

CHAPTER 1

Developing the TIMSS Mathematics and Science Achievement Instruments

Charlotte E. A. Aldrich
Allison Bookbinder
Lale Khorramdel

Introduction

TIMSS assessments are designed to provide valid measurement of mathematics and science knowledge and skills that are internationally valued and covered in the fourth- and eighth-grade curricula of participating countries. The general approach to developing the TIMSS mathematics and science achievement items to meet this goal is similar from one assessment cycle to the next. Still, each cycle has some unique characteristics that influence instrument development.

The TIMSS assessments evolve each cycle while maintaining continuity with past assessments to measure trends. Although most achievement items are brought forward from previous assessment cycles for this purpose, developing new mathematics and science items for each new TIMSS cycle requires substantial effort. Every four years, TIMSS assesses both subjects at two grades; therefore, the equivalent of four different assessments must be prepared for each TIMSS cycle.

This chapter describes the general approach to developing the TIMSS mathematics and science achievement instruments for each assessment cycle. It provides an overview of each step within the development process and highlights the roles and responsibilities of organizations and individuals involved. The end of the chapter focuses on instrument development specific to TIMSS 2023.

The TIMSS Approach to Measuring Trends

TIMSS is designed to measure trends in educational achievement over time. Consequently, the mathematics and science assessment content at each grade must stay to a large extent the same from one cycle to the next. TIMSS is based on a well-known premise for designing trend assessments (ascribed to John Tukey and Albert Beaton):

“If you want to measure change, do not change the measure.”

However, the achievement items in mathematics and science must be updated for each cycle to present students with items that include contemporary contexts that reflect current learning goals and curricula for the relevant content domains. TIMSS assessments are developed to reflect the cutting edge of science and mathematics teaching and to be presented to students in ways familiar to them from everyday instruction and experience.

TIMSS has a specific design for rotating items out of the assessment after each cycle and replacing them with newly developed items. Items retained after each cycle are referred to as “trend items” and are kept confidential to be readministered in subsequent cycles. With this design, each TIMSS assessment includes items developed during three different cycles—items newly developed for the current cycle and items from the previous two cycles.

The TIMSS’ transition to digital assessment was completed in the 2023 cycle, marking the first cycle that new assessment content was developed only for computer-based assessment, with no new paper items introduced. The transition from paper-and-pencil to computer-based assessments began in TIMSS 2019 (Fishbein et al., 2018; von Davier et al., 2020). To strengthen the link between the paper and digital assessments and maintain trend measurements, TIMSS 2019 paid particular attention to converting its paper trend items (items used in TIMSS 2015 and retained in the 2019 assessment) to the TIMSS computer-based item formats to ensure a high level of comparability between response modes. TIMSS 2023 continued efforts to advance computer-based assessment in TIMSS, including further development of a variety of complex and innovative digital item formats. TIMSS 2019 also introduced Problem-Solving and Inquiry (PSI) tasks as an extension of the main assessment (Mullis et al., 2021). PSI tasks are based on visually attractive, interactive scenarios that allow students to follow a series of items supported by a shared context or goal. The role of PSI tasks in measuring higher-order cognitive processes was further developed and more fully integrated into the TIMSS 2023 design with newly-developed assessment content.

Development Process for TIMSS Mathematics and Science Achievement Instruments

In each cycle of TIMSS, the development of new test content adheres to best practices in assessment design, including clearly defining the target construct to be measured, specifying the items needed to measure it, establishing standards for constructing items and test forms, and ensuring that the assessments meet the test specifications delineated in the Mathematics and Science Assessment Frameworks and the Assessment Design Framework. The validity of this process is supported by following the principles of the evidence-centered design framework (ECD; Mislevy et al., 2003) and meeting recognized standards for test design and development and instrument validation (AERA, APA, & NCME, 2014; International Testing Commission, 2019). This process follows several broad steps that remain at the core of TIMSS development each cycle to maintain coherence between the assessment goals and the data:

- updating the assessment frameworks to identify and prioritize the mathematics and science content that the assessment will measure

- developing new achievement items as well as scoring guides for constructed-response items to meet the assessment specifications delineated in the frameworks
- conducting a full-scale field test to evaluate the measurement properties of the item pool and practice the data collection and scoring procedures
- selecting new items from the field-tested items to meet the assessment specifications based on the field test results and trend items from previous cycles
- conducting training on how to reliably score students' responses to human-scored constructed-response items to ensure high-quality data

Roles of Organizations and Individuals in TIMSS Instrument Development

The TIMSS mathematics and science instrument development process is directed and managed by assessment development experts who are part of the TIMSS & PIRLS International Study Center staff. The study center's assessment development experts are responsible for overseeing development throughout all steps of the process and for working closely with development partners, including TIMSS National Research Coordinators (NRCs), the TIMSS Science and Mathematics Item Review Committee (SMIRC), and digital assessment platform providers to help ensure the items are of high quality, have adequate psychometric properties, and are aligned with the frameworks and goals of the assessment.

TIMSS National Research Coordinators (NRCs) are designated by their countries to be responsible for the complex tasks involved in implementing TIMSS in their countries and play a key role in achievement item development. The TIMSS & PIRLS International Study Center works with the NRCs and experts from the participating countries to update the assessment frameworks and develop the new achievement items, including the scoring guides for constructed-response items. The NRCs review the items before the field test and help select the items for the assessment afterward.

The TIMSS SMIRC provides additional advice and guidance through periodic reviews. The SMIRC typically comprises experts in mathematics and science education and assessment in these subject domains. To ensure a broad representation of expertise, SMIRC members for each TIMSS cycle are nominated by countries participating in TIMSS and selected by the TIMSS & PIRLS International Study Center. The SMIRC assists in guiding updates to the TIMSS frameworks and reviewing the content of the TIMSS assessment instruments throughout the development process.

SMIRC members meet several times during each cycle to review mathematics and science frameworks, review draft items and scoring guides, and help select items for the final assessment following the field test. Based on the final assessment results, the SMIRC helps conduct a scale anchoring analysis to describe student competencies at the TIMSS International Benchmarks of Mathematics and Science Achievement.

Development partners from the international assessment community with experience in test development for student assessments may help to meet particular development goals for each

cycle. Cognitive laboratories are sometimes conducted to gain insight into students' interactions with novel item types or changes to the TIMSS digital user interface.

A significant portion of the instrument development process is dedicated to ensuring digital achievement items are fully functional and can be properly translated and localized into the various languages of the TIMSS participating countries. Staff from the TIMSS & PIRLS International Study Center and IEA work closely with a digital assessment platform provider to build and prepare the digital assessment instruments for delivery to students via computers and tablets. Stand-alone items are built in the international source language (American English) by TIMSS & PIRLS International Study Center staff within the platform's system through a web-based application for creating digital achievement items, and the items undergo several rounds of quality control review. Because of their complexity, PSI tasks are built by the assessment platform provider.

Updating the TIMSS Assessment Frameworks

The first step in developing the TIMSS achievement instruments is to define and prioritize the mathematics and science content and skills that the assessment will measure. The assessment frameworks cannot drastically change from cycle to cycle but are routinely updated to keep up with fresh ideas and current information about curricula, standards, and instruction in mathematics and science education worldwide.

For each TIMSS cycle, the Mathematics and Science Assessment Frameworks are updated using an iterative process involving input from the SMIRC and NRCs from participating countries. Recommendations for updating content and cognitive domains from one cycle to the next may involve modifying descriptions of the content domains and their priorities or weightings in relation to the other domains (by no more than 5%); adding, deleting, or modifying topic areas within content domains to keep current with what is covered in curricula; ensuring that the number of topic areas reflects the content domain weighting; reviewing topic descriptions to improve clarity for item writers; and combining topics to reduce redundancy.

Writing and Reviewing Field Test Items and Scoring Guides

The TIMSS & PIRLS International Study Center uses a collaborative process involving the participating countries to develop a substantial number of new items needed for the field test. The field test conducted for each TIMSS cycle includes approximately one and a half times the number of achievement items needed for data collection to ensure sufficient numbers of high-quality items for the main assessment. Development partners support the item development activities by drafting items, particularly the more complex [PSI tasks](#) and other targeted development for an assessment cycle based on specific needs. At an item writing workshop, TIMSS NRCs, experienced item writers from participating countries, and staff from the TIMSS & PIRLS International Study Center draft items for the mathematics and science field tests. All items requiring students to construct a response are accompanied by scoring guides that describe the key characteristics of a full-, partial- (where applicable), or no-credit response.

In preparation for the item writing workshop, TIMSS & PIRLS International Study Center staff identify the total number of items needed based on the weight assigned to a particular topic in the Mathematics and Science Assessment Frameworks and how many trend items measuring each topic were retained from previous assessments. The TIMSS & PIRLS International Study Center also updates the TIMSS Item Writing Guidelines, an item writing manual specifically developed for each TIMSS assessment. The manual provides descriptions of the digital item formats with details about their functionality and examples to encourage their use. The manual presents general procedures for obtaining good measurement of mathematics and science achievement (e.g., items must be independent and not provide clues to the correct responses of other items), as well as specific information on how to write items that facilitate comparable translations (e.g., using TIMSS’ fictitious unit of currency, the “zed,” for items involving money instead of allowing currency conversions). The manual also describes the steps for developing scoring guides for constructed-response items and provides checklists for reviewing TIMSS items.

Following the item writing workshop, assessment development experts at the TIMSS & PIRLS International Study Center review each item in light of the framework specifications and select an optimal group of items for further review and revision. Expert consultants and development partners may help draft additional items to improve coverage of areas of the frameworks that are especially challenging to measure.

The TIMSS & PIRLS International Study Center then works with the digital assessment platform provider to prepare an international version of all TIMSS achievement items in English. After the SMIRC reviews the items, they are revised again, and the NRCs review the complete set of draft field test items. The security of the TIMSS achievement items is paramount. All SMIRC members and national center staff must have signed confidentiality agreements to protect the validity of the inferences made based on the TIMSS achievement test and its results.

Preparing the Field Test Instruments

In preparation for the main data collection, TIMSS routinely conducts a full-scale field test to evaluate the measurement properties of the item pool and provide an opportunity for countries to practice the data collection and scoring procedures. All instruments and operational procedures are field tested with samples of students selected according to rigorous sampling procedures (see [Chapter 3](#)). The field test in each country is designed to yield at least 200 student responses to each mathematics and science item.

TIMSS & PIRLS International Study Center staff work with the assessment delivery platform specialists to implement the final suggested revisions from NRCs and provide the international versions of the field test instruments to the NRCs. Subsequently, the items are translated by participating countries into their languages of instruction to create internationally comparable translations that are appropriately adapted for the national context. Therefore, a significant portion of the development and review effort by NRCs is dedicated to ensuring that the achievement items can be translated accurately.

After translating and adapting the international instruments, the national instruments undergo translation verification and localization (see [Chapter 5](#)). Then, countries check the functionality of their national instruments, prepare digital devices for the field test, and check the compatibility of the assessment software with the devices before beginning test administration.

Developing Materials for Field Test Scoring Training

To ensure the quality of the TIMSS assessment results, it is critical that students' responses to the achievement items demonstrate the knowledge, application, or reasoning in mathematics or science required by the items to receive credit. It is particularly important that students' responses to constructed-response items are scored consistently to enable comparisons of students' mathematics and science achievement across countries and over time. For these reasons, TIMSS expends considerable effort to ensure the validity and reliability of the scores assigned to students' responses to the TIMSS achievement items (see [Chapter 7](#)).

In addition to developing a unique scoring guide for each constructed-response item, the TIMSS & PIRLS International Study Center provides training for NRCs and their scoring supervisors to ensure that the scoring guides for all human-scored constructed-response items are applied consistently within and across countries. The TIMSS training materials consist of sets of student responses for a selected group of items with the most complicated scoring guides. For each item, the training set consists of 8 to 12 student responses illustrating the codes in the scoring guide (example responses), followed by 8 to 12 student responses without pre-assigned score codes (practice responses).

To allow student responses that require human scoring to be scored immediately upon completion of the field test, scoring training materials for the newly developed constructed-response items must be prepared in advance of the field test administration. To provide a basis for these materials, several English-speaking countries are asked to pilot-test a selection of the newly developed constructed-response field test items in several classrooms with English-speaking students before the field test.

TIMSS & PIRLS International Study Center staff review responses collected in the pilot test and create the training materials jointly with SMIRC members. At the training sessions for NRCs and scoring supervisors, the trainers describe each item and explain its purpose. The trainer then describes the scoring guide, explaining each category and the rationale for the score given to each example response. The country representatives are then given time to score the practice responses to apply the scoring guides and learn to make distinctions among score categories. The correct codes for each practice response are then reviewed, any inconsistencies in scoring are discussed, and, as necessary, the scoring guides are clarified, and sometimes scoring categories are revised.

Finalizing the Achievement Instruments

After the field test, the TIMSS & PIRLS International Study Center analyzes the TIMSS field test data and selects the items to be combined into new item blocks, which are finally combined

with the trend item blocks according to the assessment design for data collection. When selecting items, both the individual items' measurement properties (item statistics) and the overall content and cognitive domain coverage of the group of items are considered to ensure that the final achievement instruments meet the specifications in the Mathematics and Science Assessment Frameworks. Item blocks are assembled according to the difficulty targets defined in the Assessment Design Framework.

Staff at the TIMSS & PIRLS International Study Center review the field test data to make an initial judgment about the quality of each item based on its measurement properties. To assess the quality of the field test items, achievement data almanacs are produced, which contain summary statistics for each field test item. The achievement data almanacs display the following information for each item and each country: the number of students to whom the item was administered, the item difficulty and discrimination, the percentage of students answering each response option (if a selected-response item) or in each score category (if a constructed-response item), the point-biserial correlation for each selected-response option or constructed-response category, and the degree of scoring agreement for human-scored constructed-response items.

First, the TIMSS & PIRLS International Study Center conducts an item-by-item review of the field test data to eliminate any items from further consideration if they have poor measurement properties, such as being too difficult or too easy or having low discrimination. PSI items are also reviewed individually at this stage, with problematic items considered for removal. If all items have suitable measurement properties and can be combined into a block according to the difficulty requirements, PSI tasks are selected to proceed to the main data collection.

After the item-by-item review, TIMSS & PIRLS International Study Center staff prepare draft item blocks using automated test assembly and then collaborate with SMIRC members to refine the automated sets into recommended blocks of achievement items. The group reviews the viable field test items for each content domain topic to select a coherent group of items for each topic, then verifies that the items are appropriately distributed across the cognitive domains, item formats, and target block difficulties, also considering the distribution of the trend items. The SMIRC scrutinizes the recommendations for the newly developed achievement item blocks, reviewing the items and scoring guides for content accuracy, clarity, and adherence to the frameworks.

TIMSS & PIRLS International Study Center staff implement the SMIRC's recommendations for the NRCs' penultimate review. NRCs are asked to review the recommended assessment blocks in light of the framework specifications and field test results. Each country is also asked to check any unusual national results indicative of translation or other issues or to recommend revisions to the national version.

Following the final review, the newly developed and existing trend item blocks are arranged into booklets according to the TIMSS assessment design for national instrument preparation.

TIMSS 2023 Achievement Instrument Development Process

The item development process for the TIMSS 2023 assessments followed a similar approach to all previous cycles of TIMSS while also advancing to keep up with current mathematics and science education and advancements in assessment technology. In addition to continuing the TIMSS trend measure that has monitored changes in education achievement since 1995, mathematics and science instrument development for TIMSS 2023 had some unique aspects.

- It was the first TIMSS assessment in which new assessment material was developed for digital administration only. Paper instruments containing trend material brought forward from TIMSS 2019 were administered in a few countries not yet prepared to administer a large-scale digital assessment. TIMSS 2023 also included an additional paper-based national comparison study option for countries participating in the full digital administration.
- TIMSS 2023 implemented a new group adaptive assessment design to accommodate a range of achievement distributions across countries. Newly developed item blocks were assembled according to targets specified by the [TIMSS 2023 Assessment Design Framework](#), in addition to adhering to the targets of the [Mathematics Assessment Framework](#) and [Science Assessment Framework](#).

Much of the TIMSS 2023 assessment development cycle occurred under unprecedented circumstances brought on by the COVID-19 pandemic starting in Spring 2020. Therefore, some adaptation to the original schedule for development activities was required. The first year of the TIMSS 2023 assessment cycle was committed to updating the assessment frameworks to guide item writing, developing field test items, and preparing the digital assessment interface. This first year included numerous in-person meetings; in the 2023 cycle, these were converted to virtual meetings. These efforts continued into the second year when countries administered the TIMSS 2023 Field Test starting in March 2022 and received training for scoring human-scored items therein. The field test results were carefully reviewed, and the TIMSS 2023 Data Collection materials were finalized by October 2022. Northern Hemisphere countries began the TIMSS 2023 Data Collection in March 2023, with Southern Hemisphere countries beginning in September 2023.

The TIMSS 2023 assessments were prepared and delivered to students in participating countries through a digital assessment platform owned by RM Results. Most assessment items were prepared in RM's Assessment Master, a web-based application for creating digital achievement items and instruments for delivery to students via computers and tablets. The TIMSS 2023 item-building interface included templates for various item formats and a variety of tools for designing the items, such as features for uploading and adding text to images, creating tables, and previewing items as they would appear to students during the test. Furthermore, to best match the devices students use internationally, the orientation of the item interface was changed from the portrait orientation of eTIMSS 2019 to a landscape orientation for TIMSS 2023.

Both newly-developed items administered in the field test as well as the trend items brought forward from TIMSS 2019 were entered into the Assessment Master using one of several item

formats, described in Exhibit 1.1. Many PSI items also used these formats, but additionally included unique item types built specifically for PSIs as they required unique customization.

Exhibit 1.1: TIMSS 2023 Digital Item Types

Item Type	Description
Single selection	Students select one of several given response options in the form of words, phrases, or images.
Multiple selection	Students select two or more of several given response options in the form of words, phrases, or images.
Compound selection	Students respond to multiple single-selection item parts by selecting one of several given response options common across the item parts. Item parts can be shown in a matrix format with rows for each part and common answer options presented in columns or using drop-down menus.
Matching pairs	Students draw lines to match each of several prompts given in one column with a set of response options in a parallel column. Prompts and response options can be in the form of words or phrases.
Drag and drop	Students drag objects in the form of words, phrases, or images and place them in droppable areas or positions.
Type text	Students type words, phrases, or sentences.
Number pad	Students enter a numerical response using a number keypad consisting of the digits 0–9, a decimal period or comma, a negative sign, and a fraction format.
Drawing tool	Students freely draw or write a response.
Graphing tool	Students draw lines or plot points on a coordinate area.
Chart tool	Students complete bar graphs or line graphs with single clicks.

Organizations and Individuals Involved in TIMSS 2023 Development

TIMSS 2023 item development unfolded as a close collaboration between the TIMSS & PIRLS International Study Center, the TIMSS 2023 Science and Mathematics Item Review Committee (SMIRC), the National Research Coordinators (NRCs), and technical specialists from the digital assessment platform provider, RM Results.

Content experts from the participating countries on the TIMSS 2023 Science and Mathematics Item Review Committee (SMIRC) worked with staff at the TIMSS & PIRLS International Study Center to update the assessment frameworks and develop the new achievement items, including the scoring guides for constructed-response items. The SMIRC reviewed draft items for the field test and selected items for the data collection item blocks assembled by the TIMSS & PIRLS International Study Center. The TIMSS 2023 SMIRC comprised 15 members nominated by the participating countries: 7 experts in mathematics and mathematics education and 8 in science and science education. Exhibit 1.2 lists the TIMSS 2023 SMIRC members.

Exhibit 1.2: TIMSS 2023 Science and Mathematics Item Review Committee (SMIRC)

Mathematics	
<p>Ray Philpot Australian Council for Educational Research (ACER) Australia</p> <p>Kiril Bankov Faculty of Mathematics and Informatics, University of Sofia Bulgaria</p> <p>Franck Salles Department of Evaluation, Forecasting, and Performance, French Ministry of Education and Youth France</p> <p>Khattab Mohammad Ahmad Abulibdeh National Center for Human Resources Development Jordan</p>	<p>Hege Kaarstein Department of Teacher Education and School Research, University of Oslo Norway</p> <p>Cheow Kian Soh Ministry of Education, Curriculum Planning, and Development Division, Mathematics Branch Singapore</p> <p>Mary Lindquist Professor Emeritus Mathematics Education, Columbus State University United States</p>
Science	
<p>Yun-Ping Ge International Master's Program of Learning and Instruction, National Taipei University of Education Chinese Taipei</p> <p>Svatava Janoušková Science Faculty Department of Teaching and Didactics of Chemistry, Charles University Prague Czech Republic</p> <p>Christian Christrup Kjeldsen Danish School of Education, Aarhus University Denmark</p> <p>Ute Harms Department of Biology Education, IPN – Leibniz Institute for Science and Mathematics Education Germany</p>	<p>Berenice Michels Faculty of Science, Freudenthal Institute for Science and Mathematics Education The Netherlands</p> <p>Galina Kovaleva (through 2021) Federal Institute for Strategy of Education Development of the Russian Academy of Education Center for Evaluating the Quality of Education Russian Federation</p> <p>Jens Anker-Hansen Department of Curricula, Swedish National Agency for Education Sweden</p> <p>Alicia Alonzo Department of Teacher Education, Michigan State University United States</p>

SMIRC members met three times for TIMSS 2023, twice in virtual meetings and once in person. At the 1st TIMSS 2023 SMIRC meeting (October 2021), the committee reviewed the Mathematics and Science Assessment Frameworks and initial drafts of the mathematics and science PSIs. In place of an online 2nd meeting, SMIRC members conducted an asynchronous rolling review of potential field test items between April and August 2022. At the 3rd meeting held virtually in September 2022, the SMIRC reviewed field test results and recommended the items to be included in the TIMSS 2023 mathematics and science assessments. At the final meeting in June 2024, the TIMSS 2023 scale anchoring process was conducted (see [Chapter 14](#)).

Developing technology-enhanced achievement items to meet the development goals for TIMSS 2023 necessitated support from development partners, collaboration with SMIRC members, and development efforts by RM Results to build functional items and item templates in the digital assessment system. The Australian Council for Educational Research (ACER) assisted in developing mathematics items. The Leibniz Institute for Science and Mathematics Education (IPN) assisted in developing science items. Several SMIRC members worked closely with the TIMSS & PIRLS International Study Center throughout the development process to achieve the development goals and as part of working groups for scoring training preparation. Mary Lindquist and Ray Philpot provided additional subject-matter expertise and support for mathematics. Berenice Michels served in this capacity for science.

The TIMSS & PIRLS International Study Center also worked with NRCs designated by their countries to oversee the complex tasks relating to the national implementation of TIMSS, beginning with the development of the *TIMSS 2023 Assessment Frameworks* and test content. NRCs provided feedback on the updates to the assessment frameworks and were invited to participate in the item writing workshop at the 2nd NRC meeting in February 2021 or nominate another representative from the national center to attend the workshop and draft items for consideration in the field test. The NRCs also reviewed the items before the field test at the 3rd NRC meeting in October 2021. Similarly, NRCs reviewed the draft items and were able to voice concerns regarding selected items or block composition for the TIMSS 2023 Data Collection at the 5th NRC meeting in September 2022.

Technical specialists from RM Results worked with the TIMSS & PIRLS International Study Center to develop item templates for stand-alone item creation and build each of the PSI tasks individually.

Updating the Assessment Frameworks for TIMSS 2023

Updating the frameworks for mathematics and science at the fourth and eighth grades was the first step in developing TIMSS 2023. Led by experts at the TIMSS & PIRLS International Study Center, initial updates to the TIMSS 2019 frameworks were proposed based on descriptions of mathematics and science curricula provided by participating countries in the [TIMSS 2019 Encyclopedia](#). The updates also considered teachers' responses to items in the TIMSS 2019 Teacher Context Questionnaire about topics taught to students based on the *TIMSS 2019 Assessment Frameworks*. The SMIRC reviewed and updated the initial draft, which was then shared with NRCs.

NRCs were asked to respond to a topic-by-topic survey of the draft frameworks online. This survey aimed to ensure that the proposed topics of the *TIMSS 2023 Assessment Frameworks* reflected the mathematics and science standards of countries participating in TIMSS 2023. For each component of the frameworks, NRCs were asked to choose from among the following options for each topic: Keep, Delete, or Modify, and propose changes if choosing the latter. TIMSS & PIRLS International Study Center experts and SMIRC consultants revised the framework drafts based on the survey responses and presented that version to the NRCs at the 1st/2nd NRC meeting held virtually in February 2021.

The 1st and 2nd NRC meetings were held online consecutively due to schedule restrictions necessary to adapt to the COVID-19 pandemic. At the meeting, the TIMSS & PIRLS International Study Center summarized the changes reflected in the recent version. While no major revisions were made to the mathematics or science frameworks in the interest of trend measurement, key updates included text recognizing TIMSS 2023 as the first fully digital assessment cycle, describing the integration of Problem-Solving and Inquiry Tasks (PSIs), and describing the expectations for a range of problem-solving contexts in the assessment. Updates to the TIMSS 2023 Mathematics Framework included clarifying language to distinguish between content and cognitive domains and reducing the number of Knowing and Reasoning cognitive subdomains. Updates to the TIMSS 2023 Science Framework included clarifying descriptions of the content domains, combining or revising objectives within several topic areas, expanding descriptions of processes underlying the Reasoning cognitive domain, and emphasizing that science practices are not linear. The frameworks were updated a final time before publication.

Writing and Reviewing the TIMSS 2023 Field Test Items and Scoring Guides

The TIMSS & PIRLS International Study Center oversaw a collaborative item writing process wherein many new items were written by representatives from participating countries. Development for the TIMSS 2023 Field Test required approximately one and a half times the number of achievement items needed for data collection. Nearly 600 items were newly developed for the field test across all grades and subjects, and all new items were designed for digital administration. The majority of the 2nd NRC meeting held in February 2021 was dedicated to a virtual item writing workshop.

Before drafting items, experts at the TIMSS & PIRLS International Study Center established the scope of the TIMSS 2023 development needs to ensure a broad assessment framework coverage in the TIMSS 2023 Data Collection. An added dimension of complexity in the TIMSS 2023 development cycle was specifying item development needs by item and block difficulty. Considerations of the total number of items required to fulfill the weights described in the *TIMSS 2023 Assessment Frameworks* were taken alongside the number of items from TIMSS 2019 that would be included in TIMSS 2023 as trend items. To fulfill the requirements introduced by the group-adaptive assessment design, a certain number of easy, medium, and difficult items had to be developed within each content domain to maintain framework coverage.

To support quality item writing that adhered to the content and design specifications from the TIMSS 2023 frameworks, the TIMSS & PIRLS International Study Center updated the TIMSS Item Writing Guidelines from previous cycles, adding in descriptions of new, innovative digital item formats and features of the RM system templates to be leveraged during development.

Developing the PSI tasks was particularly intensive and time-consuming because of the complexities of establishing a narrative principle to the task, defining digital interactivity beyond what is possible in normal items, and the need to keep students engaged as they work through all the items. Therefore, experts at the TIMSS & PIRLS International Study Center began working with consultants from the SMIRC early in January 2020 to draft possible new PSI tasks. Even with the reduction of meetings resulting from the onset of the COVID-19 pandemic, all new PSI tasks were reviewed iteratively by the SMIRC and NRCs asynchronously and during meeting discussions. Furthermore, two PSI tasks were subject to cognitive laboratory testing, and all PSI items were scrutinized based on performance in the field test.

Cognitive laboratories have been implemented in previous cycles of TIMSS & PIRLS to gauge the appropriateness of assessment content and evaluate students' engagement with digital assessment tools. For TIMSS 2023, the American Institutes of Research (AIR) conducted cognitive labs in December 2020 with 13 students in the greater Washington, D.C. area. Due to the COVID-19 pandemic during the 2020–2021 school year, it was necessary to adapt the cognitive laboratory protocol to an online format. Following the reports, AIR prepared an extensive report that informed important changes to the draft PSI tasks and provided instructive feedback to other PSI tasks under development.

The 2nd TIMSS 2023 NRC meeting was designated as an item writing workshop to draft items for possible inclusion in the TIMSS 2023 Field Test. About 120 participants from 50 countries were assigned to groups based on grade/subject expertise; each group worked together (via virtual breakout rooms) to draft items to pre-assigned framework topics. Participants drafted about 1,000 mathematics and science items for consideration for the field test. Exhibit 1.3 summarizes the participants and approximate output of the TIMSS 2023 Item Writing Workshop.

Exhibit 1.3: Summary of TIMSS 2023 Item Writing Workshop

Participants	
Number of Countries and Benchmarking Entities	50
Number of Country Representatives	120
Approximate Number of Field Test Items Written at Item Writing Workshop	
Grade 4 Mathematics	300
Grade 4 Science	200
Grade 8 Mathematics	300
Grade 8 Science	200

Following the NRC meeting, the TIMSS & PIRLS International Study Center organized the items submitted by item writing groups and identified items to be further refined and reviewed by the SMIRC in a rolling review from April through July 2021. As necessary, consultants from the SMIRC drafted additional mathematics and science items to improve coverage of framework topics.

Preparing the TIMSS 2023 Field Test Instruments

In preparation for the TIMSS 2023 Field Test, items were reviewed for their content several times in paper formats (i.e., in shareable text documents), at which point considerations for any intended functionality were also planned. After drafting and reviewing the items on paper, TIMSS & PIRLS International Study Center staff entered them into RM’s Assessment Master, where they were subjected to several rounds of quality control review. At this stage, any planned functionality was evaluated based on the implementation into Assessment Master as reviews and adjustments continued. Because of the complexity designed into the PSI tasks, RM programmed those items separately outside of the item-building interface, although similar reviews continued during their development.

Hundreds of items were field-tested in the participating countries to have a sufficient number of new items for the new item blocks used in TIMSS 2023. The TIMSS 2023 fourth-grade assessment required 137 mathematics and 126 science items to be field-tested. Among these items were four mathematics Problem-Solving and Inquiry (PSI) tasks and one science PSI. The TIMSS 2023 eighth-grade assessment required 163 mathematics items and 170 science items to be field-tested, including four mathematics PSI tasks and two science PSI tasks. Exhibit 1.4 shows the total number of items included in the TIMSS 2023 Field Test by grade and subject and the number of participants. Exhibits 1.5 through 1.8 detail the composition of the field test items by content domain and cognitive domain. Because of rounding, percentages may not sum to 100.

Exhibit 1.4: TIMSS 2023 Field Test Total Number of Items and Participants

	Grade 4	Grade 8
Items in Field Test		
Mathematics	137	163
Science	126	170
Total	263	333
Responses per Item per Country (approx.)	200	200
Participants		
Countries	51	39
Benchmarking Entities	6	4
Students	61,508	51,422
Teachers	4,187	7,319
Schools	1,735	1,205

Exhibit 1.5: TIMSS 2023 Field Test Items by Content Domain Item Format – Grade 4

Content Domain	Number of Items	Percentage of Total Items
Mathematics		
Number	64	47%
Measurement and Geometry	42	31%
Data	31	23%
Total	137	
Science		
Life Science	55	44%
Physical Science	40	32%
Earth Science	31	25%
Total	126	

Exhibit 1.6: TIMSS 2023 Field Test Items by Cognitive Domain and Item Format – Grade 4

Cognitive Domain	Number of Items	Percentage of Total Items
Mathematics		
Knowing	49	36%
Applying	57	42%
Reasoning	31	23%
Total	137	
Science		
Knowing	50	40%
Applying	53	42%
Reasoning	23	18%
Total	126	

Exhibit 1.7: TIMSS 2023 Field Test Items by Content Domain and Item Format – Grade 8

Content Domain	Total Number of Items	Percentage of Total Items
Mathematics		
Number	54	33%
Algebra	38	23%
Geometry and Measurement	36	22%
Data and Probability	35	21%
Total	163	
Science		
Biology	65	38%
Chemistry	32	19%
Physics	37	22%
Earth Science	36	21%
Total	170	

Exhibit 1.8: TIMSS 2023 Field Test Items by Cognitive Domain and Item Format – Grade 8

Cognitive Domain	Total Number of Items	Percentage of Total Items
Mathematics		
Knowing	51	31%
Applying	70	43%
Reasoning	42	26%
Total	163	
Science		
Knowing	57	34%
Applying	79	46%
Reasoning	34	20%
Total	170	

Developing Materials for TIMSS 2023 Field Test Scoring Training

The TIMSS & PIRLS International Study Center prepared all scoring guides before countries administered the TIMSS 2023 Field Test so that scoring could begin immediately after the field test data were collected. Student responses for the TIMSS 2023 Field Test scoring training materials were sourced from a small constructed-response pilot conducted by English-speaking countries in November 2021. Even though the continued impact of the COVID-19 pandemic proved challenging to accommodate for many countries, Australia, England, and New Zealand

administered the pilot and provided numerous responses to the TIMSS & PIRLS International Study Center. Across the three countries at the fourth grade, approximately 200 responses were received across five mathematics items, and 800 responses were received across 23 science items. At the eighth grade, New Zealand collected approximately 200 responses across 16 mathematics items and 400 responses across 33 science items. The responses were reviewed and incorporated into the draft field test scoring guides and the field test scoring materials compiled by consultants and experts at the TIMSS & PIRLS International Study Center during an online meeting in January 2022. For the TIMSS 2023 Field Test, training materials were prepared for 7 items at the fourth grade and 11 items at the eighth grade.

The TIMSS 2023 NRCs and their scoring supervisors received scoring training for the field test constructed-response items in February 2022 as part of the 4th TIMSS 2023 NRC meeting held online. TIMSS & PIRLS International Study Center experts and consultants followed a consistent procedure to introduce the item and scoring guide criteria for each item subjected to training.

Finalizing the TIMSS 2023 Achievement Instruments

Following the field test, the TIMSS & PIRLS International Study Center reviewed the data and selected items for data collection based on psychometric quality and the assessment framework specifications, including the group adaptive design requirements. With a pool of items selected, items were combined into recommended blocks based on overall block difficulty, content and cognitive domain, and item type.

The SMIRC scrutinized the recommendations for the newly developed achievement items at the 3rd TIMSS 2023 SMIRC meeting, reviewing the items and scoring guides for content accuracy, clarity, and adherence to the frameworks. Block organization was finalized such that blocks generally progressed from easier items to harder items and that a block contained a variety of item formats to keep students engaged. At each grade and subject, two difficult blocks, one medium block, and three easy blocks were composed of new items administered in the TIMSS 2023 Field Test.

NRCs had the opportunity to review the recommended assessment blocks in light of the field test results. The review included a check for any unusual national results that might indicate translation errors and resulted in requests to correct the translation as necessary. Finally, the 5th TIMSS 2023 NRC meeting, held virtually in September 2022, was devoted to reviewing all the newly developed items and compiled item blocks.

Following the final review, the newly developed and existing trend item blocks were arranged into booklets and finalized in the Assessment Master by the TIMSS & PIRLS International Study Center and RM Results.

Distribution of the TIMSS 2023 Achievement Items

The final TIMSS 2023 Data Collection blocks consist of about 40% newly developed field-tested items and 60% trend items administered in previous cycles. The newly developed items were selected and combined into blocks so that, when combined with the trend blocks, the overall distribution of the TIMSS assessment reflects the specifications of content and cognitive domains defined in the assessment frameworks.

Exhibits 1.9 and 1.10 display the number of trend items and newly developed items and the number of score points in the TIMSS 2023 fourth-grade assessment by content domain and cognitive domain, respectively. Exhibits 1.11 and 1.12 present the items in TIMSS 2023 eighth-grade assessments by content and cognitive domains. PSI items administered in TIMSS 2019 are included in the trend item counts.

Exhibit 1.9: TIMSS 2023 Achievement Items by Content Domain – Grade 4

Content Domain	Trend			New			Total			Target
	Number of Items	Score Points	Percentage of Score Points	Number of Items	Score Points	Percentage of Score Points	Number of Items	Score Points	Percentage of Score Points	Percentage of Score Points
Mathematics										
Number	55	59	51%	41	41	52%	96	100	51%	50%
Measurement and Geometry	29	31	27%	21	21	27%	50	52	27%	30%
Data	23	26	22%	17	17	22%	40	43	22%	20%
Total	107	116		79	79		186	195		
Science										
Life Science	39	41	39%	40	44	52%	79	85	45%	45%
Physical Science	39	44	42%	23	23	27%	62	67	36%	35%
Earth Science	19	19	18%	16	17	20%	35	36	19%	20%
Total	97	104		79	84		176	188		

Exhibit 1.10: TIMSS 2023 Achievement Items by Cognitive Domain – Grade 4

Cognitive Domain	Trend			New			Total			Target
	Number of Items	Score Points	Percentage of Score Points	Number of Items	Score Points	Percentage of Score Points	Number of Items	Score Points	Percentage of Score Points	Percentage of Score Points
Mathematics										
Knowing	36	36	31%	23	23	29%	59	59	30%	40%
Applying	48	54	47%	38	38	48%	86	92	47%	40%
Reasoning	23	26	22%	18	18	23%	41	44	23%	20%
Total	107	116		79	79		186	195		
Science										
Knowing	39	43	41%	31	33	39%	70	76	40%	40%
Applying	36	36	35%	37	40	48%	73	76	40%	40%
Reasoning	22	25	24%	11	11	13%	33	36	19%	20%
Total	97	104		79	84		176	188		

Exhibit 1.11: TIMSS 2023 Achievement Items by Content Domain – Grade 8

Content Domain	Trend			New			Total			Target
	Number of Items	Score Points	Percentage of Score Points	Number of Items	Score Points	Percentage of Score Points	Number of Items	Score Points	Percentage of Score Points	Percentage of Score Points
Mathematics										
Number	33	35	28%	30	30	35%	63	65	31%	30%
Algebra	41	42	33%	17	17	20%	58	59	28%	30%
Geometry and Measurement	24	28	22%	18	18	21%	42	46	22%	20%
Data and Probability	21	22	17%	20	20	24%	41	42	20%	20%
Total	119	127		85	85		204	212		
Science										
Biology	40	47	33%	39	40	42%	79	87	37%	35%
Chemistry	29	33	23%	14	14	15%	43	47	20%	20%
Physics	36	38	27%	17	18	19%	53	56	24%	25%
Earth Science	22	24	17%	23	24	25%	45	48	20%	20%
Total	127	142		93	96		220	238		

Exhibit 1.12: TIMSS 2023 Achievement Items by Cognitive Domain – Grade 8

Cognitive Domain	Trend			New			Total			Target
	Number of Items	Score Points	Percentage of Score Points	Number of Items	Score Points	Percentage of Score Points	Number of Items	Score Points	Percentage of Score Points	Percentage of Score Points
Mathematics										
Knowing	32	33	26%	28	28	33%	60	61	29%	35%
Applying	54	55	43%	39	39	46%	93	94	44%	40%
Reasoning	33	39	31%	18	18	21%	51	57	27%	25%
Total	119	127		85	85		204	212		
Science										
Knowing	45	50	35%	28	28	29%	73	78	33%	35%
Applying	49	56	39%	44	46	48%	93	102	43%	35%
Reasoning	33	36	25%	21	22	23%	54	58	24%	30%
Total	127	142		93	96		220	238		

TIMSS 2023 Achievement Items by Item Format and Content and Cognitive Domains

TIMSS 2023 item development advanced the objectives of digital assessment started in TIMSS 2019 by leveraging the capabilities of a fully digital assessment. To assess the broad range of mathematics and science topics and skills described in the assessment frameworks, the TIMSS 2023 fourth- and eighth-grade assessments included a variety of item types (see earlier Exhibit 1.1).

The TIMSS assessments include two general types of selected-response items—single selection, in which students choose one of four response options, and multiple selection, in which students choose more than one option from several response options or make a series of selections to respond to a question. Most TIMSS 2023 selected-response items were worth one score point, although some multiple selection items were worth two. The two-point multiple selection items were scored as fully correct (all parts answered correctly; 2 score points), partially correct (most parts answered correctly; 1 score point), or incorrect (few or no parts answered correctly; 0 score points). Selection items were entirely machine-scored in TIMSS 2023.

Constructed-response items involving writing or typing words or numbers, drawing, or dragging and dropping were worth one or two score points depending on the degree of complexity involved. The 1-point constructed response items were scored as correct (1 score point) or incorrect (0 score points), whereas 2-point constructed response items were scored as fully correct (2 score points), partially correct (1 score point), or incorrect (0 score points). Fully correct responses show a complete or deeper understanding of a task, while partially correct responses show only partial understanding of the concepts or procedures embodied in the task. The proportion of items requiring human scoring was greatly reduced as a direct benefit of fully digital item development and computer delivery of the TIMSS 2023 assessment. As in TIMSS 2019, items requiring a simple numerical response as well as drag-and-drop items

continued to be machine-scored. Constructed-response items using graphing tool and chart tool formats also were machine-scored in TIMSS 2023. See [Chapter 7](#) of this report for more information about scoring and scoring reliability in TIMSS 2023.

It is important to verify that an assortment of selected- and constructed-response items is used to assess each domain to ensure sufficient coverage of the assessment frameworks. Exhibits 1.13 through 1.16 display the number of items (and score points) by item format for each content and cognitive domain in the fourth- and eighth-grade assessments.

Exhibit 1.13: TIMSS 2023 Achievement Items by Content Domain and Item Format – Grade 4

Content Domain	Selected-Response		Constructed-Response	
	Number of Items	Number of Score Points	Number of Items	Number of Score Points
Mathematics				
Number	37	37	59	63
Measurement and Geometry	28	28	22	24
Data	14	14	26	29
Total	79	79	107	116
Achieved Percentage of Score Points	41%		59%	
Science				
Life Science	40	41	39	44
Physical Science	39	39	23	28
Earth Science	24	24	11	12
Total	103	104	73	84
Achieved Percentage of Score Points	55%		45%	

Exhibit 1.14: TIMSS 2023 Achievement Items by Cognitive Domain and Item Format – Grade 4

Content Domain	Selected-Response		Constructed-Response	
	Number of Items	Number of Score Points	Number of Items	Number of Score Points
Mathematics				
Knowing	34	34	25	25
Applying	32	32	54	60
Reasoning	13	13	28	31
Total	79	79	107	116
Achieved Percentage of Score Points	41%		59%	
Science				
Knowing	50	51	20	25
Applying	39	39	34	37
Reasoning	14	14	19	22
Total	103	104	73	84
Achieved Percentage of Score Points	55%		45%	

Exhibit 1.15: TIMSS 2023 Achievement Items by Content Domain and Item Format – Grade 8

Content Domain	Selected-Response		Constructed-Response	
	Number of Items	Number of Score Points	Number of Items	Number of Score Points
Mathematics				
Number	17	18	46	47
Algebra	22	22	36	37
Geometry and Measurement	8	8	34	38
Data and Probability	13	14	28	28
Total	60	62	144	150
Achieved Percentage of Score Points	29%		71%	
Science				
Biology	43	45	36	42
Chemistry	21	22	22	25
Physics	33	33	20	23
Earth Science	28	29	17	19
Total	125	129	95	109
Achieved Percentage of Score Points	54%		46%	

Exhibit 1.16: TIMSS 2023 Achievement Items by Cognitive Domain and Item Format – Grade 8

Content Domain	Selected-Response		Constructed-Response	
	Number of Items	Number of Score Points	Number of Items	Number of Score Points
Mathematics				
Knowing	29	30	31	31
Applying	22	22	71	72
Reasoning	9	10	42	47
Total	60	62	144	150
Achieved Percentage of Score Points	29%		71%	
Science				
Knowing	57	59	16	19
Applying	45	47	48	55
Reasoning	23	23	31	35
Total	125	129	95	109
Achieved Percentage of Score Points	54%		46%	

References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (2014). *Standards for educational and psychological testing*. Washington DC: AERA Publications.
- Fishbein, B., Martin, M. O., Mullis, I. V. S., & Foy, P. (2018). The TIMSS 2019 item equivalence study: Examining mode effects for computer-based assessment and implications for measuring trends. *Large-scale Assessments in Education*, 6(1), 11. <https://doi.org/10.1186/s40536-018-0064-z>
- International Test Commission (ITC). (2019). ITC guidelines for the large-scale assessment of linguistically and culturally diverse populations. *International Journal of Testing*, 19(4), 301–336. <https://doi.org/10.1080/15305058.2019.1631024>
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (Research Report No. RR-03-16). Princeton, NJ: Educational Testing Service.
- Mullis, I. V. S., Martin, M. O., Fishbein, B., Foy, P., & Moncaleano, S. (2021). *Findings from the TIMSS 2019 Problem Solving and Inquiry Tasks*. Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/timss2019/psi/>
- von Davier, M., Foy, P., Martin, M. O., & Mullis, I. V. S. (2020). Examining eTIMSS country differences between eTIMSS data and bridge data: A look at country-level mode of administration effects. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report* (pp. 13.1–13.24). Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/timss2019/methods/chapter-13.html>