CHAPTER 1

Developing the TIMSS 2019 Mathematics and Science Achievement Instruments

Kerry E. Cotter
Victoria A.S. Centurino
Ina V.S. Mullis

Unique Characteristics of TIMSS 2019

The TIMSS assessments are designed to provide valid measurement of the mathematics and science content and skills that are valued by the international education community and included in the 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, but each cycle has some unique characteristics that influence instrument development. Besides providing measures on another cycle for the TIMSS trend lines monitoring changes in educational achievement since 1995, TIMSS 2019 also was remarkable for several reasons.

- TIMSS 2019 marked the beginning of the transition to eTIMSS—a digital version of TIMSS designed for computer- and tablet-based administration. eTIMSS offered an engaging, interactive, and visually attractive assessment that enabled TIMSS 2019 to better assess complex areas of the mathematics and science frameworks and increase operational efficiency in translation, assessment delivery, data entry, and scoring.
- As a part of the transition to digital assessment, eTIMSS 2019 included a series of extended Problem Solving and Inquiry (PSI) tasks in mathematics and science at both the fourth and the eighth grades. The eTIMSS PSIs were designed to simulate real world or laboratory situations in which students could integrate and apply process skills and content knowledge to solve mathematics problems or conduct virtual scientific experiments and investigations.
- Building on the success of TIMSS Numeracy 2015, the TIMSS 2019 assessment design was expanded to support a less difficult version of the fourth grade mathematics assessment that had some blocks of items in common with the regular fourth grade mathematics assessment. The less difficult version enabled TIMSS 2019 to provide improved measurement for participating countries where fourth grade students were still developing fundamental mathematics skills.

Results for the two versions of the fourth grade mathematics assessment were linked through the common items and reported on the same TIMSS achievement scale.

Transitioning TIMSS to eTIMSS

Transitioning to digital assessment is important to "keep up with the times" and to increase both construct representation and data utility. Because not all TIMSS countries were prepared to conduct digital assessments, IEA decided to implement the transition over two assessment cycles—TIMSS 2019 and TIMSS 2023. More than half of the 64 countries participating in TIMSS 2019 elected to administer the "e" version of the assessments, while the rest of the countries administered TIMSS in paper-and-pencil format, as in previous assessment cycles (paperTIMSS).

The eTIMSS 2019 assessments included a variety of technology-enhanced item formats, with colorful graphics and interactive features. These features extended coverage of the mathematics and science frameworks and promoted student engagement. The digital mode of administration also allowed for a substantial proportion of the eTIMSS mathematics items to be machine scored.

The eTIMSS 2019 assessments were created and administered using IEA's eAssessment System, which houses a collection of online tools used for instrument creation, translation and adaptation, verification, delivery to students, scoring, and data entry. The digital mode of administration allowed eTIMSS to collect information about how students work through the items, such as screen-by-screen timing data and additional process variables that can be analyzed to study students' interactions with the achievement items.

The eTIMSS 2019 PSIs, designed exclusively for eTIMSS, were a new and pioneering effort to improve measurement of higher-order mathematics and science skills by capitalizing on the digital mode of administration. Each PSI consisted of a sequence of 4 to 16 items that were set in a cohesive context and addressed a range of topics from the *TIMSS 2019 Assessment Frameworks* (Mullis & Martin, 2017), such as solving a series of mathematics problems to plan a school event or conducting a virtual scientific experiment to study plant growth. The items within these situational tasks included a broader array of innovative digital features than the regular eTIMSS achievement items and provided scaffolding for complex mathematics problems and science investigations.

Less Difficult Mathematics at the Fourth Grade

For a variety of reasons, there are some countries where most children in the fourth grade are still developing fundamental mathematics skills. To offer countries the most effective assessment of fourth grade mathematics, IEA offers options for matching the TIMSS fourth grade mathematics assessment to the country's educational development and students' mathematics proficiency. For some countries, the less difficult version of the TIMSS 2019 fourth grade mathematics assessment was a better match with students' learning.

The TIMSS 2019 fourth grade assessment with less difficult mathematics was developed together with the regular fourth grade mathematics assessment and reflected the mathematics described in the TIMSS 2019 Mathematics Framework (Lindquist, Philpot, Mullis, & Cotter, 2017). The regular and less difficult versions of the assessment were equivalent in scope, and about one-third of the items were the same between the two versions. The other two-thirds of the items in each version of the assessment addressed the same areas of the mathematics framework, but the items in the less difficult version involved less complex numbers and situations. The items in common between the two versions of the fourth grade mathematics assessment enabled the two assessments to be linked so that the results could be reported together and directly compared. Expert committees reviewed both the regular and less difficult mathematics items together at each phase of development.

The design of the TIMSS 2019 less difficult mathematics assessment improved upon the design of TIMSS Numeracy 2015, which was a stand-alone mathematics assessment that did not include science. For TIMSS 2019 at the fourth grade, countries could opt for either regular or less difficult mathematics, together with science. A substantial portion of the items in the less difficult version of the TIMSS 2019 mathematics assessment was carried forward from TIMSS Numeracy 2015, which enabled trend measurement for countries that participated in TIMSS Numeracy 2015.

The TIMSS Approach to Measuring Trends

Because TIMSS is designed to measure trends, the assessments of mathematics and science cannot change dramatically from cycle to cycle. That is, 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 also need to be updated with each cycle to prevent the assessments from becoming dated and no longer relevant to current learning goals and policy issues. It is important that TIMSS reflects the most recent discoveries in the field and is presented in ways consistent with students' instructional and everyday experiences.

To maintain continuity with past assessments while keeping up with current topics and technology, the TIMSS assessments evolve with each cycle. TIMSS has a specific design for rotating items out of the assessment after each cycle and replacing them with newly developed items for the following cycle. The remaining achievement items, referred to as "trend items," are kept secure to be re-administered in subsequent cycles. With this design for item replacement, each TIMSS assessment includes items from three cycles—newly developed items, items from the previous cycle, and items from two cycles before.

Overview of the TIMSS 2019 Achievement Items

Although the majority of the TIMSS achievement items are carried forward from the previous assessment cycle to measure trends, the task of updating the instruments for each new cycle—every four years since 1995—is a substantial undertaking. Because TIMSS assesses two subjects at two grades, it actually encompasses five different assessments of achievement—regular and less difficult mathematics at the fourth grade, mathematics at the eighth grade, and science at the fourth and eighth grades.

The TIMSS 2019 fourth grade assessments required developing and field testing 261 new mathematics and science items in both digital and paper formats as well as 66 new paper-based items for the less difficult version of the mathematics assessment. The TIMSS 2019 eighth grade assessments required developing and field testing 325 new mathematics and science items in both digital and paper formats. For eTIMSS 2019, the field test also included eight mathematics and science PSI tasks at the fourth grade and seven mathematics and science PSI tasks at the eighth grade.

Since the beginning in 1995, the TIMSS assessments have included two general item formats: selected response (i.e., questions to which students choose their answer(s) from a set of options) and constructed response (i.e., questions to which students construct their own responses). For each constructed response item, a unique scoring guide is developed along with the item with clear distinctions among correct, incorrect, and, if applicable, partially correct answers. The format of each item is chosen based on the mathematics or science content and cognitive domain being assessed.

The Item Development Process

The TIMSS & PIRLS International Study Center at Boston College employs a collaborative process inspired by the principles of the evidence-centered design framework (ECD; Mislevy, Almond, & Lukas, 2003) to develop the new achievement items needed for each TIMSS cycle. With this approach, validity is supported by adhering to best practices in assessment design throughout the development process—namely, clearly defining the target construct to be measured, specifying the items needed to measure it, establishing standards for items and test forms, and ensuring that the assessments meet the test specifications. A broad overview of this process to support coherence between the assessment goals and data includes:

- Updating the assessment frameworks to identify and prioritize the mathematics and science content and skills that the assessment will measure
- Developing 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 the new items to meet the assessment specifications based on the field test results and trend items from previous cycles
- Conducting training in how to reliably score students' responses to constructed response items to ensure the quality of the data

The development process is directed and managed by the staff of the TIMSS & PIRLS International Study Center, who collectively have considerable experience in the measurement and assessment of mathematics and science achievement. For TIMSS 2019, Executive Director, Ina Mullis, and Assistant Director of Mathematics, Kerry Cotter, managed the mathematics assessment development. Executive Director, Michael Martin, and Assistant Director of Science, Victoria Centurino, managed the science assessment development.

Also playing a key role in achievement item development were the TIMSS 2019 National Research Coordinators (NRCs) designated by their countries to be responsible for the complex tasks involved in implementing TIMSS in their countries. The TIMSS & PIRLS International Study Center worked 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 reviewed the items prior to the field test and helped select the items for the assessment after the field test.

The TIMSS & PIRLS International Study Center prepared an international version of all the TIMSS achievement items in English. Subsequently, the items were translated by participating countries into their languages of instruction with the goal of creating high quality translations that were appropriately adapted for the national context and at the same time remained internationally comparable. Therefore, a significant portion of the NRCs' development and review effort was dedicated to ensuring that the achievement items could be translated accurately.

Additional advice and guidance was provided through periodic reviews by the TIMSS 2019 Science and Mathematics Item Review Committee (SMIRC). SMIRC members for each TIMSS cycle are nominated by countries participating in TIMSS and provide guidance in developing the TIMSS assessments. The TIMSS 2019 SMIRC consisted of 13 members: 7 experts in mathematics and mathematics education and 6 experts in science and science education.

SMIRC members met four times for TIMSS 2019. At the 1st TIMSS 2019 SMIRC meeting in Amsterdam, The Netherlands (April 2017), the committee reviewed the mathematics and science content frameworks and initial drafts of the mathematics and science PSIs. At the 2nd meeting in Windsor, England (September 2017), SMIRC reviewed draft field test items, together with the scoring guides for constructed response items. At the 3rd meeting in Tromsø, Norway (July 2018), SMIRC reviewed field test results and made recommendations regarding the items to include in the TIMSS 2019 mathematics and science assessments. At the final meeting in Singapore (May 2020), SMIRC conducted the TIMSS 2019 scale anchoring process (see Using Scale Anchoring to Interpret the TIMSS 2019 Achievement Scales). Exhibit 1.1 lists the TIMSS 2019 SMIRC members.

Exhibit 1.1: TIMSS 2019 Science and Mathematics Item Review Committee (SMIRC)

Mathematics

Ray Philpot
Australian Council for Educational Research (ACER)
Australia

Kiril Bankov
Faculty of Mathematics and Informatics, University of Sofia
Bulgaria

Khattab Mohammad Ahmad Abulibdeh
National Center for Human Resources Development
Jordan

Arne Hole
Department of Teacher Education and School Research, University of Oslo
Norway

Cheow Kian Soh
Ministry of Education, Curriculum Planning, and Development Division, Mathematics Branch
Singapore

Mary Lindquist
Professor Emeritus Mathematics Education, Columbus State University
United States

Linda Hall
Mathematics Consultant
United States

Science

Svatava Janoušková
Science Faculty Department of Teaching and Didactics of Chemistry, Charles University Prague
Czech Republic

Emily Jones
National Foundation of Educational Research (NFER)
England

Jouni Viiri
Department of Teacher Education, University of Jyväskylä
Finland

Berenice Michels
Faculty of Science, Freudenthal Institute for Science and Mathematics Education
The Netherlands

Galina Kovaleva
Federal Institute for Strategy of Education Development of the Russian Academy of Education Center for Evaluating the Quality of Education
Russian Federation

Christopher Lazzaro
The College Board
United States

Developing the PSIs and technology-enhanced achievement items to meet the ambitious development goals for eTIMSS 2019 necessitated even more expert review and collaboration than previous TIMSS cycles. Several SMIRC members worked closely with staff at the TIMSS & PIRLS International Study Center throughout the development process to achieve these goals. For mathematics, Mary Lindquist and Ray Philpot provided additional subject-matter expertise and support. For science, Emily Jones, Christopher Lazzaro, and Berenice Michels served in this capacity.

The TIMSS 2019 Development Schedule

In preparation for the transition to eTIMSS, development work for TIMSS 2019 began over three years before the TIMSS 2019 Field Test and included a series of novel activities to develop the eTIMSS user interface, eAssessment System, and PSIs. Essentially, the first two years were devoted to updating the assessment frameworks and pilot testing the mathematics and science PSIs and trend items in digital format. The third year was dedicated to writing new achievement items in both digital and paper format, continuing to refine the PSIs, and testing components of the eAssessment System to ensure successful delivery of eTIMSS across a variety of digital devices and testing conditions.

The TIMSS 2019 Field Test was conducted from March through May 2018. After a thorough review of the results, the materials for data collection were finalized in August 2018. TIMSS 2019 Data Collection began in the Southern Hemisphere in September 2018 and continued in the Northern Hemisphere through May 2019.

Exhibit 1.2 shows the TIMSS 2019 development schedule for the achievement items beginning with initial work on the eAssessment System through TIMSS 2019 Data Collection.

Exhibit 1.2: TIMSS 2019 Development Schedule for Achievement Items

Date(s)		Group and Activity
January	2015	TIMSS & PIRLS International Study Center and IEA Hamburg began designing the eTIMSS assessment system, user interface, and digital item types, including the PSIs, in preparation for the transition to eTIMSS
March	2015	TIMSS & PIRLS International Study Center began work with members of the Science and Mathematics Item Review Committee (SMIRC), other external expert consultants, and IEA Hamburg to design and operationalize prototype PSIs
August	2015	Consultants and staff at the TIMSS & PIRLS International Study Center began drafting additional PSIs for both the fourth and eighth grade assessments (Boston, USA)
August	2015	American Institutes for Research (AIR) conducted cognitive laboratories for two prototype PSIs (one fourth grade mathematics and one eighth grade science) and a sample of TIMSS trend items converted to digital format
October	2015	Consultants and staff at the TIMSS & PIRLS International Study Center reviewed the results of the cognitive laboratories, continued revising the draft PSIs, and drafted new PSIs (Boston, USA)
June	2016	TIMSS & PIRLS International Study Center presented an informational video introducing the features of the eTIMSS assessments and debuting the PSIs to National Research Coordinators (NRCs) (8th NRC meeting—Quebec, Canada)
June–September	2016	TIMSS & PIRLS International Study Center conducted content analysis of the curricular topics described in the *TIMSS 2015 Encyclopedia* and proposed updates to the mathematics and science frameworks for TIMSS 2019
September	2016	SMIRC reviewed the draft *TIMSS 2019 Assessment Frameworks* and provided feedback to staff at the TIMSS & PIRLS International Study Center. The staff then met with SMRIC consultants to incorporate SMIRC's comments (Boston, USA)

Exhibit 1.2: TIMSS 2019 Development Schedule for Achievement Items (continued)

Date(s)		Group and Activity
October	2016	Australia, Canada, and Singapore administered the eTIMSS prePilot, which included a sample of trend items converted to digital format and draft PSIs
November	2016	Consultants and staff at the TIMSS & PIRLS International Study Center reviewed the results of the eTIMSS prePilot and revised the PSIs and user interface specifications based on these results. The group also drafted one additional PSI for each grade, fulfilling the development requirements for the eTIMSS 2019 Field Test (Boston, USA)
February	2017	NRCs reviewed the draft *TIMSS 2019 Assessment Frameworks* (1st NRC meeting—Hamburg, Germany). Following the meeting, NRCs completed an online survey to provide feedback as to whether each topic area should be kept as is, modified, or deleted
March–April	2017	TIMSS & PIRLS International Study Center prepared draft *TIMSS 2019 Item Writing Guidelines*, including specific guidelines for the enhanced item formats available for eTIMSS. Staff also revised the draft *TIMSS 2019 Assessment Frameworks* based on feedback from NRCs
April	2017	SMIRC reviewed the draft *TIMSS 2019 Assessment Frameworks*, *TIMSS 2019 Item Writing Guidelines*, and PSIs (1st TIMSS 2019 SMIRC meeting—Amsterdam, The Netherlands)
May	2017	NRCs reviewed the *TIMSS 2019 Assessment Frameworks* and developed draft field test items and scoring guides using the *TIMSS 2019 Item Writing Guidelines* (2nd NRC meeting—Hamburg, Germany)
May	2017	The eTIMSS Pilot/Item Equivalence Study, designed to investigate mode effects for the TIMSS trend items, was conducted to provide information about the robustness of the eAssessment System and countries' readiness to conduct a digital assessment
July	2017	Consultants and staff at the TIMSS & PIRLS International Study Center reviewed and revised draft field test items and scoring guides, including PSIs (Boston, USA)
September	2017	SMIRC reviewed the draft field test items and scoring guides, including PSIs (2nd SMIRC meeting—Windsor, England)
September	2017	Consultants and staff at the TIMSS & PIRLS International Study Center reviewed the updated field test items and PSIs and refined the scoring guides with special attention to machine scoring (Boston, USA)
November	2017	NRCs reviewed and approved the TIMSS 2019 Field Test instruments (3rd NRC meeting—Melbourne, Australia)
December	2017	TIMSS & PIRLS International Study Center and IEA Hamburg assembled all TIMSS 2019 Field Test instruments and released the international instruments to countries for translation
January–March	2018	TIMSS & PIRLS International Study Center and IEA Hamburg collaborated to establish specifications for eTIMSS data capture and machine-scored constructed response items
January	2018	TIMSS & PIRLS International Study Center collected student responses to constructed response items from English-speaking countries to develop scoring training materials for the field test
January	2018	Consultants and staff at the TIMSS & PIRLS International Study Center reviewed the field test scoring guides and prepared scorer training materials (Boston, USA)
March–May	2018	Countries conducted the TIMSS 2019 Field Test

Exhibit 1.2: TIMSS 2019 Development Schedule for Achievement Items (continued)

Date(s)		Group and Activity
March	2018	NRCs received scoring training for constructed response field test items (4th NRC meeting—Madrid, Spain)
May	2018	Countries submitted TIMSS 2019 Field Test achievement data for analysis and review
May	2018	NRCs provided feedback to the TIMSS & PIRLS International Study Center about the field-tested PSIs. Based on the NRC's evaluations, the TIMSS & PIRLS International Study Center selected the PSIs to move forward to eTIMSS 2019 Data Collection and began editing the tasks based on NRC feedback
June	2018	IEA Hamburg completed data processing and TIMSS & PIRLS International Study Center completed scoring of machine-scored items
June	2018	TIMSS & PIRLS International Study Center reviewed the field test item statistics and assembled sets of proposed items for data collection
July	2018	SMIRC reviewed the proposed items for data collection in conjunction with the field test results (3rd SMIRC meeting—Tromsø, Norway)
August	2018	NRCs reviewed and approved the proposed item blocks for TIMSS 2019 Data Collection (5th NRC meeting—Stockholm, Sweden)
September	2018	TIMSS & PIRLS International Study Center and IEA Hamburg finalized all TIMSS 2019 Data Collection instruments and released the international instruments to countries for translation
September–December	2018	Southern Hemisphere countries conducted TIMSS 2019 data collection
September	2018	Consultants and staff at the TIMSS & PIRLS International Study Center reviewed and updated scoring guides and scorer training materials (Boston, USA)
November	2018	NRCs from Southern Hemisphere countries received scoring training for constructed response items (Cape Town, South Africa)
November	2018	TIMSS & PIRLS International Study Center finalized scoring guides and training materials for constructed response items and distributed them to NRCs from Southern Hemisphere countries
March	2019	NRCs from Northern Hemisphere countries received scoring training for constructed response items (Limassol, Cyprus)
March–June	2019	Northern Hemisphere countries conducted TIMSS 2019 data collection

Updating the Assessment Frameworks for TIMSS 2019

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 around the world. The first two chapters of the *TIMSS 2019 Assessment Frameworks* (Mullis & Martin, 2017), respectively, describe the mathematics and science frameworks in detail.

Writing and Reviewing the TIMSS 2019 Field Test Items and Scoring Guides

The TIMSS 2019 Field Test included approximately one and a half times the number of achievement items needed for data collection, to ensure a sufficient number of high quality items for the TIMSS 2019 assessments. In all, about 800 items were field tested. With the exception of the PSIs (eTIMSS only) and less difficult mathematics items (paper only), all items were prepared and administered in both digital and paper format. These items were designed to be identical in content across eTIMSS and paperTIMSS, with the only difference being the response mode (e.g., a drag and drop item in eTIMSS may be a matching item in paperTIMSS).

The TIMSS & PIRLS International Study Center uses a collaborative process involving the participating countries to develop the substantial number of new items and scoring guides needed for the field test. Most of the 2nd TIMSS 2019 NRC meeting in Hamburg, Germany was devoted to a workshop for developing the field test items. The NRCs, together with experienced item writers from participating countries and staff from the TIMSS & PIRLS International Study Center, drafted the majority of the new items for the mathematics and science field tests during this workshop.

In preparation for the item writing workshop, staff at the TIMSS & PIRLS International Study Center identified the scope of the item writing task for the field test. Considerations included the total items needed based on the weight assigned to a particular topic in the *TIMSS 2019 Assessment Frameworks* (Mullis & Martin, 2017), as well as how many items existed from previous assessments. The TIMSS & PIRLS International Study Center also updated the item writing manual specifically developed for TIMSS assessments. The manual contains general information about 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 deal with translation and comparability issues (e.g., using TIMSS' fictitious unit of currency, the "zed," for items involving money). The manual also includes the necessary steps for developing scoring guides for constructed response items, as well as checklists for reviewing TIMSS items.

Updated for the transition to eTIMSS, the *TIMSS 2019 Item Writing Guidelines* provided additional instructions for taking advantage of the technology-enhanced item formats—drag and drop, sorting, selection, drop-down menus, and a line drawing tool. These guidelines included examples of how each enhanced item format might be used (e.g., using drag and drop for adding labels to graphs or diagrams) and some details about the functionality of the formats (e.g., the maximum number of "draggable" parts available in a drag and drop item).

At the TIMSS 2019 Item Writing Workshop, country representatives were divided into teams and given specific item writing assignments based on their areas of expertise to ensure that enough field test items were developed for each of the content areas and cognitive processes specified in the frameworks.

Staff from the TIMSS & PIRLS International Study Center used the *TIMSS 2019 Item Writing Guidelines* to provide training to the teams on item writing procedures. The teams were asked to provide a complete draft of each item they developed, including the content topic and cognitive area from the framework that the item addressed and the information needed to score the item (i.e., an answer key for selected response items or scoring guide for constructed response items). Once teams had completed their own item writing assignments, they reviewed the items drafted by other teams. In addition, some teams continued to send items to the TIMSS & PIRLS International Study Center for several weeks after the Item Writing Workshop.

Exhibit 1.3 shows the number of participants in the TIMSS 2019 Item Writing Workshop and the approximate number of items written.

Exhibit 1.3: TIMSS 2019 Item Writing Workshop to Develop Field Test Items

Participants	
Number of Countries and Benchmarking Entities	53
Number of Country Representatives	118
Approximate Number of Field Test Items Written at Item Writing Workshop	
Fourth Grade Mathematics	300
Fourth Grade Science	200
Eighth Grade Mathematics	300
Eighth Grade Science	200

Following the item writing workshop, staff at the TIMSS & PIRLS International Study Center reviewed each item in light of the framework specifications and selected an optimal group of items for further review and revision. Consultants from the Australian Council for Educational Research (ACER) and the National Foundation of Educational Research (NFER) drafted additional mathematics and science items, respectively, to improve coverage of areas of the frameworks that are especially challenging to measure.

In July 2017, several SMIRC members with particular item writing skills met with staff from the TIMSS & PIRLS International Study Center to continue revising the draft field test items. SMIRC then reviewed all of the proposed draft field test items at the 2nd TIMSS 2019 SMIRC meeting. After SMIRC's review, the items were revised again, and the NRCs reviewed the complete set of draft field test items at the 3rd TIMSS 2019 NRC meeting in Melbourne, Australia. Following this meeting, staff at the TIMSS & PIRLS International Study Center implemented the final suggested revisions and provided the international versions of the field test instruments in digital or paper format to the NRCs so that they could begin translating the field test materials into their languages of instruction.

Preparing eTIMSS Field Test Items for Digital Delivery

Preparing the eTIMSS field test items for digital delivery required the additional step of entering each item into IEA's Item Builder, a web-based application for creating digital achievement items and instruments for delivery to students via computers and tablets. For eTIMSS 2019, the Item Builder included templates for both traditional (e.g., standard multiple-choice) and enhanced (e.g., drag and drop) item formats as well as 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 field test. After drafting and reviewing the field test items on paper, staff at the TIMSS & PIRLS International Study Center entered all eTIMSS items into the Item Builder and collaborated with IEA Hamburg to conduct extensive quality control tests to ensure each item would appear and function as intended for students.

Developing Problem Solving and Inquiry (PSIs) Tasks for eTIMSS

In many ways, PSI development work followed the standard TIMSS procedures for ensuring that the items provide valid measurement of the *TIMSS 2019 Assessment Frameworks* (Mullis & Martin, 2017). However, because the PSIs involved a new and more innovative approach to assessing mathematics and science achievement in a digital environment, PSI development required additional efforts.

Developing engaging problem contexts with cohesive sets of achievement items necessitated even more rounds of expert review than is typical for TIMSS items. Staff at the TIMSS & PIRLS International Study Center began collaborating with SMIRC members to develop the PSIs in March 2015, nearly two years before item writing for the rest of the TIMSS 2019 items began. Several SMIRC members worked closely with TIMSS staff to develop the PSIs, which included providing initial ideas for the tasks and participating in a series of meetings to develop and refine the problem contexts, items, and scoring guides.

Leading up to the field test, several SMIRC members and staff at the TIMSS & PIRLS International Study Center met a total of five times at Boston College and conducted many online reviews to refine the PSIs. SMIRC as a whole conducted its first in-depth review of the PSIs at the 1st TIMSS 2019 SMIRC meeting, which focused on the alignment between the tasks and the frameworks, the extent to which the technology in the tasks supported the intended response processes, and the cross-cultural appropriateness of the problem scenarios. The NRCs reviewed the PSIs prior to the field test at the 3rd TIMSS 2019 NRC meeting.

In addition to extensive expert review, cognitive laboratories and a pilot test were conducted in several eTIMSS countries in advance of the field test to gain insight into students' interactions with the PSIs and to test the functionality of the eAssessment System. This strand of development work provided critical information about the usability of innovative item types and the eTIMSS interface, the amount of time it took students to complete each task, and the approximate difficulty of the tasks. Following each

outing and review, improvements were made to both the PSIs and their software with the aim of eliciting the intended types of responses from students.

eTIMSS Cognitive Laboratories

Staff at the TIMSS & PIRLS International Study Center partnered with the American Institutes for Research (AIR) to conduct cognitive laboratories in the very early stages of the transition to eTIMSS (August 2015). The goal of this study was to investigate two aspects of digital assessment that would inform next steps in eTIMSS development: students' interactions with drafts of the first PSIs, and students' experiences with the eTIMSS interface.

The TIMSS & PIRLS International Study Center prepared two prototype PSIs and a set of TIMSS trend items in digital format at each grade, along with a list of research questions, from which AIR developed interview protocols. During the interviews, students explained their thoughts while engaging with the items on tablets, providing insight into how the PSI format and eTIMSS interface could be improved.

AIR conducted the interviews with a purposive sample of 32 fourth and eighth grade students from the greater Washington, D.C. area. Following the interviews, AIR prepared a report to address each of the TIMSS & PIRLS International Center's research questions. The reports from the cognitive laboratories prompted substantial revisions to the PSI item format and the eTIMSS interface. In particular, the students reported difficulties in using a stylus to write or draw, so the device keyboards were enabled for items requiring a written response and a new tool for drawing lines was developed.

eTIMSS prePilot

The eTIMSS prePilot was conducted in September 2016 to collect more information on students' interactions with the draft PSIs and eTIMSS interface in a standard testing situation. The prePilot instruments included a total of 12 PSI tasks across both subjects and grades and incorporated a broader variety of interactive features and item types than the first prototypes. The instruments also were designed to be administered on both computers and tablets to accommodate a wider range of devices and support more countries' participation in eTIMSS.

The eTIMSS prePilot was conducted in three English-speaking countries with experience in conducting digital assessments: Australia, Canada, and Singapore. Each country selected two to four classes at each grade to participate and made efforts to include students with a range of mathematics and science ability. This sample yielded approximately 100 responses per item at both the fourth and the eighth grade.

Students' responses to the draft PSIs and participating countries' reports on their experiences carrying out the study prompted additional changes to both the PSIs and their software before the field test.

The TIMSS 2019 Field Test

In preparation for data collection, TIMSS routinely conducts a full-scale field test for the purposes of evaluating the measurement properties of the item pool and practicing the data collection and scoring procedures. For TIMSS 2019, the field test was a particularly critical "dress rehearsal" because it was the first large-scale administration of eTIMSS on computers and tablets. In addition to providing important information about how well each prospective item and PSI functioned, the field test results prompted a number of improvements to the components in the eAssessment System as well as to the directions for test administrators and students.

All eTIMSS and paperTIMSS materials and operational procedures were field tested with samples of students selected according to rigorous sampling procedures. The field test in each country was designed to be conducted in approximately 30 schools and yield at least 200 student responses to each mathematics and science item. The school samples for the TIMSS 2019 Field Test and Data Collection were drawn simultaneously, using the same random sampling procedures. This ensured that the field test samples closely approximated the data collection samples, and that a school was selected for either the field test or data collection, but not both. For example, if a country needed 150 schools for data collection and another 30 for the field test, then a larger sample of 180 schools was selected and a systematic sample of 30 schools was selected for the field test from the 180 schools. See Chapter 3 for details about the school and classroom sampling techniques used in TIMSS 2019.

Preparing for the eTIMSS 2019 Field Test was quite complicated and involved several additional steps beyond those included in paperTIMSS. After translating and adapting the international instruments in IEA's online translation system, countries checked the functionality of their national instruments, loaded the eTIMSS Player software onto each computer or tablet to be used in the field test, and checked the compatibility of the software with the devices. Following each testing session, test administrators uploaded students' responses to IEA's servers.

Exhibit 1.4 shows the total number of items in each fourth and eighth grade field test, as well as the number of students, teachers, and schools that participated. Exhibits 1.5 through 1.8 provide a detailed summary of the number of field test items in the eTIMSS and paperTIMSS field tests by format, content domain, and cognitive domain.

Exhibit 1.4: Overview of the TIMSS 2019 Field Test

	Fourth Grade			Eighth Grade	
	eTIMSS	paperTIMSS	Less Difficult Mathematics	eTIMSS	paperTIMSS
Items in Field Test					
Mathematics	174	127	130	201	158
Science	164	134	134	212	167
Total	**338**	**261**	**264**	**413**	**325**
Responses per Item per Country (approx.)	200	200	200	200	200
Participants					
Countries	31	18	7	22	14
Benchmarking Entities	6	–	–	5	–
Students	50,158	19,656	8,128	37,512	16,225
Teachers	3,337	1,176	471	5,009	1,826
Schools	1,340	526	203	852	342

Counts for eTIMSS include the items from the PSI tasks.
Five item blocks (64 items) were common to both the regular and less difficult fourth grade mathematics assessment.

Exhibit 1.5: TIMSS 2019 Number of Field Test Items by Content Domain and Item Format – Fourth Grade

Content Domain	Number of Selected Response Items	Number of Constructed Response Items	Total Number of Items	Percentage of Total Items
Mathematics – eTIMSS and paperTIMSS				
Number	27	25	52	40%
Measurement and Geometry	22	20	42	32%
Data	15	22	37	28%
Total	**64**	**67**	**131**	
Mathematics – Less Difficult				
Number	31	24	55	42%
Measurement and Geometry	21	17	38	29%
Data	17	20	37	28%
Total	**69**	**61**	**130**	
Science – eTIMSS and paperTIMSS				
Life Science	40	22	62	46%
Physical Science	28	13	41	31%
Earth Science	18	13	31	23%
Total	**86**	**48**	**134**	

Four mathematics items were only field tested in eTIMSS and four items were only field tested in paperTIMSS. Counts include all eight of these items.
Five item blocks (64 items) were common to both the regular and less difficult fourth grade mathematics assessments.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.6: TIMSS 2019 Number of Field Test Items by Cognitive Domain and Item Format – Fourth Grade

Cognitive Domain	Number of Selected Response Items	Number of Constructed Response Items	Total Number of Items	Percentage of Total Items
Mathematics – eTIMSS and paperTIMSS				
Knowing	29	14	43	33%
Applying	27	38	65	50%
Reasoning	8	15	23	18%
Total	**64**	**67**	**131**	
Mathematics – Less Difficult				
Knowing	36	11	47	36%
Applying	25	30	55	42%
Reasoning	8	20	28	22%
Total	**69**	**61**	**130**	
Science – eTIMSS and paperTIMSS				
Knowing	42	16	58	43%
Applying	28	17	45	34%
Reasoning	16	15	31	23%
Total	**86**	**48**	**134**	

Four mathematics items were only field tested in eTIMSS and four items were only field tested in paperTIMSS. Counts include all eight of these items.
Five item blocks (64 items) were common to both the regular and less difficult fourth grade mathematics assessments.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.7: TIMSS 2019 Number of Field Test Items by Content Domain and Item Format – Eighth Grade

Content Domain	Number of Selected Response Items	Number of Constructed Response Items	Total Number of Items	Percentage of Total Items
Mathematics – eTIMSS and paperTIMSS				
Number	18	28	46	29%
Algebra	25	28	53	34%
Geometry	9	22	31	20%
Data and Probability	14	14	28	18%
Total	**66**	**92**	**158**	
Science – eTIMSS and paperTIMSS				
Biology	42	19	61	36%
Chemistry	22	16	38	23%
Physics	24	13	37	22%
Earth Science	24	7	31	19%
Total	**112**	**55**	**167**	

Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.8: TIMSS 2019 Number of Field Test Items by Cognitive Domain and Item Format – Eighth Grade

Cognitive Domain	Number of Selected Response Items	Number of Constructed Response Items	Total Number of Items	Percentage of Total Items
Mathematics – eTIMSS and paperTIMSS				
Knowing	28	19	47	30%
Applying	32	46	78	49%
Reasoning	6	27	33	21%
Total	**66**	**92**	**158**	
Science – eTIMSS and paperTIMSS				
Knowing	46	12	58	35%
Applying	39	23	62	37%
Reasoning	27	20	47	28%
Total	**112**	**55**	**167**	

Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

The eTIMSS 2019 Field Test also included eight mathematics and science PSI tasks at the fourth grade, comprising 72 items, and seven mathematics and science PSI tasks at the eighth grade, comprising 83 items. Because the PSIs were designed with the distinct goals of increasing coverage of traditionally difficult to measure areas of the mathematics and science frameworks in the applying and reasoning cognitive domains by capitalizing on technology, choices about the content topics to assess with each task were largely guided by the problem contexts and potential uses of technology to enhance measurement. Following the field test, two-thirds of the PSI tasks were selected for data collection (see Exhibit 1.18 for a description of the selected tasks).

Developing the Materials for TIMSS 2019 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 item to receive credit. It also is critical that students' responses are evaluated 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.

In addition to developing a unique scoring guide for each constructed response item, the TIMSS & PIRLS International Study Center provided training for the NRCs and their scoring supervisors to ensure that the scoring guides for all human-scored constructed response items were applied consistently within and across countries. The TIMSS 2019 training materials consisted of sets of student responses for a selected group of items with the most complicated scoring guides. For each item, the training set consisted 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 for field test scoring to begin immediately upon completion of data collection, it was necessary to prepare scoring training materials for the newly developed constructed response items in advance of the field test. To provide "grist" for these materials, Australia, England, and Ireland pilot tested a selection of the newly developed constructed response field test items in several classrooms with English-speaking students in January 2018. Because students may express their answers in different ways when typing versus writing by hand, both typed and handwritten responses were collected for the all items in both the eTIMSS and paperTIMSS assessments.

Exhibit 1.9 provides the number of items included in the pilot test and the number of student responses collected. Only a small number of mathematics items required scoring training, so the majority of the items in the pilot were in science.

Exhibit 1.9: Pilot Test Student Responses for Field Test Scoring Training Materials Development

	Number of Items	Approximate Number of Responses	
		eTIMSS	paperTIMSS
Fourth Grade			
Mathematics	5	93	96
Science	21	93	96
Countries		England	Australia and Ireland
Eighth Grade			
Mathematics	6	80	43
Science	19	80	43
Countries		England	Ireland

Consultants and staff at the TIMSS & PIRLS International Study Center met in January 2018 to review responses collected in the pilot test and create the training materials. For the TIMSS 2019 Field Test, training sets of example and practice responses were created for a total of 23 fourth grade items and 30 eighth grade items. These sets included both typed and handwritten responses to prepare scorers to score student responses in both modes of administration.

The TIMSS 2019 NRCs and their scoring supervisors received scoring training for the field test in March 2018 in Madrid, Spain, as part of the 4th TIMSS 2019 NRC meeting. At the training sessions, the trainers explained the purpose of each item and read it aloud. The trainer then described the scoring guide, explaining each category and the rationale for the score given to each example paper. The country representatives were then given time to score the practice papers so they could apply the scoring guides and learn how to make distinctions among categories. The correct codes for each practice paper were then reviewed, any inconsistencies in scoring were discussed, and, as necessary, the scoring guides were clarified and sometimes categories were revised.

Finalizing the TIMSS 2019 Achievement Instruments

Subsequent to the field test, the TIMSS & PIRLS International Study Center analyzed the TIMSS field test data and selected the new items to be combined with the trend items for data collection. When selecting the items, both the measurement properties (item statistics) of the individual items and the overall content and cognitive domain coverage of the group of items were considered to ensure that the final achievement instruments met the assessment specifications in the frameworks.

To review the measurement properties of the field test items, staff at the TIMSS & PIRLS International Study Center prepared almanacs containing summary item statistics for each field test item. The achievement data almanacs displayed for each item, row by row for each country: the number

of students to whom the item was administered, the item difficulty and discrimination, the percentage of students answering each option (selected response) or in each score category (constructed response), 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. The field test data were used by the TIMSS & PIRLS International Study Center, expert committees, and NRCs to assess the quality of the field test items.

First, staff at the TIMSS & PIRLS International Study Center reviewed the field test data to make an initial judgment about the quality of each item based on its measurement properties. Items were eliminated from further consideration if they had poor measurement properties, such as being too difficult or too easy or having low discrimination. Particular attention was paid to unusual item statistics in individual countries because these could indicate errors in translation.

After the item-by-item review, staff at the TIMSS & PIRLS International Study Center collaborated with a subset of SMIRC members to choose a set of recommended achievement items. The group reviewed the viable field test items for each content domain topic in relation to the trend items to select a coherent group of items for each topic, then verified that the set of items were appropriately distributed across the cognitive domains and item formats. SMIRC scrutinized the recommendations for the newly developed achievement items at the 3rd TIMSS 2019 SMIRC meeting, reviewing the items and scoring guides for content accuracy, clarity, and adherence to the frameworks.

To allow for any major revisions to the PSIs to be completed in time for data collection, the NRCs were asked to provide feedback on the PSIs when they submitted their field test data. Staff at the TIMSS & PIRLS International Study Center reviewed all NRC comments in conjunction with the data, selected the PSIs for the eTIMSS 2019 assessments based on the NRCs' recommendations, and began editing the selected tasks in June 2018. SMIRC also reviewed the PSIs at their 3rd meeting.

Next, staff at the TIMSS & PIRLS International Study Center implemented SMIRC's recommendations and assembled the items into assessment blocks for the NRCs' penultimate review. The NRCs had the opportunity to review the recommended assessment blocks in light of the field test results and within the security of their own countries. Each country also could check any unusual national results that might be indicative of translation errors and correct the translation as necessary or recommend revisions to better accommodate translation. Finally, the 5th TIMSS 2019 NRC meeting held in Stockholm, Sweden, in August 2018 was devoted to reviewing all the newly developed items.

Following the final review, the newly developed item blocks and existing trend item blocks were arranged into digital block combinations for eTIMSS and booklets for paperTIMSS according to the TIMSS 2019 Assessment Design (Martin, Mullis & Foy, 2017). For eTIMSS, the trend item blocks were converted from paper to digital format to be administered via the eAssessment System along with the new item blocks. The results of the TIMSS 2019 Item Equivalence Study (Fishbein, Martin, Mullis, &

Foy, 2018), a pilot conducted in 25 eTIMSS countries to investigate potential differences in student achievement on the trend items between the paper and digital modes of administration, provided evidence that the mathematics and science constructs assessed by the trend items were mostly unaffected in the transition to eTIMSS at both grades. Still, to ensure that the eTIMSS and paperTIMSS results could be reported on the same achievement scale, eTIMSS 2019 countries that had participated in TIMSS 2015 also re-administered the trend items in paper booklets to a nationally representative sample of students during data collection to provide a "bridge" between paperTIMSS and eTIMSS (see Chapter 12 for additional details).

Distribution of the TIMSS 2019 Achievement Items

It is critical to document the coherence between the assessment frameworks and achievement instruments to ensure that an assessment measures what it is intended to measure and provide evidence for the validity of the assessment results. Because the TIMSS assessments encompass two domains (content and cognitive) and include both trend and newly developed items in a variety of formats, it is necessary to demonstrate the alignment between the items and assessment specifications from multiple perspectives.

Achievement Items by Content and Cognitive Domain

The TIMSS 2019 assessments consisted of approximately 40 percent new items and 60 percent trend items, which were used to continue trend measurement from the previous assessment cycles. Therefore, it is important to confirm that the distribution of both the trend and new items across the content and cognitive domains reflects the specifications described in the assessment frameworks. The distribution of the trend items typically varies from the target specifications because the assessment frameworks are updated with each cycle and items are "retired" from the assessment, so the new items are selected to ensure the final assessments are aligned with the frameworks.

Exhibits 1.10 and 1.11 present the number of trend and newly developed items as well as the number of score points in the TIMSS 2019 fourth grade assessments by content domain and cognitive domain, respectively. The number of items represents the number of distinct questions in the assessment, while the number of score points represents the complexity and weight given to each item. Exhibits 1.12 and 1.13 present the TIMSS 2019 eighth grade assessments by content and cognitive domain.

Exhibit 1.10: TIMSS 2019 Achievement Items by Content Domain – Fourth Grade

Content Domain	Trend		New		Total		Target Percentage of Score Points
	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	
Mathematics – eTIMSS and paperTIMSS							
Number	55 (59)	61%	29 (30)	32%	84 (89)	47%	50%
Measurement and Geometry	26 (27)	28%	27 (31)	33%	53 (58)	31%	30%
Data	11 (11)	11%	27 (32)	34%	38 (43)	23%	20%
Total	**92 (97)**		**83 (93)**		**175 (190)**		
Mathematics – Less Difficult							
Number	67 (68)	59%	29 (32)	42%	96 (100)	52%	50%
Measurement and Geometry	31 (34)	29%	20 (21)	28%	51 (55)	29%	30%
Data	13 (14)	12%	19 (23)	30%	32 (37)	19%	20%
Total	**111 (116)**		**68 (76)**		**179 (192)**		
Science – eTIMSS and paperTIMSS							
Life Science	44 (47)	46%	34 (36)	46%	78 (83)	46%	45%
Physical Science	36 (37)	36%	26 (26)	33%	62 (63)	35%	35%
Earth Science	18 (18)	18%	17 (17)	22%	35 (35)	19%	20%
Total	**98 (102)**		**77 (79)**		**175 (181)**		

Score points are shown in parentheses.
Two mathematics items involving an on-screen ruler tool were only included in eTIMSS assessment.
Four item blocks (48 items) were common to both the regular and less difficult fourth grade mathematics assessments.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.11: TIMSS 2019 Achievement Items by Cognitive Domain – Fourth Grade

Cognitive Domain	Trend		New		Total		Target Percentage of Score Points
	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	
Mathematics – eTIMSS and paperTIMSS							
Knowing	34 (34)	35%	29 (29)	31%	63 (63)	33%	40%
Applying	40 (42)	43%	34 (39)	42%	74 (81)	43%	40%
Reasoning	18 (21)	22%	20 (25)	27%	38 (46)	24%	20%
Total	**92 (97)**		**83 (93)**		**175 (190)**		
Mathematics – Less Difficult							
Knowing	56 (56)	48%	25 (26)	34%	81 (82)	43%	40%
Applying	39 (40)	34%	27 (32)	42%	66 (72)	38%	40%
Reasoning	16 (20)	17%	16 (18)	24%	32 (38)	20%	20%
Total	**111 (116)**		**68 (76)**		**179 (192)**		
Science – eTIMSS and paperTIMSS							
Knowing	42 (45)	44%	31 (32)	41%	73 (77)	43%	40%
Applying	35 (36)	35%	30 (30)	38%	65 (66)	36%	40%
Reasoning	21 (21)	21%	16 (17)	22%	37 (38)	21%	20%
Total	**98 (102)**		**77 (79)**		**175 (181)**		

Score points are shown in parentheses.
Two mathematics items involving an on-screen ruler tool were only included in eTIMSS assessment.
Four item blocks (48 items) were common to both the regular and less difficult fourth grade mathematics assessments.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.12: TIMSS 2019 Achievement Items by Content Domain – Eighth Grade

Content Domain	Trend		New		Total		Target Percentage of Score Points
	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	
Mathematics – eTIMSS and paperTIMSS							
Number	36 (37)	30%	28 (30)	30%	64 (67)	30%	30%
Algebra	31 (32)	26%	31 (32)	32%	62 (64)	29%	30%
Geometry	25 (28)	22%	18 (20)	20%	43 (48)	21%	20%
Data and Probability	25 (28)	22%	17 (17)	17%	42 (45)	20%	20%
Total	**117 (125)**		**94 (99)**		**211 (224)**		
Science – eTIMSS and paperTIMSS							
Biology	39 (48)	32%	38 (43)	39%	77 (91)	35%	35%
Chemistry	22 (23)	18%	22 (25)	22%	44 (48)	20%	20%
Physics	33 (33)	27%	22 (25)	22%	55 (58)	25%	25%
Earth Science	28 (29)	23%	16 (17)	16%	44 (46)	20%	20%
Total	**122 (133)**		**98 (110)**		**220 (243)**		

Score points are shown in parentheses.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.13: TIMSS 2019 Achievement Items by Cognitive Domain – Eighth Grade

Cognitive Domain	Trend		New		Total		Target Percentage of Score Points
	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	
Mathematics – eTIMSS and paperTIMSS							
Knowing	35 (35)	28%	30 (32)	32%	65 (67)	30%	35%
Applying	58 (61)	49%	39 (40)	40%	97 (101)	45%	40%
Reasoning	24 (29)	23%	25 (27)	27%	49 (56)	25%	25%
Total	**117 (125)**		**94 (99)**		**211 (224)**		
Science – eTIMSS and paperTIMSS							
Knowing	45 (50)	37%	35 (36)	36%	80 (86)	36%	35%
Applying	46 (50)	38%	36 (44)	37%	82 (94)	37%	35%
Reasoning	31 (33)	25%	27 (30)	28%	58 (63)	26%	30%
Total	**122 (133)**		**98 (110)**		**220 (243)**		

Score points are shown in parentheses.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Achievement Items by Item Formats within Content and Cognitive Domains

To assess the broad range of mathematics and science topics and skills described in the assessment frameworks, the TIMSS 2019 fourth and eighth grade assessments included a wide variety of selected response and constructed response items. Both the digital and paper versions of the TIMSS 2019 assessments included two general types of selected response items—single selection, in which students choose one of four response options, and multiple selection, in which students chose more than one option from a number of response options or made a series of selections to respond to a question. In eTIMSS, the answer options for some selected response items were presented in drop-down menus or as clickable pictures or words. Most TIMSS 2019 selected response items were worth one score point, although some multiple selection items were worth two score points. The 2-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).

Constructed response items, which involve writing or typing words or numbers, drawing, or dragging and dropping for eTIMSS, 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 demonstrate only a partial understanding of the concepts or procedures embodied in the task.

To ensure sufficient coverage of the assessment frameworks, it is important to verify that an assortment of selected and constructed response items are used to assess each domain. Exhibits 1.14 through 1.17 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 1.14: TIMSS 2019 Achievement Items by Content Domain and Item Format – Fourth Grade

Content Domain	Selected Response Items		Constructed Response Items		Total Items	Percentage of Score Points
	Single Selection	Multiple Selection	1 Point	2 Points		
Mathematics – eTIMSS and paperTIMSS						
Number	39 (39)	7 (7)	33 (33)	5 (10)	84 (89)	47%
Measurement and Geometry	25 (25)	6 (6)	17 (17)	5 (10)	53 (58)	31%
Data	8 (8)	3 (3)	22 (22)	5 (10)	38 (43)	23%
Total	**72 (72)**	**16 (16)**	**72 (72)**	**15 (30)**	**175 (190)**	
Achieved Percentage of Score Points	**46%**		**54%**			
Mathematics – Less Difficult						
Number	46 (46)	1 (1)	45 (45)	4 (8)	96 (100)	52%
Measurement and Geometry	26 (26)	2 (2)	19 (19)	4 (8)	51 (55)	29%
Data	10 (10)	2 (3)	16 (16)	4 (8)	32 (37)	19%
Total	**82 (82)**	**5 (6)**	**80 (80)**	**12 (24)**	**179 (192)**	
Achieved Percentage of Score Points	**46%**		**54%**			
Science – eTIMSS and paperTIMSS						
Life Science	35 (35)	6 (7)	33 (33)	4 (8)	78 (83)	46%
Physical Science	35 (35)	5 (5)	21 (21)	1 (2)	62 (63)	35%
Earth Science	24 (24)	4 (4)	7 (7)	– –	35 (35)	19%
Total	**94 (94)**	**15 (16)**	**61 (61)**	**5 (10)**	**175 (181)**	
Achieved Percentage of Score Points	**61%**		**39%**			

Score points are shown in parentheses.
Two fourth grade mathematics items involving an on-screen ruler tool were only included in eTIMSS assessment.
Four item blocks (48 items) were common to both the regular and less difficult fourth grade mathematics assessments.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.15: TIMSS 2019 Achievement Items by Cognitive Domain and Item Format – Fourth Grade

Cognitive Domain	Selected Response Items		Constructed Response Items		Total Items	Percentage of Score Points
	Single Selection	Multiple Selection	1 Point	2 Points		
Mathematics – eTIMSS and paperTIMSS						
Knowing	33 (33)	12 (12)	18 (18)	– –	63 (63)	33%
Applying	25 (25)	2 (2)	40 (40)	7 (14)	74 (81)	43%
Reasoning	14 (14)	2 (2)	14 (14)	8 (16)	38 (46)	24%
Total	**72 (72)**	**16 (16)**	**72 (72)**	**15 (30)**	**175 (190)**	
Achieved Percentage of Score Points	**46%**		**54%**			
Mathematics – Less Difficult						
Knowing	46 (46)	2 (2)	32 (32)	1 (2)	81 (82)	43%
Applying	25 (25)	2 (3)	34 (34)	5 (10)	66 (72)	38%
Reasoning	11 (11)	1 (1)	14 (14)	6 (12)	32 (38)	20%
Total	**82 (82)**	**5 (6)**	**80 (80)**	**12 (24)**	**179 (192)**	
Achieved Percentage of Score Points	**46%**		**54%**			
Science – eTIMSS and paperTIMSS						
Knowing	42 (42)	10 (11)	18 (18)	3 (6)	73 (77)	43%
Applying	33 (33)	3 (3)	28 (28)	1 (2)	65 (66)	36%
Reasoning	19 (19)	2 (2)	15 (15)	1 (2)	37 (38)	21%
Total	**94 (94)**	**15 (16)**	**61 (61)**	**5 (10)**	**175 (181)**	
Achieved Percentage of Score Points	**61%**		**39%**			

Score points are shown in parentheses.
Two fourth grade mathematics items involving an on-screen ruler tool were only included in eTIMSS assessment.
Four item blocks (48 items) were common to both the regular and less difficult fourth grade mathematics assessments.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.16: TIMSS 2019 Achievement Items by Content Domain and Item Format – Eighth Grade

Content Domain	Selected Response Items		Constructed Response Items		Total Items	Percentage of Score Points
	Single Selection	Multiple Selection	1 Point	2 Points		
Mathematics – eTIMSS and paperTIMSS						
Number	27 (27)	4 (6)	32 (32)	1 (2)	64 (67)	30%
Algebra	32 (32)	1 (1)	27 (27)	2 (4)	62 (64)	29%
Geometry	15 (15)	2 (2)	21 (21)	5 (10)	43 (48)	21%
Data and Probability	18 (18)	5 (7)	18 (18)	1 (2)	42 (45)	20%
Total	**92 (92)**	**12 (16)**	**98 (98)**	**9 (18)**	**211 (224)**	
Achieved Percentage of Score Points	**48%**		**52%**			
Science – eTIMSS and paperTIMSS						
Biology	37 (37)	9 (12)	20 (20)	11 (22)	77 (91)	37%
Chemistry	19 (19)	4 (5)	18 (18)	3 (6)	44 (48)	20%
Physics	29 (29)	7 (7)	16 (16)	3 (6)	55 (58)	24%
Earth Science	30 (30)	4 (6)	10 (10)	– –	44 (46)	19%
Total	**115 (115)**	**24 (30)**	**64 (64)**	**17 (34)**	**220 (243)**	
Achieved Percentage of Score Points	**60%**		**40%**			

Score points are shown in parentheses.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.17: TIMSS 2019 Achievement Items by Cognitive Domain and Item Format – Eighth Grade

Cognitive Domain	Selected Response Items		Constructed Response Items		Total Items	Percentage of Score Points
	Single Selection	Multiple Selection	1 Point	2 Points		
Mathematics – eTIMSS and paperTIMSS						
Knowing	41 (41)	5 (7)	19 (19)	– –	65 (67)	30%
Applying	40 (40)	4 (4)	49 (49)	4 (8)	97 (101)	45%
Reasoning	11 (11)	3 (5)	30 (30)	5 (10)	49 (56)	25%
Total	**92 (92)**	**12 (16)**	**98 (98)**	**9 (18)**	**211 (224)**	
Achieved Percentage of Score Points	**48%**		**52%**			
Science – eTIMSS and paperTIMSS						
Knowing	56 (56)	11 (14)	10 (10)	3 (6)	80 (86)	35%
Applying	38 (38)	8 (10)	26 (26)	10 (20)	82 (94)	39%
Reasoning	21 (21)	5 (6)	28 (28)	4 (8)	58 (63)	26%
Total	**115 (115)**	**24 (30)**	**64 (64)**	**17 (34)**	**220 (243)**	
Achieved Percentage of Score Points	**60%**		**40%**			

Score points are shown in parentheses.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

eTIMSS Problem Solving and Inquiry Tasks (PSIs) by Content and Cognitive Domain

Exhibit 1.18 provides a brief description of the eTIMSS 2019 PSI problem scenarios and the total number of items (and score points) in each task. The tasks covered a range of mathematics and science content domain topics and, consistent with the goal of the PSIs to assess higher-order skills, the majority of the items in the PSIs involved applying and reasoning.

Exhibit 1.18: eTIMSS 2019 Mathematics and Science Problem Solving and Inquiry Tasks (PSIs)

Fourth Grade PSIs	Total Items
Mathematics	
School Party – Students plan a party for a school by determining the price for tickets and the amount of food, drinks, and decorations to purchase for the party	**11 (14)**
Robots – Students use a robot that can follow input-output rules to solve mathematics problems and determine the robot's rules	**6 (7)**
Little Penguins – Students add information to a website about Little Penguins by solving a series of mathematics problems involving facts about penguins	**12 (14)**
Science	
Farm Investigation – Students carry out a virtual investigation to identify the farm animal responsible for eating garden plants	**10 (16)**
Sugar Experiment – Students design and carry out a virtual experiment to test which of three types of sugar dissolves fastest in water	**9 (13)**
Eighth Grade PSIs	**Total Items**
Mathematics	
Dinosaur Speed – Students use the relationships between foot length, leg height, and stride length to estimate how fast a dinosaur could run	**12 (13)**
Building – Students determine the dimensions of a shed to store equipment, including a barrel to collect rainwater	**9 (11)**
Robots – Students determine functions using a robot that applies a function to determine y for any given value of x	**4 (4)**
Science	
Sunken Ship – Students carry out a virtual investigation into the circumstances that resulted in the sinking of a ship	**16 (17)**
Pepper Plants – Students design and carry out a virtual experiment to test the effects of two fertilizers on the growth and development of pepper plants	**13 (18)**

Score points are shown in parentheses.

The addition of the PSIs for eTIMSS resulted in a slight increase in coverage of the applying and reasoning cognitive domains at both the fourth and the eighth grade. However, comprising only a small part of the whole assessment (approximately 12 percent), the PSIs did not substantially alter the framework coverage provided by the eTIMSS assessments. The pie charts in Exhibits 1.19 and 1.20 show the percentage of assessment score points in each content and cognitive domain in the eTIMSS 2019 assessments, both with and without the PSIs included, compared to the target percentage of testing time allocated to each domain.

Exhibit 1.19: Comparison of Target and Achieved Percentages of Domain Coverage in the eTIMSS 2019 Mathematics and Science Assessments – Fourth Grade

- Target percentage of testing time specified in the framework
- Achieved percentage of score points from regular items
- Achieved percentage of score points from regular and PSI items

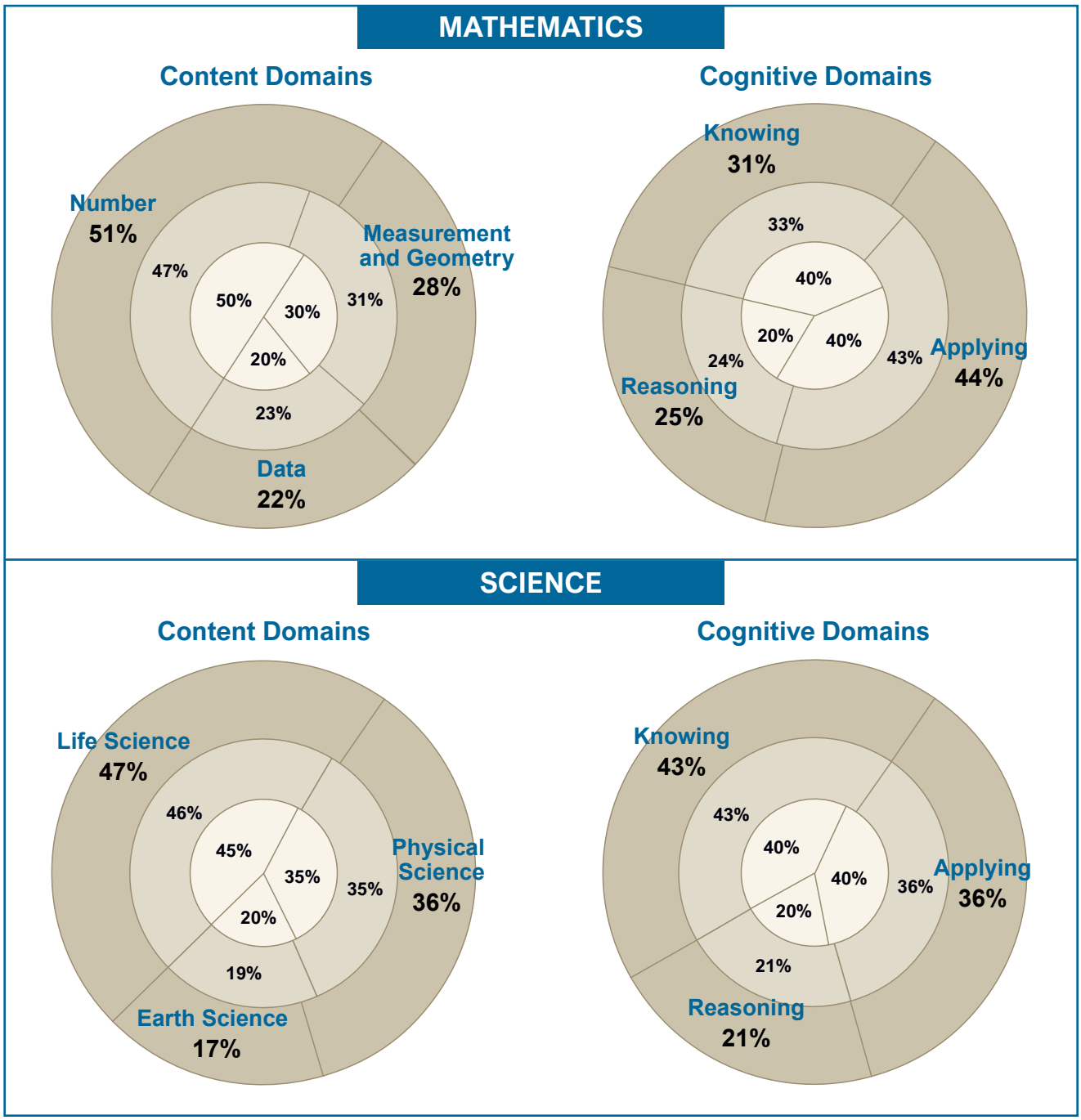
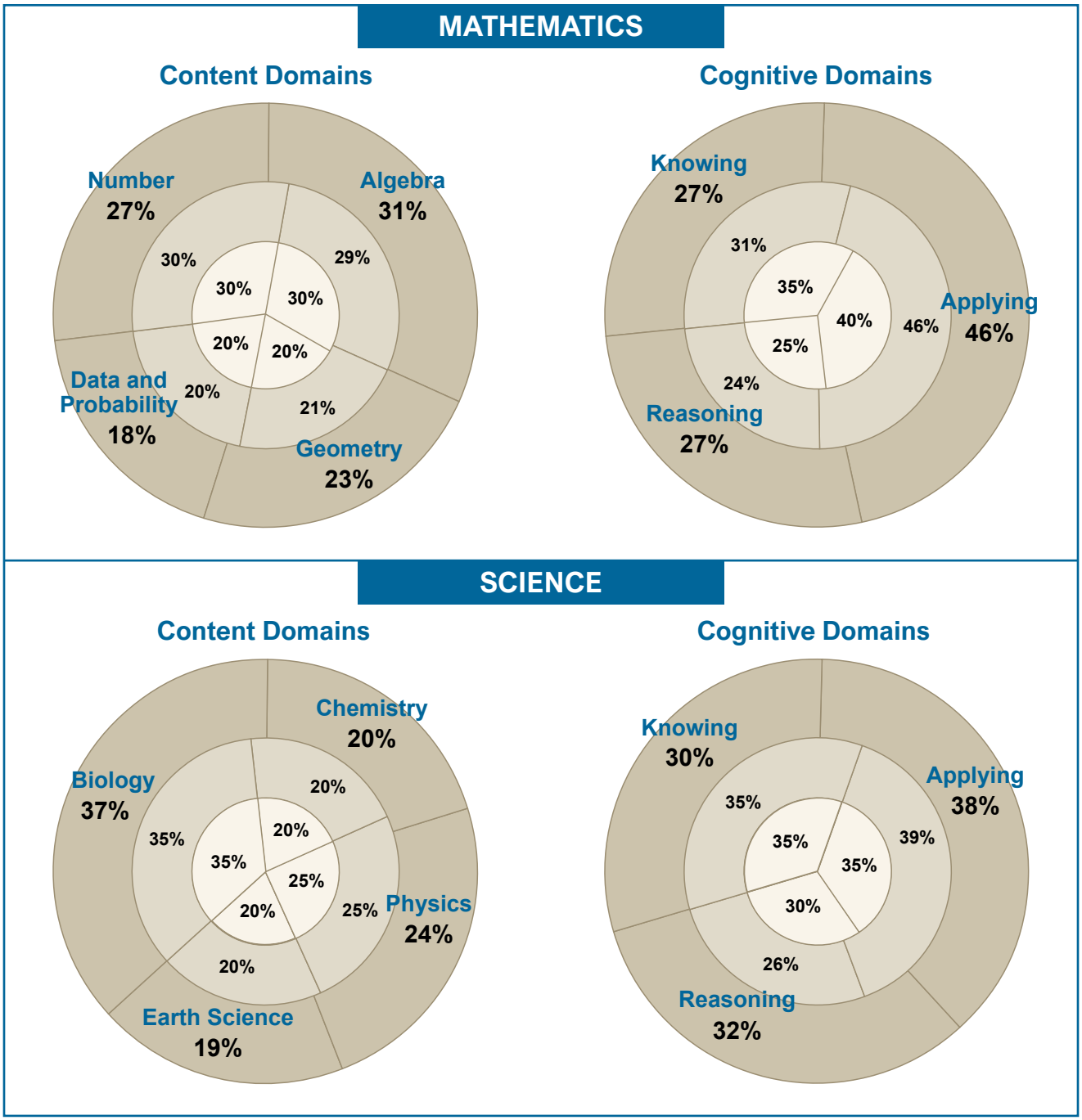


Exhibit 1.20: Comparison of Target and Achieved Percentages of Domain Coverage in the eTIMSS 2019 Mathematics and Science Assessments – Eighth Grade

- Target percentage of testing time specified in the framework
- Achieved percentage of score points from regular items
- Achieved percentage of score points from regular and PSI items



TIMSS 2019 Constructed Response Scoring Training

In preparation for the main data collection scoring training, some TIMSS 2019 scoring guides were further refined or clarified based on the results of the field test. This also included a thorough review of the field test scoring training materials to ensure that the student responses were still suitable for the updated scoring guides. In some cases, example and practice sets used in the field test were expanded to further illustrate particular aspects of a scoring guide.

The TIMSS 2019 scoring training materials also included the training sets for the trend items used in TIMSS 2015. These training materials were updated for TIMSS 2019 to include both typed and handwritten responses. In all, the TIMSS 2019 scoring training materials included sets of example and practice responses for a total of 26 fourth grade items and 27 eighth grade items.

To provide scoring training for all the countries participating in TIMSS 2019, the TIMSS & PIRLS International Study Center conducted two training sessions. First, the NRCs for Southern Hemisphere countries and their scoring supervisors received scoring training in November 2018 in Cape Town, South Africa. The NRCs for Northern Hemisphere countries and their scoring supervisors received scoring training in March 2019 in Limassol, Cyprus as part of the 6th TIMSS 2019 NRC meeting. Exhibit 1.21 shows the number of participants in the two scoring training sessions.

Exhibit 1.21: TIMSS 2019 Scoring Training Participation

Participants	Southern Hemisphere	Northern Hemisphere
Number of Countries	7	52
Number of Benchmarking Entities	–	5
Number of Country Representatives	24	150

After participating in scoring training, the NRCs and their scoring supervisors organized and carried out scoring activities in their respective countries. In addition to scoring the student responses, all countries participated in several supplementary scoring activities to document the scoring reliability of the human-scored items. The procedures used to establish scoring reliability within each country, across countries, and across assessment cycles are described in Survey Operations Procedures for TIMSS 2019.

The Process Following Instrument Development

After the participating countries received the international version of the achievement instruments, they began the process of translation and cultural adaptation (some adaptation to local usage typically is necessary even in English-speaking countries) and production of the materials needed to administer the assessment. The tasks involved in producing the materials differed depending on whether eTIMSS or paperTIMSS was being administered. At the same time, countries made final arrangements for data collection, including the host of activities necessary to obtain school participation and implement test administration.

References

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.

Lindquist, M., Philpot, R., Mullis, I. V. S., & Cotter, K. E. (2017). TIMSS 2019 mathematics framework. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 assessment frameworks* (pp. 11-25). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Martin, M. O., Mullis, I. V. S., & Foy, P. (2017). TIMSS 2019 assessment design. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 assessment frameworks* (pp. 79-91). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

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. (Eds.). (2017). *TIMSS 2019 assessment frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/