

TRENDS IN INTERNATIONAL MATHEMATICS AND SCIENCE STUDY

TIMSS

2011 ITEM WRITING GUIDELINES

Ina V.S. Mullis and Michael O. Martin
Executive Directors



TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

© IEA, 2011

Table of Contents

Summary of TIMSS 2011 Item Writing Process and Guidelines.....	2
Introduction	4
About the TIMSS Tests	4
The Purpose of the Guidelines	5
General Issues in Writing Items for TIMSS	5
Alignment with the Frameworks	6
Types of Items	6
Testing Time.....	7
Grade Appropriate Language and Context.....	7
Item Difficulty.....	8
Avoiding Bias	8
Facilitating Comparable Translation	9
Problems Involving Money	9
Graphics.....	10
Copyright.....	10
Algebra Pattern Items—Mathematics ONLY	10
Use of Calculators—Eighth-Grade ONLY	11
Writing Multiple-Choice Items	11
The Stem	12
Structure of the Response Options (or Alternatives)	13
Plausibility of Distracters	14
Writing Constructed-Response Items and Scoring Guides	15
Guidelines for Writing Constructed-Response Items	16
Writing Scoring Guides	17
Documenting the TIMSS 2011 Items	24
Reviewing Items and Scoring Guides	24
Reviewing Multiple-choice Items	25
Reviewing Constructed-response Items.....	26
Appendix A: Multiple-choice Item Review Checklist	27
Appendix B: Constructed-response Item and Scoring Guide Review Checklist	28

Summary of TIMSS 2011 Item Writing Process and Guidelines

The TIMSS 2011 item-writing will be conducted in four major areas, with approximately a quarter of the participants working in each area – mathematics fourth and eighth grades and science fourth and eighth grades. Typically, participants will work in groups of two or three. Each group will be assigned specific content areas. Participants will be writing items in English and saving them as Microsoft Word files that will be collected at the end of each day.

When writing items, PLEASE:

1. Address the TIMSS 2011 Framework. Write questions that match the topics in each content domain, and pay particular attention to writing questions that cover the range of the three cognitive domains. In accordance with the TIMSS 2011 Frameworks, write questions that address the applying and reasoning domains, as well as the knowing domain.
2. Consider the best item format for the question. About half of the items you develop should be multiple-choice and the other half should be constructed-response items worth 1 or 2 score points.
3. For the content domain(s) you are assigned, write at least 10 to 12 items each day.
4. For each item, consider the timing, grade appropriateness, difficulty level, potential sources of bias (cultural, gender, or geographical), and ease of translation. Make sure that item validity is not affected by factors that unnecessarily increase the difficulty of the item, such as unfamiliar or

overly difficult vocabulary, grammar, directions, contexts, or stimulus materials.

5. For multiple-choice items, keep the guidelines for writing multiple-choice questions in mind. In particular—ask a direct question, make sure there is one and only one correct answer, and provide plausible distracters.
6. For constructed-response questions, write a full-credit answer to the question in terms of the language, knowledge, and skills that a good fourth- or eighth-grade student could be expected to possess. This tests the clarity of the question and also provides guidance about whether to allocate 1 or 2 score points to the item.
7. Develop a specific scoring guide for each constructed-response item.

HAVE A GOOD TIME!

Introduction

About the TIMSS Tests

TIMSS, IEA's Trends in International Mathematics and Science Study, assesses mathematics and science achievement at the fourth and eighth grades across a large number of countries, cultures, and languages. TIMSS is designed to help countries improve the teaching and learning of mathematics and science.

The TIMSS tests provide two major types of data:

- Comprehensive and internationally comparable data about the mathematics and science students have learned by the fourth and eighth grades
- Trends in mathematics and science learning across time for students at the fourth and eighth grades.

To ensure a sufficient number of high-quality items for the TIMSS 2011 tests in mathematics and science, there will be an ambitious field test during March-April 2010. TIMSS will field test approximately twice as many items as needed to complete item development for the final tests. At each grade, TIMSS released 6 out of the 14 blocks assessed in TIMSS 2007, and these 6 blocks need to be replaced. Thus, $6 \times 2 = 12$ blocks of items will be field tested for each grade and subject. More specifically, the field test will include:

For mathematics,

- Fourth grade – 12 blocks containing approximately 140 items in total
- Eighth grade – 12 blocks containing approximately 170 items in total

For science,

- Fourth grade – 12 blocks containing approximately 140 items in total
- Eighth grade – 12 blocks containing approximately 170 items in total

The Purpose of the Guidelines

These guidelines are to help ensure that the best possible items are developed for TIMSS 2011. The TIMSS & PIRLS International Study Center has developed these guidelines for writing and reviewing items and scoring guides to facilitate the success of item development. It is important to follow some basic procedures so that the TIMSS test is uniform in approach and format. During the item-writing sessions, please ask staff or consult these guidelines if you have any questions.

General Issues in Writing Items for TIMSS

Item writing is a task that requires imagination and creativity, but at the same time demands considerable discipline in working within the assessment framework and following the guidelines for item construction provided in this manual. These guidelines pertain to good item and test development practices in general, and have been collected from a number of sources. They are designed to help produce items that measure achievement in mathematics and science fairly and reliably, and that enhance the validity of the TIMSS tests. All of the following issues must be considered in judging the quality and suitability of an item for TIMSS 2011. Items must meet these guidelines to be considered for inclusion in the field test for TIMSS in 2010.

Alignment with the Frameworks

The TIMSS assessment frameworks in mathematics and science describe those outcomes generally regarded as important at the fourth and eighth grades. It is fundamental that every item written for mathematics or science measures one of the content topics and one of the cognitive domains described in the TIMSS 2011 frameworks. In preparing to produce an item for either fourth or eighth grade, the first step is to focus on the content topic to be assessed. In writing each item, remember that it also contributes to a measure of proficiency in a cognitive domain. Keep in mind that TIMSS is assessing student learning in particular topics. Think:

- What should the student know?
- What should the student be able to do?

That is, what knowledge does this item allow a student to show? What cognitive processes does this item require a student to demonstrate?

Types of Items

TIMSS includes two types of items: multiple-choice items where the student chooses the correct answer from four response options, and constructed-response items where the student is required to provide a written response.

PLEASE keep item format in mind. About half of the items you develop should be multiple-choice and half should be constructed-response.

- **Multiple-choice items** allow valid, reliable, and economical measurement of a wide range of content in a relatively short testing time.

- **Constructed-response items** allow students to provide explanations, support an answer with reasons or numerical evidence, draw diagrams, or display data.

If you think of another item type, it may be used as long as it provides valid measurement and is feasible to administer and to score reliably.

Testing Time

In developing items, it is important to consider the time required for students to complete the required task. The amount of time required to complete an item should be consistent with the time allotment for items in the overall test design. As a general rule, a multiple-choice item is expected to require about 1 minute or less to complete, and constructed-response items are allocated 1-3 minutes. Items should be designed to require the appropriate amount of time. It can be noted that extended constructed-response tasks typically consist of several items.

Grade-Appropriate Language and Context

The language, style, and reading level used in items must be appropriate to the age and experiences of the students in the target grades. Keep the language as simple as possible, and take care to use grade-appropriate vocabulary and terms. The reading level of items should be at an elementary level for the target grade. Write questions in the active voice (i.e., doer of action (subject) before action (verb)) and avoid conditional words, clauses, and tenses (e.g., if, suppose, when).

The context for the item may relate only to the discipline of mathematics or science, or to aspects of those subjects encountered in everyday life. However, if the item involves a “real-world” setting make sure the setting is one that is familiar to students. An

unnecessarily complicated item context may artificially increase the difficulty of the item and pose a threat to item validity.

Item Difficulty

Information from individual TIMSS items should provide valuable insights into student learning. Additionally, each of the items needs to contribute to the overall mathematics or science test. It is desirable that there be some relatively easy items and some challenging items. However, items that almost all students or almost no students are able to answer correctly reduce the effectiveness of the test to discriminate between groups with high achievement and groups with low achievement. Typically, the majority of items used in the final test will be ones that are answered correctly by 30 to 70 percent of the students.

Avoiding Bias

In preparing test items, be sensitive to the possibility of unintentionally placing particular groups of students at an unfair disadvantage. An international study requires special attention to the diversity of environments, backgrounds, beliefs, and mores among students in the participating countries.

Considering National Contexts

Be particularly aware of issues related to nationality, culture, ethnicity, and geographic location. Items requiring background knowledge confined to a subset of participating countries are unlikely to be suitable.

Geographic location has an effect on the learning experiences students are exposed to, as aspects of the local environment have an impact on schooling. Even though television and the Internet can provide students with some knowledge of remote places, firsthand experience of some

phenomena enhances understanding and can give some students an advantage over others.

Gender

A gender-related context included in an item may distract some students from the purpose of the item. Situations in which stereotypical roles or attitudes are unnecessarily attributed to males or females, or in which there is implicit disparagement of either gender, are not acceptable.

Facilitating Comparable Translation

The international version of items will be in American English. Therefore, items developed at this meeting must be submitted in English. Keep in mind, however, that after review and revision, the items selected for the field test and main data collection will be translated from English into the languages of instruction of the countries in the study.

Please be sensitive to issues that might affect how well items can be translated to produce internationally comparable items. The TIMSS translation procedures do allow names and places to be changed to what is appropriate for a country, provided the essential nature and difficulty of the item are not altered. Idioms and expressions that defy translation must be avoided.

Problems Involving Money

Problems involving computations with money, especially those set in “real life” contexts, are problematic for international studies. The cost of a common article in one country may be a fraction of the base unit of currency, while the same article in another country may cost thousands of the base unit. In some countries, the cost of an article may never include a decimal point. If the inclusion of

costs is an essential part of a problem, use “Zeds”. This is the TIMSS fictitious unit of currency, which enables each country to work with the same numbers.

Graphics

Take special care to ensure that diagrams and graphs are drawn accurately (to scale unless otherwise noted), and are correctly and fully labeled. Any graphics included in an item should be necessary in order to solve the problem and should be adequately explained and referred to directly within the item. Graphics for items may be submitted as a paper document.

Copyright

All of the items developed for TIMSS 2011 will be copyrighted by the IEA. For this copyright to be valid, it is important that TIMSS items do not infringe on other copyrights. All of the items used in TIMSS 2011 must be specifically developed for TIMSS 2011 and never used in other tests. Also, in developing items for TIMSS, any copyrighted stimulus material must be acknowledged appropriately. For example, statistical graphs from publications or extracts from articles in publications that are used in an item must be identified appropriately, and full details about the sources must be submitted with the item.

Algebra Pattern Items—Mathematics ONLY

In developing algebra items assessing the objective “extend patterns or find missing terms in patterns”, the patterns must be well defined in the question stem. The stem needs to describe that the pattern “repeats every four shapes” or “increases by the same amount from one number to the next.” For example, the stem might state: “Ellen made a number pattern using the rule add 4.” For geometric pattern items, it is possible to say “the same rule is used to get from one figure to the next” (if it is the same, e.g., adding 2 circles and a square

to each figure). Often the algebra pattern items should be in the constructed-response format asking students to justify or explain the rule for the pattern.

Use of Calculators—Eighth-Grade ONLY

In TIMSS 2011, calculators will be permitted, but not required, at the eighth grade. In developing new eighth-grade mathematics and science items, every effort should be made to ensure that the items do not advantage or disadvantage students either way – with or without calculators. Calculators are not permitted at the fourth grade.

Writing Multiple-choice Items

A multiple-choice item asks a question or establishes the situation for a response. For TIMSS 2011, this type of item provides four response choices, or options, from which the correct answer is selected. A multiple-choice item is characterized by the following components:

- The *stem* is the initial part of the item in which the task is defined.
- The *options* refer to the entire set of labeled response choices presented under the stem.
- The *key* is the correct response option.
- The *distracters* are the incorrect response options.

At least half of the items developed for TIMSS 2011 will be multiple-choice items. The next sections present guidelines specific to multiple-choice items, including writing the stem, structuring the response options, and developing plausible distracters.

PLEASE keep the guidelines for writing multiple-choice questions in mind. In particular, ask a direct question, make sure there is one and only one correct answer, and provide plausible distracters.

The Stem

For TIMSS 2011, since clarity is of vital importance, please phrase all stems as a **direct question**.

Example of a stem formulated as a question:

Which group of numbers is ordered from LARGEST to SMALLEST?

- (A) 10,011; 10,110; 11,001; 11,100
- (B) 10,110; 10,011; 11,100; 11,001
- (C) 11,001; 11,100; 10,110; 10,011
- (D) 11,100; 11,001; 10,110; 10,011

M042003

1. Provide sufficient information in the stem to make the question clear and unambiguous to students. In nearly all cases, the question must be able to stand alone, and be answerable without the response options. An exception would be items asking which is the best estimate of a quantity.
2. The stem should not include extraneous information. Extraneous information is liable to confuse students who otherwise would have determined the correct answer.
3. Avoid using negative stems—those containing words such as NOT, LEAST, WORST, EXCEPT, etc. If it is absolutely necessary to use a negative stem, highlight the negative word, (e.g., capitalize, underline, or put in bold type so that it stands out for the student). If the stem is negative, use only positive response options —do not use double negatives.

4. If there is not one universally agreed upon answer to the question, it is best to include “of the following” or some similar qualifying phrase in the stem.
5. Avoid questions for which a wrong method yields the correct answer (e.g., a question about a circle with a radius of 2, since $2r=r^2$ students computing either the area or the circumference get 4π).

Structure of the Response Options (or Alternatives)

1. Write multiple-choice items with **four** response options, labeled A–D (as shown in the example item about ordered numbers).
2. Make sure that one of the four response options or alternatives is the **key or correct answer**. Make sure there is only **one** correct or best answer.
3. Make sure that the four response options are independent. For example, response options should not represent subsets of other options. Also, do not include pairs of response options that constitute an inclusive set of circumstances (e.g., day or night, does or does not).
4. Make sure that the grammatical structure of all response options “fit” the stem. Inconsistent grammar can provide clues to the key or eliminate incorrect response options. Avoid writing items where the options complete a sentence begun in the stem, because these can cause problems with translation.
5. Make sure all (or sets) of the response options are parallel in length, level of complexity, and grammatical structure. Avoid the tendency to include more details or

qualifications in the correct response, thus making it stand out. If the best options are not parallel in length, please order the options short to long if at all possible.

6. Do not use words or phrases in the stem that are repeated in one of the response options and, therefore, act as a clue to the correct response.
7. Do **NOT** use “none of these” and “all of these” as response options.
8. Arrange the response options in a logical order if this makes sense and saves the student time in reading the options (e.g., years in chronological order, numbers from least to greatest).
9. Avoid writing items where students can work backwards from the response options to find the correct answer (e.g., solving for x in an equation). Sometimes described as “plug and chug” items, such questions or problems will not be included in TIMSS 2011. In such cases, a constructed-response item may be more appropriate than a multiple-choice item.

Plausibility of Distracters

Use plausible distracters (incorrect response options) that are based on likely student errors or misconceptions. This reduces the likelihood of students arriving at the correct response by eliminating other choices and, equally important, may allow identification of widespread student misunderstandings or tendencies that could lead to curricular or instructional improvements. If there are no plausible errors or misconceptions, still make the options “reasonable.” For example, they should be from the same area of content. However, avoid the use of “trick” distracters.

Writing Constructed-response Items and Scoring Guides

For some desired outcomes of mathematics and science education, constructed-response items provide more valid measures of achievement than do multiple-choice items. The quality of constructed-response items depends largely on the ability of scorers to assign scores consistently and reliably within and across countries. Thus, it is essential that each constructed-response item and its scoring guide be developed together.

PLEASE keep the guidelines for writing constructed-response questions in mind. In particular, ask a clear question, and develop a scoring guide for the question at the same time as the question is developed.

Constructed-response items usually require students to give a numerical result, provide a short explanation or description given in one or two phrases or sentences, complete a table, or provide a sketch. They are scored as either 1 or 2 points for fully-correct answers.

- 1-point constructed-response items are scored as correct (1 score point) or incorrect (0 score points).
- 2-point constructed-response items are scored as fully correct (2 score points), partially correct (1 score point), or incorrect (0 score points). For example, a response demonstrating thorough understanding of concepts and processes will receive full credit (2-points). These responses show a complete or deeper understanding than a response that will receive partial credit (1-point). (Developing scoring guides is explained in the next section.)

Constructed-response items should be used when it is desirable that the student be required to think of an answer without the possibility of being cued by an option in a multiple-choice item. If too few plausible distracters are available for a multiple-choice item, it may be better framed as a constructed-response item.

Developing a constructed-response item accurately targeted on the ability to be assessed, along with the accompanying scoring guide, is not a straightforward task. Care in writing constructed-response items is especially important for two reasons. First, if the task is not well specified students may interpret the task in different ways and respond to different questions. Second, a constructed-response item may carry more score points than a multiple-choice item.

Guidelines for Writing Constructed-response Items

1. Write questions in language appropriate to the age and experience of the target population. If students are uncertain about what they are to do the difficulty of the item will increase, and this may be falsely attributed to lack of student proficiency.
2. Make what is expected of students as clear as possible without compromising the intent of the item. Students will not be allowed to ask the test administrator for clarification. Use words such as “explain” or “describe” to focus students on the task rather than “fuzzy” words such as “discuss” or “comment” that can lead to wide variation in the content of responses.
3. Students should be able to complete the task in the time allocated for each constructed-response item, that is, 1–3 minutes.

4. Select real life problem settings that are likely to be “real” to students at the target grade levels, and that involve quantities that are realistic for the situations.
5. Avoid asking questions that could give rise to answers that cannot be scored strictly in terms of accuracy of mathematical or scientific understanding (e.g., “What are satellites used for?”).
6. Give an indication, where appropriate, of the extent, or level of detail, of the expected answer (e.g., “Give three reasons ...” rather than “Give some reasons ...” and “Draw a labeled diagram illustrating the water cycle” rather than “What is meant by the term ‘water cycle?’”).
7. Write an appropriate answer to the question in terms of the language, knowledge, and skills that a good student at the target grade could be expected to possess. This tests the clarity of the question and is also an essential first step in producing a scoring guide for the item. It is also helpful for those who are reviewing the question.
8. Produce a scoring guide (see below). This action usually results in amendments to the item to clarify its purpose and improve the quality of information that can be obtained from student responses.

Writing Scoring Guides

To ensure reliability, constructed-response items need scoring guides with well-defined categories for allocating score points. It also is important to collect information of value for educational improvement. Students’ answers can provide insights into what they know and are able to do, including common misconceptions.

The TIMSS Generalized Scoring Guidelines

The generalized scoring guidelines used for 1- and 2-point constructed-response items are described in Table 1.

Table 1: TIMSS Generalized Scoring Guidelines for Constructed-response Items

Score Points for 1-point Items
1 Point A one-point response is correct. The response indicates that the student has completed the task correctly.
0 Points A zero-point response is completely incorrect, irrelevant, or incoherent.
Score Points for 2-point Items
2 Points (Full credit) A two-point response is complete and correct. The response demonstrates a thorough understanding of the concepts and/or procedures embodied in the task. <ul style="list-style-type: none">• Indicates that the student has completed all aspects of the task, showing the correct application of concepts and/or procedures• Contains clear, complete explanations, supporting work, or evidence when required
1 Point (Partial Credit) A one-point response is only partially correct. The response demonstrates only a partial understanding of the concepts and/or procedures embodied in the task. <ul style="list-style-type: none">• Addresses some elements of the task correctly but may be incomplete• May contain a correct answer but with an incomplete explanation when required• May contain an incorrect answer but with an explanation or supporting work indicating a correct understanding of the concepts
0 Points A zero-point response is inaccurate or inadequate, irrelevant, or incoherent.

The TIMSS Two-digit Diagnostic Scoring System

The TIMSS diagnostic scoring system uses two digits. For example, 10, 11, or 20.

The **first digit** is the score indicating the degree of correctness of the response as described in the generalized scoring guidelines.

The **second digit** is used to classify the method used in solving a problem, or perhaps to track common errors or misconceptions. The information from the second digit addresses questions such as: Do approaches that lead to correct responses to the item vary across countries? Is there one approach that students have more success with than others? What are the common misconceptions that students have about the matter being tested? What common errors are made?

The First Digit

The **first digit for correct or partially correct responses** signifies the number of score points given to the response. Thus:

The **first digit for correct responses** is **1** for one point or **2** for two points.

When TIMSS started in the early 1990s, it was decided not to use 0 for the first digit. Thus:

The **first digit for incorrect responses** is **7**.

The **first digit for a blank response** is **9**.

The Second Digit

The **second digit for correct or incorrect responses** provides diagnostic information. Thus:

The **second digits** used for diagnostic purposes with either **correct or incorrect responses** can be **0 through 5**

(codes **20–25**, **10–15**, and **70–75**).

However, it is unusual for an item to give rise to more than two correct methods, or more than one common error or misconception. Frequently no more than one or two categories are required. In other words, the specific diagnostic codes should capture only the predominant correct and incorrect approaches/strategies used by students. Scoring of constructed-response items is a significant cost factor for national centers, so care should be taken not to provide codes for response types that do not have apparent value for educational improvement.

Since not all incorrect student responses should be categorized into pre-defined categories, for codes with a first digit of 7 the second digit of 9 is used to designate a response that is “other” than any specific diagnostic codes included in the guide. Thus, an incorrect response not fitting a pre-defined incorrect code is given a 79 for “other incorrect.” If no diagnostic categories are defined, all incorrect responses are coded 79.

The **second digits** of 7 and 8 are reserved for **National Options**.

The second digits 7 and 8 (e.g., 17, 18, or 27, 28, or 77, 78) have been reserved for countries that wish to track particular responses frequently occurring in that country and that have no pre-defined category in the international scoring guide. If a country wishes to define these codes to obtain information pertinent to its needs, it may do so for local processing and purposes. However, for the purposes of international comparisons, all such codes not specifically identified in the scoring guides will be recoded to 19, 29, or 79 codes in the TIMSS international database.

Code 99 means a completely **BLANK** response.

Examples of Scoring Guides

The following examples are given to illustrate the diagnostic scoring guides used in TIMSS.

Grade 8 mathematics item (1 point):

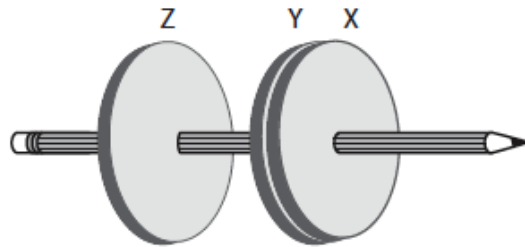
Joe knows that a pen costs 1 zed more than a pencil.
 His friend bought 2 pens and 3 pencils for 17 zeds.
 How many zeds will Joe need to buy 1 pen and 2 pencils?

Show your work.

M042263

Code	Response	Item: M042263
	Correct Response	
10	10 zeds and equation(s) shown. Equations should involve the use of letter(s) as variable(s), e.g., $2y + 3x = 17$.	
11	10 zeds and other work shown, e.g., pen = pencil + 1	
	Incorrect Response	
70	10 zeds, no work shown	
79	Other incorrect (including crossed out, erased, stray marks, illegible, or off task)	
	Nonresponse	
99	Blank	

Grade 8 science item (2 points):



The diagram shows what happens to three magnets when they are placed close together on a pencil.

Magnets X and Y move until they touch each other, but magnets Y and Z remain separated.

1. Explain why magnets X and Y touch each other.

2. Explain why magnets Y and Z remain separated.

S042292

Code	Response	Item: S042292
Correct Response		
20	Refers to unlike poles (NS or SN) AND like poles (NN or SS). Attraction and repulsion may or may not be mentioned. <i>Examples:</i> 1. Magnets X and Y touch each other because the north and south poles were facing. 2. Magnets Y and Z remain separated because they may have had south and south or north and north facing each other. 1. Magnets X and Y attract each other as they have unlike poles facing each other. 2. Magnets Y and Z repel as they have like poles facing each other.	
Partially Correct Response		
10	Refers to unlike poles (NS or SN) OR like poles (NN or SS). Attraction and repulsion may or may not be mentioned. <i>Examples:</i> 1. Side Y is facing side X (incorrect). 2. The side of Y facing Z has the same, e.g. south and south (correct). 1. Magnets X and Y attract because they aren't the same (correct). 2. Magnets Y and Z remain separated because they are north and south (incorrect).	
11	Mentions attraction and/or repulsion but does not explain why. <i>Examples:</i> 1. Magnets X and Y touch each other because the magnetic forces pull the magnets together. 2. Magnets Y and Z remain separated because the magnetic forces push apart. 1. Magnets X and Y touch each other because specific sides of a magnet attract. 2. Magnets Y and Z remain separated because specific sides of a magnet resist.	
Incorrect Response		
70	Responses reverse the the poles (i.e., like poles attract and unlike poles repel). <i>Examples:</i> 1. Magnets X and Y touched together because they found south and south together. 2. Magnets Y and Z remained separate because they were north and south.	
71	Responses refer to negative and positive sides. <i>Examples:</i> 1. Magnets X and Y have opposite sides that faced each other so they connected. 2. Magnets Y and Z have the same side, either negative or positive facing each other so they didn't connect.	
79	Other incorrect (including crossed out, erased, stray marks, illegible, or off task)	
Nonresponse		
99	Blank	

Note: If the diagram is correctly labeled with N and S, but the explanation is minimal or not provided, code as 20.

Documenting the TIMSS 2011 Items

During the item-writing sessions, teams will be writing items on computers using Microsoft Word. At the end of each day, the TIMSS & PIRLS International Study Center will collect the files from each team.

When entering the TIMSS items, please use the template that has been provided and complete the necessary documentation as described below.

Filename: Subject, grade, and team number (to be assigned)

For each individual item, provide:

1. The TIMSS Content Domain, topic area, and objective the item measures
2. The TIMSS Cognitive Domain and sub-area the item addresses
3. The item number (1, 2, 3, etc.)
4. The key (multiple-choice items only) or
5. The scoring guide

Reviewing Items and Scoring Guides

Items selected for inclusion in the TIMSS field test or main data collection will go through a thorough review process involving the TIMSS & PIRLS International Study Center staff, the mathematics and science coordinators and consultants, the Science and Mathematics Item Review Committee (SMIRC), and the National Research Coordinators. The first step in this item review process begins with you. Item writers are expected to review and revise their own items in accordance with the procedures outlined here and

presented in the item-writing sessions. In addition, depending on the time available, the items will be reviewed by other item writing teams.

If it happens that items are written after the NRC meeting, the item writers are expected to arrange to have their items reviewed by at least one independent reviewer in their own country. Any problems with items and/or scoring guides detected in the course of this review should be corrected prior to submitting items to the TIMSS & PIRLS International Study Center.

Item writers and item reviewers must be very critical when reviewing items, and the item writers should expect to have to defend their items. The earlier that necessary changes are made to items, the better. Last minute changes to items to remove errors often result in other flaws being introduced.

The following sections provide guidelines for the review of multiple-choice items and constructed-response items together with their scoring guides and are to be used by item writers and reviewers. To facilitate item review, item review checklists for multiple-choice and constructed-response items are provided in Appendix A and Appendix B, respectively, of this manual.

Reviewing Multiple-choice Items

In reviewing each multiple-choice item, item reviewers should:

1. Identify what they consider to be the (only) correct response and compare this with that originally identified by the item writer.
2. Check that their judgments of the TIMSS classifications correspond with those indicated by the item-writing team.
3. Check the item against each of the entries in the Multiple-choice Item Review Checklist (see Appendix A).
4. Identify and note any problems with the item.

Reviewing Constructed-response Items

In reviewing each constructed-response item, item reviewers should:

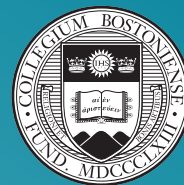
1. Check that their judgments of the TIMSS classification correspond with those indicated by the item-writing team.
2. Check the item against each of the entries in the Constructed-response Item and Scoring Guide Review Checklist (See Appendix B).
3. Write an outline of what s/he believes would be a good response to the item for a student at the target grade. Review the scoring guide for the item, comparing it with your response, to make sure that you agree with the number of score points allocated and the clarity of the distinction made between the levels. Also, see if the most likely types of student responses have been categorized.
4. Identify and note any problems with the item and/or scoring guide.

Appendix A: Multiple-choice Item Review Checklist

Item Characteristic	Yes	No
Is the mathematics/science correct?		
Task clear to students?		
Free of cultural, gender, or geographical bias?		
Seems to be OK for translation?		
No unfamiliar factors contributing to difficulty?		
Negative stem avoided (or negative word highlighted if used)?		
One (only) correct response?		
Distracters plausible but demonstrably incorrect?		
Options parallel in structure?		
Words in stem NOT repeated in options?		
Content classification correct?		
Cognitive classification correct?		

Appendix B: Constructed-response Item and Scoring Guide Review Checklist

Item Characteristic	Yes	No
Is the mathematics/science correct?		
Task clear to students?		
Free of cultural, gender, or geographical bias?		
Seems to be OK for translation?		
No unfamiliar factors contributing to difficulty?		
Clear expectations for full-credit response?		
Task can be completed in a reasonable time?		
Scoring guide has appropriate number of score points?		
Scoring guide descriptors clear?		
Content classification correct?		
Cognitive classification correct?		



BOSTON
COLLEGE

timssandpirls.bc.edu



© IEA, 2011
International Association
for the Evaluation of
Educational Achievement