

**TIMSS**  
**2019**

# **TIMSS 2019**

# **Item Writing Guidelines**

Ina V.S. Mullis  
Michael O. Martin  
Kerry E. Cotter  
Victoria A.S. Centurino



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

## Table of Contents

Summary .....	1
Introduction .....	2
General Guidelines for Writing Items for the TIMSS 2019 Assessments .....	3
Item Writing Template .....	8
Writing Multiple-Choice Items .....	12
Writing Constructed Response Items and Scoring Guides .....	18
Writing Items for eTIMSS 2019 .....	26
Reviewing Items and Scoring Guides .....	32
Appendix A: Multiple-Choice Item and Response Rationale Review Checklist .....	34
Appendix B: Constructed Response Item and Scoring Guide Review Checklist .....	35
References .....	36

## Summary

Typically, participants will work in groups of two or three. Each group will be assigned specific content areas. All items will be written in English. During the item-writing sessions, teams will use the TIMSS 2019 Item Template to record their items. The basic requirements for writing items are summarized below.

### When writing items, please:

1. Address the [TIMSS 2019 Assessment Frameworks](#) (Mullis & Martin, 2017). 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 frameworks, write questions that address the applying and reasoning domains, as well as the knowing domain.
2. Consider the **best** item format and design 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 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.
4. 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. For each distracter, write a rationale, such as a common misconception the response option represents or the process by which a student could have arrived at that incorrect answer.
5. For constructed response items, write a full-credit answer to the question in terms of the language, knowledge, and skills that a student in the target grade 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. For each constructed response item, develop a scoring guide with well-defined categories for allocating score points to student responses.
6. Consider if the item is suitable for using the eTIMSS item components (e.g., drop-down menus, drag and drop). In particular, those components are designed to enable computer scoring of constructed response formats.

## Introduction

The TIMSS 2019 assessment cycle may be the most challenging to date because it is the first step in transitioning to eTIMSS. Because not all countries are ready to move to eTIMSS in TIMSS 2019, about half the countries will continue with paperTIMSS while the other half make the transition to eTIMSS. Also, to facilitate the transition to eTIMSS for as many countries as possible, TIMSS is working hard to accommodate a wide range of platforms, including PC/Windows and a variety of Android tablets.

TIMSS 2019 has twin goals for developing new achievement items:

- The need to maintain continuity to measure trends means comparability with paperTIMSS, which links to the past.
- The need to incorporate innovations to be relevant and take advantage of new technologies means building links to the future with eTIMSS.

To report the TIMSS 2019 results, the paperTIMSS and eTIMSS achievement scales need to be linked to maintain trends and to meet the goal of making cross national comparisons in TIMSS 2019. For the scales to be linked, the overlap between eTIMSS and paperTIMSS items is crucial. A substantial number of items (at least half), need to match to enable comparisons across countries and over time. Thus, the priorities for this item writing workshop are:

FIRST priority

- To develop good mathematics and science items that measure the TIMSS frameworks and can be used in both paperTIMSS and eTIMSS.

SECOND priority

- Having developed a good mathematics or science item, consider taking advantage of the eTIMSS item components described on pages 26 to 31.

These guidelines are to help ensure that the best possible items are developed for TIMSS 2019, including mathematics and science, fourth and eighth grades, as well as paperTIMSS and eTIMSS. The TIMSS & PIRLS International Study Center has developed these guidelines for writing and reviewing items and scoring guides to facilitate successful item development. It is important to follow these basic procedures so that the TIMSS assessments are uniform in approach and format. During the item-writing sessions, please ask staff or consult these guidelines if you have any questions.

## General Guidelines for Writing Items for the TIMSS 2019 Assessments

Item writing is a task that requires imagination and creativity, but at the same time demands considerable discipline in working within the assessment frameworks and following the guidelines for item design 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 assessments. All of the following issues must be considered in judging the quality and suitability of an item for inclusion in the TIMSS field tests.

### Alignment with the Frameworks

Consistent with the principles of evidence-centered design (e.g., Mislevy, Almond, & Lukas, 2003), the TIMSS 2019 assessments are based on:

- Detailed content and cognitive domain descriptions organized into frameworks for each assessment
- Items aligned with the content topics and cognitive domains and designed to collect evidence about what students know and are able to do
- Scoring guides with well-defined categories and a detailed description of the kind of responses that belong in each category

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 the two things described in the TIMSS 2019 frameworks:

- One of the **content topics** (mathematics) or **objectives** (science)
- One of the **cognitive domains**

When preparing to produce an item for fourth or eighth grade, the first step is to focus on the content topic or objective to be assessed. When writing each item, remember that it also contributes to a measure of proficiency in a cognitive domain. These two elements together provide evidence about what students know and are able to do. That is, what content knowledge does this item allow a student to show? What cognitive processes does this item require a student to demonstrate? What task best allows the student to demonstrate this knowledge or ability?

Keep in mind that TIMSS 2019 needs to assess student learning in particular topics. Think about the topic:

- What should the student know?
- What should the student be able to do?
- What kind of evidence best demonstrates this knowledge or ability?

The TIMSS 2019 science framework includes a section on science practices fundamental to all science disciplines. Some of the items developed for TIMSS 2019 fourth and eighth grade science should not only focus on a specific objective and cognitive domain, but should also produce evidence that a student can employ skills associated with practicing science.

### Types of Items

TIMSS includes two general types of items—multiple-choice items where the student chooses the correct answer from given response options, and constructed response items where the student is required to construct his/her own answer.

- **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 or results of calculations, support an answer with reasons or numerical evidence, draw diagrams, display data, or create ordered lists.

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.

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

### Testing Time

When 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 for TIMSS 2019 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.

## Grade Appropriate Language and Context

In keeping with the principles of universal design (e.g., Dolan & Hall, 2007) for assessment items and tasks, the language, style, and reading level used in items must be accessible to a range of 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. In general, the amount of reading should be kept to a minimum, given the context of the problem. 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).

If the item involves a “real world” setting make sure the setting is familiar to students. Avoid using context specific vocabulary that may not be familiar to all students. An unnecessarily complicated item context or unfamiliar context-specific vocabulary may artificially increase the difficulty of the item and pose a threat to item validity.

## Item Difficulty

Each item needs to contribute to the overall measurement of mathematics or science achievement. 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 assessment to discriminate between groups with high achievement and groups with low achievement. Typically, the majority of items used in the final assessment will be ones that are answered correctly by 30 to 70 percent of the students on average, internationally.

## Avoiding Bias

When preparing assessment 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 cultures 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 students’ learning experiences, 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.

## Considering 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, all items 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. Do not use idioms and expressions that defy translation across countries. The TIMSS 2019 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.

## 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. If the inclusion of costs is an essential part of a problem, use “Zeds”. This is the TIMSS fictitious unit of currency, which enables students in 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 or to answer the question and should be adequately explained and referred to directly within the item, as indicated by the principles of universal design for assessment items. In particular, visual elements should:

- Correspond to the wording and task presented in the item text
- Depict only information necessary to solve the problem or answer the question so as not to distract or confuse students
- Be included to emphasize an important part of an item if its inclusion makes the item accessible for more students
- Be labeled clearly



During initial item development, digital images or hand drawn elements may be included as draft graphics. Please keep in mind that all graphics must be able to be viewed equally well on a computer screen and on a printed page. In particular,

- When using color or greyscale, choose images with few colors and a limited amount of shading
- Do not reference specific colors in item prompts (e.g., the blue line on the graph represents...)
- Use high quality, high resolution images and .png, .gif, or .jpg file formats

### Copyright

All of the items developed for the TIMSS 2019 assessments will be copyrighted by 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 2019 must be specifically developed for TIMSS 2019 and never used in other assessments. Also, in developing items for TIMSS 2019, 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. All graphics also must be identified properly and complete source information must be provided.

### Use of Calculators—TIMSS 2019 Eighth Grade ONLY

As in the past, eighth grade students participating in paperTIMSS may bring their own calculators to the assessment. However, devices that can connect to the internet are prohibited (i.e., students may not use phones as calculators). For eTIMSS, eighth grade students will have access to an on-screen calculator for the entire assessment and may **not** bring their own external calculators. For both paperTIMSS and eTIMSS, every effort should be made to ensure that the mathematics items are calculator neutral—do not advantage or disadvantage students either way, with or without calculators. As in previous cycles, students participating at the fourth grade are not permitted to use calculators on either form of the assessment.

## Item Writing Template

During the item-writing sessions, teams will use the TIMSS 2019 Item Template to record their items. Exhibit 1 contains the template. The template is provided as a separate word document as a part of the meeting materials for the 2<sup>nd</sup> NRC meeting.

To maintain future trends, most of the items developed for TIMSS 2019 need to be suitable to be administered via computer, if not in this cycle then in the next cycle. As a general guideline, even if your country is participating in paperTIMSS, please avoid developing items that are too long to fit on a single screen on a tablet or PC. The TIMSS 2019 Item Template indicates the space available for an item on a tablet or PC, which tends to be smaller than the size of pages in TIMSS paper booklets.

Typically the smaller size will not be a problem. For example, in transferring the TIMSS trend items from previous assessments to eTIMSS, approximately 80% fit “as is” and most of the rest were rearranged only slightly. Also, eTIMSS students will be able to scroll so it is possible to use 2-3 centimeters (about 1 inch) beyond the indicated space and still have the item used in both paperTIMSS and eTIMSS.

**Exhibit 1: TIMSS 2019 Item Template**

<b>Item Number:</b> <input style="width: 100%;" type="text"/>	<b>TIMSS 2019 Item Template</b>				
<b>Framework Classification</b>					
<b>Subject:</b> Select one	<b>Content Domain:</b>		<b>Cognitive Domain:</b> Select one		
<b>Grade:</b> Select one	<b>Topic Area:</b>		<b>Cognitive Area:</b>		
	<b>Topic:</b>				
<b>Item Type:</b> Select one	<b>Science Objective:</b>	<b>Science Practice:</b>			
<b>eTIMSS Item Features</b>					
<b>Tablet/PC Tools</b>	<input type="checkbox"/> Ruler <input type="checkbox"/> Calculator ( 8th grade only)				
<b>Item Components</b>	<input type="checkbox"/> Drag and Drop	<input type="checkbox"/> Sorting	<input type="checkbox"/> Selection	<input type="checkbox"/> Drop-down	<input type="checkbox"/> Line Tool
<b>Item</b>					
The box indicates the approximate screen size for an item without scrolling.					

**Exhibit 1: TIMSS 2019 Item Template (Continued)**

**Scoring**

**Multiple-Choice Items: Response Rationale**

Please write "KEY" by the correct response option and for more difficult items provide supporting calculations for or an explanation of why the answer is correct. For each incorrect response option, describe the common misconception or computation error the option represents.

Option	Rationale
A	
B	
C	
D	

**Compound-Multiple Choice Items: Scoring Rules**

Please write the number of parts correct that will be awarded 1 point or given no points (e.g., 1 Point: All 5 parts correct; 0 Points: 4 or fewer parts correct). In the space next to "Key" list the correct answers for each part (e.g., A, B, B, A, A).

	Scoring Rules
1 Point	
0 Points	
Key	

**Exhibit 1: TIMSS 2019 Item Template (Continued)**

**Constructed Response Items: Scoring Guides**

Please provide the student response or responses that will be awarded full credit and examples of each. For 2-point items also provide the student responses that will be awarded partial credit (1 of 2 points). When applicable, diagnostic codes (e.g., 21, 11, 70) may be added to capture common errors or misconceptions.

**Scoring Guide – 1-point**

Code	Response
	<b>Correct Response</b>
10	
	<i>Example Response:</i>
	<b>Incorrect Response</b>
70	
79	Other incorrect (including crossed out, erased, stray marks, illegible, or off task)
	<b>Nonresponse</b>
99	Blank

**Scoring Guide – 2-points**

Code	Response
	<b>Correct Response</b>
20	
	<i>Example Response:</i>
	<b>Partially Correct Response</b>
10	
	<i>Example Response:</i>
	<b>Incorrect Response</b>
70	
79	Other incorrect (including crossed out, erased, stray marks, illegible, or off task)
	<b>Nonresponse</b>
99	Blank

For each individual item, fill out one TIMSS 2019 Item Template, including:

1. Item Number is the team number (e.g, M01, S02) followed by a 3-digit number beginning with 001 (e.g., M01\_001, M01\_002)
2. TIMSS Subject (select Mathematics or Science)
3. TIMSS Grade (select 4 or 8)
4. TIMSS Item Type (select Multiple-Choice or Constructed Response)
5. TIMSS Content Domain, Topic Area, and Topic (as well as Science Objective and Science Practice, if applicable)
6. TIMSS Cognitive Domain and Cognitive Area
7. eTIMSS Item Features incorporated into the item
8. Complete the item as much as possible, including clear guidance about stimulus materials and citation information. Do not spend time on producing perfect graphics
9. Response rationale for multiple-choice item; scoring rules for compound multiple-choice item
10. Scoring guide for constructed response item

## Writing Multiple-Choice Items

A multiple-choice item asks a question or establishes the situation for a response. For the TIMSS 2019 assessments, this type of item provides response choices, or options, from which the correct answer is selected. A multiple-choice item is characterized by the following components:

- The stimulus presents the contextual information relevant to the item
- The stem is the initial part of the item in which the task is defined
- The response 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 2019 will be multiple-choice items. The next sections present guidelines specific to multiple-choice items, and include information about writing the stem, structuring the response options, developing plausible distracters, and providing response rationales.

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 the TIMSS 2019 assessments, since clarity is of vital importance, please phrase all stems as a **direct question**. An example of a stem phrased as a direct question is provided in Exhibit 2.

### Exhibit 2: Example of a Stem that Asks a Direct Question – Grade 4 Mathematics

Jan plants 8 trees in each of 5 rows. How many trees does he plant altogether?

- (A) 13
- (B) 32
- (C) 35
- (D) 40

M041023

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.

1. The stem should not include extraneous information. Extraneous information is liable to confuse students who might otherwise have determined the correct answer.
2. Avoid using negative stems—those containing words such as “not,” “least,” “worst,” “except,” etc. If it is absolutely necessary to use a negative stem, put the word 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.
3. 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.
4. 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  $4\pi$ ).

## Structure of the Response Options

1. Write multiple-choice items with four response options, labeled A–D (as shown in the example item about planting trees in Exhibit 2).
2. Make sure that one of the four response options is the correct answer (key). 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 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 to save 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 any of the TIMSS 2019 assessments. In such cases, a constructed response item may be more appropriate than a multiple-choice item.

## Compound Multiple–Choice Items






Some mathematics and science concepts lend themselves to measurement using compound multiple-choice items. Exhibit 3 shows an example of a compound multiple-choice item.



**Exhibit 3: Example of a Compound Multiple-Choice Item – Grade 4 Science**

Which objects shown below conduct electricity?

Fill in one circle next to each object.

		Conducts Electricity	
		Yes	No
	wooden spoon	----- (A) -----	----- (B) -----
	plastic comb	----- (A) -----	----- (B) -----
	silver chain	----- (A) -----	----- (B) -----
	rubber ball	----- (A) -----	----- (B) -----
	iron key	----- (A) -----	----- (B) -----

S051121

	Scoring Rules	Item: S051121
<b>1 Point</b>	All 5 parts correct	
<b>0 Points</b>	4 or fewer parts correct	
<b>Key</b>	B, B, A, B, A	

The individual multiple-choice elements included in a compound multiple-choice item each address a particular aspect of a mathematics or science concept. Taken together, the group of multiple-choice elements provides better measurement than any single element. Generally, compound multiple-choice items are scored as fully correct (1 point) if all parts are answered correctly. For all compound multiple-choice items, please use the template to provide the rules for scoring and key for each part as shown in Exhibit 3.

### Plausibility of Distracters

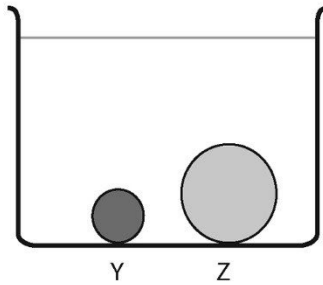
Use plausible incorrect response options (distracters) that are based on likely student errors or misconceptions. This reduces the likelihood of students arriving at the correct response by eliminating other choices and, equally important, may allow identification of widespread student misunderstandings 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.

### Response Option Analysis

For all multiple-choice items, please use the template to include a brief analysis of each response option or rationale for inclusion of specific response options (one sentence at the most for each response option). Label the correct response option as the KEY and provide the supporting calculations for the answer or an explanation of why it is correct. For each incorrect response option, describe the common misconception the option represents or the process by which a student could have arrived at that incorrect answer. Exhibits 4 and 5 show multiple-choice items together with their response option analyses.

**Exhibit 4: Example Multiple-Choice Item and Response Rationales – Grade 4 Science**

Andrea has two objects made of different materials. They have the same shape, but Y is smaller than Z. She puts them in a container of water and observes that Y and Z both sink.



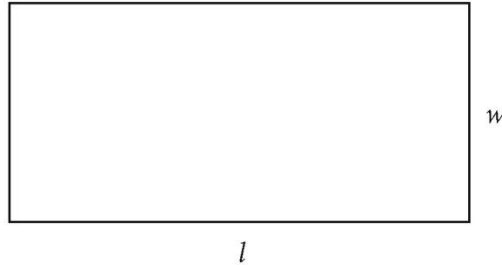
What can she tell about the weight of objects Y and Z?

- (A) Y is heavier than Z.
- (B) Z is heavier than Y.
- (C) Y and Z have the same weight.
- (D) It is not possible to tell without weighing the objects.

S041080

Option	Rationale	Item: S041080
A	There is no evidence to support that Y is heavier than Z.	
B	Misconception that the larger the object, the more it weighs.	
C	Misconception that if both objects sink they must have the same weight.	
D	KEY	

**Exhibit 5: Example Multiple-Choice Item and Response Rationales – Grade 8 Mathematics**



The shape above is a rectangle, with length  $l$ , and width  $w$ .

If the length is doubled and the width stays the same, which formula gives the area ( $A$ ) of the new rectangle?

- (A)  $A = 2l + 2w$
- (B)  $A = 2l + 4w$
- (C)  $A = 2lw$
- (D)  $A = 4lw$

M052066

Option	Rationale	Item: M052066
A	A formula for the perimeter of the original rectangle.	
B	A formula for the perimeter of a rectangle with the original width doubled.	
C	KEY	
D	A formula for the area of a rectangle with the original length and width doubled.	

## 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, create a list, 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 possible cues provided 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 easily accessible language appropriate to the age and experience of the target population. Use simple vocabulary and sentence structure, and avoid using complicated names for the subjects in the item.
2. Make what is expected of students as clear as possible without compromising the intent of the item. 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?’”). 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.
3. 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?”).

4. Students should be able to complete the task in the time allocated for each constructed response item, that is, a maximum of 3 minutes.
5. Write an appropriate answer to the question in terms of the language, knowledge, and skills that a 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.
6. 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.

### Response Types for Constructed Response Items

It is important to select the best response mode for each constructed response item. For most constructed response items, students will write the results of calculations, descriptions, or explanations. It is also possible to ask students to draw graphs or diagrams.

### 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.

### General Scoring Guidelines for TIMSS Constructed Response Items

General scoring guidelines used for 1-point and 2-point constructed response items are described in Exhibit 6.

#### Exhibit 6: TIMSS General Scoring Guidelines for Constructed Response Items

##### Scoring 1-Point Items

###### 1 Point (Full Credit)

A 1-point response is correct. The response indicates that the student has completed the task correctly.

###### 0 Points (No Credit)

A 0-point response is incorrect, irrelevant, or incoherent.

**Exhibit 6: TIMSS General Scoring Guidelines for Constructed Response Items (Continued)**

**Scoring 2-Point Items**

**2 Points (Full Credit)**

A 2-point response is complete and correct. The response demonstrates a thorough understanding of the concepts and/or procedures embodied in the task.

- Indicates that the student has completed all aspects of the task, showing correct application of concepts and/or procedures
- Contains clear, complete explanations, supporting work, or evidence when required

**1 Point (Partial Credit)**

A 1-point response is only partially correct. The response demonstrates only a partial understanding of the concepts and/or procedures embodied in the task.

- Addresses some elements of the task correctly, but may be incomplete
- May contain a correct answer, but an incomplete explanation when required
- May contain an incorrect answer with an explanation or supporting work indicating a correct understanding of the concepts

**0 Points (No Credit)**

A 0-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, 70, or 79.

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 a correct response is 1 for 1-point items or 2 for 2-point items

However, for an **incorrect or blank** response, the first digit is not zero. Instead:

- The first digit for an incorrect response 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 2 (codes 20–22, 10–12, and 70–72)
- 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 receive **code 79**.
- **Code 99** means a **completely** blank response.

It is unusual for an item to give rise to more than two commonly used 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 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.

## Writing Scoring Guides

It is essential that item writers provide scoring guides that include:

- The student response or responses that will be awarded full credit (1 or 2 points)
- For 2-point items, the student response or responses that will be awarded partial credit (1 of 2 points)
- If applicable, and absolutely necessary, diagnostic codes
- Example student responses for each code used

The example constructed response items and their scoring guides presented in Exhibits 7 and 8 illustrate the two-digit scoring system used in TIMSS and include the required information described above.



**Exhibit 7: Example Constructed Response Item and Scoring Guide – Grade 4 Mathematics**

M061272

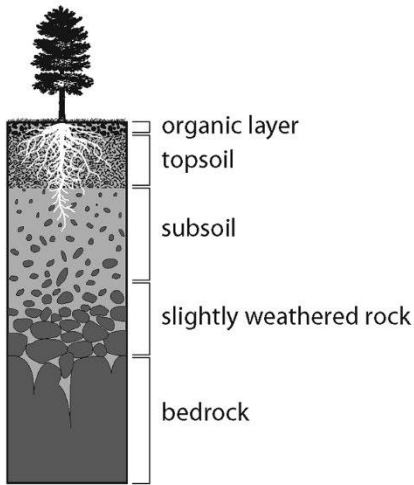
43 ÷ 5 =

Answer: \_\_\_\_\_

Code	Response	Item: M061272
<b>Correct Response</b>		
10	8 with a remainder of 3 OR $8\frac{3}{5}$ or equivalent	
<b>Incorrect Response</b>		
70	8 (does not include remainder)	
79	Other incorrect (including crossed out, erased, stray marks, illegible, or off task)	
<b>Nonresponse</b>		
99	Blank	

**Exhibit 8: Example Constructed Response Item and Scoring Guide – Grade 8 Science**

The diagram shows layers in the soil.



Most plants have roots that grow in the topsoil, but some have roots that reach into the subsoil.

Write two advantages for a plant to have long roots that go down into the subsoil.

1.

2.

S052116

**Exhibit 8: Example Constructed Response Item and Scoring Guide – Grade 8 Science (Continued)**

Code	Response	Item: S052116
	<b>Correct Response</b>	
20	<p>Identifies any <b>two of these acceptable</b> advantages.</p> <ul style="list-style-type: none"> <li>• Long roots can reach more nutrients (minerals). <i>Examples:</i> <i>The plant will be able to receive more nutrients that are hidden in the subsoil.</i> <i>More soil nutrients can be reached</i></li> <li>• Long roots can reach water (when topsoil is dry or frozen). <i>Examples:</i> <i>They can reach deep underground water.</i> <i>Will get more water.</i> <i>So that in time of drought they can reach the subsoil water.</i></li> <li>• Long roots can anchor the plant better. <i>Examples:</i> <i>They are not easily blown over by strong winds.</i> <i>Will make the hold stronger in the ground.</i> <i>It holds the stem of the tree firmly.</i></li> </ul>	
	<b>Partially Correct Response</b>	
10	Identifies <b>one</b> acceptable advantage only. The second advantage may be incorrect or missing.	
	<b>Incorrect Response</b>	
79	Incorrect (including crossed out, erased, stray marks, illegible, or off task)	
	<b>Nonresponse</b>	
99	Blank	

## Writing Items for eTIMSS 2019

### eTIMSS Tools for Students

Two eTIMSS tools for students can be incorporated into the design of an item.



The **ruler** allows students to measure lengths of objects on the screen.

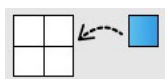


The on-screen **calculator** is available for all eTIMSS eighth grade mathematics items. This calculator includes a decimal point, four functions (+, −, ×, ÷), a key to change the sign of the number displayed ( $\pm$ ), and a key to take the square root ( $\sqrt{\quad}$ ).

### eTIMSS Components for Item Writers

Five additional response modes, referred to as eTIMSS components, are available for item writers: drag and drop, sorting, selection, drop-down menus, and a line drawing tool. In brief, the first three components—drag and drop, sorting, and selection—make it possible in eTIMSS to machine score several constructed response item types included in previous TIMSS assessments. Keep in mind that items using eTIMSS components need scoring guides just like their originally hand-scored counterparts. The drop-down component is a variant on multiple-choice or compound multiple-choice formats that can be used to optimize the space available on the screen. Items using the drop-down component need a key and/or an explanation of the number of points to be awarded. Lastly, the line tool is used in mathematics items asking students to create drawings on a square grid. Each of the components is described in the following sections.

#### Drag and Drop




The drag and drop response option enables students to drag images or text (referred to as “draggable elements”) to specific areas on the screen (drop zones). The drop zone has an underlying square grid that the square draggable elements snap into. Based on constructed response items that have been hand-scored in previous assessments, several construct relevant uses of drag and drop are:

- **Matching**—assesses the association between elements in two lists, such as a list of images to a list of text descriptions as shown in Exhibit 9, or a list of images to another list of images.
- **Completing systems**—assesses understanding of systems (e.g., water cycle, food chain, electric circuit). Systems are shown with parts missing and draggable images can be used to complete the systems.
- **Labeling graphs or completing pictographs** (e.g., adding labels to sections of a pie chart or symbols to rows of a pictograph).

**Exhibit 9: Example of Drag and Drop to Match Animals with Ecosystems**

**1** Drag each animal to the space under the name of the ecosystem in which it is most likely to be found.

camel monkey zebra whale

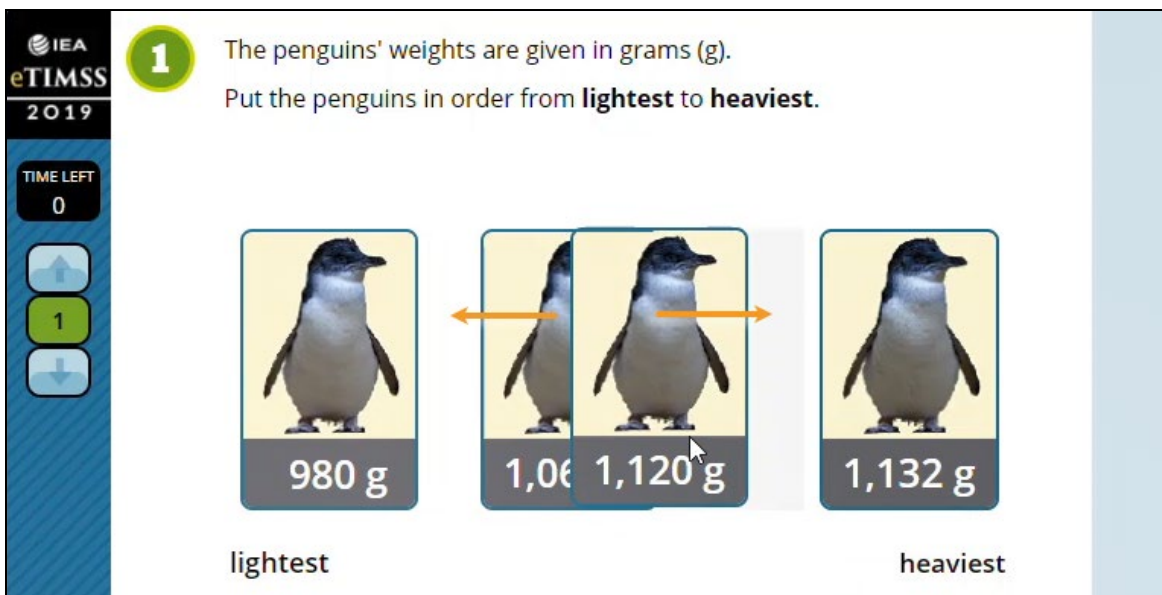
Ecosystem			
Tropical rain forest	Desert	Ocean	Grassland
			

Please note that the square grid underlying the drop zone can be at most 6 rows by 6 columns (for reference, the Ecosystem drop zone in Exhibit 9 is 1 row by 4 columns). The draggable elements and areas of the drop zone can have short labels, but the text must be kept to a minimum to facilitate translation. In developing drag and drop items, limit the number of elements that need to be dragged for a correct response to typically 4 or 5, and no more than 6. When possible, avoid a 1 to 1 correspondence to reduce “clues.” For example, provide more draggable elements than available spaces in the drop zone. You may also have a stack of the same image (e.g., symbols for a pictograph). Finally, make sure that drag and drop is the most efficient response mode (e.g., assigning animals to groups of vertebrates and non-vertebrates is more efficient as a compound multiple-choice item).





**Sorting** 

The sorting component is a special case of drag and drop specifically designed for asking students to order, sequence, or prioritize. Students are presented with vertical or horizontal lists of text and/or images and asked to reorder the lists according to some criteria. Again, please keep the text that needs to be translated to a minimum so that there will be enough space for the various TIMSS languages (e.g., German takes about twice as much space as English). As shown in Exhibit 10, the sorting component can include images with minimal text that can be translated across languages in the space available. Sometimes the sorting component can be used as an alternative response option to drag and drop (e.g., for cycles or for plants and animals that make up a food chain).

**Exhibit 10: Example of Sorting to Order Penguins by Weight**



**1** The penguins' weights are given in grams (g).  
Put the penguins in order from **lightest** to **heaviest**.

 980 g	 	 1,132 g
lightest		heaviest

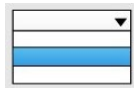


**Selection**

The selection component allows students to click areas of the response space to indicate their answers. Students can be asked to demonstrate their understanding of subsets of larger entities, such as shown in Exhibit 11. This item asks students to select all the factors of 12. The scoring guide could have a diagnostic code 70 for students that did not select 12 but were otherwise correct. In science, the selection component could be used to ask students to identify various types of subsets such as climates on a map (e.g., tropical rain forests) or certain elements in the periodic table. When students click a response option it will become highlighted and can be un-selected by clicking it again. Response options can be text, images, or a combination of both, presented as a list or a grid. Exhibit 11 shows the response options in a grid with 2 rows and 6 columns. The maximum grid size is 6 rows by 6 columns. However, most grids should be smaller.

**Exhibit 11: Example of Selection to Indicate the Factors of 12**

## Drop-Down



When response options are organized using the drop-down component, students first click the drop-down menu to see the responses and then click one response option to answer the item. The drop-down component is beneficial for optimizing screen space. Also, drop-down response options can be useful for items requiring multiple short answers (more complicated cases of compound multiple-choice). Exhibit 12 shows an example of an item that includes drop-down responses to predict color changes when an indicator is added to acidic, basic, and neutral solutions.

**Exhibit 12: Example of Drop-Down Menus to Predict Color Changes**

IEA  
eTIMSS  
2019

TIME LEFT  
0

↑  
1  
↓

**1** The juice of red cabbage is a natural pH indicator. The color of the juice is purple. When a few drops of the juice are added to:

- an **acid** solution, the color changes to **red**.
- a **base** solution, the color changes to **blue**.
- a **neutral** solution, the color remains **purple**.

What color is each solution after a few drops of red cabbage juice have been added?  
Use the drop-down menus to choose the color for each solution.

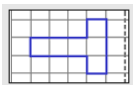
Solution	Color
Distilled water	purple
Lemon juice	red
Vinegar	red
Baking soda solution	Choose one

red

blue

purple





**Line Tool**

The line tool can be used to create drawings on a square grid. To draw a line, students click a point on the grid then hold and drag to the end point. All lines will snap to the grid, either along the grid lines or diagonally across squares. Several examples of uses of the line tool are:

- Drawing a shape with given properties (e.g., number of sides, lines of symmetry), as shown in Exhibit 13.
- Translating or reflecting a figure (e.g., over a line of symmetry, about a point).
- Completing a bar or line graph.

**Exhibit 13: Example of Line Tool to Draw a Shape with Given Properties**

**IEA eTIMSS 2019**

**1** Draw a shape that has:

- 5 sides
- at least one line of symmetry

TIME LEFT: 0

1

1 cm

1 cm

## Reviewing Items and Scoring Guides

Items selected for inclusion in the TIMSS 2019 field test or main data collection will go through a thorough review process involving TIMSS & PIRLS International Study Center staff, Science and Mathematics Item Review Committee (SMIRC), and National Research Coordinators (NRCs). 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 concerns 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 explain their items. The earlier 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 2019 content and cognitive 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 concerns with the item.

## Reviewing Constructed Response Items

In reviewing each constructed response item, item reviewers should:

1. Check that their judgments of the TIMSS 2019 content and cognitive 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 they believe 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 concerns with the item and/or scoring guide.

## Appendix A: Multiple-Choice Item and Response Rationale Review Checklist

Item Characteristic	Yes	No
Does the item address a topic/objective in the framework?		
Is the mathematics/science correct?		
Is the task clear to students?		
Is there only one correct response option?		
Is the language in the item appropriate for students in this grade?		
Is there a plausible rationale for each distracter?		
Is the stem worded positively (or are negative words in <b>bold</b> if used)?		
Can the question be answered without reading the answer options?		
Do the answer options have parallel structure?		
Is the language in the item suitable for translation?		
Is the item free of cultural, gender, or geographical bias?		
Is the context of the item familiar for students at this grade level?		

## Appendix B: Constructed Response Item and Scoring Guide Review Checklist

Item Characteristic	Yes	No
Does the item address a topic/objective in the framework?		
Is the mathematics/science correct?		
Is the task clear to students?		
Are the expectations for a full credit response clear?		
Can the task be completed in a reasonable amount of time?		
Does the scoring guide have clear distinctions between correct and incorrect responses?		
Are the response categories in the scoring guide mutually exclusive (i.e., is it clear how to assign a score code to all possible student responses)?		
Is the language in the item appropriate for students in this grade?		
Is the stem worded positively (or are negative words in <b>bold</b> if used)?		
Is the language in the item suitable for translation?		
Is the item free of cultural, gender, or geographical bias?		
Is the context of the item familiar for students at this grade level?		

## References

- Dolan, R.P., & Hall, T.E. (2007). Developing accessible tests with universal design and digital technologies: Ensuring we standardize the right things. In L. L. Cook & C. C. Cahalan (Eds.), *Large-scale assessment and accommodations: What works* (pp. 95–111). Arlington, VA: Council for Exceptional Children.
- Mislevy, R.J., Almond, R.G., & Lukas, J.F. (2003). *A brief introduction to evidence-centered design*. ETS (Research Report RR-03-16). Princeton, NJ: Educational Testing Service.
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 assessment frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/timss2019/frameworks/>





BOSTON  
COLLEGE

[timss.bc.edu](http://timss.bc.edu)



**IEA**

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

© IEA, 2017