of the TIMSS and PISA 2003 assessment items and the role that problem-solving tasks play in them, this report first describes each assessment's framework and the role problem solving plays in the framework. The mathematics tasks containing problem solving from the general TIMSS and PISA assessments are then analyzed and compared. In a like manner, the science tasks containing problem solving from the two assessments are analyzed and compared. Finally, an analysis and comparison are made of the problem-solving items from the special PISA C-D problem-solving assessment with subsets of the TIMSS mathematics and science items that were identified as PSI items.

Section II: The TIMSS and PISA Assessments

PISA is an international assessment program administered by the Organization for Economic Cooperation and Development (OECD) to measure 15-year-old students' mastery of processes, understanding of concepts, and levels of reading literacy, mathematical literacy, and scientific literacy. PISA was first administered in 2000 and is currently being readministered every 3 years with the focal point of the assessment rotating between reading literacy (2000, 2009,…), mathematical literacy (2003, 2012,…), and scientific literacy (2006, 2015,…).[1] Forty-one countries participated in PISA 2003. PISA's approach to literacy assessment involves tasks that assess knowledge and skills required to meet real-life challenges, rather than sampling topics students master as part of their study of specific subject matter in their school's curriculum. In PISA 2003, problem solving was directly incorporated into the general mathematical and scientific literacy assessments, as well as featured in a special cross-disciplinary (C-D) problem-solving study separate from the other literacy assessments (OECD 2003). Students were to answer questions designed specifically to measure their problem-solving skills. These C-D questions appeared in a section of the PISA assessment that was separate from the general mathematical and scientific literacy sections that contained items measuring mathematical or scientific literacy.

TIMSS is conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA) to measure trends in the performance of fourth- and eighth-grade students in content found in school mathematics and science in participating countries. Within TIMSS, there is a particular emphasis on examining the links between the curriculum intended for students, the curriculum implemented in classrooms, and the curriculum expectations attained by students. The assessment results serve as measures of the attained curriculum. The first TIMSS was administered in 1995; the second in 1999; and the most recent in 2003. Some 45 countries participated in TIMSS 2003 at the eighth-grade level. In TIMSS 2003, problem solving was directly incorporated into the mathematics and science assessments (Mullis et al. 2003) through traditional problem-solving items and special sets of items contained in problem-solving and inquiry (PSI) units. These PSI units were sets of related questions incorporated into the general mathematics and science assessments. In order to make as relevant as possible comparisons to PISA findings, which assesses 15-year-olds, this report considers only results from the TIMSS eighth-grade assessment (students who were, on average, between 13 and 14 years old).

As described in section I, the definition of problem solving in this report contains three conditions: a *problem* exists when (1) the student is engaged; (2) the student does not have a known strategy to apply in order to answer the item; and (3) the situation (i.e., the item) calls for a solution. Two methods were employed in order to determine whether the assessment items from the TIMSS and PISA assessments met this definition of problem solving. First, the field-test data for the relevant portions of the TIMSS and PISA assessments were examined to see if the same proportions of students were answering questions identified as problem-solving items as were answering items not judged to be measuring problem solving. The results indicate that the students had both the motivation and testing time to work on and solve the items identified

[1] Note that scientific literacy was not a major domain in the PISA 2003 assessment. There were only 35 items included to measure trends in this area. A major assessment of scientific literacy in PISA takes place in 2006.

by the authors as problem solving. The response rates and student performance evidence suggest that students were engaged with the TIMSS and PISA assessment items. Second, the items contained in the relevant sections of the TIMSS and PISA assessments were examined from the perspective of whether students had known strategies to apply upon first reading the items or whether the items presented problems (i.e., they were not immediately solvable). The following section describes how problem solving was incorporated into the TIMSS and PISA assessment frameworks and what characteristics identify a problem-solving item.

Assessment Frameworks

New frameworks were created by the sponsoring organizations for both the TIMSS and PISA 2003 assessments (Mullis et al. 2003; OECD 2003). Each framework served as a blueprint for the construction of new assessment items. These frameworks guided the selection of old items and the development of new items to measure and describe student performance and trends in student capabilities. Unreleased, secure items from previous assessments were selected to assist in measuring trends in student performance over time. New items were added to replace items released with results from earlier assessments. In addition, both assessment programs incorporated new items specifically designed to measure problem solving. TIMSS added PSI items embedded within the mathematics and science assessments. PISA added a broad range of new mathematics items, as PISA 2003 marked the first time mathematics had been the major domain in a PISA assessment. In addition, PISA created a separate subtest of C-D problem-solving items designed to assess problem solving in contexts that fell outside the boundaries of a single curricular area. These C-D items were presented to students in separate test blocks rotated among the other mathematics and science item blocks. The mathematics, science, reading, and C-D items were not identified as such in the assessment booklets, however.

As detailed in exhibit 1, the TIMSS 2003 mathematics and science assessments were organized around *content* and *cognitive process domains.* The content domains can generally be understood as dealing with subject-matter content (facts, concepts, theories,…), while cognitive domains can be seen as sets of cognitive processing behaviors (knowing, identifying, applying, solving,…) expected of students engaged in mathematics and science tasks.

The features of the PISA assessment frameworks are shown in exhibit 1 in a similar way. The first organizing feature is *context/setting*. This feature is not shown in exhibit 1; rather, it is described below. The second feature is *mathematical literacy content areas* and *scientific literacy themes.* This feature is comparable to the content domains for the TIMSS assessment. The third organizing feature of the PISA framework is referred to as *process competency clusters* in the mathematics and science frameworks. These are comparable to the *cognitive domains* for TIMSS. More detailed overviews can be found in the PISA 2003 and TIMSS assessment frameworks (OECD 2003; Mullis et al. 2003).

Because of PISA's specific focus on student ability to perform real-life tasks, the *context,* or setting, of each particular item is an important element of PISA's organizational structure and central to the OECD's concept of literacy:

> *In PISA, literacy is regarded as knowledge and skills for adult life … The three [major PISA assessments of mathematical literacy, reading literacy, and scientific literacy] therefore*

> *emphasize the ability to undertake a number of fundamental processes in a range of situations, backed by a broad understanding of key concepts, rather than the possession of specific knowledge.* (OECD 2000, p. 7)

As a result, care was given in considering the context, or setting, in which students might encounter mathematics, science, or problem solving in their lives. The PISA contexts for mathematics are *personal; social, work, and leisure; community and society;* and *scientific*. The TIMSS frameworks have no comparable categorization method to PISA's categorization of context/setting discussed here. While the PISA program coded each of its items with a context/setting category, little is done with this in structuring the assessments or in the analysis of student performance.

Mathematics

TIMSS 2003 grade 8 mathematics items are categorized within the following content domains: *number, measurement, geometry*, *data,* and *algebra.* These content domains link well with the PISA content areas (exhibit 1). They also link well with the National Assessment of Educational Progress (NAEP) categories, which facilitates the interpretation of results with respect to U.S. curricular patterns and trends (see Neidorf et al. 2006; see also National Assessment Governing Board [NAGB] 1999, 2002, 2004).

TIMSS 2003 grade 8 mathematics items are categorized with respect to the following four cognitive domains: *knowing facts and procedures, using concepts, solving routine problems,* and *reasoning*. The cognitive domain of *reasoning* applies to nonroutine situations that call for the integration of varied types of mathematical knowledge and skills, along with critical thinking. While "reasoning" is commonly associated with abstract mathematics and the solution of puzzle-type problems, proof and reasoning also occur in problem-solving situations such as in programming a VCR. The resolution of such a real-life problem calls for an analysis of the situation, the conjecturing of what may be the cause of the problem, an evaluation of the structure of the VCR system itself, the selection of an appropriate tool or method to deal with the problem, the application of that tool or method, and the analysis of the progress or outcome in light of the actions taken. Thus, of the TIMSS mathematical cognitive processes, *reasoning* comes closest to satisfying the problem-solving definition given in this report.

Similar to the TIMSS categorization of *content domains*, PISA categorizes items within *content areas*. However, PISA content areas are different from TIMSS content domains in that they are less comparable to curricular subject areas. The PISA mathematics content structure employs four overarching ideas that categorize mathematical content along broader concepts than those tied to more common curricular topics. These content areas were selected more for their value in assessing literacy as opposed to being a common denominator of national curricula or for serving as a base for advanced study of mathematics in higher levels of education. Consistent with the overall focus of the PISA assessments, the purpose was not to test the depth of students' curricular-related understanding, but to assess students' capabilities of using the cognitive processes.

Exhibit 1. Content and cognitive domains included in the TIMSS 2003 and PISA 2003 frameworks

TIMSS framework	PISA framework
Mathematics content domains	**Mathematical literacy content areas**
Number	Quantity
Measurement Geometry	Shape and space
Algebra	Change and relationships
Data	Uncertainty
Mathematical cognitive process domains	**Mathematical process competency clusters**
Knowing facts and procedures	Reproduction
Using concepts	Connections
Solving routine problems	Reflection
Reasoning	
Science content domains	**Scientific literacy themes**
Life science	Form and function Human Biology Physiological change Biodiversity Genetic control
Chemistry	Chemical and physical changes Structure and properties of matter
Physics	Forces and movement Energy transformations
Earth science	Earth and its place in the universe Geographical change
Environmental science	Atmospheric change Ecosystems
Science cognitive process domains	**Scientific process competency clusters**
Factual knowledge	Describing, explaining, and predicting scientific phenomena
Conceptual understanding	Understanding scientific investigation
Reasoning and analysis	Interpreting scientific evidence and conclusions

NOTE: Content and process categories are ordered and comparable categories shaded to indicate similar categories in the two assessments. Most, if not all, of the scientific literacy themes in the PISA 2003 framework are subsumed under the broader TIMSS 2003 science content domains. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; this table pertains to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.
SOURCE: Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J., and O'Connor, K.M. (2003). *TIMSS Assessment Frameworks and Specifications: 2003* (2nd ed.). Chestnut Hill, MA: International Study Center, Boston College. Organization for Economic Cooperation and Development. (2003). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving.* Paris, France: OECD.

The four overarching ideas for mathematics content in PISA are *quantity, space and shape, uncertainty,* and *change and relationships* (OECD 2003; Steen 1990). *Quantity* deals with number sense, estimation, and basic computations and operations. *Space and shape* includes those topics where recognition and use of shape and form play a major role. *Uncertainty* includes those topics normally considered in the domains of probability and statistics, or chance. *Change and relationships* includes pattern finding and related algebraic notions such as how one term or a glimpse of a pattern changes into the next term or pattern.

PISA categorizes cognitive processes into *competency* clusters. These clusters are composed of varied combinations of competency levels or skills necessary for mathematics, science, and problem-solving tasks. The three competency clusters in the PISA mathematical literacy framework build on the work of northern European mathematics educators, particularly that of Danish mathematics educator Mogens Niss (Niss 1999; Neubrand et al. 2001). The *reproduction* cluster applies to the reproduction of practical knowledge in dealing with very common demands on mathematical knowledge, such as those found in school, and everyday applications in personal life, such as recalling facts and performing routine calculations. The *connections* cluster focuses on problem solving in nonroutine, but familiar settings. Assessment items in this cluster call for students to integrate, connect, and make slight extensions of practiced knowledge and skills. The *reflection* cluster focuses on students' capabilities to reflect on solution strategies or use them in settings that call for more innovative approaches than those the student has typically practiced. Assessment items associated with the *reflection* cluster call for advanced reasoning, argumentation, abstraction, generalization, and model building.

Both TIMSS and PISA include the concept of problem solving in their 2003 mathematics frameworks. While TIMSS has no explicit problem-solving framework, the mathematics cognitive domains of *solving routine problems* and *reasoning* include the term "problem solving" in their descriptions. However, in the former domain, the role of problem solving is relegated to situations usually referred to as exercises: that is, the use of problem-like contexts to provide additional opportunities for students to work with practiced concepts, strategies, and problem-solving skills. In most of these cases, the focus is on consolidating knowledge and skills while helping students correctly recognize when and how to use them appropriately and effectively. A few of the items so classified include problem solving (e.g., when students have had little opportunity to learn the content or when the context involved requires students to make a number of connections in order to use their knowledge to resolve the problem presented). Many of the items in TIMSS in the domain of reasoning call on students to apply cognitive processes and skills to solve nonroutine problems. The science framework for the TIMSS assessment contains a brief description of the role of scientific inquiry, but does not elaborate on this area as it does for its content or cognitive domains.

The PISA mathematical literacy framework includes an examination of problem solving in its *connections* and *reflections* competency clusters. While the *reproduction* cluster mentions problem solving, it does so in the consideration of review or practice problems, as in the TIMSS domain of *solving routine problems*. In the *connections* competency cluster, the PISA framework includes situations where students have to frame and structure solution strategies for problems that reside at the edge of routine and nonroutine settings. Items classified as measuring *reflection* processes are most likely problems, according to the definition used in this report.[2]

Science

TIMSS 2003 grade 8 science items are categorized within the following content domains: *life science, chemistry, physics, earth science,* and *environmental science*. As with mathematics,

[2] One has to consider the opportunity-to-learn factor, which differs across nations, states, and local school districts. If a student has sufficient practice with a setting or particular combinations of concepts and skills, he or she may make routine the processes involved in dealing with such a situation in the *reflection* cluster.

these align well with the NAEP framework for grade 8 (NAGB 1999), with one major exception: the TIMSS 2003 framework includes more chemistry concepts than the corresponding NAEP framework (Neidorf, Binkley, and Stephens 2006).

The TIMSS 2003 grade 8 cognitive domains for science are *factual knowledge, conceptual understanding*, and *reasoning and analysis* (Mullis et al. 2003). The *reasoning and analysis* domain focuses on students' abilities to operate at a more abstract level, utilizing cognitive processes to critically reason through problems and hypothetical issues. These cognitive processes include the ability to collect and analyze data, plan experiments, and solve problems. Such actions call for students to integrate and synthesize information, form hypotheses, abstract patterns, and generalize from the results. Perhaps most important to the conduct of science is the process of controlling a variety of factors while simultaneously drawing conclusions based on the data available and justifying these conclusions. *Reasoning and analysis* calls on students to consider the available information and the desired outcome and to use reasoning and analysis to connect the known knowledge with the desired outcomes.

Like the PISA mathematical literacy framework, the 2003 scientific literacy framework in PISA considers the contexts, or settings, in which science takes place in students' lives. The science contexts are *life and health*, *earth and environment*, and *technology*. TIMSS does not have a feature in its framework linking items to the contexts in which they may appear in students' lives. While the PISA program coded each of its items with a context/setting category, little is done with this in the analysis of student performance or structuring of the assessments themselves.

The application of the definition of scientific literacy leads to a listing of 13 scientific literacy themes that bound the science items found in the PISA 2003 assessment instruments. These literacy themes are

- *form and function;*
- *human biology;*
- *physiological change;*
- *biodiversity;*
- *genetic control;*
- *chemical and physical changes;*
- *structure and properties of matter;*
- *forces and movement;*
- *energy transformations;*
- *the earth and its place in the universe;*
- *atmospheric change;*
- *geographical change;* and
- *ecosystems.*

While these scientific literacy themes are named by recognizable content categories, the PISA assessment examines such knowledge as it appears in settings that relate to the context areas mentioned above. It should be noted that the PISA 2003 science framework was revised for the 2006 assessment, in which scientific literacy is the major domain to be assessed.

In addition to the content themes, PISA 2003's science framework, like the mathematics framework, includes three science competency clusters. Like the corresponding competency clusters in mathematics, the PISA science competency clusters focus on decisionmaking and analysis themes that can easily be related to "doing science" in a real-life setting. The first cluster is *describing, explaining, and predicting scientific phenomena.* Student cognition associated with this cluster relates to describing or explaining scientific events or to predicting outcomes to science-related situations. Definitions, properties, and scientific principles are the major focus of activities associated with this process. The second cluster is *understanding scientific investigation.* Student cognition associated with this second process focuses on recognizing and communicating aspects of questions investigated using scientific inquiry. Assessment items associated with this process do not depend on a high level of knowledge of scientific definitions or principles, but rather on knowing how they are applied. The third cluster is *interpreting scientific evidence and conclusions.* Student cognition associated with this process relates to capabilities associated with developing and communicating findings from scientific investigations. Assessment items associated with this process are intended to determine students' ability to draw conclusions and communicate outcomes from scientific experiments or evaluate observations and discuss whether they support a generalization.

Table 1 provides a quantitative comparison of the various features of the TIMSS and PISA mathematics and science assessments, as well as of the PISA C-D assessment.

Problem Solving

The final PISA 2003 framework of interest is that for C-D problem solving. This framework guides the special study on problem solving included as part of the PISA 2003 assessment. This special assessment focuses on real-life aspects of problem solving not assessed as part of the mathematical, scientific, or reading literacy assessments. PISA 2003 C-D problem solving examines students' capabilities of understanding, structuring, solving, and communicating solutions of problems in real-life contexts.

Like the mathematics and scientific contexts, the problem-solving contexts range from personal life to community and society. In the special C-D problem-solving study in PISA, the contexts are *personal; school, work and leisure;* and *community and society*.

However, in order to grasp the nature of students' problem-solving capabilities, the PISA C-D study framework limited the breadth of problem-solving situations to just three areas of problem solving: *decisionmaking*, *system analysis and design*, and *troubleshooting*. *Decisionmaking* problems describe situations where an individual has to understand a situation, identify the relevant alternatives and constraints, and select among them to reach the best decision. *System analysis and design* problems characterize situations that require a student to consider and dissect a complex situation or set of requirements and the myriad relationships existing among them. *Troubleshooting* problems require a student to understand the main features of a malfunctioning system or device, eliminate possibilities that might explain the difficulty, and devise an explanation for the perceived difficulty (see exhibit 2).

Table 1. Distribution of items by content categories for TIMSS grade 8 and PISA assessments: 2003

Mathematics

TIMSS Content category	Number	Percent	PISA Content category	Number	Percent
Total	194	100	Total	85	100
Number	57	29	Quantity	23	27
Measurement	31	16	Space and shape	20	24
Geometry	31	16			
Algebra	47	24	Change and relationships	22	26
Data	28	14	Uncertainty	20	24

Science

TIMSS Content category	Number	Percent	PISA Content category	Number	Percent
Total	189	100	Total	35	100
Life science	56	30	Form and function	3	9
			Human biology Physiological change Biodiversity	4	11
			Genetic control	2	6
Chemistry	31	16	Chemical and physical changes	1	3
			Structure and properties of matter	6	17
Physics	45	24	Forces and movement	1	3
			Energy transformations	4	11
Earth science	31	16	Earth and its place in the universe	7	20
			Geographical change	1	3
Environmental science	26	14	Atmospheric change	3	9
			Ecosystems	3	9

TIMSS			PISA cross-disciplinary problem solving Content category	Number	Percent
Not applicable			Total	19	100
			Decisionmaking	7	37
			System analysis and design	7	37
			Troubleshooting	5	26
Total TIMSS Items	383	100	Total PISA Items	139	100
Total TIMSS problem-solving items	123	32	Total PISA problem-solving items	73	53

NOTE: The content categories and the number of items that fall into each category in mathematics, science, and cross-disciplinary problem solving are based on the item classifications as determined by the IEA and OECD. The classification of items as problem solving, as shown as totals at the bottom of the table, is based on the classifications of the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. Shading indicates content categories that most closely map across TIMSS and PISA. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003; and unpublished tabulations.

Exhibit 2. Content and problem-solving attributes included in the PISA 2003 cross-disciplinary problem-solving framework

PISA cross-disciplinary problem-solving framework
Problem-solving areas
Decision making
System analysis and design
Troubleshooting
Problem-solving processes
Understanding the problem
Characterizing the problem
Representing the problem
Solving the problem
Reflecting on the solution
Communicating the problem solution

NOTE: PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.
SOURCE: Organization for Economic Cooperation and Development. (2003). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving.* Paris, France: Author.

Unlike the broader competency clusters established to discuss cognitive processes in the PISA mathematical and scientific literacy frameworks, the competency clusters in C-D problem solving focus more directly on the stages of processing in which students use their reasoning skills and associated knowledge to solve problems. These competency clusters are *understanding the problem, characterizing the problem, representing the problem, solving the problem, reflecting on the problem,* and *communicating the problem solution*.

TIMSS 2003 had no separate framework for problem solving. Part of the design of the TIMSS mathematics and science frameworks called for the development of special sets of items that would focus on problem solving and inquiry within the content areas themselves. The PSI sets consist of items related to a common context. These item sets probe student knowledge in the content domains to a greater depth than individual items normally do. The focus is on students' ability to solve a problem and to provide justification for their answers. The PSI items are not separated from the content items or analyzed as a separate substudy, but they are distinguished by their development and their presentation as a set of successive items tied to the same contextual theme.

Methodology

The methodology employed in this study consisted of a targeted coding of the items in the assessments with respect to the original assessment designs and with respect to a number of item characteristics developed especially for use in this study. These item characteristics focus on the content, context, specific knowledge and skills, cognitive processes, and difficulty loads associated with the individual assessment items and the assessments as a whole. As noted earlier, only TIMSS eighth-grade items and PISA items were used in the analysis.

While there are a number of variables that can be used to examine the nature of the items in large-scale assessments such as TIMSS and PISA, an attempt has been made to restrict analyses to variables that can be coded as meeting or not meeting criteria in order to improve the reliability of the judgments made. The variables chosen and described below were ones that have a direct and easily interpretable relationship to the findings of the study from a practical and applicable standpoint. Using the central feature of the definition of problem solving as involving a situation where students do not have a known strategy to immediately apply, as well as the nature of the cognitive behaviors associated with problem solving, the coding process chosen for the analyses was as follows:

- Using the definition of problem solving, items were coded as problem-solving items if they required students to resolve a situation that, most likely, had not been explicitly studied or for which the student would not have a ready procedure.

- Interrater reliabilities were coded for the author's individual codings of items as satisfying the definition of problem solving prior to any discussion to adjudicate differences. These values are reported in appendix F. The interrater reliability coefficients were obtained using Craig's generalized value for Scott's π coefficient (Craig 1981; Scott 1955). The values of the coefficient for the first codings of the mathematics items were 0.74 for the TIMSS mathematics items and 0.84 for the PISA mathematics items. These were considered as good to substantial and excellent to almost perfect ratings, respectively, using established criteria for interpretation of such coefficients (Von Eye and Mun 2005).

- Items collectively judged as problem-solving items were then coded for values on a number of variables related to the task situation posed to students. These coding variables are detailed below. In some categories, items could be coded in multiple ways and in other categories only a single code was allowed. These differences are covered for each coding made.

Content Coverage

Items were first coded as belonging to a single content class by the authors using the PISA and TIMSS 2003 frameworks (see exhibit 1).

Cognitive Processes

Items were then coded as belonging to a single cognitive process/competency class by the authors in terms of the cognitive processes detailed in the PISA and TIMSS 2003 frameworks (see exhibit 1).

Problem-Solving Attributes

Items were also coded with respect to various problem-solving skills, including *identify variables or relationships*, *critically evaluate information*, *justify/prove solution*, *generalize or predict applicability*, and *communicate solution*. A given problem could be coded with one or more of these attributes according to what it required from the student. The results of the various item codings of problems indicated differences or significant factors with each content area or with problem solving in general. These differences then serve as a basis for the comparisons made between the assessments.

Item Format

Both TIMSS and PISA assessments included a variety of item formats. See appendix B for examples of these item formats. *Simple multiple-choice* items asked students to select from a list of alternatives (for an example of a simple multiple-choice item, see item 1, appendix B). *Complex multiple-choice* items asked students to respond to a series of "true/false" or "yes/no" items (for an example of a complex multiple-choice item, see item 2, appendix B).

In addition to multiple-choice items, varieties of other forms of items are used that call upon students to construct and communicate their responses. *Short constructed response (SCR)* items call for a computational or a short verbal response. If the item has one possible response or solution method, the item is a *closed SCR* item (for examples of closed SCR items, see items 3 and 4, appendix B). An *open SCR* item is one that allows different answers or has the possibility of many different ways of arriving at the solution (for an example of an open SCR item, see item 5, appendix B). Both open and closed SCR items allow for the possibility of partial credit.

An *extended constructed response (ECR)* item is one that requires several steps and a more lengthy response to explain the answer. *Scaffolded ECR* items are presented as a number of smaller questions that provide structure for students' responses and direct the approach taken to some degree. The students are led via a series of questions, often labeled a, b,… , to answer several parts of an extended question. As such, the students are guided to a solution using a specific problem-solving approach (for an example of a scaffolded ECR item, see item 6, appendix B). *Open ECR* items tend to ask one large question in which the solution strategy and nature and structure of the response are left open to the student (for an example of an open ECR item, see item 7, appendix B). Each item was coded into one and only one of the item format classes. Any one of these item format classes, with the exception of the *open ECR*, class, could contain either exercises or problems.

Computational Aspects of Items

Given that many problem-solving items require the determination of a value or some comparative measure, computation can play a significant role in problem-solving situations. Hence, items were coded with respect to the computational load they placed on the problem solver. An item was judged to have a computational load beyond basic if it required computations beyond straightforward work with whole numbers, fractions, and decimals or the solution of a simple linear equation involving whole numbers or integers. Otherwise, the item was coded as not having a computational requirement. This attribute of items is discussed further in the mathematics analyses later in this report. Examples of exercises with no computational requirements are items 1 and 2 in appendix A. These items require only the solution of a simple linear equation involving integer values or recall of definitions.

Translations of Representations

Part of successful problem solving involves recognizing the nature of the information provided in a problem and working with that information in another form. This may involve taking information from a table or chart and calculating percentages, or it may involve examining the spatial arrangement of objects relative to a particular object and determining the degree to which the position of a given object affects the positioning of other objects. As a result, items were coded as having one or more of the following translation of representation features:

- developing a drawing or sketch;
- interpreting a figural representation;
- interpreting a graphical representation;
- interpreting a statistical representation;
- interpreting a functional representation; and
- interpreting a tabular representation.

Summary of Methodology

The variables employed in coding items resulted from an analysis of characteristics discussed in the problem-solving literature and models used in other analyses of items in international assessments conducted by NCES (Nohara 2001; Neidorf et al. 2006). Several versions of each possible variable were considered. The coding of items by the three report authors was checked, where possible, against each assessment's original design features and categorization of items. After preliminary coding and an examination of the results within each content area and each assessment, sets of variables and categories within each variable appropriate to each content area and assessment emerged. The three authors then individually coded the items and submitted their codings. These coders' first rating values were then analyzed for agreement. In cases where the three authors agreed, or two of the three agreed, the rating of the agreeing authors was used. In the few cases where all three authors disagreed on an item's code, the item was discussed in order to arrive at a mutually agreeable coding. In all cases, differences about any item were communicated to all three authors so that any minority opinion could be stated and discussed prior to the use of any code for an item in further analyses. See appendix F for more information on the methods used in this report and interrater reliability coefficients.

The focus of the analyses in this report is on those items judged to be problem-solving items, as defined earlier. Thus, the analyses focus on a subset of items from the total pool of items included in the PISA and TIMSS 2003 assessments. To aid the reader in interpreting the frequencies and percentages shown, the total number of items included in the assessment and the total number of items judged to be problem-solving items are provided in each of the tables in this report. The reader should be aware that many comparisons in this report between subsets of items drawn from the respective assessments involve disproportionate numbers of items, which sometimes are quite small. Thus, a higher percentage of items does not always correspond to a greater number of items. It is for this reason that both numbers and percentages are presented in the data tables supporting the analyses.

Statistical analyses for significance ($\alpha = 0.05$) of the resulting comparisons were conducted using χ^2 analyses, corrected with Yates' correction for continuity, for 2 x 2 comparisons (Fleiss 1981; Yates 1934) and using the G^2 likelihood-ratio form of the chi-square test for R x C comparisons (Agresti 1996). See appendix F for more details.

Section III: Problem Solving in the TIMSS and PISA Mathematics Assessment Items

The results of the coding of items in the mathematics assessments indicate that over one-third of the items (74 of the 194 items) in the TIMSS 2003 mathematics assessment at grade 8 were classified as problem-solving items. In the PISA 2003 mathematical literacy assessment (administered to 15-year-old students), nearly half of the mathematical literacy items (41 of the 85 items) were classified as problem-solving items (table 2). While it appears that PISA had a larger percentage of its mathematics items judged as problem solving items than TIMSS, the difference was not statistically significant ($\chi^2 = 2.08$, df = 1, $p > 0.05$). The mathematics items that were considered to be problem-solving items (74 for TIMSS, 41 for PISA) were then evaluated with respect to the characteristics described previously, including content coverage, cognitive processes, problem-solving attributes, item format, computational aspects, and translation of representations.

Table 2. Distribution of TIMSS and PISA mathematics problem-solving items, by content category: 2003

<table>
<tr><th colspan="3">TIMSS</th><th colspan="3">PISA</th></tr>
<tr><th>Content category</th><th>Number</th><th>Percent</th><th>Content category</th><th>Number</th><th>Percent</th></tr>
<tr><td>Total mathematics items</td><td>194</td><td>100</td><td>Total mathematics items</td><td>85</td><td>100</td></tr>
<tr><td>Total mathematics items identified as problem solving</td><td>74</td><td>38</td><td>Total mathematics items identified as problem solving</td><td>41</td><td>48</td></tr>
<tr><td>Number</td><td>15</td><td>20</td><td>Quantity</td><td>8</td><td>20</td></tr>
<tr><td>Measurement</td><td>15</td><td>20</td><td rowspan="2">Space and shape</td><td rowspan="2">10</td><td rowspan="2">24</td></tr>
<tr><td>Geometry</td><td>17</td><td>23</td></tr>
<tr><td>Algebra</td><td>16</td><td>22</td><td>Change and relationships</td><td>12</td><td>29</td></tr>
<tr><td>Data</td><td>11</td><td>15</td><td>Uncertainty</td><td>11</td><td>27</td></tr>
</table>

NOTE: Of the 194 TIMSS mathematics items and 85 PISA mathematics items, 74 and 41 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. Shading indicates content categories that most closely map across TIMSS and PISA. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003; and unpublished tabulations.

Content Coverage

In order to determine whether some content areas are more likely to be associated with problem-solving situations within the respective assessments, the problem-solving items' locations within the content categories of the two assessments were identified (table 2).[3] These codings reflect the ways items are categorized according to the TIMSS and PISA frameworks. The shaded horizontally grouped content categories in table 2 indicate categories that are nominally related. An analysis of the 2 x 4 table showing the examinations crossed with PISA item classification

[3] While these percentages deal with the assessments as a whole, readers should be aware that many comparisons between subsets of items drawn from the respective assessments in this report involve disproportionate numbers of items, which sometimes are quite small. Thus, a higher percentage of items does not always correspond to a greater number of items. It is for this reason that numbers, percentages, and their statistical significance are presented in the data tables and text supporting the analyses.

indicates there were no significant differences in the ways in which the items in the two exams were distributed across the four PISA content categories ($G^2 = 5.25$, df = 3, $p > 0.05$).

An alternative method of comparing the distribution of the two assessments' problem-solving items to content categories can be made by mapping the PISA problem solving items onto the TIMSS framework (table 3). Table 3 provides a direct comparison of the relative weighting of item content using the language of the TIMSS content categories. These data present a different picture of the relative weightings given the varied content categories across the two assessments.

Table 3. Distribution of PISA mathematics problem-solving items within TIMSS categories, by content: 2003

PISA content categories	Number of items by PISA content area	TIMSS content domains: Number	Geometry and measurement	Algebra	Data
Total PISA mathematics items	85	23	20	22	20
Total PISA mathematics items identified as problem solving	41	10	10	7	14
Quantity	8	7	0	0	1
Space and shape	10	0	10	0	0
Change and relationships	12	2	0	7	3
Uncertainty	11	1	0	0	10
Percentage of PISA problem-solving items in TIMSS content category	100	24	24	17	34
Percentage of TIMSS problem-solving items in TIMSS content category	100	20	43	22	15

NOTE: There are 41 PISA mathematics items classified as problem-solving items by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. The second row from the top of the table gives the number of PISA mathematics problem-solving items mapped to each category of the TIMSS framework. The second to last row in the table provides the percentage of PISA problem-solving items mapped to each of the TIMSS content domains. In TIMSS, geometry and measurement are separate content domains. They are combined in this table to better match the PISA items. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003; unpublished tabulations.

An analysis of the overall structure of the data in the table found no statistical difference between the allocation of the PISA mathematics problem-solving items, classified into the TIMSS categories, with the proportion of TIMSS mathematics problem-solving items in those same categories ($G^2 = 7.5$, df = 3, $p > 0.05$). Thus, while there were no significant differences found between the allocations of mathematics problem-solving items in TIMSS and PISA based on the TIMSS content categories, a numerical comparison of the percent of items allocated to the subject content categories suggests that TIMSS may emphasize *geometry*, *measurement*, and *algebra* items, while the PISA assessment may emphasize *number* and *data* items.

Cognitive Processes

The analysis of the measurement of students' cognitive processes requires a matching of the TIMSS problem-solving cognitive domains with the PISA competency clusters. The TIMSS domain of *knowing facts and procedures* corresponds with the PISA cluster of *reproduction*, while the TIMSS domains *using concepts* and *solving routine problems* matches with the PISA cluster *connections*. Finally, the TIMSS domain of *reasoning* relates to the PISA cluster *reflection*. Table 4 details the number and percentage of problem-solving items in the corresponding TIMSS cognitive domains and PISA competency clusters.

Table 4. Distribution of TIMSS and PISA mathematics problem-solving items, by cognitive process domains (TIMSS) and competency clusters (PISA): 2003

TIMSS			PISA		
Cognitive process domain	Number	Percent	Competency cluster	Number	Percent
Total mathematics items	194	100	Total mathematics items	85	100
Total mathematics items identified as problem solving	74	38	Total mathematics items identified as problem solving	41	48
Knowing facts and procedures	4	5	Reproduction	8	20
Using concepts	11	15	Connections	20	49
Solving routine problems	28	38			
Reasoning	31	42	Reflection	13	32

NOTE: Of the 194 TIMSS and 85 PISA mathematics items, 74 and 41 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. Shading indicates content categories that most closely map across TIMSS and PISA. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003; unpublished tabulations.

The comparisons of the item data for cognitive domains and competency clusters show that the majority of problem-solving items in TIMSS were found in the *reasoning* (42 percent) and *solving routine problems* (38 percent) domains, while the largest percentage of PISA items were found in the competency cluster *connections* (49 percent) (table 4). Comparing the allocations of items to matching categories showed no measurable differences existed in the proportions of items apportioned to the varied process categories shown in table 4 ($G^2 = 5.57$, df = 2, $p > 0.05$). Though a direct comparison of the percentages of the PISA mathematics problem-solving items in the *reproduction* cluster and the TIMSS corresponding domain *knowing facts and procedures* did not result in any significant differences, the distribution of items in these cognitive domains/competency clusters may have been a result of the literacy focus of the PISA assessment versus the curriculum focus of the TIMSS assessment, since the *reproduction* cluster often includes items that deal with very common, everyday demands of mathematical knowledge. A comparable proportion of items were allocated to TIMSS' *using concepts* and *solving routine problems* (53 percent) and PISA's *connections* cluster (49 percent). Finally, the TIMSS allocation of items to *reasoning* (42 percent) was greater than the proportion found in PISA's *reflection* cluster (32 percent). This result may follow from the earlier difference between the percentages of items allotted to *knowing facts and procedures* in TIMSS and *reproduction* in PISA mentioned above, as the percentages allotted to the cognitive categories must add to 100.

Problem-Solving Attributes

The mathematics problem-solving items were next coded with respect to which of the various attributes of the problem-solving process they exemplified. Table 5 contains the data resulting from this coding. Note that items could be classified as having more than one attribute. *Identify variables or relationships* indicates whether students have to be cognizant of variables or relationships in solving a given problem (for an example of a mathematics item with the attribute *identify variables or relationships*, see item 8, appendix C). For example, does the item require students to sort out important variables or establish relationships among variables? The attribute *critically evaluate information* indicates whether students have to compare and contrast information or carefully examine information concerning the problem constraints (for an example of a mathematics item with the attribute *critically evaluate information*, see item 10, appendix C). The attribute *justify/prove solution* refers to the level of explanation required of a student by a given problem (for an example of a science item with the attribute *justify/prove solution*, see item 11 in appendix C). The attribute *generalize or predict applicability* refers to whether an item calls on students to hypothesize or generalize in responding to an item (for an example of a mathematics item requiring *generalize or predict applicability* behaviors, see item 12, appendix C). The analysis of the differences in problem solving attributes present in the mathematics items between the two assessments found only one significant difference: the TIMSS assessment placed significantly more emphasis on *identify variables or relationships* (91 versus 63 percent) in its mathematics items than the PISA assessment ($\chi^2 = 10.86$, df = 1, $p < 0.05$).

Table 5. Distribution of TIMSS and PISA mathematics problem-solving items, by problem-solving attributes: 2003

	TIMSS		PISA	
Problem-solving attribute	Number	Percent	Number	Percent
Total mathematics items	194	100	85	100
Total mathematics items identified as problem solving	74	38	41	48
Identify variable or relationship	67	91 *	26	63 *
Critically evaluate information	17	23	16	39
Justify/prove solution	1	1	3	7
Generalize or predict applicability	5	7	0	0

*p<.05. Denotes a significant difference between assessments in this category.

NOTE: Of the 194 TIMSS and 85 PISA mathematics items, 74 and 41 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. Items could be classified into more than one attribute category. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003. Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003.

Though the data seem to show that TIMSS had a larger proportion of *generalize or predict applicability* items (7 versus 0 percent) than the PISA assessment ($\chi^2 = 1.50$, df = 1, $p > 0.05$), and PISA had a larger proportion of items calling on students' capabilities to *critically evaluate information* (39 versus 23 percent) and *justify/prove solution* (7 versus 1 percent) than the TIMSS assessment, these differences were not found to be significant ($\chi^2 = 2.58$, df = 1, $p > 0.05$ and $\chi^2 = 1.30$, df = 1, $p > 0.05$, respectively). Though not significant, the distribution of items among the problem-solving attributes appear to be related to the main purposes of the two assessment programs. The processes emphasized in TIMSS, namely, working with variables, relationships, and generalizing, are in many instances tied to the performance of algorithmic procedures and reflect the curriculum-based focus of the TIMSS assessment. The critical evaluation of a situation and the justification or communication of a solution relate more directly to using problem-solving skills in real-life settings, the focus of PISA.

Item Formats

There were significant differences between TIMSS and PISA in the mixture of mathematics item formats used to assess students' problem-solving abilities. Table 6 details the usage of item

Table 6. Distribution of item formats in TIMSS and PISA mathematics problem-solving items, by survey: 2003

	TIMSS			PISA		
Item format	Number	Percent		Number	Percent	
Total mathematics items	194	100		85	100	
Total mathematics items identified as problem solving	74	38		41	48	
Multiple choice	42	57	*	12	29	*
Closed short constructed response	16	22	*	18	44	*
Open short constructed response	11	15		9	22	
Scaffolded extended constructed response	5	7		0	0	
Open extended constructed response	0	0		2	5	

*p<.05. Denotes a significant difference between assessments in this category.

NOTE: Of the 194 TIMSS and 85 PISA mathematics items, 74 and 41 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds. Detail may not sum to totals because of rounding.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003; unpublished tabulations.

response formats of TIMSS and PISA mathematics problem-solving items. A comparison of the percentage of multiple choice items indicated a significant difference ($\chi^2 = 6.94$, df = 1, $p < 0.05$), as did the proportion of closed short constructed response items ($\chi^2 = 5.26$, df = 1, $p < 0.05$). These significant differences in format usage suggest TIMSS provided a greater focus on finding correct answers through the use of multiple choice items, while PISA did the same through asking students to provide these answers in a constrained open response setting. The analysis results for the proportion of open short constructed response items ($\chi^2 = 0.49$, df = 1, $p > 0.05$), scaffolded extended constructed response items ($\chi^2 = 1.50$, df = 1, $p > 0.05$), and open extended constructed response items ($\chi^2 = 1.37$, df = 1, $p > 0.05$) showed no measurable differences between the two assessments' usage of these formats for assessing mathematics knowledge.

Computational Aspects of Items

The problem-solving items were coded with respect to whether they required students to perform calculations that go beyond basic whole number, fraction, or decimal operations or whether they required students to deal with comparisons beyond those of simple ratios, percentages, and proportions. Algebraic equations were considered basic if the work involved was essentially working with whole numbers. Under these constraints, none of the problem-solving items in TIMSS exceeded the basic level of demand. In PISA, only 1 out of 41 problem-solving items (2 percent) exceeded the basic level of computational demand. The one item involved converting a measurement to an equivalent measurement where several conversion rates were involved (see table 7). There was no significant difference found in the proportion of computational demand on the two assessments ($\chi^2 = 0.09$, df = 1, $p > 0.05$).

Translations of Representations

Another characteristic often associated with problem-solving items is the requirement to combine information from multiple sources to derive a solution. Such requirements cause students to have to translate from one representational form to another. For example, data from a graph has to be interpreted numerically for use in a calculation. Other examples of problem-solving items that involve translations of representations in the mathematics assessments are found in items 1–4 of appendix D. Table 7 displays the data about translations required to shift information from one representational form to another to solve problems presented by the items in the two assessments.

Table 7. Distribution of TIMSS and PISA mathematics problem-solving items requiring computational aspects of items and translations of representations, by skill required: 2003

	TIMSS			PISA		
Skill	Number	Percent		Number	Percent	
Total mathematics items	194	100		85	100	
Total mathematics items classified as problem solving	74	38		41	48	
Computational aspects of items						
Requires computations beyond basics	0	0		1	2	
Translations of representations						
Requires drawing or sketch	15	20	*	0	0	*
Interpret figural representation	33	45		13	32	
Interpret statistical representation	2	3	*	7	17	*
Interpret functional representation	5	7		1	2	
Interpret tabular representation	3	4		4	10	

*p<.05. Denotes a significant difference between assessments in this category.

NOTE: Of the 194 TIMSS and 85 PISA mathematics items, 74 and 41 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003; unpublished tabulations.

The analysis indicated that the TIMSS 2003 mathematics assessment required students to *draw or sketch a representation* more often than did the PISA 2003 assessment (20 percent versus 0 percent). This difference in proportions was significant ($\chi^2 = 7.85$, df = 1, $p < 0.05$). The PISA assessment was found to have a significantly larger proportion of items requiring students to *interpret statistical representation* (i.e., interpret information in statistical graph or chart formats) (17 versus 3 percent; $\chi^2 = 5.69$, df = 1, $p < 0.05$). Though TIMSS would appear to have a larger percentage of mathematics items requiring students to *interpret a figural representation* and to *interpret a functional representation* (i.e., a graph of distance versus time) than the PISA 2003 assessment, these were not found to be significant ($\chi^2 = 1.33$, df = 1, $p > 0.05$ and $\chi^2 = 0.31$, df = 1, $p > 0.05$, respectively). The remaining representational translation, that of *interpreting tabular representations* (i.e., reading a table) was not found to differ between the PISA and TIMSS assessments either ($\chi^2 = 0.67$, df = 1, $p > 0.05$).

One aspect that stands out in these particular comparisons is the interaction of the translations of representations with the content categories measured in the problem-solving items. Many of the items calling for sketching or interpreting figures are *geometry* or *measurement* items. The interpretation of statistical graphs or table results, more often than not, is related to *data* or *uncertainty* items. The overall contents of the item pool indicate that items with translations were often found in the content area of *data*. This, in turn, is strongly related to the curricular focus of TIMSS versus the real-life literacy focus of PISA. The analysis of the effect of the interaction on students would require an in-depth study focusing on individual student performance on such items.

Summary of Problem Solving in the TIMSS and PISA Mathematics Assessments

While the proportion of mathematics items judged to measure problem solving appeared smaller in TIMSS than in PISA (38 versus 48 percent), the difference was not found to be significant (table 2). Across the mathematics content areas, the two assessments tended to have relatively equivalent proportions of items devoted to problem solving in the *number* domain (termed *quantity* in PISA; table 3). Though it would appear that TIMSS emphasized problem-solving items dealing with *measurement* and *geometry* (*space and shape* in PISA) and *algebra* (*change and relationships* in PISA) while the PISA assessment had a larger percentage of items focusing on problem solving in *quantity* and *uncertainty* (*number* and *data* in TIMSS), none of these differences in proportions reached the level of being statistically significant.

A comparison of the cognitive processes associated with the problem-solving items indicated that although the two assessments emphases in number of items differed by process category, these differences were not found to be statistically significant (table 4).

The data in table 5 provide evidence that the TIMSS mathematics problem-solving items were more likely to require students to *identify variables and relationships* than the PISA mathematics problem-solving items. Among the other attributes studied—*critically evaluate information*, *justify/prove solution*, and *generalize or predict applicability*—there were no measurable differences found between the two assessments.

Finally, considering the demands made on students relative to representational modes for the given information and how that information has to be handled in response to the items' demands, the information in table 7 indicates that TIMSS 2003 mathematics items were more likely to *require drawing or sketching* than the PISA mathematics items. On the other hand, a larger percentage of the PISA mathematics problem-solving items than the TIMSS items required students to *interpret statistical information*. There were no significant differences found between the TIMSS and PISA mathematics items in requiring students to *interpret a figural representation*, *interpret a functional representation*, or *interpret a tabular representation*.

Section IV: Problem Solving in the TIMSS and PISA Science Assessment Items

As with the mathematics items, all science items were first examined to determine whether they measured problem solving as defined in this report. In the TIMSS 2003 grade 8 science assessment there were 189 items. Sixteen of the 189 items were grouped into three problem-solving and inquiry (PSI) units. Fourteen of the 16 TIMSS items in the three PSI units were determined to measure problem-solving skills (table 8).[4] Besides these 14 problem-solving items in the PSI units, an additional 35 individual TIMSS science items were considered problem solving. This totaled to 49 problem-solving items out of 189 TIMSS science items (26 percent).

As with the mathematics portion of the tests, values for Craig's generalization of Scott's π coefficient for interrater reliability were calculated for the varied tests involving science or general problem solving processes (Craig 1981; Scott 1955). The values of the coefficient for the first codings were 0.71 for the TIMSS science items and 0.63 for the PISA science items. These were considered as good to substantial ratings using established criteria for interpretation of such coefficients (Von Eye and Mun 2005). A full description of this process is given in appendix F.

A chi-square comparison of the proportions of science problem-solving items in the two assessments shows that the PISA assessment had a statistically greater proportion of items measuring problem solving in science than did the TIMSS assessment (table 8; $\chi^2 = 1.55$, df = 1, $p < 0.05$).

In the PISA 2003 scientific literacy assessment, there were 35 items. This small number of items reflects the fact that science was a minor domain in the PISA 2003 assessment. In 2006, scientific literacy is the focus of the PISA assessment and, thus, there are many more science items in the assessment. Thirty-four of these 35 items (97 percent) were grouped into 12 sets of related items tied to the same overall theme. Seventeen (49 percent) of the science items in PISA were classified as problem-solving items.[5] All of the PISA items identified as problem-solving items were contained in one of the sets of related items. The remaining items, including the stand-alone item, measured conceptual knowledge of science or student comprehension of the knowledge presented. A comparison of the proportion of science problem-solving items presented as grouped sets of items showed that the larger proportion of items presented in grouped sets in the PISA assessment was greater than the proportion of problem solving items appearing in grouped sets having a common context in the TIMSS assessment ($\chi^2 = 23.06$, df = 1, $p < 0.05$).

---

[4] Not all PSI items directly measured problem solving, as defined in this report. Some items in these categories measured whether students understood the problem situation as a precursor to solving a problem, such as assessing the understanding of a problem or performing a calculation to provide a basis for addressing the problem in a related follow-up item.

[5] While these percentages deal with the assessments as a whole, readers should be aware that many comparisons in this report between subsets of items drawn from the respective assessments involve disproportionate numbers of items, which sometimes are quite small. Thus, a higher percentage of items does not always correspond to a greater number of items. It is for this reason that both numbers and percents are presented in the data tables supporting the analyses.

None of the PISA special cross-disciplinary (C-D) problem-solving assessment items were included in the comparisons in this section. They were compared with the PSI science items considered in section V of this report.

Table 8. Distribution of science items within sets, by survey: 2003

Item	TIMSS		PISA	
	Number	Percent	Number	Percent
Total science items	189	100	35	100
Total science items identified as problem solving	49	26 *	17	49 *
Number of problem-solving item sets	3	†	12	†
Total number of items in sets	16	8 *	34	97 *
Problem-solving items in sets	14	29	17	100
Stand-alone problem-solving items	35	71	0	0

†Not applicable.

*p<.05. Denotes a significant difference between assessments in this category.

NOTE: Of the 189 TIMSS and 35 PISA science items, 49 and 26 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003; unpublished tabulations.

Content Coverage

In both assessments, the items that were judged to assess problem-solving skills were classified by content domain. As was done for the mathematics items, the PISA science problem-solving items were classified as belonging to the five TIMSS science content domains. However, for purposes of reporting, several of the TIMSS grade 8 science content areas were combined when categorizing the PISA science items: *chemistry* and *physics*, and *earth science* and *environmental science* (see table 9). These categories for PISA correspond to the TIMSS content categories, allowing a comparison of the percentages of the TIMSS and PISA items. Table 9 indicates the distribution of items judged as problem-solving science items in the various TIMSS content domains along with the merged PISA domains. The items in the 13 PISA science categories listed in exhibit 1 collapse easily within the three content categories to match with the content in the five TIMSS categories (table 9). In TIMSS, the content areas of *life science* and *environmental science* contained the same percentage of science items measuring problem-solving skills (27 percent). Slightly fewer items were classified as *physics* (22 percent) and *chemistry* (18 percent), and few were classified as *earth science* (6 percent). In PISA, the items in *earth and environmental science* accounted for nearly one-half of the science problem-solving items identified (47 percent). Items in *life science* accounted for slightly more than one-quarter of the PISA problem-solving items (29 percent), while the percentage of items in *chemistry/physics* in PISA was somewhat less (24 percent). The overall analysis of the differences in proportions of items allotted to the categories indicated no significant differences between the two assessments ($G^2 = 1.86$, df = 1, $p > 0.05$).

Table 9. Distribution of TIMSS and PISA science problem-solving items by content category: 2003

TIMSS				PISA			
Content category	Number	Percent		Content category	Number	Percent	
Total science items	189	100		Total science items	35	100	
Total science items identified as problem solving	49	26	*	Total science items identified as problem solving	17	49	*
Life science	13	27		Life science	5	29	
Chemistry	9	18		Chemistry/physics	4	24	
Physics	11	22					
Earth science	3	6		Earth and environmental science	8	47	
Environmental science	13	27					

**p*<.05. Denotes a significant difference between assessments in this category.
NOTE: Of the 189 TIMSS and 35 PISA science items, 49 and 26 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. Shading indicates content categories that most closely map across TIMSS and PISA. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003; unpublished tabulations.

Cognitive Processes

As detailed in section II, TIMSS and PISA 2003 items were classified by cognitive processes into cognitive domains and competency clusters, respectively. The TIMSS cognitive domains of *factual knowledge* and *conceptual understanding* correspond with the PISA competency cluster of *describing, explaining, and predicting scientific phenomena*. The *reasoning and analysis* domain in TIMSS encompasses *understanding scientific investigation* and *interpreting scientific evidence and conclusions* clusters in PISA. Table 10 displays the number and percentage of problem-solving items in each cognitive processes classification.

As table 10 shows, of the 49 science problem-solving items in TIMSS 2003, none were classified as *factual knowledge*, while 29 percent were classified as *conceptual understanding*. Of the 17 science problem-solving items in PISA 2003, 35 percent were classified as *describing, explaining, and predicting scientific phenomena*. Examination of these PISA items suggests that they corresponded to the TIMSS *conceptual understanding* domain rather than the *factual knowledge* domain. Thus, the percentage of problem-solving items addressing the TIMSS cognitive domains *factual knowledge* and *conceptual understanding* together appears to have been slightly lower than the percentage of items in the PISA competency cluster of *describing, explaining and predicting scientific phenomena* (29 versus 35 percent), though the difference was not found to be significant. The percentage of science problem-solving items in the TIMSS cognitive domain of *reasoning and analysis* appears to have been slightly greater than the percentage of items in the comparable PISA clusters of *understanding scientific investigation* and *interpreting scientific evidence and conclusions* (71 versus 65 percent). However, as with the other categories, this difference in proportion was not found to be statistically significant (χ^2 = 0.05, df = 1, $p > 0.05$).

Table 10. Distribution of TIMSS and PISA science problem-solving items, by cognitive process domain (TIMSS) and competency cluster (PISA): 2003

TIMSS				PISA			
Cognitive process domain	Number	Percent		Competency cluster	Number	Percent	
Total science items	189	100		Total science items	35	100	
Total science items identified as problem solving	49	26	*	Total science items identified as problem solving	17	49	*
Factual knowledge	0	0		Describing, explaining, and predicting scientific phenomena	6	35	
Conceptual understanding	14	29					
Reasoning and analysis	35	71		Understanding scientific investigation	3	18	
				Interpreting scientific evidence and conclusions	8	47	

**p*<.05. Denotes a significant difference between assessments in this category.

NOTE: Of the 189 TIMSS and 35 PISA science items, 49 and 26 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. Shading indicates content categories that most closely map across TIMSS and PISA. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003; unpublished tabulations.

Problem-Solving Attributes

As described earlier, problem-solving items were also classified according to various attributes of the problem-solving process. The results for the science items are shown in table 11. See appendix C for examples of items classified by the various problem-solving attributes.

Two of the areas that appeared to be stressed more in PISA science than in TIMSS science were *identify variables or relationships* (100 versus 80 percent) and *critically evaluate information* (65 versus 31 percent). The apparent difference between the two assessments in requiring students to *identify variables or relationships* was not found to be statistically significantly ($\chi^2 = 2.65$, df = 1, $p > 0.05$), while the difference between the two assessments in items that required students to *critically evaluate information* was found to be statistically significant ($\chi^2 = 4.79$, df = 1, $p < 0.05$). For an example of a science item with the attribute *identify variables or relationships,* see item 2, appendix C. These two attributes are especially important when measuring scientific literacy. This was true in PISA, where many items called for students to find information in graphs and tables and then build a case based on that knowledge. Items that stressed the problem-solving attribute *communicate solution*, which required students to communicate their answers beyond a single-word response, appeared to be more prevalent in TIMSS than in PISA (76 versus 65 percent). However, this difference was not found to be statistically significant ($\chi^2 = 0.29$, df = 1, $p > 0.05$). For an example of a science item with the attribute *communicate solution*, see item 6, appendix C. In addition, 80 percent of the science items in TIMSS and 35 percent of those in PISA included aspects of the attribute *requires science knowledge*, a significant difference ($\chi^2 = 9.46$, df = 1, $p < 0.05$). These items required students to have some science knowledge and an understanding beyond what was provided in the item to successfully respond to the item. For an example of a science item with the attribute *requires science knowledge*, see item 7, appendix C.

Table 11. Distribution of TIMSS and PISA science problem-solving items, by problem-solving attributes: 2003

Attribute	TIMSS Number	TIMSS Percent		PISA Number	PISA Percent	
Total science items	189	100		35	100	
Total science items identified as problem solving	49	26	*	17	49	*
Identify variable or relationship	39	80		17	100	
Critically evaluate information	15	31	*	11	65	*
Communicate solution	37	76		11	65	
Requires science knowledge	39	80	*	6	35	*

*$p<.05$. Denotes a significant difference between assessments in this category.

NOTE: Of the 189 TIMSS and 35 PISA science items, 49 and 26 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. Items could be included in more than one attribute category. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003; unpublished tabulations.

Item Formats

Table 12 presents a summary of the item formats used in TIMSS and PISA science problem-solving items. A full description of item formats is found in section II. A comparison of the two assessments' use of item formats shows both similarities and differences. Two types of multiple-choice items were found in the science assessments (for examples of multiple-choice items, see items 1 and 2, appendix B). All of the problem-solving multiple-choice items in TIMSS were of the simple form that required the selection of one alternative from a group of choices. The multiple-choice items in PISA included both simple and complex forms. The latter involve a series of related true-false or "select all that apply" items. The apparent difference in the proportions of complex multiple choice items between the two assessments was not found to be statistically significant ($\chi^2 = 2.62$, df = 1, $p > 0.05$).

In addition to the multiple-choice format, students were presented with closed short constructed response (SCR) items. These items are essentially multiple-choice items without the choices: there was one response or figure that would answer the item and students either knew the answer or not. Due to the nature of this item format, there was little room for innovation. There was no significant difference found between the two assessments in the proportion of closed SCR science items ($\chi^2 = 0.39$, df = 1, $p > 0.05$). Multiple-choice and closed SCR items tend to measure whether students can find the right answer among a set of possible alternatives or whether they can generate the correct answer in a constrained setting. A comparison of the multiple-choice (simple and complex) and closed SCR item percentage totals showed an apparently lower proportion of "closed items" (i.e., the total of simple multiple-choice, complex multiple-choice, and closed SCR questions; 26 versus 30 percent) for TIMSS than for PISA (table 12). However, this apparent difference in proportions was not statistically significant.

Table 12. Distribution of item formats in TIMSS and PISA mathematics problem-solving items, by survey: 2003

Item format	TIMSS Number	TIMSS Percent		PISA Number	PISA Percent	
Total science items	189	100		35	100	
Total science items identified as problem solving	49	26	*	17	49	*
Simple multiple choice	9	18		3	18	
Complex multiple choice	0	0		2	12	
Closed short constructed response	4	8		0	0	
Open short constructed response	24	49		12	71	
Scaffolded extended constructed response	9	18		0	0	
Open extended constructed response	3	6		0	0	

*p<.05. Denotes a significant difference between assessments in this category.
NOTE: Of the 189 TIMSS and 35 PISA science items, 49 and 26 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds. Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment; unpublished tabulations.

The TIMSS science problem-solving items employed all three of the open constructed response item formats: open SCR, scaffolded extended constructed response (ECR), and open ECR. For examples of open SCR and ECR items, see items 3–7 of appendix B. In contrast, PISA science items used only one format—open SCR. The comparison of the more open items—open SCR, scaffolded ECR, and open ECR—appeared to show a slightly higher percentage of these "open format" items in the TIMSS science assessment than in the PISA science assessment (73 versus 71 percent). However, there were no significant differences found between the two assessments in this regard: open SCR ($\chi^2 = 1.59$, df = 1, $p > 0.05$), scaffolded ECR ($\chi^2 = 2.22$, df = 1, $p > 0.05$), and open ECR ($\chi^2 = 0.15$, df = 1, $p > 0.05$). These apparent differences in item format usages between the mathematics and science assessments seem to show internal differences in the construction of the respective assessments (see table 6). However, these apparent differences did not reach the level of statistical significance.

Computational Aspects of Items

There were several problem-solving items in TIMSS science that required some computation; however, no items were found to go beyond basic calculations (table 13). No items in the PISA scientific literacy assessment required any computation.

Translations of Representations

Understanding and using graphs, tables, and figures is an important part of most content domains. Many items in TIMSS and PISA made use of these types of stimuli as a stepping-stone to ascertain relationships between variables, to analyze and interpret data, or to aid in the design of investigations. Many sets of science items in PISA were also centered on understanding the textual information presented in addition to figures, graphs, and tables. Table 13 shows the number and percentage of problem-solving items in the TIMSS and PISA science assessments that required either a drawing or sketch or that contained figures, graphs, or tables to be interpreted. For examples of problem-solving items that involve such translations of representations, see appendix D.

Table 13. Distribution of TIMSS and PISA science problem-solving items requiring computational aspects of items and translations of representations, by skill required:2003

	TIMSS			PISA		
Skill	Number	Percent		Number	Percent	
Total science items	189	100		35	100	
Total science items identified as problem solving	49	26	*	17	49	*
Computational aspects of items						
Requires computations beyond basics	0	0		0	0	
Translations of representations						
Requires drawing or sketch	2	4		0	0	
Interpret figural representation	22	45		5	29	
Interpret graphical representation	6	12		2	12	
Interpret tabular representation	10	20		6	35	

*p<.05. Denotes a significant difference between assessments in this category.

NOTE: Of the 189 TIMSS and 35 PISA science items, 49 and 26 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. Results do not total to 100 because some items did not require computation beyond basics or translation of representations. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment; unpublished tabulations.

As items could be coded as requiring more than one of these translations of representations for their solution, the significance of the translational demands were tested individually. Though there appeared to be differences in the percentage of science items in TIMSS and PISA that required students to *interpret a figural representation* and *interpret a tabular representation*, no significant differences were found ($\chi^2 = 0.69$, df = 1, $p > 0.05$ and $\chi^2 = 0.82$, df = 1, $p > 0.05$, respectively). The percentages of problem-solving science items that *require a drawing or sketch* or required a student to *interpret a graphical representation* between the two assessment were also not found to differ ($\chi^2 = 0.00$, df = 1, $p > 0.05$ and $\chi^2 = 0.14$, df = 1, $p > 0.05$, respectively).

Only one science item was categorized as requiring more than one translation of representation. This TIMSS item required both the interpretation of a figural diagram and the analysis of a tabular display. In both the TIMSS and PISA assessments, items were presented with opening passages presenting a context for the items that follow. While an investigation of the impact of this is beyond the scope of this report, it is important to note that students' reading comprehension abilities could play a critical role in measuring their problem-solving abilities.

Summary of Problem Solving in the TIMSS Science and PISA Science Assessments

The analysis of TIMSS and PISA problem-solving items in science shows both similarities and differences between the assessments. The differences appear to align with the goals of each assessment: in TIMSS, to measure science achievement across different science domains; in PISA, to measure the capacity to use and understand scientific concepts in order to "identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world…" (OECD 2003).

Significantly more of the items in PISA scientific literacy assessment were judged to measure problem solving compared to the items in TIMSS science assessment (although there are 2.8 times as many science problem-solving items in TIMSS as in PISA; see table 8). In both the TIMSS and PISA science assessments, the overall patterns of allocation of items to the content categories appeared similar. Examining the cognitive processes demanded by the items showed that there were no significant differences found in the percentage of science items addressing the corresponding cognitive domains and competency clusters between the two assessments (table 10). An analysis of problem-solving item attributes showed a significantly higher usage of items demanding students to *critically evaluate information* in the PISA science assessment than in the TIMSS assessment, while there was a significantly greater percentage of TIMSS science items than PISA science items that *required science knowledge* (table 11).

Comparisons of the types of item formats used in the science problem-solving items in PISA and TIMSS indicated no significant differences (table 12). This was also the case in the examination of the kinds of translations of representations demanded by the TIMSS and PISA science problem-solving items (table 13).

Section V: Problem Solving in the PISA Cross-Disciplinary Study and TIMSS Problem-Solving and Inquiry Items

Special portions of both the PISA and TIMSS 2003 assessments focused on problem solving. The PISA 2003 assessment included a separate subtest of 19 cross-disciplinary (C-D) problem-solving items. These items were directed toward assessing students' capabilities to solve problems independent of an assessment of students' capabilities related to a particular curricular domain. These 19 items were in a separate section of the assessment and were not part of the mathematics or science assessments discussed previously. TIMSS incorporated problem-solving and inquiry (PSI) items within the mathematics and science assessments. These PSI items have been analyzed as part of the previous discussions as they were specifically designed to be either mathematics or science items. The PISA C-D items occasionally draw on concepts and problem-solving strategies taught in either mathematics or science classes; however, they were designed to be independent of the curricular areas of mathematics and science.

The 19 PISA C-D problem-solving items were grouped into nine different sets, ranging in length from one to three items. Each set was constructed around a single contextual setting followed by an associated question or questions related to the setting. This approach was similar to the thematic approach taken in the 1996 NAEP assessment, where more than one problem-solving item in mathematics and science drew on the same context for information (Mitchell et al. 1999).

TIMSS PSI items were included within both the mathematics and science assessments. They are reviewed here as a comparison to the PISA C-D items. Seven sets of PSI items were included in the 2003 TIMSS assessment. Four of the sets were in mathematics and three were in science. The four mathematics sets contained 17 items, and the three science sets contained 16 items.

Both the PISA C-D and the TIMSS PSI items included opening passages that presented the context for the items that followed. Although beyond the scope of this report, the potential impact of students' reading abilities in the measurement of their problem-solving proficiency is an area of further study.

As with the mathematics and science items, the first step was to consider whether or not the PISA C-D and TIMSS PSI items measured problem solving. Using the definition of problem solving employed in this report, 15 of the 19 PISA 2003 C-D items were classified as problem-solving items (table 14). The remaining four items measured whether students *understand* a problem, which is the first step toward problem solving. These four items were the first items in the stimulus materials for a C-D set of items. As such, they were positioned in the assessment to ensure that students understood the data and were able to interpret the data correctly as a basis for attacking or addressing the problems that followed.

When the 33 TIMSS 2003 PSI items were examined, 23 were found to be problem-solving items. The remaining items, like the PISA items, dealt either with assessing the understanding of a problem or performing a calculation to provide a basis for attacking or addressing the problem in a related follow-up item. Though the data in table 14 appear to show that a larger percentage of the PISA C-D problems were coded as problem-solving items than TIMSS PSI items (79 vs.

70 percent), this difference was not found to be statistically significant ($\chi^2 = 0.00$, df = 1, $p > 0.05$).

Table 14. Distribution of items classified as problem-solving items, by survey component: 2003

Survey component	Total number of items	Number of items classified as problem solving	Percent
TIMSS problem solving and inquiry (PSI)	33	23	70
PISA cross disciplinary (C-D)	19	15	79

NOTE: Of the 33 TIMSS PSI and 19 PISA C-D items, 23 and 15 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment; unpublished tabulations.

Problem-Solving Attributes

Almost all items in both the TIMSS and PISA assessments (96 percent and 93 percent, respectively) called on students to *synthesize or integrate information* (table 15). A larger percentage of TIMSS PSI items than PISA C-D items required students to *identify variables or relationships* (83 versus 40 percent; $\chi^2 = 5.55$, df = 1, $p < 0.05$). Though it would appear that a higher percentage of items in TIMSS than in PISA required students to *communicate their solution* (61 versus 53 percent) and a higher percentage of items in PISA than in TIMSS measured students' ability to *critically evaluate information* (33 versus 22 percent), these differences were not found to be significant ($\chi^2 = 0.02$, df = 1, $p > 0.05$ and $\chi^2 = 0.17$, df = 1, $p > 0.05$, respectively). See appendix C for examples of items classified by these problem-solving attributes.

Table 15. Distribution of TIMSS problem-solving and inquiry (PSI) items and PISA cross-disciplinary (C-D) items, by problem-solving attributes: 2003

Attribute	TIMSS		PISA	
	Number	Percent	Number	Percent
Total PSI and C-D items	33	100	19	100
Total PSI and C-D items identified as problem solving	23	70	15	79
Identify variable or relationship	19	83 *	6	40 *
Critically evaluate Information	5	22	5	33
Synthesize or integrate information	22	96	14	93
Communicate solution	14	61	8	53

**p*<.05. Denotes a significant difference between assessments in this category.
NOTE: Of the 33 TIMSS PSI and 19 PISA C-D items, 23 and 15 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments.* Items could be included in more than one attribute category. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment; unpublished tabulations.

Item Formats

Table 16 provides a profile of the item formats used in the PISA 2003 C-D assessment and in the TIMSS 2003 PSI item sets. None of the problem-solving items in the TIMSS PSI clusters were multiple-choice or complex multiple-choice items, while 33 percent of the items in PISA C-D clusters were. However, given the small numbers of such items, neither format for multiple choice items showed a difference between usage in the two assessments: multiple choice (χ^2 = 1.12, df = 1, $p > 0.05$) and complex multiple choice (χ^2 = 2.62, df = 1, $p > 0.05$). Both assessments made use of short constructed response (SCR) items and extended constructed response (ECR) items—both of which required students to construct their own responses in order to answer the given problem. Each assessment had approximately the same percentage of closed SCR items (χ^2 = 0.11, df = 1, $p > 0.05$). However, there was a greater percentage of open SCR items in the TIMSS PSI items than in the PISA C-D items (48 versus 7 percent; χ^2 = 5.34, df = 1, $p < 0.05$). The two assessments appeared to differ in their usage of ECR items as well. TIMSS had 26 percent of its items in this category (scaffolded and open ECR, combined), while PISA had 33 percent. However, these differences were not found to be statistically significant (χ^2 = 0.05, df = 1, $p > 0.05$ and χ^2 = 0.99, df = 1, $p > 0.05$, respectively). Looking at the distribution of item types within ECR, one sees that the TIMSS and PISA distributed ECR items between scaffolded and open ECR items in slightly different proportions. Thus, the items in the TIMSS PSI clusters identified as problem-solving items tended to provide students more support in framing their answers in open ECR items than the PISA C-D items identified as problem-solving items.

Table 16. Distribution of item formats in TIMSS problem-solving and inquiry (PSI) items and PISA cross-disciplinary (C-D) items, by survey: 2003

Item format	TIMSS			PISA		
	Number	Percent		Number	Percent	
Total PSI and C-D items	33	100		19	100	
Total PSI and C-D items identified as problem solving	23	70		15	79	
Simple multiple choice	0	0		2	13	
Complex multiple choice	0	0		3	20	
Closed short constructed response	6	26		4	27	
Open short constructed response	11	48	*	1	7	*
Scaffolded extended constructed response	5	22		2	13	
Open extended constructed response	1	4		3	20	

*p<.05. Denotes a significant difference between assessments in this category.
NOTE: Of the 33 TIMSS PSI and 19 PISA C-D items, 23 and 15 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. Results do not total to 100 because some items did not require computation beyond basics or translation of representations. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment; unpublished tabulations.

Computational Aspects of Items

An analysis of the mathematics required to solve the problems posed in either the TIMSS PSI or PISA C-D items identified as problem-solving items showed that no mathematical procedural skills beyond those considered basic were required (table 17). Hence, there was no statistical difference in the proportions of items requiring computational demand between the assessments.

Table 17. Distribution of TIMSS problem-solving and inquiry (PSI) items and PISA cross-disciplinary (C-D) items requiring computational aspects of items and translations of representations, by skill required: 2003

Skill	TIMSS			PISA		
	Number	Percent		Number	Percent	
Total PSI and C-D items	33	100		19	100	
Total PSI and C-D items identified as problem solving	23	70		15	79	
Computational aspects of items						
Requires computations beyond basics	0	0		0	0	
Translations of representations						
Requires drawing or sketch	4	17		3	20	
Interpret figural representation	9	39		9	60	
Interpret graphical representation	3	13		0	0	
Interpret tabular representation	5	22		8	53	
Interpret information from a reading passage	8	35	*	13	87	*

*$p<.05$. Denotes a significant difference between assessments in this category.

NOTE: Of the 33 TIMSS PSI and 19 PISA C-D items, 23 and 15 are classified as problem-solving items, respectively, by the authors of this report, *Problem Solving in the PISA and TIMSS 2003 Assessments*. Items could be classified under more than one skill category. TIMSS assesses the mathematics and science knowledge and abilities of fourth- and eighth-graders; the data in the table pertain to the grade 8 assessment only. PISA assesses the reading, mathematics, and scientific literacy of 15-year-olds. Detail may not sum to totals because of rounding.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment; unpublished tabulations.

Translations of Representations

Though it would appear that a higher percentage of PISA 2003 C-D items than TIMSS 2003 PSI items required students to translate representations (table 17), none of these differences was found to be significant: *require a drawing or sketch* ($\chi^2 = 0.05$, df = 1, $p > 0.05$), *interpret figural representation* ($\chi^2 = 0.86$, df = 1, $p > 0.05$), *interpret graphical representation* ($\chi^2 = 0.71$, df = 1, $p > 0.05$), or *interpret tabular representation* ($\chi^2 = 2.75$, df = 1, $p > 0.05$). The only significant difference found was in the proportion of items that required students to *interpret information from a reading passage*, in favor of the PISA C-D items ($\chi^2 = 7.90$, df = 1, $p < 0.05$). Eighty-seven percent of the PISA C-D items required such an interpretation, compared with 35 percent of the TIMSS PSI items. For examples of problem-solving items that involved translations of representations, see appendix E.

Summary of the Comparison of the Special PISA Problem-Solving Study With the TIMSS PSI Items

A comparison of the sets of items comprising the PSI items in TIMSS with the C-D problem-solving study in PISA showed different emphases in both the items and the nature of the assessments. Some of these differences result from the frameworks for the assessments themselves (OECD 2003; Mullis et al. 2003), and some result from the choice of items and approaches to assessing students in problem solving.

A comparison of the percentage of items that actually measure problem solving as defined in this study showed a relatively equal focus on problem solving in the special items sets of the two assessments (table 14). However, the TIMSS PSI problem-solving items placed a significantly greater emphasis than the PISA C-D items on students' capabilities to *identify variables or relationships* (83 vs. 40 percent; table 15). Expectations of students' capabilities to *critically evaluate information*, *synthesize or integrate information*, and *communicate solutions* differed in their proportions of items, but these differences were not found to be significant.

The two assessments made relatively equivalent use of closed student response formats in eliciting specific information about students' understanding and ability to discern outcomes related to certain aspects of the problems. The TIMSS 2003 PSI sets contained no multiple-choice items, but 26 percent of the items were in the closed SCR format (table 16). PISA 2003 C-D had 33 percent of its items as either simple or complex multiple-choice items and another 27 percent as closed SCR items. Combining these for a measure of closed response—that is, items where students have to give or select a specific answer—26 percent of TIMSS PSI problem-solving items were either multiple choice or closed short response, compared to 60 percent of the PISA C-D problem-solving items (but, note that the actual number of items in both assessments was relatively small). The converse was true when one considers the use of items where students had to structure some form of an open response showing their own thoughts. A larger percentage of the TIMSS PSI items than the PISA C-D items were designed as short open constructed response items (48 vs. 7 percent). Overall, 40 percent of the PISA C-D items judged to be problem solving were designed as open response items, 33 percent of which were ECR items, while 74 percent of the TIMSS PSI items judged to be problem solving were designed in a similar format.

Finally, examination of the problem-solving attributes that surface in these two specialized sets of items revealed a single significant difference: A larger percentage of problem-solving items in the PISA C-D clusters than in TIMSS PSI sets required students to *interpret information from a reading passage* (87 vs. 35 percent; table 17).

Section VI: Summary and Conclusions

In summary, the TIMSS 2003 and PISA 2003 assessment frameworks differed in their focus on problem solving in mathematics and science. TIMSS 2003 focused on what eighth-grade students had achieved as a result of their schooling. PISA 2003, in contrast, focused on problem solving in a real-life context and targeted a population consisting of 15-year-olds. The proportions of problem-solving items also varied across the two assessments ($\chi^2 = 17.07$, df = 1, $p < 0.05$). In PISA 2003, 53 percent of all items (in mathematics and science, as well as the special C-D clusters) were found to fall under the definition of problem solving used in this report (73 of 139 total items; table 1). In TIMSS 2003, 32 percent of all items (in mathematics and science, which included the PSI items) were judged to be problem-solving items (123 of 383 total items). Of these, 38 percent of the TIMSS mathematics assessment items and 48 percent of the PISA mathematical literacy items were found to measure problem solving (table 2). In science, 26 percent of the TIMSS assessment items and 49 percent of the PISA scientific literacy assessment items were found to measure problem solving (table 8). As expected, a large percentage of the items in the TIMSS PSI clusters and the PISA C-D problem-solving assessment were found to measure problem solving as defined by this report (70 percent and 79 percent, respectively; table 14).

In terms of the content areas, the mathematics portions of the TIMSS and PISA 2003 assessments were not found to differ. TIMSS had 43 percent of mathematics problem-solving items devoted to *measurement/geometry* compared to the 24 percent of mathematics problem-solving items in PISA devoted to the comparable category of *space and shape*. PISA had 27 percent of the mathematics problem-solving items focused on *uncertainty* compared to 15 percent of the TIMSS mathematics problem-solving items focused on *data* (table 2). It was also found that the distribution of science items into the content areas in the TIMSS and PISA 2003 assessment did not differ. Thus, though 40 percent of TIMSS science problem-solving items focused on *chemistry* and *physics* compared to 24 percent of the PISA science problem solving items, and 47 percent of the PISA science items identified as problem solving were related to *earth and environmental science* compared to 33 percent of TIMSS science items, these apparent differences were not significant (table 9).

The comparison of the cognitive processes associated with the problem-solving items in mathematics and science in the two assessments indicated no significant differences in emphasis (tables 4 and 10). In mathematics, around half of the problem-solving items in both assessments focused on *using concepts* and *solving routine problems* (*connections* in PISA; table 4). In science, at least two-thirds of the problem-solving items in both assessments focused on *reasoning and analysis* (*understanding scientific investigation* and *interpreting scientific evidence and conclusions* in PISA; table 10).

With respect to the format of items, both assessments used a range of item formats (tables 6 and 12). The only significant differences between item formats in both assessments was in mathematics, where the problem-solving items in PISA were more likely to be in a closed short constructed response format and the TIMSS problem-solving items were more likely to be in a multiple choice format (table 6).

In the mathematics assessments, it is interesting to note the interaction of translation of representations and content dimensions. For example, the TIMSS problem-solving items in mathematics were more likely to *require a drawing or sketch* than the PISA problem-solving items in mathematics (table 7). This may be related to the large percentage of TIMSS mathematics problem-solving items related to *geometry and measurement* (table 2). The PISA problem-solving items in mathematics were more likely than the TIMSS items to require students to *interpret a statistical representation* (table 7). This may be related to the percentage of PISA mathematics items related to *uncertainty* (table 2). This type of interaction was not noted in the science comparisons or the special problem-solving comparisons.

In conclusion, PISA and TIMSS provide cross-sectional views of students' problem solving through the items created for each assessment. With a focus on grade 8 students in TIMSS and 15-year-olds in PISA, data from the two assessments provide a glimpse of problem-solving abilities in early adolescence. However, the two programs have separate orientations. TIMSS measures students' learning of material presented in the classroom. The TIMSS problem-solving items presented a mixture of items set in real-life contexts as well as others that measured classroom-related mathematics and science skills and knowledge. PISA measures students' mathematical and scientific literacy separately from instructional and curricular influences. The PISA problem-solving items presented a series of problems set in real-life settings that had little or no relationship to content studied in school.

While it is possible to compare and contrast the problem-solving items contained in each assessment, not surprisingly, the findings reflect the goals of the assessment programs themselves. That is, the TIMSS 2003 assessment items tended to focus on students' knowledge and ability to perform particular skills or procedures, and the PISA 2003 assessment items tended to focus on broader interpretive and application outcomes. The specific analyses of items by content coverage, cognitive processes, item formats employed, and problem-solving attributes provided evidence of this tendency. Where the purposes of assessment in PISA and TIMSS were somewhat more similar—in the problem-solving items in the special studies areas (TIMSS PSI and PISA C-D)—some differences also existed in terms of item format and problem-solving attributes and skills.

The analysis of the problem-solving items in the special studies areas—TIMSS PSI and PISA C-D—also indicates a need for further research on the role of reading skills in the measurement of problem-solving performance, a topic that is beyond the scope of this report. In both assessments, items were presented with opening passages presenting a context for items that follow. These passages' location, length, and relationship to the individual problem-solving items relied in many cases on students' critical reading and comprehension capabilities, and could have played an important role in the measurement of problem-solving abilities.

Finally, both TIMSS and PISA will revise the frameworks used to guide future assessments. Therefore, the degree to which one assessment incorporates problem solving into the items compared to the other may shift. The PISA C-D assessment will not be repeated, as it was a one-time special study as part of the PISA 2003 assessment. However, aspects of the C-D items may find their way into future PISA assessment items.

References

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. New York: John Wiley & Sons.

Bloom, B.S., Hasting, J.T., and Madaus, G.F. (1971). *Handbook on Formative and Summative Evaluation of Student Learning.* New York: McGraw-Hill.

Bransford, J.D., Brown, A.L., and Cocking, R.R. (Eds.). (1999). *How People Learn: Brain, Mind, Experience, and School.* Washington, DC: National Academy Press.

Case, R. (1985). *Intellectual Development: Birth to Adulthood*. Orlando, FL: Academic Press.

Charles, R.I., and Silver, E.A. (Eds.). (1988). *The Teaching and Assessing of Mathematical Problem Solving.* Reston, VA: National Council of Teachers of Mathematics.

Collis, K.F., Romberg, T.A., and Jurdak, M.E. (1986). A Technique for Assessing Mathematical Problem Solving Ability. *Journal for Research in Mathematics Education*, *17*(3): 206–221.

Craig, R. T. (1981). Generalization of Scott's Index of Intercoder Agreement. *The Public Opinion Quarterly, 45*(2): 260-264.

Csapó, B. (1997). The Development of Inductive Reasoning: Cross-Sectional Assessments in an Educational Context. *International Journal of Behavioral Development*, *20*(4): 609–626.

English, L.D. (Ed.). (2002). *Handbook of International Research in Mathematics Education.* Mahwah, NJ: Lawrence Erlbaum.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions* (2nd edition). New York: John Wiley and Sons.

Henderson, K.B., and Pingry, R.E. (1953). Problem Solving in Mathematics. In Howard F. Fehr (Ed.), *The Learning of Mathematics: Its Theory and Practice* (pp. 228–270). Washington, DC: National Council of Teachers of Mathematics.

Holsti, O. R. (1969). *Content Analysis for the Social Sciences and Humanities.* Reading, MA: Addison-Wesley.

Kilpatrick, J., Swafford, J.O., and Findell, B. (2001). *Adding It Up: Helping Children Learn Mathematics.* Washington, DC: National Academy Press.

Landis, J. R., and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*: 159-174.

Lester, F.K., Jr. (1980). Research on Mathematical Problem Solving. In R.J. Shumway (Ed.), *Research in Mathematics Education* (pp. 286–323). Reston, VA: National Council of Teachers of Mathematics.

Mayer, R.E. (1985). Implications of Cognitive Psychology for Instruction in Mathematical Problem Solving. In E.A. Silver (Ed.), *Teaching and Learning Mathematical Problem Solving: Multiple Research Perspectives* (pp. 123–138). Hillsdale, NJ: Erlbaum.

Mayer, R.E. (1992). *Thinking, Problem Solving, Cognition* (2nd ed.). New York: Freeman.

Mitchell, J.H., Hawkins, E.F., Stancavage, F.B., and Dossey, J.A. (1999). *Estimation, Skills, Mathematics-in-Context, and Advanced Skills in Mathematics* (NCES 2000-451). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., González, E.J., Chrostowski, S.J., and O'Connor, K.M. (2003). *TIMSS Assessment Frameworks and Specifications: 2003 (2nd ed.).* Chestnut Hill, MA: Boston College.

National Assessment Governing Board. (1999). *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*. Washington, DC: Author.

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

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

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

Neidorf, T.S., Binkley, M., and Stephens, M. (2006). *Comparing Science Content in the National Assessment of Educational Progress (NAEP) 2000 and Trends in International Mathematics and Science Study (TIMSS) 2003 Assessments* (NCES 2006-026). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Neubrand, M., Biehler, R., Blum, W., Cohors-Fresenborg, E., Flade, L., Knoche, N., Lind, D., Löding, W., Möller, G., and Wynands, A. (2001). *Grundlagen der Ergänzung des Internationalen OECD/PISA-Mathematik-Tests in der Deutschen Zusatzerhebung. Zentralblatt für Didaktik der Mathematik, 33*(2): 33–45.

Niss, M. (1999). Kompetencer og Uddannelsesbeskrivelse. *Uddanneise*, 9: 21–29.

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

Organization for Economic Cooperation and Development. (2000).*Measuring Student Knowledge and Skills: The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy.* Paris, France: Author.

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

Pellegrino, J.W., Chudowsky, N., and Glaser, R. (Eds.). (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.

Polya, G. (1945). *How to Solve It*. Princeton, NJ: Princeton University Press.

Polya, G. (1962, 1965). *Mathematical Discovery: On Understanding, Learning, and Teaching Problem Solving* (two volumes). New York: John Wiley & Sons.

Resnick, L.B., and Ford, W.W. (1981). *The Psychology of Mathematics for Instruction*. Hillsdale, NJ: Erlbaum.

Schoenfeld, A.H. (1992). Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics. In D.A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 334–370). New York: Macmillan.

Scott, W. A. (1955). Reliability of Content Analysis: The Case of Nominal Coding. *The Public Opinion Quarterly 19*(3): 321-25.

Silver, E.A. (Ed.). (1985). *Teaching and Learning Mathematical Problem Solving: Multiple Research Perspectives*. Hillsdale, NJ: Erlbaum.

Steen, L.A. (Ed.). (1990). *On the Shoulders of Giants: New Approaches to Numeracy*. Washington, DC: National Academy Press.

Von Eye, A., and Mun, E. Y. (2005). *Analyzing Rater Agreement: Manifest Variable Methods*. Mahwah, NJ: Lawrence Erlbaum.

Vosniadou, S., and Ortony, A. (1989). *Similarity and Analogical Reasoning*. New York: Cambridge University Press.

Yates, F. (1934). Contingency Tables Involving Small Numbers and the χ^2 Test. *Journal of the Royal Statistical Society Supplement, 1*: 217.

Appendix A: Examples of Problem-Solving and Exercise Items

Item 1: Exercise from mathematics (TIMSS)

If $n = -3$, what is the value of $-3x$?

Ⓐ -9

Ⓑ -6

Ⓒ -1

Ⓓ 1

Ⓔ 9

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

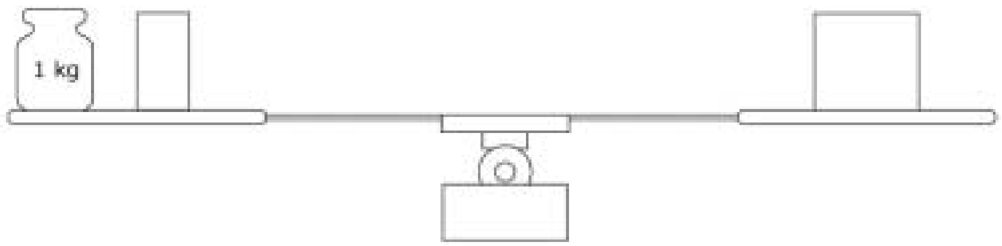
Item 2: Exercise from science (TIMSS)

A powder made up of both white specks and black specks is likely to be
- A. a solution
- B. a pure compound
- C. a mixture
- D. an element

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

Item 3: Problem-solving item from mathematics (TIMSS)

The objects on the scale make it balance exactly. On the left pan there is a 1 kg weight (mass) and half a brick. On the right pan there is one brick.



What is the weight (mass) of one brick?

Ⓐ 0.5 kg

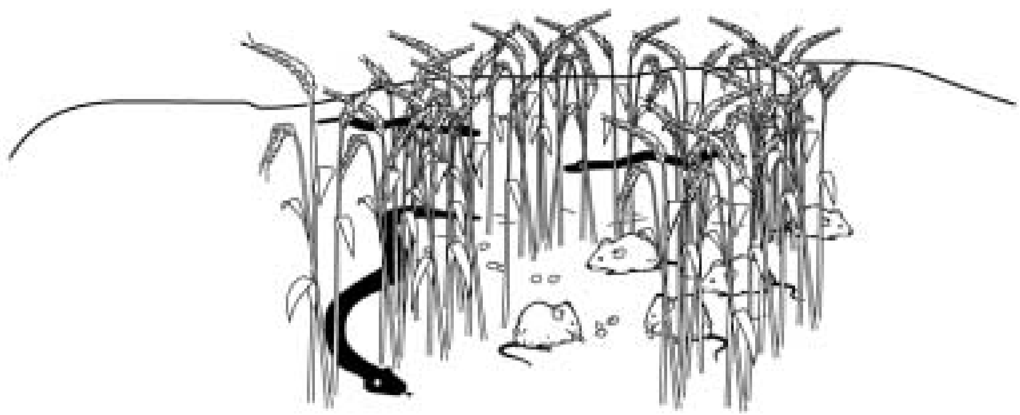
Ⓑ 1 kg

Ⓒ 2 kg

Ⓓ 3 kg

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

Item 4: Problem-solving item from science (TIMSS)



The diagram above shows a community consisting of mice, snakes and wheat plants.

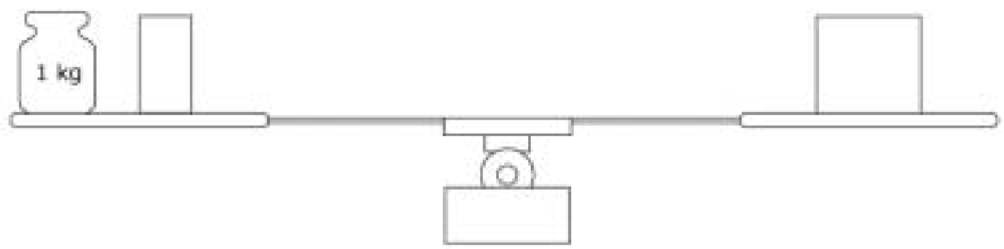
What would happen to this community if people killed the snakes?

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

Appendix B: Item Format Examples

Item 1: Simple multiple choice (TIMSS)

The objects on the scale make it balance exactly. On the left pan there is a 1 kg weight (mass) and half a brick. On the right pan there is one brick.



What is the weight (mass) of one brick?

(A) 0.5 kg

(B) 1 kg

(C) 2 kg

(D) 3 kg

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

Item 2: Complex multiple choice (PISA)

Jane bought a new cabinet-type freezer. The manual gave the following instructions:

- Connect the appliance to the power and switch the appliance on.
    - You will hear the motor running now.
    - A red warning light (LED) on the display will light up.
- Turn the temperature control to the desired position. Position 2 is normal.

Position	Temperature
1	–15°C
2	–18°C
3	–21°C
4	–25°C
5	–32°C

-
    - The red warning light will stay on until the freezer temperature is low enough. This will take 1 - 3 hours, depending on the temperature you set.
- Load the freezer with food after four hours.

Jane followed these instructions, but she set the temperature control to position 4. After 4 hours, she loaded the freezer with food.

After 8 hours, the red warning light was still on, although the motor was running and it felt cold in the freezer.

Jane wondered whether the warning light was functioning properly. Which of the following actions and observations would suggest that the light was working properly?

Circle "Yes" or "No" for each of the three cases.

Action and Observation	Does the observation suggest that the warning light was working properly?
She put the control to position 5 and the red light went off.	Yes / No
She put the control to position 1 and the red light went off.	Yes / No
She put the control to position 1 and the red light stayed on.	Yes / No

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), Cross-Disciplinary Assessment, 2003.

Item 3: Closed short constructed response (TIMSS)

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

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

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

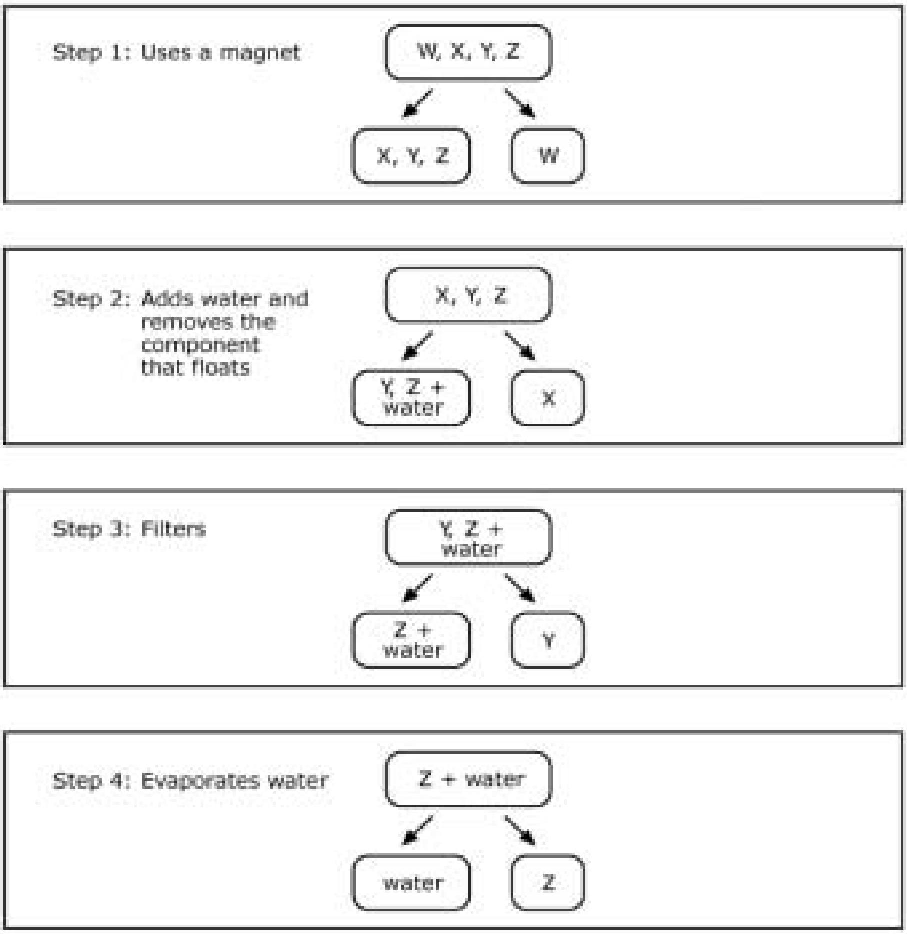
Darlene signed up for the *Plan B*, and the cost of one month of service was 75 zeds. How many minutes did she talk that month? Show your work.

Minutes talked ______________

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

Item 4: Closed short constructed response (TIMSS)

Teresa is given a mixture of salt, sand, iron filings, and small pieces of cork. She separates the mixture using a 4-step procedure as shown in the diagram. The letters W, X, Y, and Z are used to stand for the four components but do not indicate which letter stands for which component.



Identify what each component is by writing *salt*, *sand*, *iron*, or *cork* in the correct spaces below

Component W is: ______________________________

Component X is: ______________________________

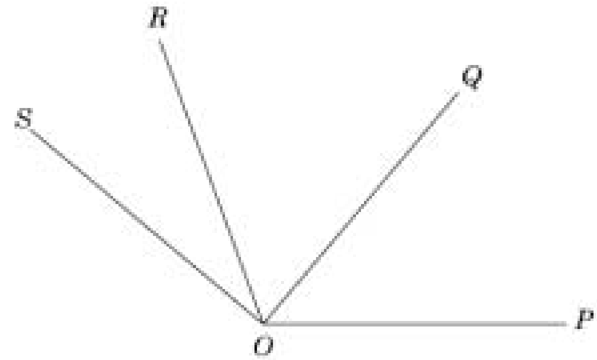
Component Y is: ______________________________

Component Z is: ______________________________

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

Item 5: Open short constructed response (TIMSS)

In the figure, the measure of $\angle POR$ is 110°, the measure of $\angle QOS$ is 90°, and the measure of $\angle POS$ is 140°.



What is the measure of $\angle QOR$?

Answer: ______________

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

Item 6: Scaffolded extended constructed response (TIMSS)

The three figures below are divided into small congruent triangles.

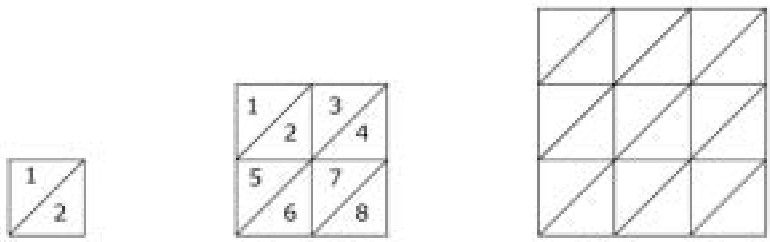


Figure 1 Figure 2 Figure 3

A. Complete the table below. First, fill in how many small triangles make up Figure 3. Then, find the number of small triangles that would be needed for the 4th figure if the sequence of figures is extended.

Figure	Number of Small Triangles
1	2
2	8
3	
4	

B. The sequence of figures is extended to the 7th figure. How many small triangles would be needed for Figure 7?

Answer: _______________

C. The sequence of figures is extended to the 50th figure. Explain a way to find the number of small triangles in the 50th figure that does not involve drawing it and counting the number of triangles.

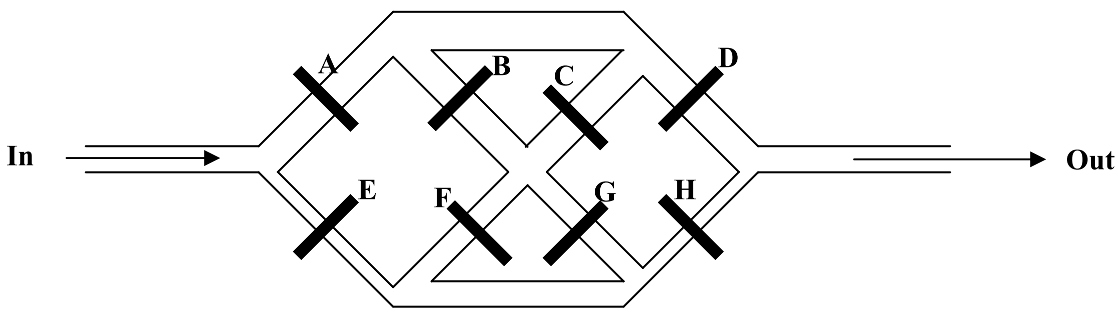
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Item 7: Open extended constructed response (PISA)

Below is a diagram of a system of irrigation channels for watering sections of crops. The gates A to H can be opened and closed to let the water go where it is needed. When a gate is closed no water can pass through it.

This is a problem about finding a gate which is stuck closed, preventing water from flowing through the system of channels.

Figure 1: A system of irrigation channels



Michael notices that the water is not always going where it is supposed to.

He thinks that one of the gates is stuck closed, so that when it is switched to “open”, it does not open.

Michael wants to be able to test whether **gate D** is stuck closed.

In the following table, show settings for the gates to test whether **gate D** is stuck closed when it is set to “open”.

Settings for gates (each one “open” or “closed”)

A	B	C	D	E	F	G	H

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), Cross-Disciplinary Assessment, 2003.

Appendix C: Examples of Problem-Solving Attributes in Items

Item 1: Identify variable or relationship (TIMSS)

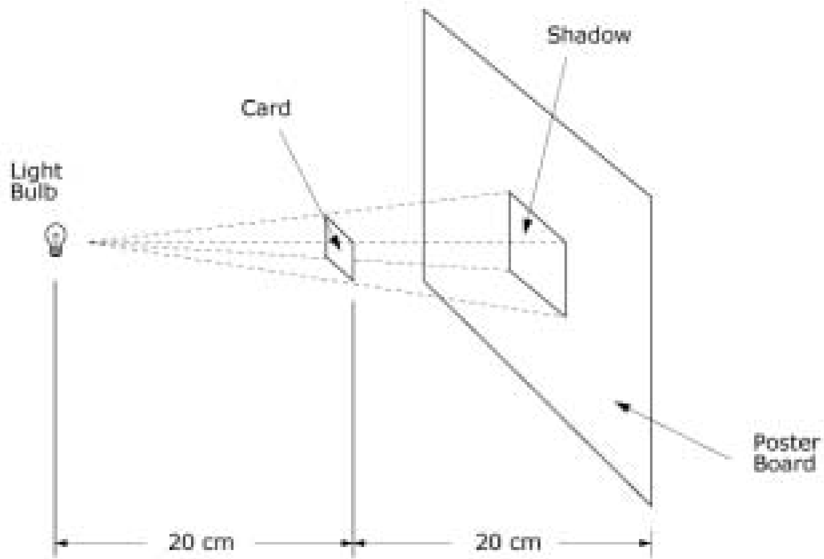
A thin wire 20 centimeters long is formed into a rectangle. If the width of this rectangle is 4 centimeters, what is its length?

- (A) 5 centimeters
- (B) 6 centimeters
- (C) 12 centimeters
- (D) 16 centimeters

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

Item 2: Identify variable or relationship (TIMSS)

A tiny light bulb is held 20 centimeters to the left of a square card, which is in turn held 20 centimeters to the left of a poster board, as shown. The shadow of the card on the poster board has a side of 10 centimeters.



If the poster board is moved 40 cm further to the right so that it is 80 cm from the light, what will be the new side of the card's shadow on the poster board?

Ⓐ 5 cm

Ⓑ 10 cm

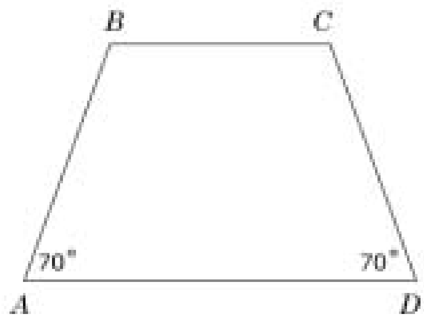
Ⓒ 15 cm

Ⓓ 20 cm

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

Item 3: Critically evaluate information (TIMSS)

ABCD is a trapezoid.



Another trapezoid, *GHIJ* (not shown), is congruent (the same size and shape) to *ABCD*. Angles *G* and *J* each measure 70°. Which of these could be true?

(A) $GH = AB$

(B) Angle *H* is a right angle.

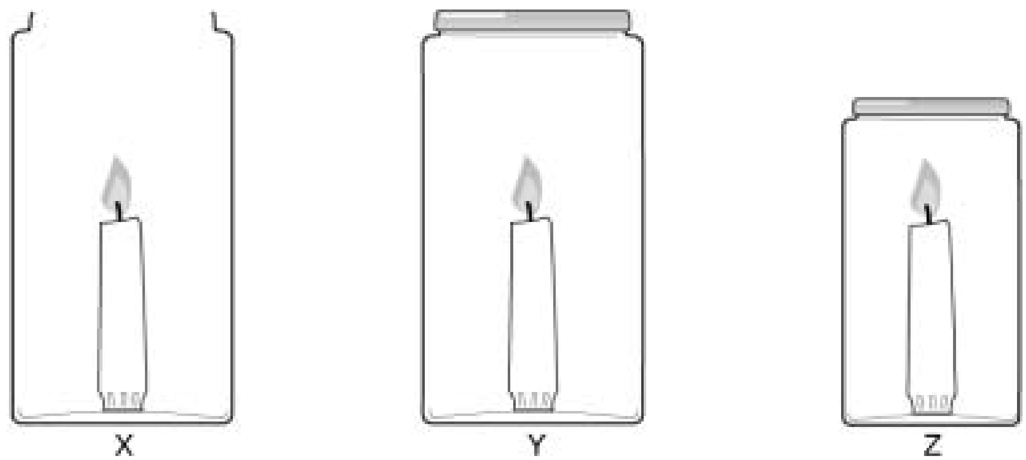
(C) All sides of *GHIJ* are the same length.

(D) The perimeter of *GHIJ* is 3 times the perimeter of *ABCD*.

(E) The area of *GHIJ* is less than the area of *ABCD*.

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

Item 4: Justify/prove solution (TIMSS)



Three identical candles are placed in the three jars shown above and lit at the same time. Jars Y and Z are then sealed with lids, and Jar X is left open.

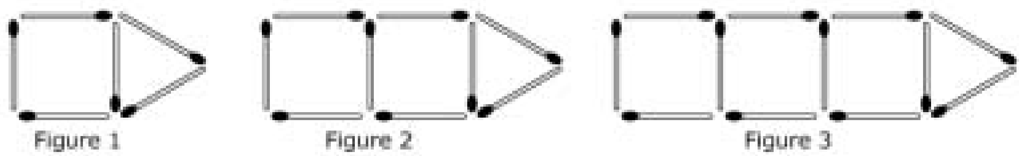
Which candle flame will go out first (X, Y, or Z)? ______________

Explain your answer.

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

Item 5: Generalize or predict applicability (TIMSS)

Matchsticks are arranged as shown in the figures.



If the pattern is continued, how many matchsticks would be used to make Figure 10?

Ⓐ 30

Ⓑ 33

Ⓒ 36

Ⓓ 39

Ⓔ 42

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

Item 6: Communicate solution (TIMSS)

Which organisms that live on land most likely inhabited the Galapagos Islands first?

(Check one box.)

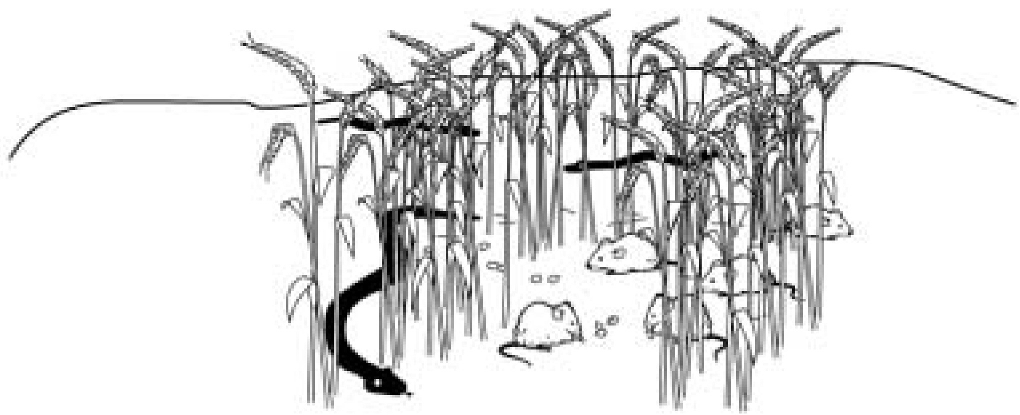
☐ Land plants

☐ Land animals

Explain your answer.

NOTE: This item is part of a problem-solving and inquiry (PSI) set.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Item 7: Requiring science knowledge (TIMSS)



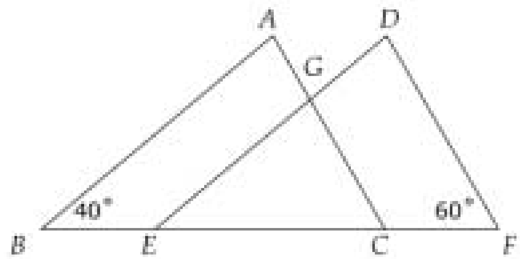
The diagram above shows a community consisting of mice, snakes and wheat plants.

What would happen to this community if people killed the snakes?

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

Item 8: Synthesize or integrate information (TIMSS)

In this figure, triangles ABC and DEF are congruent with $BC = EF$.



What is the measure of angle EGC?

Ⓐ 20°

Ⓑ 40°

Ⓒ 60°

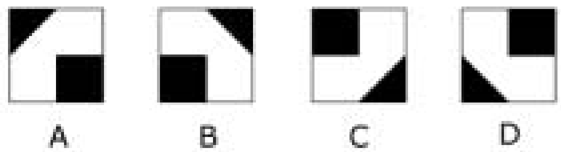
Ⓓ 80°

Ⓔ 100°

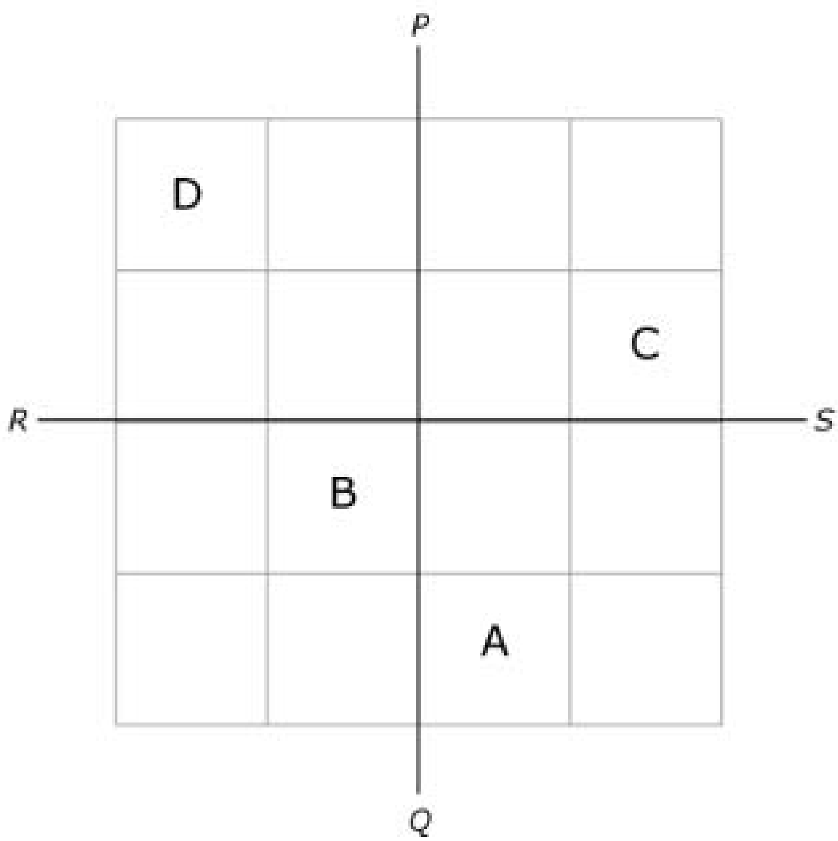
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Appendix D: Examples of Items Requiring Translations of Representations (TIMSS)

Item 1: Requires drawing or sketch

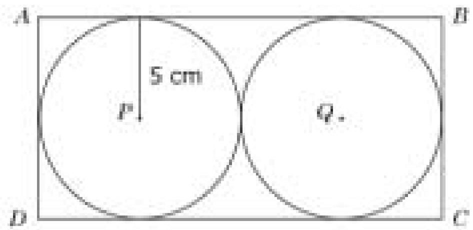


Continue to identify the tiles as shown above. On the grid below, write the letters A, B, C, or D to make a symmetrical pattern where *PQ* and *RS* would be lines of symmetry. Arrange the tiles to make a pattern.



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

Item 2: Interpret figural representation (TIMSS)



In the figure above, *ABCD* is a rectangle, and circles *P* and *Q* each have a radius of 5 cm. What is the area of the rectangle?

Ⓐ 50 cm^2

Ⓑ 60 cm^2

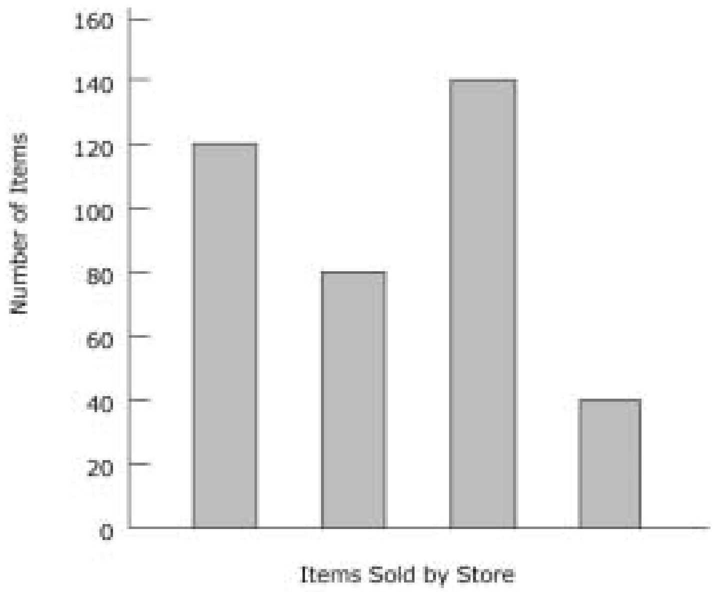
Ⓒ 100 cm^2

Ⓓ 200 cm^2

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

Item 3: Interpret statistical representation (TIMSS)

The graph shows the number of pens, pencils, rulers, and erasers sold by a store in one week.



The names of the items are missing from the graph. Pens were the item most often sold, and fewer erasers than any other item were sold. More pencils than rulers were sold. How many pencils were sold?

Ⓐ 40

Ⓑ 80

Ⓒ 120

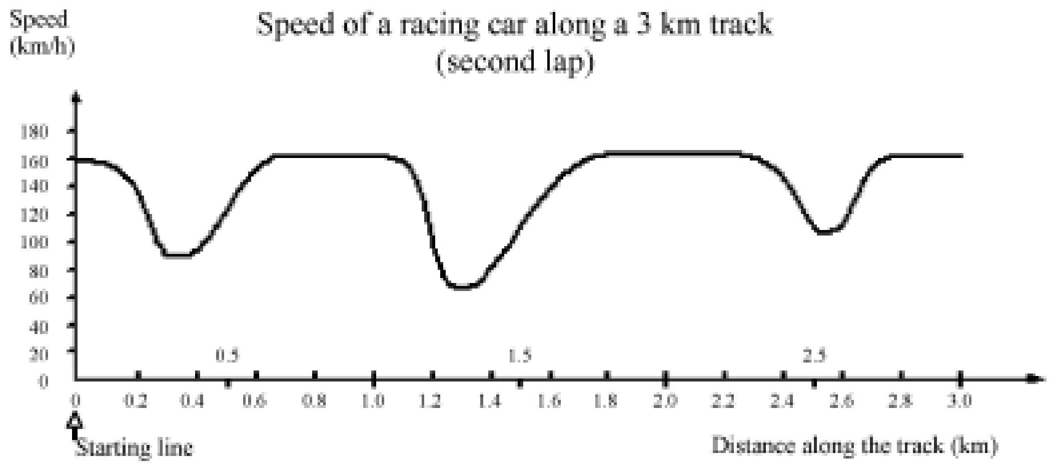
Ⓓ 140

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

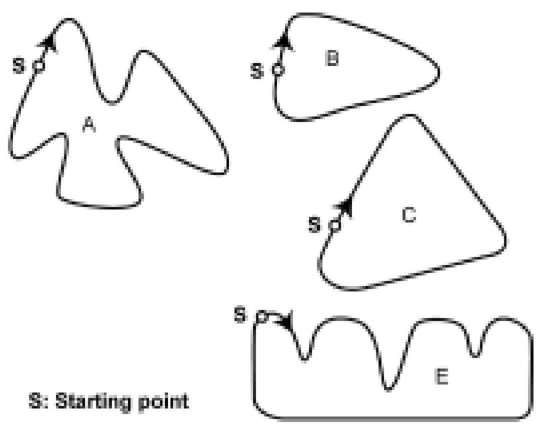
Item 4: Interpret functional representation (PISA)

SPEED OF A RACING CAR

This graph shows how the speed of a racing car varies along a flat 3 kilometer track during its second lap.



Here are pictures of five tracks. Along which one of these tracks was the car driven to produce the speed graph shown earlier?



NOTE: This is a sample PISA item released in 2000. No 2003 PISA items that required students to interpret functional representation were publicly released.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2000.

Item 5: Interpret tabular representation (PISA)

The Zedish Community Service is organizing a five-day Children's Camp. 46 children (26 girls and 20 boys) have signed up for the camp, and 8 adults (4 men and 4 women) have volunteered to attend and organize the camp.

Table 1: Adults

Mrs Madison
Mrs Carroll
Ms Grace
Ms Kelly
Mr Stevens
Mr Neill
Mr Williams
Mr Peters

Table 2: Dormitories

Name	Number of beds
Red	12
Blue	8
Green	8
Purple	8
Orange	8
Yellow	6
White	6

Dormitory rules:

1. Boys and girls must sleep in separate dormitories.
2. At least one adult must sleep in each dormitory.
3. The adult(s) in a dormitory must be of the same gender as the children.

Dormitory Allocation.

Fill the table to allocate the 46 children and 8 adults to dormitories, keeping to all the rules.

Name	Number of boys	Number of girls	Name(s) of adult(s)
Red			
Blue			
Green			
Purple			
Orange			
Yellow			
White			

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

Item 6: Interpret tabular representation (TIMSS)

The table below lists the density for different metals.

Metal	Density (g/cm^3)
Platinum	21.4
Gold	19.3
Silver	10.5
Copper	8.9
Zinc	7.1
Aluminum	2.7

B. The density of the crown was found to be 12.0 g/cm^3. What would you report to the king about what metal or mixture of metals the jeweler used to make the crown?

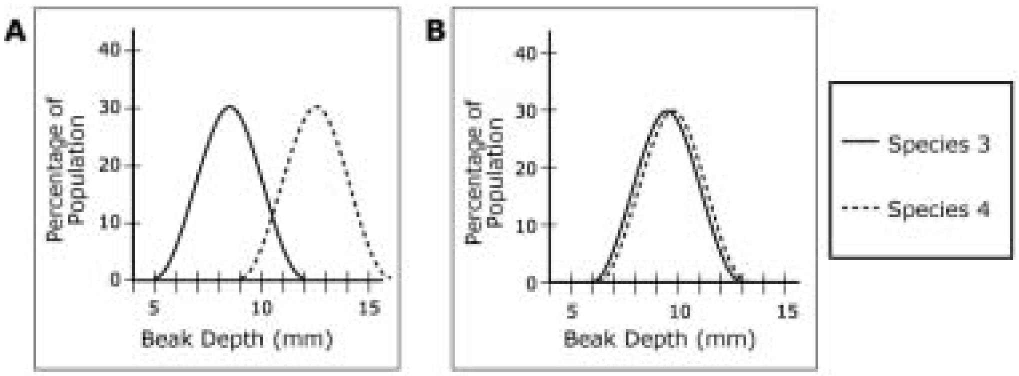
NOTE: This item is part of a problem-solving and inquiry (PSI) cluster.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Item 7: Interpret graphical representation (TIMSS)

Two other species (Species 3 and Species 4) live on Santa Maria Island, which also has a range of seed types.

Which of the following graphs shows a range of beak depths for Species 3 and Species 4 that would best insure the survival of both species on Santa Maria Island?

(Circle the letter by the correct graph.)



Explain why this range of beak depths would be best.

NOTE: This item is part of a problem-solving and inquiry (PSI) cluster.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Item 8: Interpret information from an informational passage (PISA)

This problem is about finding a suitable time and date to go to the cinema.

Isaac, a 15-year-old, wants to organise a cinema outing with two of his friends, who are of the same age, during the one-week school vacation. The vacation begins on Saturday, 24th March and ends on Sunday, 1st April.

Isaac asks his friends for suitable dates and times for the outing. The following information is what he received.

Fred: *"I've to stay home on Monday and Wednesday afternoons for music practice between 2:30 and 3:30"*

Stanley: *"I've to visit grandmother on Sundays, so it can't be Sundays. I have seen Pokamin and don't want to see it again."*

Isaac's parents insist that he only goes to movies suitable for his age and does not walk home. They will fetch the boys home at any time up to 10 p.m.

Isaac checks the movie times for the vacation week. This is the information that he finds.

TIVOLI CINEMA

Advance Booking Number: 01924 423000
24 hour phone number: 01924 420071
Bargain Day Tuesdays: All films $3

Films showing from Fri 23rd March for two weeks:

Children in the Net		**Pokamin**	
113 mins 14:00 (Mon-Fri only) 21:35 (Sat/Sun only)	Suitable only for persons of 12 years and over	105 mins 13:40 (Daily) 16:35 (Daily)	Parental Guidance. General viewing, but some scenes may be unsuitable for young children
Monsters from the Deep		**Enigma**	
164 mins 19:55 (Fri/Sat only)	Suitable only for persons of 18 years and over	144 mins 15:00 (Mon-Fri only) 18:00 (Sat/Sun only)	Suitable only for persons of 12 years and over
Carnivore		**King of the Wild**	
148 mins 18:30 (Daily)	Suitable only for persons of 18 years and over	117 mins 14:35 (Mon-Fri only) 18:50 (Sat/Sun only)	Suitable for persons of all ages

If the three boys decided on going to “Children in the Net”, which of the following dates is suitable for them?

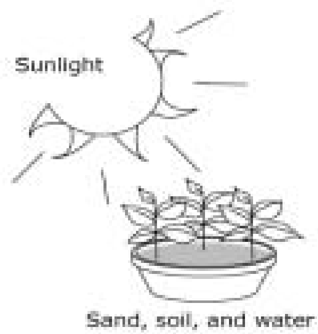
A Monday, 26th March
B Wednesday, 28th March
C Friday, 30th March
D Saturday, 31st March
E Sunday, 1st April

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

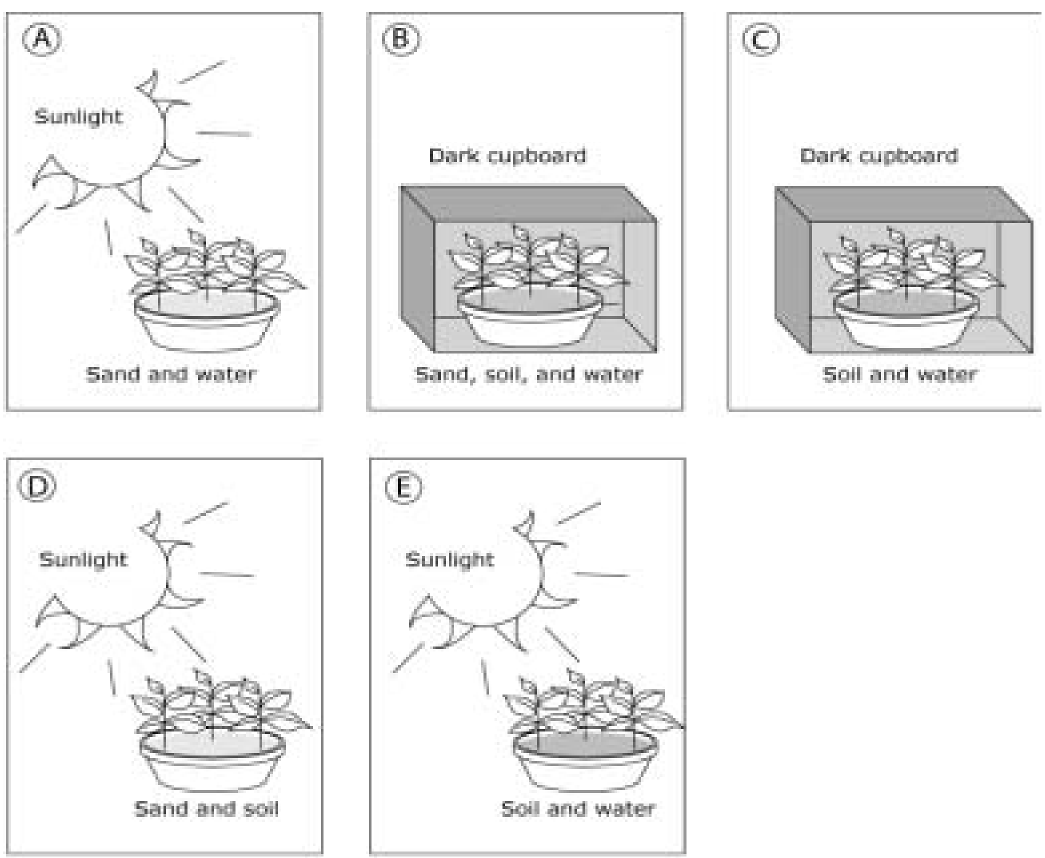
Appendix E: Scientific Inquiry Item Examples

Item 1: Multiple choice (TIMSS)

A girl has an idea that green plants need sand in the soil for healthy growth. In order to test her idea she uses two pots of plants. She sets up one pot of plants as shown below.



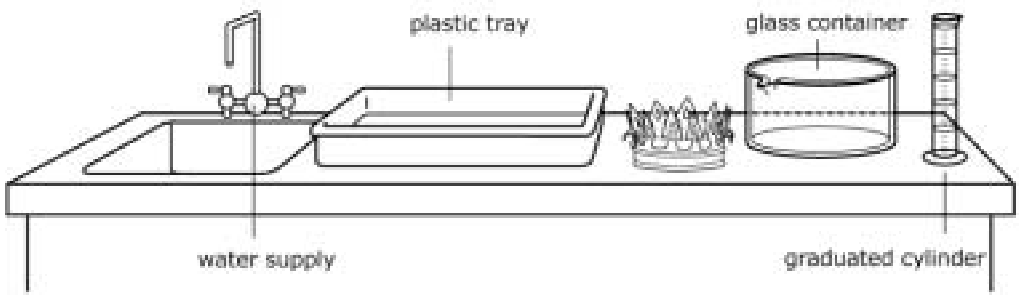
Which ONE of the following should she use for the second pot of plants?



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

Item 2: Extended constructed response (TIMSS)

The scientists then needed to find the volume of the crown in order to determine its density. The following equipment and materials were available for them to use.



Describe a procedure that the scientists could use to find the volume of the crown using some or all of the equipment and materials shown above. You may use diagrams to help explain your procedure.

NOTE: This item is part of a problem-solving and inquiry (PSI) cluster.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Appendix F: Technical Notes

The methodology employed in considering the knowledge and skills assessed by the TIMSS and PISA 2003 assessments consisted of a targeted coding of the assessment instruments with respect to the original assessments' design and with respect to a number of variables developed especially for use in this study. These variables, detailed in the body of the report, focused on the content, context, specific knowledge and skills, and cognitive processes associated with the individual assessment items, and with the assessments as a whole. As noted earlier, the analyses utilize the PISA and grade 8 TIMSS items to allow for as much comparability between the two studies as possible.

While there are a number of variables that can be used to examine the nature of the items in large-scale assessments such as TIMSS and PISA, an attempt was made to restrict analyses to variables that could be coded as meeting or not meeting criteria in order to improve the reliability of the judgments made. The variables chosen and described below and in the body of the report were ones that have a direct and easily interpretable relationship to the findings of the study from a practical and applicable standpoint.

Based on a review of the literature (see section I for a detailed discussion) and on models used in other analyses of items in international assessments conducted for NCES (see Nohara 2001; Neidorf et al. 2006; Neidorf, Binkley, and Stephens 2006), problem solving was defined as a situation where an individual's known attempts or ideas for resolving a situation do not work. In these cases, the individual must consider new vantage points or simplify the problem to a workable one. The behavior of the individual and the nature of the approaches used by that individual provide evidence that he or she is working on a problem. Thus, a problem-solving item was identified when

- the context allows students to be engaged,
- students do not have a known strategy to immediately apply, and
- the situation calls for a solution.

All 383 items included in the TIMSS 2003 grade 8 assessment and all 139 items included in the PISA 2003 assessment were reviewed. Items were coded as problem-solving items if they required students to resolve a situation that, most likely, had not been explicitly studied or for which the student would not have a ready procedure. Once identified as a problem-solving item, items were classified based on a number of variables related to the task posed to students. The coding variables are detailed below. In some cases, items could be included in more than one classification, while in others, an item could receive only one classification.

Several versions of each possible variable were considered. The coding of items by the three report authors was checked, where possible, against each assessment's original design features and categorization of items. After preliminary coding and an examination of the results within each content area and each assessment, sets of variables and categories within each variable appropriate to each content area and assessment emerged. The three authors then individually coded the items and submitted their codings. These values were then analyzed for agreement. In cases where the three authors agreed or two of the three agreed, the rating of the agreeing authors was used. In the few cases where all three authors disagreed on a code, the item was discussed to arrive at a mutually agreeable coding. In all cases, differences about any item were communicated to all three authors so that any minority opinion could be stated and discussed prior to the use of any code for an item in further analyses.

Content Coverage

Each item in PISA and TIMSS was classified by the original item developers according to its content, based on the content areas covered in the PISA and TIMSS 2003 frameworks (OECD 2003; Mullis et al. 2003). These are the classifications used in this report (see exhibit 1).

Cognitive Processes

Items were coded as belonging to a single cognitive process/competency class by the authors in terms of cognitive processes detailed in the PISA and TIMSS 2003 frameworks. Each item in PISA and TIMSS was classified by the original item developers according to its cognitive process (TIMSS) or competency requirement (PISA). These are the classifications used in this report (see exhibit 1).

Problem-Solving Attributes

Items were coded with respect to various problem-solving attributes, including

- identify variables or relationships (see items 1 and 2, appendix C),
- critically evaluate information (see item 3, appendix C),
- justify/prove solution (see item 4, appendix C),
- generalize or predict applicability (see item 5, appendix C),
- communicate solution (see item 6, appendix C),
- require science knowledge (for science items only; see item 7, appendix C), and
- integrate or synthesize information (see item 8, appendix C).

Each item could be classified as having one or more of these attributes according to what it required from the student.

Item Format

Both the TIMSS and PISA assessments included a variety of item formats. Items were coded with respect to the following item formats (see appendix B for examples of these item formats):

- *Simple multiple-choice* items ask students to select from a list of alternatives (see item 1, appendix B).

- *Complex multiple-choice* items ask students to respond to a series of “true/false” or “yes/no” items (see item 2, appendix B).

- *Short constructed response (SCR)* items call for a computational or a short verbal response. SCR items can be of the following two types:

    - *Closed SCR* items have one possible response or solution method (see items 3 and 4, appendix B).

    - *Open SCR* items allow different answers or have the possibility of many different ways of arriving at the solution (see item 5, appendix B).

- *Extended constructed response (ECR)* items require several steps and a more lengthy response to explain the answer. ECR items can be of the following two types:

    - *Scaffolded ECR* items are presented as a number of smaller questions that provide structure for students' responses and direct the approach taken to some degree. The students are led via a series of questions, often labeled a, b,… , to answer several parts of an extended question. As such, the students are guided to a solution using a specific problem-solving approach (see item 6, appendix B).

    - *Open ECR* items tend to ask one large question in which the solution strategy and the nature and structure of the response are left open to the student (for an example of an open ECR item, see item 7, appendix B).

Each item was classified as belonging to one and only one of the item format classes.

Computational Aspects of Items

Given that many problem-solving items require the determination of a value or some comparative measure, computation can play a significant role in problem-solving situations. Items were coded with respect to the computational load they placed on the problem solver.

An item was judged to have a *computational load beyond basic* if it required computations beyond straightforward work with whole numbers, fractions, and decimals or the solution of a simple linear equation involving whole numbers or integers. Otherwise, an item was coded as not having a computational requirement (see items 1 and 2, appendix A).

Translations of Representations

Part of successful problem solving involves recognizing the nature of the information provided in a problem and working with that information in another form. This may involve taking information from a table or chart and calculating percentages, or it may involve examining the spatial arrangement of objects relative to a particular object and determining the degree to which the position of a given object affects the positioning of other objects. As a result, items were coded as having one or more of the following translation of representation features:

- developing a drawing or sketch (see item 1, appendix D);
- interpreting a figural representation (see item 2, appendix D);

- interpreting a statistical representation (see item 3, appendix D);
- interpreting a functional representation (see item 4, appendix D);
- interpreting a tabular representation (see items 5 and 6, appendix D);
- interpreting a graphical representation (see item 7, appendix D); and
- interpreting an informational passage (see item 8, appendix D).

Each item was classified as belonging to one and only one of these categories.

Data Processing

After the the three authors came to agreement on the classification of items, a dataset was produced that included the classifications for all items from each assessment. The raw data containing the final classifications was used in the writing of this report.

Interrater Reliability

In the process of coding of the items in the assessments studied in this project, data was collected and analyzed to examine the consistency of the three raters in coding items as representing problem solving or not. There are several different coefficients used to provide a judgment of how consistent raters apply the criterion in selecting an item as a problem solving item or not. Perhaps the most commonly used measure of such coding for multiple raters is Craig's generalization (Craig 1981) of Scott's π coefficient (Scott 1955). Holsti (1969) suggests that Scott's approach is computationally simple and Craig notes that it is possible to extend Scott's approach to multiple raters, where Cohen's *kappa* cannot. Further, Scott's generalized formula can be applied to find reliability coefficients for *m/h* subsets such as *m* out of *h* raters. In this study, the observations of reliability were limited to cases of total agreement of 3/3 reliability among the raters.

Since there were only three raters in this study, any group of three has to have two raters in agreement one way or another in a binomial rating situation. Thus, all ratings reported below are limited to total agreements in rating of the three raters, or authors, of this report. The data for the following rating calculations consisted of a rating for each of the three raters to each of the items in each of the assessments' item sets.

As the Scott and generalized Scott π coefficients are both generalizations of a *kappa*-like measure, one can use the varied categories that have been developed to interpret the meaning of the values of π (Von Eye and Mun 2005). As Cohen's, Scott's, and Craig's generalization of Scott's measures are all conservative tests, somewhat lower criteria are used to determine the meaning of these coefficients (Landis and Koch, 177; Fleiss 1981). Landis and Koch suggest the following for *kappa*-like measures such as π:

$0.00 \leq \pi \leq 0.20$	slight
$0.21 \leq \pi \leq 0.40$	fair
$0.41 \leq \pi \leq 0.60$	moderate
$0.61 \leq \pi \leq 0.80$	substantial
$0.81 \leq \pi \leq 1.00$	almost perfect agreement.

Fleiss' coefficient interpretation chart is quite similar:

$0.00 \leq \pi \leq 0.40$	poor agreement
$0.40 \leq \pi \leq 0.75$	good, and
$0.76 \leq \pi \leq 1.00$	excellent agreement.

The following table includes the interrater reliability coefficients found for the six different item sets discussed in this report.

Table A. Interrater reliability coefficients for the PISA and TIMSS 2003 item sets

Assessment	Scott's Generalized π
TIMSS mathematics items	0.74
PISA mathematics items	0.84
TIMSS science items	0.71
PISA science items	0.63
TIMSS PSI items	0.71
PISA Cross-disciplinary items	0.69

An analysis of the values in the table shows that the coding of the PISA mathematics subtest reached the highest level of almost perfect agreement or excellent agreement. The remaining five interrater reliability coefficients were judged to such substantial or good association depending on the coding description system employed.

Given that these values for the generalized Scott π coefficient were calculated on the individual rater's first codings after the initial training session and do not involve any of the following discussion of the ratings, these values are deemed satisfactory as a basis for a first approximation of the reliability of a joint view of problem solving. The three investigators then discussed the items and moved to agreement on each item in contention.

Tests of Significance

Statistical tests of the difference in proportions of items allocated to problem solving or one of the other categories of comparisons were carried out using χ^2 analyses employing Yates' correction for continuity for 2 x 2 comparisons (Fleiss 1981) and the G^2 likelihood-ratio form of the chi-square test for R x C comparisons (Agresti 1996).

For example, comparisons of the type that could be represented as a direct comparison of two proportions, such as shown in the table below, were analyzed using the standard χ^2 analysis for the significance of the difference of two proportions from 0. Given that the number of items making up these proportions was often quite small, Yates' correction for continuity (Fleiss 1981; Yates 1934) was employed in each comparison to adjust for the discrete nature of the data being analyzed.

The analysis of two-by-two tables of the form

	TIMSS	PISA
Has property P	A	B
Lacks property P	C	D

result in the value $\chi^2 = \dfrac{N(|AD - BC| - \frac{1}{2}N)^2}{(A+B)(C+D)(A+C)(B+D)}$ where $N = A + B + C + D$. This value of χ^2 has 1 degree of freedom.

When the number of rows or columns in the comparison of differences of proportions being analyzed exceeded 2, then the G^2 likelihood-ratio form of the chi-square test for R x C comparisons (Agresti 1996) was employed where k groups are being compared. In the n_i group the items are comprised of r_i items meeting the category definition and $n_i - r_i$ not for $i = 1, 2, 3, \ldots, k$. Further, $p_i = r_i/n_i$ for $i = 1, 2, 3, \ldots, k$. Further, n equals the sum of the n_i and r equals the sum of the r_i. Then, the value of G^2 is given as follows:

$$G^2 = 2\sum_{i=1}^{k}\left[r_i \log\left(\frac{p_i}{p}\right) + (n_i - r_i)\log\left(\frac{1-p_i}{1-p}\right)\right], \text{ where } p = r/n.$$

The value of G^2 is interpreted as a having a χ^2 distribution with (R – 1)(C – 1) degrees of freedom.