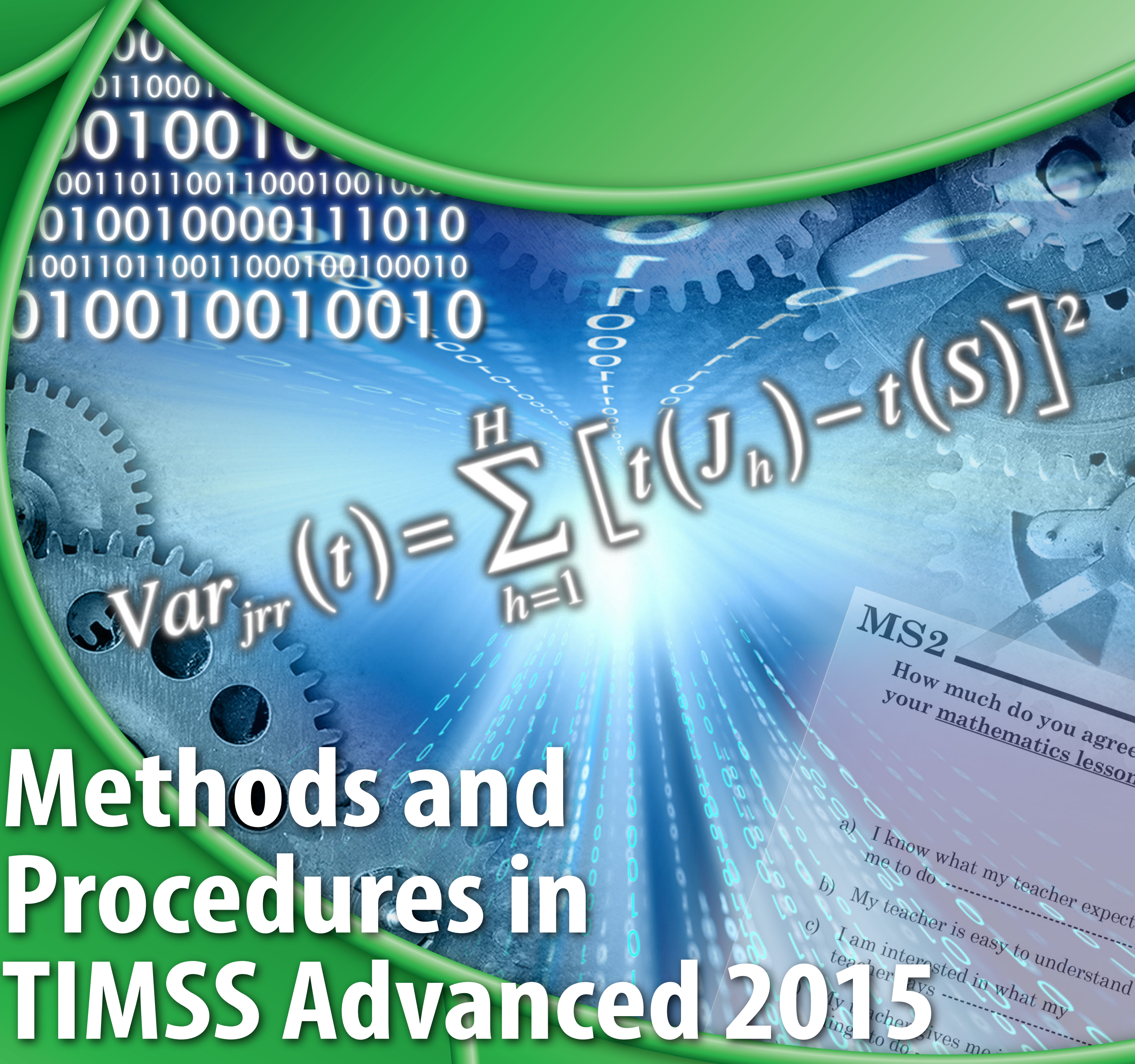
TRENDS IN INTERNATIONAL MATHEMATICS AND SCIENCE STUDY

TIMSS



Methods and Procedures in TIMSS Advanced 2015

Editors: Michael O. Martin,
Ina V.S. Mullis, and
Martin Hooper

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

TIMSS
Advanced
2015

TRENDS IN INTERNATIONAL MATHEMATICS AND SCIENCE STUDY

TIMSS

METHODS AND PROCEDURES IN TIMSS ADVANCED 2015

Edited by:
Michael O. Martin
Ina V.S. Mullis
Martin Hooper

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

METHODS AND PROCEDURES IN TIMSS Advanced 2015
Edited by: Michael O. Martin, Ina V.S. Mullis, and Martin Hooper

Publishers: TIMSS & PIRLS International Study Center,
Lynch School of Education, Boston College
and
International Association for the Evaluation of Educational Achievement (IEA)

Library of Congress Catalog Card Number: 2016919197
ISBN: 978-1-889938-34-9

For more information about TIMSS contact:
TIMSS & PIRLS International Study Center
Lynch School of Education
Boston College
Chestnut Hill, MA 02467
United States

tel: +1-617-552-1600
fax: +1-617-552-1203
e-mail: timss@bc.edu
timss.bc.edu

INSTRUMENT DEVELOPMENT

SAMPLING

DATA COLLECTION PROCEDURES

REPORTING

CHAPTER 1

Developing the TIMSS Advanced 2015 Achievement Items

Ina V.S. Mullis
Kerry E. Cotter
Bethany G. Fishbein
Victoria A.S. Centurino

Overview of TIMSS Advanced 2015

TIMSS Advanced is the only international assessment that provides essential information about achievement in advanced mathematics and physics for students in their final year of secondary school. First conducted in 1995 and again in 2008, TIMSS Advanced 2015 together with TIMSS 2015 will provide countries with a complete profile of mathematics and science learning from elementary through the end of secondary school.

The general approach to developing the TIMSS Advanced 2015 advanced mathematics and physics achievement items was similar to that used in 2008. However, in 2015 TIMSS Advanced and TIMSS were assessed together for the first time since 1995, providing 20 years of trend data for both assessments.

To provide more extensive coverage of the advanced mathematics and physics content and cognitive domains and provide an improved basis of comparison between the TIMSS and TIMSS Advanced assessments, the number of items in TIMSS Advanced 2015 was increased by approximately 30 percent from the 2008 assessments.

The TIMSS Advanced Approach to Measuring Trends

Because TIMSS Advanced is designed to measure trends, the assessments of advanced mathematics and physics cannot change dramatically from cycle to cycle. That is, TIMSS Advanced 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 tests also need to be updated with each cycle to prevent the assessments from becoming dated and no longer relevant to current learning goals. It is important for the content to "keep up with the times" and to be innovative. For example, TIMSS Advanced needs to reflect recent scientific discoveries and to be presented in situations consistent with students' instructional and everyday experiences.

To maintain continuity with past assessments while keeping up with current topics and technology, the TIMSS Advanced assessments evolve with each cycle. For assessing advanced mathematics and physics, TIMSS Advanced has a specific design for the steady release of items after each cycle and replacing them with newly developed items for the following cycle.

Overview of the TIMSS Advanced 2015 Achievement Items

Although the majority of the assessment items are carried forward from the previous assessment cycle to measure trends, the task of updating the instruments for each new cycle is a substantial undertaking. This was especially true, because the scope of the assessment was increased by almost one-third and TIMSS Advanced encompasses two different assessments of achievement, advanced mathematics and physics. The TIMSS Advanced 2015 advanced mathematics assessment required developing and field testing 132 new items and the physics assessment required developing and field testing 133 new items.

The Item Development Process

The TIMSS & PIRLS International Study Center at Boston College uses a collaborative process to develop the new items needed for each TIMSS Advanced cycle. A broad overview of the process includes:

- Updating the frameworks for the upcoming assessment
- Developing items and their scoring guides in accordance with the frameworks
- Conducting a full-scale field test
- Selecting the new assessment items based on the frameworks, field test results, and existing items from previous cycles
- Conducting training in how to reliably score responses to constructed response items (i.e., questions to which students provide a written response rather than choosing from a set of options)

The development process is directed and managed by the staff of the TIMSS & PIRLS International Study Center at Boston College, who collectively have considerable experience in the measurement and assessment of advanced mathematics and physics achievement. For TIMSS Advanced 2015, Executive Director, Ina Mullis, and Assistant Director of Mathematics, Kerry Cotter, managed the advanced mathematics assessment development. Executive Director, Michael Martin, and Associate Director of Science, Victoria Centurino, managed the physics assessment development. About half of the field test items were developed by the Australian Council for Educational Research (ACER) under the guidance of mathematics lead researcher, Ray Philpot and senior research fellow Ron Martin.

Also playing a key role in achievement item development were the National Research Coordinators (NRCs) designated by their countries to be responsible for the complex tasks involved in implementing TIMSS Advanced in their countries. The TIMSS & PIRLS International Study Center worked with the NRCs and experts from the countries to develop the new test items including the scoring guides for constructed response items. The NRCs also 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 prepares an international version of all the TIMSS Advanced assessment items in English. Subsequently, the items are translated by participating countries into their languages of instruction with the goal of creating high quality translations that are appropriately adapted for the national context and at the same time are internationally comparable. Therefore, a significant portion of the development and review effort by NRCs is dedicated to ensuring that the test items can be translated accurately.

To provide additional subject-matter expertise and support, and assist in coordinating between TIMSS and TIMSS Advanced, the same external mathematics and science specialists consulted very closely with staff on the development activities for both TIMSS and TIMSS Advanced 2015. The TIMSS/TIMSS Advanced 2015 Chief Mathematics Consultant was Liv Sissel Gronmo, University of Oslo, ILS, Norway, and the TIMSS/TIMSS Advanced 2015 Chief Science Consultant was Lee Jones, United States.

Additional advice and guidance were provided through periodic reviews by the Science and Mathematics Review Committee (SMIRC). The SMIRC members for each TIMSS cycle are nominated by countries participating in TIMSS and provide guidance in developing the TIMSS assessments. TIMSS 2015 and TIMSS Advanced 2015 had the same SMIRC, consisting of 16 members: 6 experts in mathematics and mathematics education and 10 experts in science and science education. Additional consultants, SMIRC members Mary Lindquist and Torgeir Onstad with Ray Philpot from ACER for advanced mathematics, and SMIRC member Gerald Wheeler with Ron Martin from ACER for physics, served as advisors to assist in completing specific tasks, such as drafting updated advanced mathematics and physics content frameworks and updating scoring guides after the field test.

SMIRC members met four times for TIMSS Advanced 2015. At the 1st SMIRC meeting in Oslo, Norway (April 2013), SMIRC reviewed the advanced mathematics and physics content frameworks and developed prototype field test items. At the 2nd meeting in St. Petersburg, Russia (September 2013), SMIRC reviewed draft field test items, together with their scoring guides. At the 3rd meeting in Sofia, Bulgaria (July 2014), SMIRC reviewed field test results and made recommendations to the NRCs regarding which items to include in the 2015 advanced mathematics and physics assessments. At the final meeting in Seoul, Korea (May 2016), SMIRC conducted the TIMSS Advanced 2015 scale anchoring process. Exhibit 1.1 lists the TIMSS/TIMSS Advanced 2015 SMIRC members.

Exhibit 1.1: TIMSS/TIMSS Advanced 2015 Science and Mathematics Item Review Committee (SMIRC)

Mathematics

Kiril Bankov
Faculty of Mathematics and Informatics
University of Sofia
Bulgaria

Sean Close
Educational Research Centre
St. Patrick's College
Ireland

Khattab Mohammad Ahmad Abulibdeh
National Center for Human Resources Development
Jordan

Sun Sook Noh
College for Education
Ewha Womans University
Korea

Torgeir Onstad
Department of Teacher Education and School
University of Olso, ILS
Norway

Mary Lindquist
United States

Science

Jouni Viiri
Department of Teacher Education
University of Jyväskylä
Finland

Alice Wong
Faculty of Education
University of Hong Kong
Hong Kong SAR

Berenice Michels
National Institute for Curriculum Development
The Netherlands

Newman Burdett
National Foundation for Educational Research
England

Galina Kovaleva
Institute of Content and Methods Education
Russian Academy of Education
Russian Federation

Vitaly Gribov
Physics Faculty
Moscow Lomonosov State University
Russian Federation

Gorazd Planinšič
Faculty of Mathematics and Physics
University of Ljubljana
Slovenia

Wolfgang Dietrich
National Agency for Education
Sweden

Christopher Lazzaro
The College Board
United States

Gerald Wheeler
National Science Teachers' Association
United States

Updating the Advanced Mathematics and Physics Assessment Frameworks for TIMSS Advanced 2015

Updating the TIMSS Advanced assessments for 2015 began with reviewing and modifying the assessment frameworks that specify the content to be assessed. The first two chapters of the *TIMSS Advanced 2015 Assessment Frameworks*, respectively, describe the advanced mathematics and physics frameworks in detail.

The basic structure of the TIMSS Advanced advanced mathematics and physics assessment frameworks is based on two dimensions: content and cognitive. The content domains for advanced mathematics are algebra, calculus, and geometry. For physics, the content domains are mechanics and thermodynamics, electricity and magnetism, and wave phenomena and atomic/nuclear physics.

The TIMSS Advanced advanced mathematics and physics frameworks specify several topic areas within each content domain. For example, the algebra content domain contains three topic areas: expressions and operations, equations and inequalities, and functions. The cognitive domains are the same for advanced mathematics and physics: knowing, applying, and reasoning. However, the descriptions of the cognitive skills to be assessed differ somewhat between advanced mathematics and physics.

For TIMSS Advanced 2015, the advanced mathematics and physics frameworks were updated to better reflect the curricula and standards of the countries participating in TIMSS Advanced using information from current research and initiatives in advanced mathematics and physics education. These updates were discussed by the NRCs from the participating countries at their first meeting. Following the discussion at the 1st NRC meeting, the NRCs consulted with their national experts and responded to a topic-by-topic survey about how best to update the content and cognitive domains for TIMSS Advanced 2015. Next, SMIRC reviewed and revised the frameworks. Using an iterative process, the frameworks as revised by SMIRC were once again reviewed by the TIMSS Advanced 2015 NRCs and updated a final time prior to publication.

Recommendations for updating content and cognitive domains can involve modifying content areas and their weightings (but no more than 5 percent); adding, deleting, or modifying topics within content areas to keep current with research findings and ensure that the number of topics reflects the content area weighting; rewriting to improve clarity for item writers; and perhaps combining some topic areas to reduce redundancy. New for 2015, a new section was added to the physics frameworks that describes the science practices to be addressed in physics assessments at the final year of secondary schooling or start of their STEM coursework in universities. Beyond that, there were no changes in the weighting of content areas for either advanced mathematics or physics and only minor revisions to content area topics. The TIMSS Advanced 2015 Development schedule is presented in Exhibit 1.2.

Exhibit 1.2: **TIMSS Advanced 2015 Development Schedule for Achievement Items**

Date(s)		Group and Activity
July – December	2012	TIMSS & PIRLS International Study Center conducted content analysis of the curricular topics described in the *TIMSS Advanced 2008 Assessment Frameworks* and *TIMSS Advanced 2008 International Report*
October	2012	Task Force proposed updates for the 2015 Assessment Frameworks, incorporating results from the content analysis (Boston, USA)
January	2013	TIMSS & PIRLS International Study Center compiled proposed updates to Assessment Frameworks in preparation for the 1st National Research Coordinator (NRC) meeting
February	2013	NRCs reviewed proposed updates to Assessment Frameworks at 1st NRC meeting (Hamburg, Germany)
March	2013	TIMSS & PIRLS International Study Center met with ACER representatives to discuss item development (Boston, MA)
March	2013	TIMSS & PIRLS International Study Center incorporated feedback from 1st NRC meeting to further refine the *TIMSS Advanced 2015 Assessment Frameworks* and surveyed NRCs online about proposed assessment topic areas and objectives
April	2013	Science and Mathematics Item Review Committee (SMIRC) reviewed proposed advanced mathematics and physics frameworks, developed innovative reasoning tasks and prototype items, and reviewed draft *TIMSS 2015 Item Writing Guidelines* at the 1st SMIRC meeting (Oslo, Norway)
May	2013	TIMSS & PIRLS International Study Center prepared final drafts of TIMSS Advanced 2015 advanced mathematics and physics assessment frameworks, incorporating SMIRC and NRC comments
May	2013	TIMSS & PIRLS International Study Center updated TIMSS Item Writing Guidelines for 2015
May	2013	NRCs reviewed TIMSS *Advanced 2015 Assessment Frameworks* and developed draft field test items using *TIMSS 2015 Item Writing Guidelines* at 2nd NRC meeting (Amsterdam, The Netherlands)
June – August	2013	TIMSS & PIRLS International Study Center further refined draft field test items and scoring guides and continued to develop additional items to cover frameworks
July	2013	Advanced Mathematics and Physics Task Forces reviewed and edited draft field test items and scoring guides, developed additional items to cover the frameworks, and classified items into preferred and alternate sets (Boston, USA)
September	2013	SMIRC reviewed draft field test items and scoring guides at 2nd SMIRC meeting (St. Petersburg, Russia)
September – October	2013	TIMSS & PIRLS International Study Center revised draft field test items and scoring guides to address SMIRC comments
November	2013	NRCs reviewed and approved proposed field test items at 3rd NRC meeting (Budapest, Hungary)
November – December	2013	TIMSS & PIRLS International Study Center assembled field test items into assessment blocks
December	2013	TIMSS & PIRLS International Study Center distributed field test achievement booklets to NRCs
January	2014	TIMSS & PIRLS International Study Center collected student responses to constructed response items from English-speaking countries to develop scoring training materials

Exhibit 1.2: TIMSS Advanced 2015 Development Schedule for Achievement Items (Continued)

Date(s)		Group and Activity
February	2014	Advanced Mathematics and Physics Task Forces modified scoring guides for constructed response items based on student responses and developed scoring training materials for 4th NRC meeting (Boston, USA)
March – April	2014	Countries conducted TIMSS Advanced 2015 field test
March	2014	TIMSS & PIRLS International Study Center published *TIMSS Advanced 2015 Assessment Frameworks*
March	2014	NRCs received scoring training for TIMSS Advanced 2015 constructed response field test items at 4th NRC meeting (Sydney, Australia)
April – May	2014	Countries submitted field test achievement data for analysis and review
June	2014	Advanced Mathematics and Physics Task Forces reviewed field test item statistics
June – July	2014	TIMSS & PIRLS International Study center assembled proposed item blocks in preparation for the 3rd SMIRC meeting
July	2014	SMIRC reviewed proposed item blocks in conjunction with field test results at 3rd SMIRC meeting (Sofia, Bulgaria)
August	2014	NRCs reviewed and approved item blocks for TIMSS Advanced 2015 data collection at 5th NRC meeting (Paris, France)
September	2014	TIMSS & PIRLS International Study Center distributes materials for the TIMSS Advanced pilot test for new items
October	2014	TIMSS & PIRLS International Study Center distributed TIMSS Advanced 2015 data collection achievement booklets to NRCs
January	2015	Advanced Mathematics and Physics Task Forces review scoring guides and update scoring training materials (Boston, MA)
February	2015	TIMSS & PIRLS International Study Center updated and prepared materials for TIMSS Advanced 2015 constructed response item scoring training and distributed them to NRCs in preparation for the 6th NRC meeting
March	2015	NRCs received scoring training for constructed response items at 6thNRC meeting (Prague, Czech Republic)
March	2015	TIMSS & PIRLS International Study Center distributed final TIMSS Advanced scoring guides and training materials for 2015 data collection

Writing and Reviewing the TIMSS Advanced 2015 Field Test Items and Scoring Guides

The TIMSS & PIRLS International Study Center uses a collaborative process involving the participating countries to develop test items and scoring guides for the field tests. Most of the 2nd TIMSS Advanced NRC meeting in Amsterdam was devoted to a workshop for developing the field test items. The NRCs, together with experienced item writers from participating countries and staff of the TIMSS & PIRLS International Study Center, created about half of the newly developed items for the advanced mathematics and physics field tests, and the other half of the items were created by ACER.

Prior to the workshop, TIMSS & PIRLS International Study Center staff members identified the scope of the item writing task for the field test, examining the weight given to each topic in each of the updated frameworks. Considerations included the total items needed based on the percentage of weight assigned to a particular area (for example, trigonometry) in the *TIMSS Advanced 2015 Assessment Frameworks*, and the number of topics in that area (three, for example), as well as how many items existed from previous assessments. Because the TIMSS & PIRLS International Study Center field tests twice the number of items actually required, the field test included the target number of new items needed multiplied by two. For TIMSS Advanced 2015, about 270 items were field tested (see Exhibit 1.4).

The TIMSS & PIRLS International Study Center used the updated *TIMSS 2015 Item Writing Guidelines* for the TIMSS Advanced 2015 item writing workshop. The *Item Writing Guidelines* contain general information about procedures for obtaining good measurement (for instance, items should 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 (for example, using TIMSS' fictitious unit of currency, the "zed," for any money items). The *Item Writing Guidelines* also include the necessary steps for developing scoring guides, as well as checklists for reviewing TIMSS items.

At the TIMSS Advanced item writing workshop, country representatives were divided into teams and given specific item writing assignments to ensure that enough field test items were developed in each of the content areas and cognitive processes areas specified in the frameworks. The TIMSS & PIRLS International Study Center staff and consultants used the *Item Writing Guidelines* to provide training to the teams on item writing procedures for the TIMSS Advanced assessments. Once teams had completed their item writing assignments, each team 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 Advanced 2015 item writing workshop and the number of items written.

Exhibit 1.3: TIMSS Advanced 2015 Item Writing Workshop to Develop Field Test Items

Attendees	
Number of Countries	8
Number of Country Representatives	18
Approximate Number of Field Test Items Written at Item Writing Workshop	
Advanced Mathematics	60
Physics	60

Following the item writing workshop, the draft set of field test items received a thorough review by the TIMSS & PIRLS International Study Center. Reviewers included staff, the chief consultants, and consultants experienced in developing assessment items, such as those from Educational Testing Service, the National Foundation for Educational Research in England, and the Australian Council for Educational Research, as well as SMIRC members with particular item writing skills.

Finally, the proposed field test blocks were reviewed by the TIMSS/TIMSS Advanced 2015 SMIRC and NRCs prior to field test instrument production. The TIMSS & PIRLS International Study Center implemented the suggested revisions and provided the final international version of the field test booklets to the NRCs so that they could begin translating the field test materials into their languages of instruction.

The TIMSS Advanced 2015 Field Test

The TIMSS Advanced field test followed typical TIMSS procedures, where it served as a full-scale "dress rehearsal" operationally for the assessment. That is, the data collection and scoring procedures to be employed in the assessment were practiced in the field test. In addition, the field test provided important information about how well each prospective item functioned and provided a basis for selecting items for the assessment.

The field test was designed to be conducted for approximately 30 schools in each country and yield at least 200 student responses to each advanced mathematics and physics item. Generally, the samples for the field test and the assessment are drawn simultaneously, using the same random sampling procedures. This ensures that field test samples closely approximate assessment samples, and that a school is selected for either the field test or the assessment, but not both. For example, if 150 schools are needed for the assessment and another 30 for the field test, then a larger sample of 180 schools is selected and a systematic sample of 30 schools is selected from the 180 schools.

The TIMSS Advanced 2015 field test was conducted in March–April 2014. Exhibits 1.4 through 1.6 provide a detailed summary of the field test effort, including the number of students, teachers, and schools that participated, and the number of items listed by format, content domain, and cognitive domain. Approximately 2,000 student responses from 10 countries were used to evaluate the measurement properties of each field test assessment item.

Exhibit 1.4: Overview of the TIMSS Advanced 2015 Field Test

	Advanced Mathematics	Physics
Items	**132**	**133**
Responses per Item (approx.)	200	200
Participants		
Countries	10	10
Students	9,537	8,252
Teachers	465	411
Schools	266	281

Exhibit 1.5: TIMSS Advanced 2015 Number of Field Test Items by Content Domain and Item Format

Content Domain	Number of Multiple-Choice Items	Number of Constructed Response Items	Total Number of Items	Total Number of Score Points	Percentage of Score Points
Advanced Mathematics Items					
Algebra	22	26	48	55	36%
Calculus	27	19	46	59	38%
Geometry	19	19	38	40	26%
Total	**68**	**64**	**132**	**154**	
Physics Items					
Mechanics and Thermodynamics	28	26	54	57	41%
Electricity and Magnetism	15	16	31	32	23%
Wave Phenomena and Atomic/Nuclear Physics	20	28	48	49	36%
Total	**63**	**70**	**133**	**138**	

Exhibit 1.6: TIMSS Advanced 2015 Number of Field Test Items by Cognitive Domain and Item Format

Cognitive Domain	Number of Multiple-Choice Items	Number of Constructed Response Items	Total Number of Items	Total Number of Score Points	Percentage of Score Points
Advanced Mathematics Items					
Knowing	28	9	37	38	25%
Applying	24	27	51	62	40%
Reasoning	16	28	44	54	35%
Total	**68**	**64**	**132**	**154**	
Physics Items					
Knowing	29	12	41	42	30%
Applying	16	42	58	61	44%
Reasoning	18	16	34	35	25%
Total	**63**	**70**	**133**	**138**	

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

Developing the Materials for TIMSS Advanced 2015 Field Test Scoring Training

It is necessary to prepare scoring training materials for the newly developed constructed response field test items in advance of the field test so field test scoring can occur immediately upon completion of data collection. To provide "grist" for these materials, small samples of English-speaking first year university students were given the newly developed constructed response items in the United States and Norway. Pilot materials were completed in December 2013 and responses were gathered from students in January 2014. The goal was to collect a total of at least 50 responses to each newly developed constructed response field test item to provide example student responses in the field test scoring guides and sets of training materials.

Additionally, the United States arranged for cognitive labs in Washington, D.C. and California. Each TIMSS Advanced constructed response item was presented to approximately five students, who were observed and prompted to answer questions about the clarity, difficulty, and familiarity of the item content and format. The TIMSS & PIRLS International Study Center received the cognitive lab reports in February 2014. Exhibit 1.7 provides the number of items included in the cognitive labs and the number of student responses collected.

Exhibit 1.7: Cognitive Lab Student Responses

Advanced Mathematics Items	20
Physics Items	36
Total Items	**56**
Responses per Item (approx.)	5
Number of Students (approx.)	50

The TIMSS Advanced 2015 NRCs and their scoring supervisors received scoring training for the field tested constructed response items in March 2014 in Sydney, Australia, as part of the 4th TIMSS Advanced 2015 NRC Meeting. Sets of example and practice papers were created for 19 advanced mathematics items and 24 physics items. The example and practice paper sets for each item included a scoring guide, approximately 8–10 example papers illustrating the categories in the scoring guide, and approximately 6–12 practice papers so that country representatives could practice making distinctions among categories and reach agreement about how to make consistent scoring decisions across countries.

At the scoring 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. After the country representatives scored the practice papers, any inconsistencies in scoring were discussed, and, as necessary, the field test guides were clarified and sometimes categories were revised.

Finalizing the TIMSS Advanced 2015 Achievement Items

Subsequent to the field test, the TIMSS & PIRLS International Study Center analyzed the TIMSS Advanced field test data and prepared almanacs containing summary item statistics for each field test item. The data almanac for an item contained, row by row for each country: the sample size, the item difficulty and discrimination, the percentage of students answering each option (multiple-choice) or in each score category (constructed response), the point-biserial correlation for each multiple-choice option or constructed response category, and the degree of scoring agreement for 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. The TIMSS & PIRLS International Study Center staff members, together with external consultants, first reviewed the field test data to make an initial judgment about the quality of each item based on its measurement properties (item statistics). Items were eliminated from further consideration if they had poor measurement properties, such as being too difficult or easy or having low discrimination. Particular attention was paid to unusual item statistics in individual countries since these could indicate errors in translation.

After the item-by-item review, the TIMSS & PIRLS International Study Center staff collaborated with consultants to assemble a set of recommended assessment blocks for review by the expert committee (SMIRC). SMIRC members scrutinized the recommendations for the newly developed assessment blocks, reviewing the items and scoring guides for content accuracy, clarity, and adherence to the frameworks. In addition, the newly developed items were considered in relation to the trend item blocks for overall coherence as a complete assessment.

The SMIRC's recommendations were implemented by staff, and the assessment blocks were sent to the NRCs for review. NRCs had the opportunity to review the recommended materials 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 an indication of translation errors and correct the translation as necessary or recommend revisions to accommodate translation. The 5th NRC meeting held in Paris, France in August 2014 was devoted to reviewing all the newly developed items. For several framework areas (e.g., optimization and rates of change in advanced mathematics) some items were necessarily revised to be less difficult for the TIMSS Advanced students. Because there were no TIMSS Advanced 2015 countries in the Southern Hemisphere, it was possible to pilot the revised items in the participating countries prior to incorporating them into the final assessments.

Distribution of TIMSS Advanced 2015 Items by Content and Cognitive Domains

Exhibits 1.8 and 1.9 present the number of trend and newly developed items as well as the number of score points in the TIMSS Advanced 2015 advanced mathematics and physics assessments. 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.

Exhibit 1.8: TIMSS Advanced 2015 Achievement Items by Content Domain

Content Domain	Number of Trend Items in TIMSS Advanced 2015	Percentage of Trend Score Points	Number of New Items in TIMSS Advanced 2015	Percentage of New Score Points	Total Items	Achieved Percentage of Score Points	Target Percentage of Score Points
Advanced Mathematics Items							
Algebra	12 (14)	36%	25 (29)	35%	37 (43)	35%	35%
Calculus	9 (13)	33%	25 (31)	37%	34 (44)	36%	35%
Geometry	11 (12)	31%	20 (24)	29%	31 (36)	29%	30%
Total	**32 (39)**		**70 (84)**		**102 (123)**		
Physics Items							
Mechanics and Thermodynamics	13 (15)	44%	27 (33)	40%	40 (48)	41%	40%
Electricity and Magnetism	8 (8)	24%	20 (23)	28%	28 (31)	26%	25%
Wave Phenomena and Atomic/ Nuclear Physics	10 (11)	32%	25 (27)	33%	35 (38)	33%	35%
Total	**31 (34)**		**72 (83)**		**103 (117)**		

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

Exhibit 1.9: TIMSS Advanced 2015 Achievement Items by Cognitive Domain

Cognitive Domain	Number of Trend Items in TIMSS Advanced 2015	Percentage of Trend Score Points	Number of New Items in TIMSS Advanced 2015	Percentage of New Score Points	Total Items	Achieved Percentage of Score Points	Target Percentage of Score Points
Advanced Mathematics Items							
Knowing	12 (13)	33%	21 (23)	27%	33 (36)	29%	35%
Applying	17 (22)	56%	23 (28)	33%	40 (50)	41%	35%
Reasoning	3 (4)	10%	26 (33)	39%	29 (37)	30%	30%
Total	**32 (39)**		**70 (84)**		**102 (123)**		
Physics Items							
Knowing	9 (10)	29%	22 (22)	27%	31 (32)	27%	30%
Applying	15 (16)	47%	27 (35)	42%	42 (51)	44%	40%
Reasoning	7 (8)	24%	23 (26)	31%	30 (34)	29%	30%
Total	**31 (34)**		**72 (83)**		**103 (117)**		

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

Distribution of TIMSS Advanced 2015 Item Formats within Content and Cognitive Domains

Exhibits 1.10 and 1.11 display the number of items (and score points) by item format for each content and cognitive domain. As described in the *TIMSS Advanced 2015 Assessment Frameworks*, at least half of the total number of score points represented by all the questions should come from multiple-choice items. Most TIMSS Advanced multiple-choice items are worth one score point, although some compound multiple-choice items are worth two score points. The 2-point compound multiple-choice items are scored as all parts answered correctly as fully correct (2 score points), and most parts answered correctly as partially correct (1 score point). Constructed response items generally are worth one or two score points depending on the degree of complexity involved. The 1-point constructed response items are scored as correct (1 score point) or incorrect (0 score points), whereas 2-point constructed response items are 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.

Exhibit 1.10: TIMSS Advanced 2015 Achievement Items by Content Domain and Item Format

Content Domain	Multiple-Choice Items		Constructed Response Items		Total Items	Percentage of Score Points
	Four Response Options	Compound	1 Point	2 Points		
Advanced Mathematics Items						
Algebra	18 (18)	1 (2)	13 (13)	5 (10)	37 (43)	35%
Calculus	19 (19)	2 (4)	5 (5)	8 (16)	34 (44)	36%
Geometry	19 (19)		7 (7)	5 (10)	31 (36)	29%
Total	**56 (56)**	**3 (6)**	**25 (25)**	**18 (36)**	**102 (123)**	
Achieved Percentage of Score Points	**50%**		**50%**			
Target Percentage of Score Points	**50%**		**50%**			
Physics Items						
Mechanics and Thermodynamics	24 (24)		8 (8)	8 (16)	40 (48)	41%
Electricity and Magnetism	17 (17)	1 (1)	7 (7)	3 (6)	28 (31)	26%
Wave Phenomena and Atomic/Nuclear Physics	19 (19)		13 (13)	3 (6)	35 (38)	33%
Total	**60 (60)**	**1 (1)**	**28 (28)**	**14 (28)**	**103 (117)**	
Achieved Percentage of Score Points	**52%**		**48%**			
Target Percentage of Score Points	**50%**		**50%**			

Score points are shown in parentheses.

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

Exhibit 1.11: TIMSS Advanced 2015 Achievement Items by Cognitive Domain and Item Format

Cognitive Domain	Multiple-Choice Items		Constructed Response Items		Total Items	Percentage of Score Points
	Four Response Options	Compound	1 Point	2 Points		
Advanced Mathematics Items						
Knowing	25 (25)	2 (4)	5 (5)	1 (2)	33 (36)	29%
Applying	22 (22)		8 (8)	10 (20)	40 (50)	41%
Reasoning	9 (9)	1 (2)	12 (12)	7 (14)	29 (37)	30%
Total	**56 (56)**	**3 (6)**	**25 (25)**	**18 (36)**	**102 (123)**	
Achieved Percentage of Score Points	**50%**		**50%**			
Target Percentage of Score Points	**50%**		**50%**			
Physics Items						
Knowing	24 (24)		6 (6)	1 (2)	31 (32)	27%
Applying	17 (17)	1 (1)	15 (15)	9 (18)	42 (51)	44%
Reasoning	19 (19)		7 (7)	4 (8)	30 (34)	29%
Total	**60 (60)**	**1 (1)**	**28 (28)**	**14 (28)**	**103 (117)**	
Achieved Percentage of Score Points	**52%**		**48%**			
Target Percentage of Score Points	**50%**		**50%**			

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

TIMSS Advanced 2015 Constructed Response Scoring Training

In preparation for the main data collection scoring training, some TIMSS Advanced 2015 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. Several new scoring training sets were also added to the training materials for items revised and piloted after the field test. For TIMSS Advanced 2015 scoring training the example and practice paper training sets included those used in TIMSS Advanced 2008 for the trend items and the updated training sets for the newly developed items selected for TIMSS Advanced 2015, resulting in 16 example and practice paper sets for advanced mathematics and 14 for physics.

The TIMSS Advanced NRCs and their scoring supervisors received scoring training led by the TIMSS & PIRLS International Study Center in March 2015 in Prague, Czech Republic as part of the 6th TIMSS Advanced 2015 NRC Meeting. Exhibit 1.12 shows the number of participants in the scoring training sessions.

Exhibit 1.12: TIMSS Advanced 2015 Scoring Training Participation

Participants	
Number of Countries	9
Number of Country Representatives	27

The Process Following Instrument Development

In general, after the participating countries received the international version of the assessment 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 for printing. At the same time, countries made final arrangements for data collection, including the host of activities necessary to obtain school participation, implement test administration, and score the responses to the tests and questionnaires (see following chapters).

CHAPTER 2

Developing the TIMSS Advanced 2015 Context Questionnaires

Martin Hooper
Lauren Palazzo

TIMSS Advanced assesses students in specialized advanced mathematics and physics programs during their final year of secondary school. Monitoring the achievement of these students is a high priority in many countries since these students are considered to be the future leaders in science, technology, engineering, mathematics—fields renowned for fueling economic growth and innovation.

The TIMSS Advanced context questionnaires complement the achievement data by providing key information on students' contexts for learning. Countries collect background data from students, their teachers, and their principals, and National Research Coordinators (NRCs) from participating countries also provide country-level data. These data when analyzed in relation to TIMSS Advanced achievement yield insights into factors related to student achievement that can be relevant in developing educational policies.

Development Process for the TIMSS Advanced 2015 Context Questionnaires

Developing the TIMSS Advanced 2015 context questionnaires was a collaborative process involving multiple rounds of reviews by staff at the TIMSS & PIRLS International Study Center, policy analysis experts on the TIMSS 2015 Questionnaire Item Review Committee (QIRC), and the NRCs from the participating countries. In broad strokes, the TIMSS Advanced 2015 context questionnaire development process for the student, home, school, and teacher questionnaires included:

- Writing the TIMSS Advanced context questionnaire framework
- Modifying and developing new context questionnaire items by staff at the TIMSS & PIRLS International Study Center
- Reviewing and revising the questionnaires by the QIRC and NRCs
- Administering the TIMSS Advanced 2015 field test
- Using the field test results to refine the questionnaires

Developing the Curriculum Questionnaires followed a collaborative process similar to other TIMSS Advanced questionnaires, including identifying important framework topics, developing questionnaire items, and iterative reviews by NRCs.

Exhibit 2.1 presents the TIMSS Advanced 2015 questionnaire development schedule. The development process was directed and managed by the staff of the TIMSS & PIRLS International Study Center at Boston College, including the Executive Directors Ina V.S. Mullis and Michael O. Martin, and the TIMSS Questionnaire Coordinator, Martin Hooper. NRCs had an essential role in updating the questionnaires by providing feedback and ideas at successive NRC meetings. The QIRC made major contributions in updating the TIMSS Advanced 2015 questionnaires with the 1st QIRC meeting focused on developing items/scales, and the 2nd meeting focused on refining the questionnaires in light of the field test results. Exhibit 2.2 lists the members of the QIRC.

Exhibit 2.1: TIMSS Advanced 2015 Context Questionnaire Development Schedule

Date(s)		Group and Activity
February	2013	NRCs reviewed TIMSS 2011 context questionnaires at the 1st NRC meeting (Hamburg, Germany) to inform the TIMSS Advanced Context Questionnaire Framework
May-June	2013	Staff at TIMSS & PIRLS International Study Center drafted the TIMSS Advanced Context Questionnaire Framework Chapter
June	2013	1st meeting of the Questionnaire Item Review Committee (QIRC) to review the draft TIMSS 2015 Context Questionnaire Framework chapter and the draft questionnaire items and scales (Singapore)
July–October	2013	TIMSS & PIRLS International Study Center extended the draft TIMSS 2015 questionnaires to include TIMSS Advanced-specific topics and continued drafting the TIMSS Advanced Context Questionnaire Framework chapter
November	2013	NRCs reviewed the TIMSS Advanced Context Questionnaire Framework chapter and draft context questionnaires at the 3rd NRC meeting (Budapest, Hungary). A working group of NRCs developed additional items and scales for the TIMSS Advanced questionnaires.
November–December	2013	Staff at TIMSS & PIRLS International Study Center finalized the TIMSS Advanced field test context questionnaires
December	2013	Staff at TIMSS & PIRLS International Study Center provided the TIMSS Advanced field test context questionnaires to NRCs
March	2014	Staff at TIMSS & PIRLS International Study Center published *TIMSS Advanced 2015 Assessment Frameworks*, which includes the chapter on the Context Questionnaire Framework
March–April	2014	Countries conducted TIMSS Advanced 2015 field test
April–May	2014	Countries submitted field test data for analysis and review
June	2014	TIMSS & PIRLS International Study Center conducted an internal review of field test results
July	2014	QIRC reviewed questionnaire field test data, TIMSS 2011 Curriculum Questionnaire, and the TIMSS Advanced 2011 Curriculum Questionnaire at the 2nd QIRC meeting (Muenster, Germany)

Exhibit 2.1: TIMSS Advanced 2015 Context Questionnaire Development Schedule (Continued)

Date(s)		Group and Activity
August	2014	NRCs reviewed and approved context questionnaires for TIMSS Advanced 2015 data collection at 5th NRC meeting (Paris, France)
September	2014	TIMSS & PIRLS International Study Center distributed TIMSS Advanced 2015 data collection context questionnaire instruments to NRCs for translation
January–March	2015	TIMSS & PIRLS International Study Center drafted TIMSS Advanced 2015 Curriculum Questionnaires
March	2015	NRCs approved draft TIMSS Advanced 2015 Curriculum Questionnaires at the 6th NRC meeting (Prague, Czech Republic)
March–June	2015	TIMSS Advanced 2015 data collection
April-August	2015	TIMSS Advanced 2015 Curriculum Questionnaires administered online to NRCs

Exhibit 2.2: TIMSS 2015 Questionnaire Item Review Committee (QIRC)

Sue Thomson
Australian Council for Educational Research
Australia

Josef Basl
Czech School Inspectorate
Czech Republic

Wilfried Bos
Institut für Schulentwicklungsforschung
TU Dortmund University
Germany

Martina Meelissen
Department of Research Methodology, Measurement and Data Analysis, Faculty of Behavioural Sciences
University of Twente
The Netherlands

Chew Leng Poon
Planning Division, Research and Evaluation
Ministry of Education
Singapore

Peter Nyström
National Center for Mathematics Education
University of Gothenburg
Sweden

Jack Buckley
The College Board
United States

Updating the TIMSS Advanced 2015 Context Questionnaire Framework

The TIMSS Advanced 2015 Context Questionnaire Framework, Chapter 3 of the *TIMSS Advanced 2015 Assessment Frameworks*, provided the foundation for updating the TIMSS Advanced context questionnaires for 2015. The Framework chapter presents a review of a vast array of educational research that identifies key context questionnaire topics and gives the theoretical justification for asking about these topics within the 2015 questionnaires.

NRCs described topics they thought should be covered in the TIMSS Advanced 2015 questionnaires at their 1st NRC meeting in February 2013 in Hamburg, Germany, and at the 2nd meeting in May 2013 in Amsterdam, The Netherlands. Taking into account feedback garnered at the meetings, the TIMSS Questionnaire Coordinator conducted an extensive literature review and drafted the TIMSS Advanced 2015 Questionnaire Framework chapter. Because the primary purpose of the context questionnaires is to identify factors that may contribute to differences in achievement within and between countries, the framework focuses on topics in educational research found to be related to achievement across a variety of settings and contexts.

The NRCs reviewed the draft chapter at the 3rd NRC meeting in November 2013 in Budapest, Hungary. Staff at the TIMSS & PIRLS International Study Center refined the draft based upon the recommendations received at the meeting and published the final *TIMSS Advanced 2015 Assessment Frameworks* online in March 2014, with printed copies distributed thereafter.

Field Test Questionnaire Development

Because many of the key topics for learning are the same across the TIMSS and TIMSS Advanced populations, the TIMSS & PIRLS International Study Center developed the TIMSS 2015 and TIMSS Advanced 2015 questionnaires in tandem. As outlined in Chapter 2 of the *Methods and Procedures in TIMSS 2015*, the development of the TIMSS 2015 questionnaires focused on improving and expanding the TIMSS 2011 context questionnaire scales and updating items to align with recent technological innovations. With the draft TIMSS questionnaires at hand, staff at the TIMSS & PIRLS International Study focused TIMSS Advanced questionnaire development on updating the questionnaires to be age- and context-appropriate for the TIMSS Advanced population, maintaining measurement of important components of the TIMSS Advanced 2008 questionnaires, and incorporating policies, practices, and educational contructs that are uniquely relevant to the TIMSS Advanced study.

Much of the TIMSS/TIMSS Advanced scale development work took place at the first meeting of the QIRC, which was held jointly with the first meeting of the PIRLS Questionnaire Development Group (QDG). The QIRC and QDG worked with staff from the TIMSS & PIRLS International Study center to revamp the teacher job satisfaction scale by incorporating insights gained from the *Utrecht Work Engagement Scale* (Schaufeli, Bakker, & Salanova, 2006). The questionnaire committees also revised the Confidence in Teaching Mathematics/Science scales, with item development influenced by the *Ohio State Teacher Efficacy Scale* (Tschannen-Moran & Hoy, 2001). Additional items were also included for the students engagement scales, with one item sourced from Fauth, Decristan, Rieser, Klieme, and Büttner (2014).

Updating questionnaires to "keep up with the times" was an essential part of the 2015 development process. Staff at the TIMSS & PIRLS International Study Center worked with the QIRC and QDG to ensure that the questionnaires included items on the availability of prevalent digital resources for education such as ebooks, tablets, and interactive whiteboards.

A major step of the TIMSS Advanced-specific questionnaire development process took place at the 3rd National Research Coordinators meeting in Budapest, Hungary. At the meeting, a working group composed of NRCs and other country representatives developed TIMSS Advanced-specific items focusing on topics such as educational expectations, intended areas of study, and intended future profession for these advanced students. The working group also developed the TIMSS Advanced School Supports Advanced Mathematics and Physics Education scales, and the group extended and modified existing TIMSS items and scales focusing on instructional practices, student affect, and home resources to match the age and educational context of the TIMSS Advanced populations. For example, the TIMSS Students Like Learning Mathematics/Science scales were revised and extended to better match the TIMSS Advanced population, with modifications influenced by items in the academic motivation literature (Lepper, Corpus, & Iyengar, 2005; Marsh, Craven, Hinkley, & Debus, 2003; OECD, 2003; Pierce, Stacey, & Barkatsas, 2007).

Review Field Test Results and Refine Questionnaires for Data Collection

The TIMSS Advanced 2015 countries administered a full-scale field test in the nine participating countries, eliciting questionnaire data from 11,922 students, 547 principals, and 884 teachers. Following field test administration, staff at the TIMSS & PIRLS International Study Center produced data almanacs and scale summaries to facilitate the review of the field test data:

- Data almanacs document the use of response categories for each context questionnaire item as well each item's relationship with achievement
- Scale summaries detail each scale's reliability, dimensionality, fit to the Rasch model, and relationship with achievement

In June 2014, staff at the TIMSS & PIRLS International Study Center reviewed the field test context questionnaire results, proposing revisions to the QIRC. At their 2nd meeting in July 2014, the QIRC accepted many of the recommendations and suggested some additional changes. In August 2014, NRCs reviewed the final draft questionnaires and accepted the questionnaires with a few minor revisions. Following the NRC meeting, staff at the TIMSS & PIRLS International Study Center implemented the revisions and posted the final TIMSS Advanced instruments on September 18, 2014, so that countries could begin the translation process.

Developing the TIMSS Advanced 2015 Curriculum Questionnaires

The TIMSS Advanced Curriculum Questionnaires complement the student, teacher, and school questionnaires by collecting information from NRCs about country-level contexts. The Curriculum Questionnaires cover each country's advanced mathematics and physics curricula, goals and standards for instruction, and other national or regional policies such as the preprimary education process and the teacher education process.

Similar to the other TIMSS Advanced 2015 questionnaires, the process for updating the Curriculum Questionnaires started with the TIMSS Advanced 2015 Context Questionnaire Framework. Then, the QIRC identified the information from the TIMSS 2011 and TIMSS Advanced 2008 Curriculum Questionnaires that they thought was useful to continue collecting.

Based on the framework and the QIRC feedback, staff at the TIMSS & PIRLS International Study Center updated the TIMSS Advanced 2015 Curriculum Questionnaires for review by NRCs at their 6th meeting in March 2015. Following the NRC meeting, staff at the TIMSS & PIRLS International Study Center finalized the questionnaires, incorporating the suggestions that emerged from the meeting. NRCs completed the online Curriculum Questionnaires between April 23, 2015 and August 31, 2015.

References

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, *29*, 1–9.

Lepper, M.R., Corpus, J.H., & Iyengar, S.S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology, 97*(2), 184–196.

Schaufeli, W.B., Bakker, A.B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement*, *66*(4), 701–716.

Tschannen-Moran, M. & Hoy, A.W. (2001). Teacher efficacy: Capturing an elusive construct. *Teacher and Teacher Education*, *17*, 783–805.

Organization for Economic Co-Operation and Development (OECD). (2003). *Education at a glance: OECD indicators*. Paris: Author.

Pierce, R., Stacey, K., & Barkatsas, A. (2007). A scale for monitoring students' attitudes to learning mathematics with technology. *Computers & Education, 48*, 285–300.

Marsh, H.W., Craven, R.G., Hinkley, J.W., & Debus, R.L. (2003). Evaluation of the big-two-factor theory of academic motivation orientations: An evaluation of jingle-jangle fallacies. *Multivariate Behavioral Research, 38*(2), 189–224.

CHAPTER 3

Sample Design in TIMSS Advanced 2015

Sylvie LaRoche
Pierre Foy

Introduction

TIMSS Advanced is designed to provide valid and reliable measurement of trends in student achievement in countries around the world, while keeping to a minimum the burden on schools, teachers, and students. The TIMSS Advanced program employs rigorous school and classroom sampling techniques so that achievement in the student population as a whole may be estimated accurately by assessing just a sample of students from a sample of schools. TIMSS Advanced assesses advanced mathematics and physics achievement in the final year of upper secondary schooling for students with advanced preparation in these subjects.

This chapter describes the sample design developed for TIMSS Advanced 2015. It explains how the target populations were defined in the participating countries and how the national sample designs were developed. It also explains how the sampling weights and participation rates are calculated.

National Sampling Plan

Each country participating in TIMSS Advanced needs a plan for defining its national target population and applying the TIMSS Advanced sampling methods to achieve a nationally representative sample of schools and students. The development and implementation of the national sampling plan is a collaborative exercise involving the country's National Research Coordinator (NRC) and the TIMSS Advanced sampling experts.

Statistics Canada is responsible for advising the National Research Coordinator on all sampling matters and for ensuring that the national sampling plan conforms to the TIMSS Advanced standards. In cooperation with sampling staff from the IEA Data Processing and Research Center (IEA DPC), Statistics Canada works with the National Research Coordinator to select the national school sample(s) and produce all supporting documentation for tracking the sampled schools. This includes ensuring that the school sampling frame (the school population list from which the school sample is drawn) provided by the National Research Coordinator is complete and satisfactory; checking that categories of excluded students are clearly defined, justified, and kept to a minimum; assisting the National Research Coordinator in determining the sample size and a stratification plan that will meet both international and national objectives; and drawing a national sample of schools. When sampling has been completed and all data collected, Statistics Canada documents population coverage and school and student participation rates and constructs appropriate sampling weights for use in analyzing and reporting the results.

The TIMSS & PIRLS International Study Center, in cooperation with Statistics Canada and the IEA DPC, provides National Research Coordinators with a series of manuals to guide them through the sampling process. More specifically, *TIMSS Advanced 2015 Survey Operations Procedures Unit 1: Sampling Schools and Obtaining their Cooperation* describes the steps involved in defining the national target population and selecting the school sample, and *TIMSS Advanced 2015 Survey Operations Procedures Unit 3: Contacting Schools and Sampling Classes for Data Collection* describes the procedure for sampling classes within the sampled schools and making preparations for conducting the assessments. Within-school sampling procedures for the field test are documented in *TIMSS Advanced 2015 Survey Operations Procedures Unit 2: Preparing for and Conducting the Field Test*. More information on the Survey Operations Units can be found in Chapter 6 of this volume.

The TIMSS Advanced National Research Coordinator is responsible for providing Statistics Canada with all information and documentation necessary to conduct the national sampling, and for conducting all sampling operations in the country. In particular, the NRC is expected to identify the programs, tracks, or courses that correspond to the international target population; create a sampling frame by listing all schools in the population that have classes with advanced mathematics and/or physics students in the target grade; determine national population coverage and exclusions, in accordance with the TIMSS Advanced international guidelines; work with Statistics Canada to develop a national sampling plan and identify suitable stratification variables, ensuring that these variables are present and correct for all schools; contact all sampled schools and secure their participation; keep track of school participation and the use of replacement schools; and conduct all within-school sampling of classes. Each NRC is required to complete a series of sampling forms documenting the completion of each of these tasks.

A crucial feature of each international meeting of National Research Coordinators is a one-to-one meeting between each NRC and sampling staff at Statistics Canada and the IEA DPC. At these meetings, each step of the sampling process is documented and reviewed in detail, and NRCs have the opportunity to raise issues and ask questions about their national situation and any challenges they face. Statistics Canada consults with the TIMSS & PIRLS International Study Center and the International Sampling Referee, as necessary, to resolve issues and questions. Final approval of TIMSS Advanced national sampling plans is the responsibility of the TIMSS & PIRLS International Study Center, based upon the advice of Statistics Canada and the International Sampling Referee.

Defining the TIMSS Advanced 2015 Target Populations

TIMSS Advanced 2015 measured student achievement in two student populations at the end of secondary schooling: advanced mathematics students and physics students. To allow for meaningful interpretation of the TIMSS Advanced 2015 data, and to ensure the comparability of the results across the participating countries, it was important that both target populations be accurately and consistently defined.

The TIMSS Advanced 2015 target population for advanced mathematics was defined as *the students in the final year of secondary schooling who have taken courses in advanced mathematics.* For physics, the TIMSS Advanced 2015 target population was defined as *the students in the final year of secondary schooling who have taken courses in physics.*

Courses in Advanced Mathematics and Physics

The courses that would define the target populations had to cover most, if not all, of the advanced mathematics and physics topics that were outlined in the TIMSS Advanced 2015 Assessment Frameworks. The students attending these courses were likely to be the most advanced mathematics or physics students in the final year of upper secondary schooling in the participating countries. It was the responsibility of the NRCs to identify these advanced mathematics courses and physics courses. In many cases, the courses were found in specific academic programs, or tracks, in upper secondary schools.

In the Russian Federation, TIMSS Advanced 2015 mathematics students assessed include both the Profile and Intensive streams of students. However, results also are provided separately for the students in the Intensive stream because this is the group of students assessed in TIMSS Advanced 1995 and TIMSS Advanced 2008. The results for the Intensive stream students are designated Russian Federation 6hr+.

Exhibit 3.1 and Exhibit 3.2 give an overview of the national target population definitions for advanced mathematics and physics, respectively, in terms of the courses or programs in which the eligible students would be found. In all instances, these students were in their final year of secondary schooling; although this meant the students had varied numbers of years of schooling across the participating countries and were of different average age. Exhibit M1.1 and Chapter M11 of the *TIMSS Advanced 2015 International Report in Advanced Mathematics and Physics* describe the advanced mathematics programs of the upper secondary educational systems in the participating countries, and Exhibit P1.1 and Chapter P11 provide similar descriptions for the physics programs.

Exhibit 3.1: TIMSS Advanced 2015 Advanced Mathematics Populations

Country	Advanced Mathematics Population
France	Students in the 12th grade in the scientific track
Italy	Students in the 13th grade in general schools with scientific focus on mathematics and physics (Liceo Scientifico), in general schools with a focus on science, mathematics, and physics (Liceo Scientifico opzione Scienze Applicate), or in technical institutes and receiving full-time vocational training
Lebanon	Students in the 12th grade in the general science program
Norway	Students in the 13th grade who have taken the Mathematics R2 advanced mathematics course in the academic track
Portugal	Students in the 12th grade in the Sciences and Technology or Socio-Economic programs of the academic track who have taken the Mathematica A advanced mathematics course
Russian Federation	Students in the 11th grade who have taken 4.5 hours or more per week of instruction in mathematics (Profile and Intensive streams)
Russian Federation 6hr+	Students in the 11th grade who have taken 6 hours or more per week of instruction in mathematics (Intensive stream)
Slovenia	Students in the 13th grade in general gymnasia programs
Sweden	Students in the 12th grade in the Natural Science or the Technology programs and have taken or were taking Mathematics 4 and/or Mathematics 5 course
United States	Students in the 12th grade who have taken an advanced mathematics course (AP, IB, or another advanced mathematics course specific to their state/district), in Grade 12 or in a prior grade

Exhibit 3.2: TIMSS Advanced 2015 Physics Populations

Country	Physics Population
France	Students in the 12th grade in the scientific track
Italy	Students in the 13th grade in general schools with scientific focus on mathematics and physics (Liceo Scientifico), in general schools with a focus on science, mathematics, and physics (Liceo Scientifico opzione Scienze Applicate), or in technical institutes and receiving full-time vocational training
Lebanon	Students in the 12th grade in the general science program
Norway	Students in the 13th grade of the academic track who have taken the Physics 2 course
Portugal	Students in the 12th grade in the Sciences and Technology program of the academic track who have taken physics courses
Russian Federation	Students in the 11th grade who have taken 4.5 hours or more per week of instruction in mathematics (Profile and Intensive streams)
Slovenia	Students in the 13th grade in general gymnasia programs
Sweden	Students in the 12th grade in the Natural Science or the Technology programs and have taken or were taking Mathematics 4 and/or Mathematics 5 course
United States	Students in the 12th grade who have taken an advanced mathematics course (AP, IB, or another advanced mathematics course specific to their state/district), in Grade 12 or in a prior grade

TIMSS Advanced Coverage Indices

In order to quantify the proportion of the school-leaving age cohort taking advanced mathematics and physics courses, TIMSS Advanced computed a TIMSS Advanced Mathematics Coverage Index (TAMCI) and a TIMSS Advanced Physics Coverage Index (TAPCI) for each participating country. In part, these indices reflect the overall sampling coverage of the respective populations in each country; but, more importantly, they show that only a very select group of final-year students were considered eligible for TIMSS Advanced 2015, and that the percentages of these students varied across countries.

The TIMSS Advanced coverage indices for advanced mathematics and physics were defined as follows:

$$\text{TAMCI} = \frac{\text{Estimated total number of students in the advanced mathematics population}}{\text{Total national population in the corresponding age cohort}} \times 100\%$$

$$\text{TAPCI} = \frac{\text{Estimated total number of students in the physics population}}{\text{Total national population in the corresponding age cohort}} \times 100\%$$

The numerator in each index is the total number of students eligible for TIMSS Advanced 2015, either for advanced mathematics or physics, as estimated from the weighted samples. The denominator is an estimate of the size of the eligible age cohort in 2015 corresponding to the mean age of the target population. The TIMSS Advanced coverage indices for advanced mathematics and physics are presented in Exhibits 3.3 and 3.4. The final-year age cohort for each country was defined to be the age corresponding to its average age at the time of testing, as estimated from the weighted samples, and its size was estimated from national census figures. The estimated target populations were estimated from the weighted samples.

Exhibit 3.3: Size of TIMSS Advanced 2015 Target Population for Advanced Mathematics, the Age Cohort, and the TIMSS Advanced Coverage Index for Advanced Mathematics

Country	Years of Formal Schooling*	Age Cohort Corresponding to the Final Year of Secondary School	Estimated Size of the Population of Students in the Final Year of Secondary School Taking the Advanced Mathematics Track or Program Targeted by TIMSS Advanced (Derived from TIMSS Advanced Student Sample)	Size of the Age Cohort Corresponding to the TIMSS Advanced Population Based on National Census Figures**	TIMSS Advanced Mathematics Coverage Index – the Percentage of the Entire Corresponding Age Cohort Covered by the TIMSS Advanced Target Population
France	12	18	172,178	801,889	21.5%
Italy	13	19	141,419	576,506	24.5%
Lebanon	12	18	4,457	113,204	3.9%
Norway	13	19	6,751	63,894	10.6%
Portugal	12	18	31,314	109,984	28.5%
Russian Federation	11	18	138,548	1,365,790	10.1%
Russian Federation 6hr+	11	18	25,830	1,365,790	1.9%
Slovenia	13	19	6,738	19,598	34.4%
Sweden	12	19	15,285	108,138	14.1%
United States	12	18	473,405	4,168,000	11.4%

* Represents years of schooling counting from the first year of primary or basic education (first year of ISCED Level 1).

** France: Estimate derived by dividing the population of 15–19-year-olds by 5 for the single year estimate. Data retrieved from INSEE (National Institute of Statistics and Economic Studies), Estimations de Population (résultats provisoires à fin 2015); http://www.insee.fr/fr/themes/detail.asp?reg_id=99&ref_id=estim-pop.

Italy: Value is the total population of 19-year olds in Italy in 2015. Data retrieved from ISTAT (the National Statistics Institute); http://dati.istat.it/Index.aspx?DataSetCode=DCIS_POPRES1.

Lebanon: Value is the total population of 18-year olds in Lebanon in 2015. Data retrieved from http://databank.worldbank.org/data/reports.aspx?source=health-nutrition-and-population-statistics:-population-estimates-and-projections&Type=TABLE&preview=on.

Norway: Estimate derived by dividing the population of 15–19-year-olds by 5 for the single year estimate. Data retrieved from https://stats.oecd.org/Index.aspx?DataSetCode=POP_PROJ.

Portugal: Estimate derived by dividing the 2014 population of 15–19-year-olds by 5 for the single year estimate. Data retrieved from INE (Instituto Nacional de Estatística) Annual Estimates of Resident Population;
http://www.pordata.pt/en/Portugal/Resident+population+total+and+by+age+group-10.

Russian Federation: Estimate derived by dividing the population of 15–19-year-olds by 5 for the single year estimate. Data retrieved from The Demographic Yearbook of Russia, 2015; http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1137674209312.

Slovenia: Value is the total population of 18-year olds in Slovenia as of July 1st 2015. Data retrieved from the Statistical Office of the Republic of Slovenia; http://pxweb.stat.si.

Sweden: Value is the total population of 18-year olds as of December 31, 2014 (Born 1996). Data retrieved from Statistics Sweden;
http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101A/BefolkningR1860/table/tableViewLayout1/?rxid=06695d79-5fa1-41d1-81c1-3ae51dcd09b7.

United States: Value is the total population of 18-year olds as of July 1st 2015. Data retrieved from the US Census Annual Estimates of the Resident Population by Single Year of Age and Sex for the United States: April 1, 2010 to July 1, 2013; https://www.census.gov/popest/data/national/asrh/2013/. The post-census estimates are as of July 1 of each year. For the 18 year-olds estimate in 2015, the 2015 population was projected using the year to year changes from 2010 to 2013 and extending it to 2015.

The TIMSS Advanced Mathematics Coverage Index reflects the differences across countries in the proportion of the age cohort that are enrolled in these advanced courses the final year of secondary education. In some countries, only a very select group of students was considered eligible for the study, while in others, a much larger group was included.

The TIMSS Advanced Mathematics Coverage Index (TAMCI) is defined as follows:

$$\text{TAMCI} = \frac{\text{Estimated total number of students in the advanced mathematics target population in 2015}}{\text{Total national population in the corresponding age cohort in 2015}} \times 100\%$$

The numerator is the total number of students eligible for TIMSS Advanced, estimated from the weighted sample data. These are students in the final year of secondary school taking the advanced mathematics track or program targeted by TIMSS Advanced, based on the TIMSS Advanced sample. The denominator is the size of the population age cohort corresponding to the average age of the students in the target populations and is based on national census figures.

Exhibit 3.4: Size of TIMSS Advanced 2015 Target Population for Physics, the Age Cohort, and the TIMSS Advanced Coverage Index for Physics

Country	Years of Formal Schooling*	Age Cohort Corresponding to the Final Year of Secondary School	Estimated Size of the Population of Students in the Final Year of Secondary School Taking the Physics Track or Program Targeted by TIMSS Advanced (Derived from TIMSS Advanced Student Sample)	Size of the Age Cohort Corresponding to the TIMSS Advanced Population Based on National Census Figures**	TIMSS Advanced Physics Coverage Index–the Percentage of the Entire Corresponding Age Cohort Covered by the TIMSS Advanced Target Population
France	12	18	172,178	801,889	21.5%
Italy	13	19	104,650	576,506	18.2%
Lebanon	12	18	4,464	113,204	3.9%
Norway	13	19	4,163	63,894	6.5%
Portugal	12	18	5,661	109,984	5.1%
Russian Federation	11	18	66,746	1,365,790	4.9%
Slovenia	13	19	1,491	19,598	7.6%
Sweden	12	19	15,423	108,138	14.3%
United States	12	18	199,944	4,168,000	4.8%

* Represents years of schooling counting from the first year of primary or basic education (first year of ISCED Level 1).

** France: Estimate derived by dividing the population of 15–19-year-olds by 5 for the single year estimate. Data retrieved from INSEE (National Institute of Statistics and Economic Studies), Estimations de Population (résultats provisoires à fin 2015); http://www.insee.fr/fr/themes/detail.asp?reg_id=99&ref_id=estim-pop.

Italy: Value is the total population of 19-year olds in Italy in 2015. Data retrieved from ISTAT (the National Statistics Institute); http://dati.istat.it/Index.aspx?DataSetCode=DCIS_POPRES1.

Lebanon: Value is the total population of 18-year olds in Lebanon in 2015. Data retrieved from http://databank.worldbank.org/data/reports.aspx?source=health-nutrition-and-population-statistics:-population-estimates-and-projections&Type=TABLE&preview=on.

Norway: Estimate derived by dividing the population of 15–19-year-olds by 5 for the single year estimate. Data retrieved from https://stats.oecd.org/Index.aspx?DataSetCode=POP_PROJ.

Portugal: Estimate derived by dividing the 2014 population of 15–19-year-olds by 5 for the single year estimate. Data retrieved from INE (Instituto Nacional de Estatística) Annual Estimates of Resident Population; http://www.pordata.pt/en/Portugal/Resident+population+total+and+by+age+group-10.

Russian Federation: Estimate derived by dividing the population of 15–19-year-olds by 5 for the single year estimate. Data retrieved from The Demographic Yearbook of Russia, 2015; http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1137674209312.

Slovenia: Value is the total population of 18-year olds in Slovenia as of July 1st 2015. Data retrieved from the Statistical Office of the Republic of Slovenia; http://pxweb.stat.si.

Sweden: Value is the total population of 18-year olds as of December 31, 2014 (Born 1996). Data retrieved from Statistics Sweden; http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101A/BefolkningR1860/table/tableViewLayout1/?rxid=06695d79-5fa1-41d1-81c1-3ae51dcd09b7.

United States: Value is the total population of 18-year olds as of July 1st 2015. Data retrieved from the US Census Annual Estimates of the Resident Population by Single Year of Age and Sex for the United States: April 1, 2010 to July 1, 2013; https://www.census.gov/popest/data/national/asrh/2013/. The post-census estimates are as of July 1 of each year. For the 18 year-olds estimate in 2015, the 2015 population was projected using the year to year changes from 2010 to 2013 and extending it to 2015.

The TIMSS Advanced Physics Coverage Index reflects the differences across countries in the proportion of the age cohort that are enrolled in these advanced courses in the final year of secondary education. In some countries, only a very select group of students was considered eligible for the study, while in others, a much larger group was included.

The TIMSS Advanced Physics Coverage Index (TAMCI) is defined as follows:

$$\text{TAMCI} = \frac{\text{Estimated total number of students in the physics target population in 2015}}{\text{Total national population in the corresponding age cohort in 2015}} \times 100\%$$

The numerator is the total number of students eligible for TIMSS Advanced, estimated from the weighted sample data. These are students in the final year of secondary school taking the physics track or program targeted by TIMSS Advanced, based on the TIMSS Advanced sample. The denominator is the size of the population age cohort corresponding to the average age of the students in the target populations and is based on national census figures.

National Coverage and Exclusions

TIMSS Advanced is designed to describe and summarize student achievement across the entire defined target populations, and so it is very important that national target populations aim for comprehensive coverage of eligible students. However, in some cases, political, organizational, or operational factors make complete national coverage difficult to attain. Thus, in some rare situations, certain groups of schools and students may have to be excluded from the national target population. For example, it may be that a particular geographical region, educational sub-system, or language group cannot be covered. Such exclusion of schools and students from the target population is referred to as reduced population coverage.

Even countries with complete population coverage find it necessary to exclude at least some students from the target population because they attend very small schools, have intellectual or functional disabilities, or are non-native language speakers. Such students may be excluded at the school level (i.e., the whole school is excluded) or within the school on an individual basis.

School Level Exclusions. Although it is expected that very few schools will be excluded from the national target population, NRCs are permitted to exclude schools on the following grounds when they consider it necessary:

- Inaccessibility due to their geographically remote location
- Extremely small size (e.g., four or fewer students in the target grade)
- Offering a grade structure, or curriculum, radically different from the mainstream educational system
- Providing instruction solely to students in the student-level exclusion categories listed below (i.e., catering only to special needs students)

Student Level Exclusions. As in TIMSS, students with functional or intellectual disabilities as well as non-native language speakers within each school can be excluded. However, in specialized target populations such as in TIMSS Advanced, such exclusions are uncommon.

NRCs understand that exclusion rates must be kept to a minimum in order that national samples accurately represent the national target population.

- The overall number of excluded students must not account for more than 5% of the national target population of students in a country. The overall number includes both school-level and within-school exclusions.
- The number of students excluded because they attend very small schools must not account for more than 2% of the national target population of students.

Further details on the national coverage and exclusions for each country can be found in the Characteristics of National Samples appendix to Chapter 5: Sampling Implementation.

Requirements for Sampling the Target Population

TIMSS Advanced sets high standards for sampling precision, participation rates, and sample implementation in order to achieve national samples of the highest quality and survey estimates that are unbiased, accurate, and internationally comparable.

Sampling Precision and Sample Sizes

Because TIMSS Advanced is fundamentally a study of student achievement, the precision of estimates of student achievement is of primary importance. To meet the TIMSS Advanced standards for sampling precision, national student samples should provide for a standard error no greater than .035 standard deviation units for the country's mean achievement. With a standard deviation of 100 on the TIMSS Advanced achievement scales, this standard error corresponds to a 95% confidence interval of ±7 score points for the achievement mean and of ±10 score points for the difference between achievement means from successive cycles (e.g., the difference between a country's achievement mean on TIMSS Advanced 2008 and 2015). Sample estimates of any student-level percentage estimate (e.g., a student background characteristic) should have a confidence interval of ±3.5%.

With this in mind, and taking into account the clustered nature of the samples and the added uncertainty stemming from the imputation used in scaling the achievement data (see Chapter 4), the minimum sample sizes required 4,000 tested students for advanced mathematics and 4,000 for physics, selected from a minimum of 150 schools. These minima were fixed after looking at the sample sizes and precision achieved with the TIMSS Advanced 2008 results. As these were minima, most countries increased their sample sizes to account for non-response. For the Russian Federation, the sample size was increased because of the additional clustering effect due to sampling regions before sampling schools. The selected and achieved national school sample sizes are presented in Appendix 5A: National Characteristics.

Field Test Sample

Prior to the TIMSS Advanced 2015 data collection, an extensive field test is conducted in all participating countries. The goal of this field test is to check all instruments—particularly the achievement tests—and operational procedures under conditions similar to those of the data collection. The field test sample size requirement is 200 students per field test achievement booklet. The total field test sample size is a function of the number of assessment booklets being field tested. The TIMSS Advanced 2015 field test had four assessment booklets and so required a field test sample of 800 students for each subject.

Participation Rates

To minimize the potential for non-response bias, TIMSS Advanced aims for 100% participation by sampled schools, classrooms, and students, while recognizing that some degree of non-participation may be unavoidable. For a national sample to be fully acceptable it must have either:

- A minimum school participation rate of 85%, based on originally sampled schools AND
- A minimum classroom participation rate of 95%, from originally sampled schools and replacement schools AND
- A minimum student participation rate of 85%, from sampled schools and replacement schools

OR

- A minimum combined school, classroom, and student participation rate of 75%, based on originally sampled schools (although classroom and student participation rates may include replacement schools)

Classrooms with less than 50% student participation are deemed to be not participating.

Developing and Implementing the National Sampling Plan

Although National Research Coordinators are responsible for developing and implementing national sampling plans, Statistics Canada and the IEA DPC work closely with NRCs to help ensure that these sampling plans fully meet the standards set by the TIMSS & PIRLS International Study Center, while also adapting to national circumstances and requirements. National sampling plans must be based on the international two-stage sample design (schools as the first stage and classes within schools as the second stage) and must be approved by Statistics Canada.

TIMSS Advanced 2015 proposed a uniform sample design to all participants to ensure that differences in survey results were not attributable to the use of different sampling methodologies. This uniform sample design was flexible enough to accommodate the distinctiveness of national school systems at the upper secondary level and how the target populations were defined across participating countries. All sample designs were approved by Statistics Canada.

The TIMSS Advanced Sample Design

The basic TIMSS Advanced 2015 sample design consisted of two sampling stages: schools were sampled at the first stage, and one or more intact classes of students were sampled from a list of eligible classes within a selected school at the second stage.

First Sampling Stage. Two methods were used to sample schools in TIMSS Advanced 2015. In countries where the number of schools in the population greatly exceeded the number required in the sample, a systematic probability-proportional-to-size (PPS) sampling method was used.

This method, followed by the selection of classes within the selected schools in a second sampling stage, is often referred to as systematic two-stage PPS sampling and is described in most sampling textbooks (e.g., Cochran 1977, Lohr 1999). The PPS sampling approach was used in France, Italy, Portugal, the Russian Federation, and the United States. In other countries, where the number of schools to sample from was relatively small, schools were sampled with equal probabilities. This was the case in Norway, and Sweden. In Lebanon and Slovenia, all schools were selected.

Second Sampling Stage. In all but one country, classes within selected schools were sampled using a random systematic sampling method. The only exception was in the United States where students were grouped according to whether they were in advanced mathematics and/or physics target population(s) and sampled directly using a random systematic sampling approach.

National sample designs had to take into account the expected overlap across the advanced mathematics and physics populations. In some countries, students in a specific program belonged to both advanced mathematics and physics populations. In other countries, eligible students were found in two programs: students in one program belonged to both populations, while students from the other program belonged only to the advanced mathematics population. Finally, in a third set of countries, students were free to choose the courses they took and thus the degree of overlap between the two populations could not be predicted. Thus, two principal sample designs—a single school sample and separate school samples—were developed. While countries that participated in TIMSS Advanced 2015 adopted one of these two sample designs, some opted for slight modifications to account for particular national circumstances.

Stratification

Stratification consists of arranging the schools in the target population into groups, or strata, that share common characteristics such as geographic region or school type. Examples of stratification variables used in TIMSS Advanced include region of the country (e.g., states or provinces); school type or source of funding (e.g., public or private); and school performance on national examinations.

In TIMSS Advanced, stratification is used to:

- Separate schools according to the populations found in schools—schools with advanced mathematics only, physics only, or with both populations
- Improve the efficiency of the sample design, thereby making survey estimates more reliable
- Apply different sample designs, such as disproportionate sample allocations, to specific groups of schools (e.g., those in certain states or provinces)
- Ensure proportional representation of specific groups of schools in the sample

School stratification can take two forms: explicit and implicit. In explicit stratification, a separate school list or sampling frame is constructed for each stratum and a sample of schools is drawn from that stratum. For example, the sampling frame for Norway was divided into a total of five explicit strata based on the populations present and the size of the schools.

Implicit stratification consists of sorting the schools by one or more stratification variables within each explicit stratum, or within the entire sampling frame if explicit stratification is not used. The combined use of implicit strata and systematic sampling is a very effective and simple way of ensuring a proportional sample allocation of students across all implicit strata. Implicit stratification also can lead to improved reliability of achievement estimates, provided the implicit stratification variables are correlated with student achievement.

National Research Coordinators consulted with Statistics Canada and the IEA DPC to identify the stratification variables to be included in their sampling plans. The school sampling frame was sorted by the stratification variables prior to sampling schools so that adjacent schools were as similar as possible. Regardless of any other explicit or implicit variables that may be used, the school size was always included as an implicit stratification variable.

Exhibits 3.5 and 3.6 provide the list of explicit and implicit stratification variables implemented by the participating countries for advanced mathematics and physics. Further details on the explicit and implicit stratification variables for each country can be found in Appendix 5A: Characteristics of National Samples in Chapter 5: Sampling Implementation for TIMSS Advanced 2015.

Exhibit 3.5: TIMSS Advanced 2015 Advanced Mathematics Stratification Variables

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
France	School type (2) Success rate level (5)	8	None
Italy	School type (2) Region (5)	10	None
Lebanon	School type (2)	2	Region (7)
Norway	School size (5)	5	None
Portugal	School type (2) Region (7)	8	None
Russian Federation	Region (28) Presence of advanced mathematics streams (3)	77	Region (14) Location (9)
Russian Federation 6hr+	Presence of students from two advanced mathematics streams (3) Region (27)	48	Region (14) Location (9)
Slovenia	Presence of students from the two study populations (2) Percentage of mathematics experts in school	4	None
Sweden	Programs offered (3) School size (3)	9	School type (2)
United States	Presence of advanced program (2) School type (2) Census Region (4)	9	Urbanization (4) Ethnicity status (2)

Exhibit 3.6: TIMSS Advanced 2015 Physics Stratification Variables

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
France	School type (2) Success rate level (5)	8	None
Italy	Region (5)	5	None
Lebanon	School type (2)	2	Region (7)
Norway	School size (4)	4	None
Portugal	School type (2) Region (7)	8	None
Russian Federation	Region (all certainty, sampled) (28)	29	Region (14) Location (9)
Slovenia	Percentage of mathematics experts in school (3)	3	None
Sweden	Programs offered (3) School size (3)	9	School type (2)
United States	Presence of advanced program (2) School type (2) Census Region (4)	9	Urbanization (4) Ethnicity status (2)

Sample Design for Completely Overlapping Populations

This sample design was implemented in countries where there was complete overlap of both the advanced mathematics and physics populations and consisted of selecting a single sample of schools and one or more classes for both populations. Students in each sampled class were randomly assigned an advanced mathematics booklet or a physics booklet. Consequently, about half of the students received an advanced mathematics booklet while the other half received a physics booklet. France and Lebanon implemented this design.

Sample Design for Partially Overlapping Populations

This sample design was implemented in countries where students belonged to either, or even both, populations with no discernible pattern as students were free to choose which courses they would attend. In order to streamline the within-school operations and avoid testing students twice, this sample design consisted of selecting two separate school samples whenever possible. Both samples of schools were selected simultaneously to prevent overlap or were selected sequentially, while minimizing the overlap between both samples. In one school sample, only the advanced mathematics classes were listed for class sampling, and students in the sampled classes were assigned one of the six advanced mathematics booklets. In the other school sample, only physics classes were listed for class sampling, and students in the sampled classes were assigned one of the six physics booklets. Two separate samples were selected for Norway, Sweden, and the United States. In Norway and Sweden, both school samples were selected simultaneously while they were selected sequentially in the Russian Federation. In Italy, Portugal, Slovenia, and the United States, it was not possible to select separate school samples for each subject and the special adaptations made to the two principal sample designs for these countries are described below.

Special Adaptation for the Russian Federation

In the Russian Federation, a sample of regions was selected prior to the sampling of schools. Approximately half of the regions were sampled. Regions were selected with probability proportional to size, the largest regions being sampled with certainty. Thus, the sample of regions consisted of a group of certainty regions and a group of sampled regions. In a second stage, the school sample for the advanced mathematics population was selected.

From the group of certainty regions, all schools were grouped into three explicit strata, regardless of region, according to the type of students found in the schools: schools with the Intensive stream classes only, with the Profile stream classes only, or schools with classes from both streams. Regions were used as implicit strata within each explicit stratum. The sample of schools for advanced mathematics was selected among the three different strata.

For the group of sampled regions, the sampling of schools was done within each sampled region individually—regions being the primary sample units—and schools within each sampled region were split into the same three strata by the type of classes found in the schools, as was done in schools from the certainty regions. The sample of schools for advanced mathematics was selected among the three different strata from each sampled region.

Following the selection of the advanced mathematics sample, the next step was to select the physics school sample while minimizing the overlap with the advanced mathematics school sample. The overlap control was done using the technique described in Chowdhury, Chu, and Kaufman (2000)—more information on this technique can be found in Appendix 3B of *Methods and Procedures in TIMSS 2015*. Schools from the certainty regions were grouped to form one large stratum and implicitly sorted by region. A sample of schools for physics was selected from that stratum, minimizing the overlap with the advanced mathematics school sample.

For the group of sampled regions, the sampling of schools for physics was done within each sampled region individually, regions being the primary sample units. A sample of schools for physics was selected from each sampled region, minimizing the overlap with the advanced mathematics school sample.

Special Adaptation for Italy

In Italy, the structure of advanced mathematics and physics education and the sample size restrictions required a combination of the two established sample designs. Courses of interest were found in two types of schools: Technical Institutes with advanced mathematics classes only and Scientific Lyceum in which both subjects are compulsory. Schools were split in two groups according to their type and the sample selection was done simultaneously for both subjects. For the advanced mathematics assessment, the sample was composed of schools from both groups while for the physics assessment, the sample was drawn only from the Scientific Lyceum group. To reach the required sample size for each target population, all schools sampled from the Scientific Lyceum group were assigned to the physics assessment while only a sub-sample of these schools were randomly selected for the advanced mathematics assessment. In schools selected for both subjects, students in each sampled class were randomly assigned an advanced mathematics booklet or a physics booklet. Consequently, about half of the students received an advanced mathematics test booklet while the other half received a physics test booklet. In the other schools, students were only sampled for one subject, and as such only this subjects booklets were rotated within classes.

Special Adaptation for Portugal

In Portugal, school sample size restrictions and the structure of advanced mathematics and physics education also required a combination of the two established sample designs. Two groups of schools were identified based on the information provided on the sampling frame: schools with advanced mathematics students only and schools with a mixture of students (advanced mathematics students only and advanced mathematics and physics students). The sample selection was done simultaneously for both subjects. For the advanced mathematics assessment, the sample was composed of schools from both groups while for the physics assessment, the sample was drawn only from the latter group. All schools selected from the advanced mathematics and physics group were selected for both subjects.

In schools sampled for advanced mathematics, the regular approach of rotating the advanced mathematics booklets within the sampled classes was used. In schools selected for both subjects, two groups of classes were identified: one group of classes with advanced mathematics only students and one group of classes with advanced mathematics and physics students. In classes selected from the first class group, students were assessed only for advanced mathematics. In classes selected from the second class group, booklets from both subjects were randomly assigned to each student. To meet the sample size requirements for each subject and to preserve the proportion of students belonging to each group within school, one out of six students were randomly assigned an advanced mathematics booklet while the remaining students received a physics booklet.

The classification of schools to advanced mathematics only or to both subjects was done using a school sampling frame from a previous school year. During data collection, physics students were found in a number of schools assigned to the advanced mathematics only group. Also, some schools assigned to the advanced mathematics and physics assessments did not have any physics students. Therefore, all sampled schools were considered eligible for both populations regardless of their initial classification to one group or the other.

Special Adaptation for Slovenia

In Slovenia, the relatively small student populations made it impossible to meet the TIMSS Advanced 2015 student sample size requirements with either of the two standard sample designs. In particular, all physics students in the country had to be selected. Moreover, all schools were selected for both subjects given the small number of schools in the country.

In each school, the advanced mathematics classes and the physics classes were listed separately. A sample of classes was drawn from the list of advanced mathematics classes while all classes from the list of physics classes were selected. Since some students in the selected physics classes could have been sampled for advanced mathematics as well, some students were assessed for both subjects. The order in which the two assessments was administered was determined randomly in each school.

Special Adaptation for the United States

In the United States, the structure of advanced mathematics and physics education required a direct student sampling approach. Within sampled schools, students were assigned to one of three groups: the advanced mathematics group only, the physics group only, or the advanced mathematics and physics group. The advanced mathematics sample was composed of students sampled from the first and third group while the physics sample was composed of students sampled from the second and third group. Students selected from the advanced mathematics group were randomly assigned an advanced mathematics booklet. Students selected from the physics group were randomly assigned a physics booklet. Students selected from the advanced mathematics and physics group were randomly assigned an advanced mathematics booklet or a physics booklet. Consequently, about half of the students from this third group received an advanced mathematics booklet while the other half received a physics booklet.

Further details on the sample design for each country can be found in Appendix 5A: Characteristics of National Samples in Chapter 5: Sampling Implementation for TIMSS Advanced 2015.

Replacement Schools

Ideally, all schools sampled for TIMSS Advanced should participate in the assessments, and NRCs work hard to achieve this goal. Nevertheless, it is anticipated that a 100 percent participation rate may not be possible in all countries. To avoid sample size losses, the sampling plan identifies, a priori, specific replacement schools for each sampled school. Each originally sampled school has two pre-assigned replacement schools, usually the school immediately preceding the originally sampled school on the school sampling frame and the one immediately following it. Replacement schools always belong to the same explicit stratum as the original but may come from different implicit strata if the school they are replacing is either the first or last school of an implicit stratum.

The main justification for replacement schools in TIMSS Advanced is to ensure adequate sample sizes for analysis of subpopulation differences. Although the use of replacement schools does not eliminate the risk of bias due to school nonparticipation, employing implicit stratification and ordering the school sampling frame by school size increases the chances that a sampled school's replacements would have similar characteristics. This approach maintains the desired sample size while restricting replacement schools to strata where nonresponse occurs. Since the school frame is ordered by school size, replacement schools also tend to be similar in size to the school they are designated to replace.

NRCs understand that they should make every effort to secure the participation of all of the sampled schools. Only after all attempts to persuade a sampled school to participate have failed is the use of its replacement school considered.

This strategy was implemented in France, Italy, Portugal, the Russian Federation, and the United States. In Lebanon and Slovenia, there were no replacement schools, as all eligible schools were in the sample for both populations. In Norway and Sweden, since all schools were selected for the advanced mathematics sample or for the physics sample, there were no replacement schools available either.

Calculating Sampling Weights

National student samples in TIMSS Advanced are designed to accurately represent the target populations within a specified margin of sampling error, as described previously. After the data have been collected and processed, sample statistics such as means and percentages that describe student characteristics are computed as weighted estimates of the corresponding population parameters, where the weighting factor is the sampling weight. A student's sampling weight is essentially the inverse of the student's probability of selection, with appropriate adjustments for nonresponse.

The student sampling weight in TIMSS Advanced is a combination of weighting components reflecting selection probabilities and sampling outcomes at three levels—school, class, and student. At each level, the weighting component consists of a basic weight that is the inverse of the probability of selection at that level, together with an adjustment for nonparticipation. The overall sampling weight for each student is the product of the three weighting components: school, class (within school), and student (within class). For some countries, additional adjustments were required to account for additional sampling steps.

Note that sampling weights are calculated independently for each TIMSS Advanced population and within each explicit stratum. Thus a country will have only one set of sampling weights per target population (advanced mathematics and physics).

School Weighting Component

When schools are sampled with probability proportional to school size, the basic school weight for the i^{th} sampled school of population g (where g takes the value M for advanced mathematics and P for physics) is defined as:

$$BW_{sc}^{g,i} = \frac{MOS}{n \cdot MOS_i}$$

where n was the number of sampled schools in population g, mos_i was the measure of size for the i^{th} school, and

$$MOS = \sum_{i=1}^{N} mos_i$$

where N was the total number of schools in the explicit stratum of population g.

In France, Italy, Portugal, the Russian Federation, and the United States, the school selection probabilities were proportional to school size, generally defined as the number of students in the target population.

In Norway and Sweden, equal probability sampling of schools, rather than PPS, was carried out, meaning that every school had the same measure of size ($mos_i = 1$). Thus the school weight for the i^{th} sampled school in population g in these countries was calculated as:

$$BW_{sc}^{g,i} = \frac{N}{n}$$

In Lebanon a census of schools was taken resulting in a school weight equal to unity.

Special School Weight Factor for Italy

As was mentioned earlier, special weight factors or adjustments were calculated to account for additional sampling steps introduced during school sampling and arising from special adaptations to national sample designs in some countries.

In Italy, while all 120 schools sampled from stratum of schools with the advanced mathematics and physics program (scientific lyceum school stratum) were assigned to the physics population, 88 of them were randomly sub-sampled for the advanced mathematics population. Thus, an additional weight factor for the sub-sampled schools in advanced mathematics was computed as the inverse of the probability of a school sampled from this stratum being selected for advanced mathematics, and the original school weight was multiplied by this additional weight factor.

Special Weight Factors for the Russian Federation

The sample design for the Russian Federation included a preliminary sampling stage, in which regions were sampled. Thus, the school weight also incorporated the probability of selection in this preliminary stage, and the school weight for all schools from the Russian Federation was the product of the "region" weight and the school weight described earlier. This region weight was computed in a manner similar to the school weight, with regions having selection probabilities proportional to their size.

School Nonparticipation Adjustment

If a sampled school does not participate in TIMSS Advanced and its two designated replacement schools do not participate, it is necessary to adjust the basic school weight to compensate for the reduction in sample size. The school-level nonparticipation adjustment is calculated separately for each explicit stratum and each population g, as follows:

$$A_{sc}^{g} = \frac{n_s + n_{r1} + n_{r2} + n_{nr}}{n_s + n_{r1} + n_{r2}}$$

where n_s was the number of originally sampled schools that participated; n_{r1} and n_{r2} the number of first and second replacement schools, respectively, that participated; and n_{nr} the number of schools that did not participate and were not replaced. Sampled schools that are found to be ineligible[1] are not included in the calculation of this adjustment.

Combining the basic school weight and the school nonparticipation adjustment, the final school weighting component assigned to all students in the i^{th} school of population g ($g = M$ or P), corrected for non-participating schools, becomes:

$$FW_{sc}^{g,i} = A_{sc}^{g} \cdot BW_{sc}^{g,i}$$

It should be noted that, as well as being a crucial component of the overall student weight, the final school weighting component is a sampling weight in its own right, and can be used in analyses where the school is the analytic unit.

Class Weighting Component

The class weighting component reflects the class-within-school selection probability. After a school has been sampled and has agreed to participate in TIMSS Advanced, one or two classes are sampled with equal probability from the list of all eligible classes in the school for population g. Because larger schools have more classes from which to sample than smaller schools, the probability of class selection varies with school size, with students in small schools more likely to have their class selected than students in large schools. In countries where schools were sampled with probabilities proportional to school size, this relatively greater selection probability for students in small schools offsets the lower selection probability at the first stage as sampling results in higher selection probabilities for larger schools.

1 A sampled school is ineligible if it is found to contain no eligible students (i.e., no students in the target population). Such schools usually are in the sampling frame by mistake or are schools that do not offer advanced programs anymore.

The basic class-within-school weight for a sampled class is the inverse of the probability of the class being selected from all of the classes in its school. For the i^{th} school sampled for population g, let $C^{g,i}$ be the total number of eligible classes and $c^{g,i}$ the number of sampled classes. Using equal probability sampling, the basic class weight for all sampled classes in the i^{th} school for population g is:

$$BW_{cl}^{g,i,j} = \frac{C^{g,i}}{c^{g,i}}$$

For most TIMSS Advanced participants, $c^{g,i}$ takes the values 1 or 2. Some TIMSS Advanced participants sampled all eligible classes in a selected school, in which case $c^{g,i}$ is equal to $C^{g,i}$.

In Sweden, when appropriate, classes within the sampled schools were grouped by program prior to sampling and one class (or more) was randomly selected from each class group. In Portugal, classes were also grouped in two different class groups according to the student type (advanced mathematics only or advanced mathematics and physics). As a result, the basic class weight was computed separately for each class group within the sampled schools in these two countries. For the United States, since direct student sampling was performed, the class weight was set to unity.

Class Nonparticipation Adjustment

Basic class weights are calculated for all sampled classes in the sampled and replacement schools that participate in TIMSS. A class-level nonparticipation adjustment is applied to compensate for classes that do not participate or where the student participation rate is below 50 percent. Such sampled classes are assigned a weight of zero. Class nonparticipation adjustments are applied at the explicit stratum level rather than at the school level to minimize the risk of bias. The adjustment is calculated as follows:

$$A_{cl}^{g} = \frac{\sum_{i}^{s+r1+r2} 1}{\sum_{i}^{s+r1+r2} \delta^{g,i}/c^{g,i}}$$

where $c^{g,i}$ is the number of sampled classes in the i^{th} school for population g, as defined earlier, and $\delta^{g,i}$ gives the number of participating classes in the i^{th} school of population g within the explicit stratum.

Combining the basic class weight and the class nonparticipation adjustment, the final class weighting component, assigned to all sampled classes in the ith school of population g becomes:

$$FW_{cl}^{g,i,j} = A_{cl}^{g} \cdot BW_{cl}^{g,i,j}$$

Student Weighting Component

The student weighting component represents the student-within-class selection probability. The basic student weight is the inverse of the probability of a student in a sampled class being selected.

By design, all students within selected classes are selected for the TIMSS Advanced 2015 assessments. In most cases, they are all assigned a booklet from one subject—either advanced mathematics or physics. In countries with completely overlapping populations, however, roughly half of the students in a class are assigned a booklet in one subject and the other half in the other subject.

The basic student weight for the j^{th} class in the ith school of population g is calculated as follows:

$$BW_{st}^{g,i,j} = \frac{n^{i,j}}{n^{g,i,j}}$$

where the $n^{i,j}$ is the total number of students in the j^{th} class of the ith school and $n^{g,i,j}$ is the number of students in that class selected for population g ($g = M\ or\ P$).

When classes are sampled for only one population, then $n^{g,i,j}$ is equal to $n^{i,j}$ and the probability of a student in a selected class being sampled for that population is unity. When booklets from both populations are distributed among students from a selected class, this probability is approximately one half. In both cases, the student weight is calculated separately for each selected class and for each population.

As mentioned in the sample design section, direct student sampling was used for the United States. Students were assigned to one of three groups: the advanced mathematics group only, the physics group only, or the advanced mathematics and physics group. The student weight was computed as the inverse of the probability of a student being sampled for a specific subject from each group.

Student Nonparticipation Adjustment

The student nonparticipation adjustment for the j^{th} class in the ith school of population g is calculated as:

$$A_{st}^{g,i,j} = \frac{s_{rs}^{g,i,j} + s_{nr}^{g,i,j}}{s_{rs}^{g,i,j}}$$

where $s_{rs}^{g,i,j}$ is the number of participating students (i.e., students that participated in TIMSS Advanced and have assessment scores for their assigned population g) in the j^{th} class of the i^{th} school, and $s_{nr}^{g,i,j}$ is the number of students sampled in this class who were expected to have assessment scores for their assigned population g but did not participate in the assessment. This adjustment is calculated in the same manner, regardless whether a class was selected for a single population or for both populations.

The final student weighting component for students selected for population g in the j^{th} class of the i^{th} school is:

$$FW_{st}^{g,i,j} = A_{st}^{g,i,j} \cdot BW_{st}^{g,i,j}$$

Overall Student Sampling Weight

The overall student sampling weight is the product of the final weighting components for schools, classes, and students, as follows:

$$W^{g,i,j} = FW_{sc}^{g,i} \cdot FW_{cl}^{g,i,j} \cdot FW_{st}^{g,i,j}$$

Overall student sampling weights are only attributed to participating students, with non-participants weighted at 0. All student data reported in the TIMSS Advanced international reports are weighted by the overall student sampling weight, known as TOTWGT in the TIMSS Advanced international databases.

Participation Rates

Because nonparticipation can result in sample bias and misleading results, it is important that the schools, classes, and students that are sampled to participate in TIMSS Advanced actually take part in the assessments. To show the level of sampling participation in each country, TIMSS Advanced calculates both unweighted participation rates (i.e., based on simple counts of schools, classes, and students) and weighted participation rates based on the sampling weights described in the previous section. Unweighted participation rates provide a preliminary indicator that may be used to monitor progress in securing the participation of schools and classes, whereas weighted participation rates are the ultimate measure of sampling participation.

TIMSS Advanced reports weighted and unweighted participation rates for schools, classes, and students, as well as overall participation rates that are a combination of all three. To distinguish between participation based solely on originally sampled schools and participation that also relies on replacement schools, school and overall participation rates are computed separately for originally sampled schools only and for originally sampled together with replacement schools.

Unweighted School Participation Rates

The unweighted school participation rate is the ratio of the number of participating schools to the number of originally sampled schools, excluding any sampled schools found to be ineligible. A school is considered to be a "participating school" if at least one of its sampled classes has a student participation rate of at least 50 percent. The two unweighted school participation rates are calculated for each population as follows:

$R_{unw}^{g,sc-s} =$ unweighted school participation rate for originally sampled schools only

$R_{unw}^{g,sc-r} =$ unweighted school participation rate,including originally sampled and first and second replacement schools

$$R_{unw}^{g,sc-s} = \frac{n_s}{n_s + n_{r1} + n_{r2} + n_{nr}}$$

$$R_{unw}^{g,sc-r} = \frac{n_s + n_{r1} + n_{r2}}{n_s + n_{r1} + n_{r2} + n_{nr}}$$

Unweighted Class Participation Rates

The unweighted class participation rate is the ratio of the number of sampled classes that participated to the number of classes sampled, as follows:

$$R_{unw}^{g,cl} = \frac{\sum_{i}^{s+r1+r2} c_{*}^{g,i}}{\sum_{i}^{s+r1+r2} c^{g,i}}$$

where $c^{g,i}$ is the number of sampled classes in the i^{th} school, and $c_{*}^{g,i}$ is the number of participating classes in the i^{th} school of population g. Both summations are across all participating schools.

Unweighted Student Participation Rates

The unweighted student participation rate is the ratio of the number of selected students that participated in TIMSS Advanced to the total number of selected students that should have been assessed in the participating schools and classes. Classes where less than 50 percent of the students participate are considered to be not participating, and so students in such classes also are considered to be nonparticipants. The unweighted student participation rate for population g is computed as follows:

$$R_{unw}^{g,st} = \frac{\sum_{i}^{s+r1+r2} s_{rs}^{g,i,j}}{\sum_{i}^{s+r1+r2} (s_{rs}^{g,i,j} + s_{nr}^{g,i,j})}$$

Overall Unweighted Participation Rates

The overall unweighted participation rate is the product of the unweighted school, class, and student participation rates. Because TIMSS Advanced computes two versions of the unweighted school participation rate, one based on originally sampled schools only and the other including replacements as well as originally sampled schools, there also are two overall unweighted participation rates:

$R_{unw}^{g,ov-s}$ = unweighted overall participation rate for originally sampled schools only

$R_{unw}^{g,ov-r}$ = unweighted overall participation rate, including originally sampled and first and second replacement schools

$$R_{unw}^{g,ov-s} = R_{unw}^{g,sc-s} \cdot R_{unw}^{g,cl} \cdot R_{unw}^{g,st}$$

$$R_{unw}^{g,ov-r} = R_{unw}^{g,sc-r} \cdot R_{unw}^{g,cl} \cdot R_{unw}^{g,st}$$

Weighted School Participation Rates

The weighted school participation rate is the ratio of two estimates of the size of the target student population. The numerator is derived from the measure of size of those sampled schools that participated in TIMSS Advanced and the denominator is the weighted estimate of the total student enrollment in the population. Weighted school participation rates are computed for originally sampled schools and for originally sampled and replacement schools combined, as follows:

$R_{wtd}^{g,sc-s} =$ weighted school participation rate for originally sampled schools only

$R_{wtd}^{g,sc-r} =$ weighted school participation rate, including originally sampled and first and second replacement schools

$$R_{wtd}^{g,sc-s} = \frac{\sum_{i,j}^{s} BW_{sc}^{g,i} \cdot FW_{cl}^{g,i,j} \cdot FW_{st}^{g,i,j}}{\sum_{i,j}^{s+r1+r2} FW_{sc}^{g,i} \cdot FW_{cl}^{g,i,j} \cdot FW_{st}^{g,i,j}}$$

$$R_{wtd}^{g,sc-r} = \frac{\sum_{i,j}^{s+r1+r2} BW_{sc}^{g,i} \cdot FW_{cl}^{g,i,j} \cdot FW_{st}^{g,i,j}}{\sum_{i,j}^{s+r1+r2} FW_{sc}^{g,i} \cdot FW_{cl}^{g,i,j} \cdot FW_{st}^{g,i,j}}$$

Summations in both the numerator and denominator are over all responding students and include appropriate class and student sampling weights. Note that the basic school weight appears in the numerator, whereas the final school weight appears in the denominator.

Weighted Class Participation Rates

The weighted class participation rate for each population is computed as follows:

$$R_{wtd}^{g,cl} = \frac{\sum_{i,j}^{s+r1+r2} BW_{sc}^{g,i} \cdot BW_{cl}^{g,i,j} \cdot FW_{st}^{g,i,j}}{\sum_{i,j}^{s+r1+r2} BW_{sc}^{g,i} \cdot FW_{cl}^{g,i,j} \cdot FW_{st}^{g,i,j}}$$

where both the numerator and denominator are summations over all responding students from classes with at least 50 percent of their students participating in the study, and the appropriate student-level sampling weights are used. In this formula, the basic class weight appears in the numerator, whereas the final class weight appears in the denominator. And, the denominator in this formula is the same quantity that appears in the numerator of the weighted school participation rate for all schools, whether originally sampled or replacement.

Weighted Student Participation Rates

The weighted student participation rates for each population *g* is computed as follows:

$$R_{wtd}^{g,st} = \frac{\sum_{i,j}^{s+r1+r2} BW_{sc}^{g,i} \cdot BW_{cl}^{g,i,j} \cdot BW_{st}^{g,i,j}}{\sum_{i,j}^{s+r1+r2} BW_{sc}^{g,i} \cdot BW_{cl}^{g,i,j} \cdot FW_{st}^{g,i,j}}$$

where both the numerator and denominator are summations over all responding students from participating schools. In this formula, the basic student weight appears in the numerator, whereas the final student weight appears in the denominator. Also, the denominator in this formula is the same quantity that appears in the numerator of the weighted class participation rate for all participating schools, whether originally sampled or replacement.

Overall Weighted Participation Rates

The overall weighted participation rate is the product of the weighted school, class, and student participation rates. Because there are two versions of the weighted school participation rate, one based on originally sampled schools only and the other including replacement as well as originally sampled schools, there also are two overall weighted participation rates:

$R_{wtd}^{g,ov-s}$ = weighted overall participation rate from sampled schools only

$R_{wtd}^{g,ov-r}$ = weighted overall participation rate from sampled and replacement schools

$$R_{wtd}^{g,ov-s} = R_{wtd}^{g,sc-s} \cdot R_{wtd}^{g,cl} \cdot R_{wtd}^{g,st}$$

$$R_{wtd}^{g,ov-r} = R_{wtd}^{g,sc-r} \cdot R_{wtd}^{g,cl} \cdot R_{wtd}^{g,st}$$

Weighted school, class, student, and overall participation rates are computed for each TIMSS Advanced participant, for each population *g*, using these procedures.

References

Chowdhury, S., Chu, A., & Kaufman, S. (2000). Minimizing overlap in NCES surveys. *Proceedings of the Survey Methods Research Section, American Statistical Association*, 174-179. Retrieved from http://www.amstat.org/sections/srms/Proceedings/papers/2000_025.pdf

Cochran, W.G. (1977). *Sampling techniques*, New York: John Wiley.

Lohr, S. (1999). *Sampling: Design and Analysis*, Pacific Grove: Duxbury Press.

CHAPTER 4

Estimating Standard Errors in the TIMSS Advanced 2015 Results

Pierre Foy
Sylvie LaRoche

To obtain estimates of students' proficiency in advanced mathematics and physics that are both accurate and cost-effective, TIMSS Advanced 2015 made extensive use of probability sampling techniques to sample students from student populations in their final year of secondary schooling, and applied matrix-sampling assessment designs to target individual students with a subset of the complete pool of assessment items. This approach made efficient use of resources, in particular keeping student response burden to a minimum, but at a cost of some additional variance or uncertainty in the reported statistics, such as the means and percentages computed to estimate population parameters.

To quantify this uncertainty, each statistic in the *TIMSS Advanced 2015 International Results in Advanced Mathematics and Physics* report is accompanied by an estimate of its standard error. For statistics reporting student achievement, which are based on plausible values, standard errors have two components. The first reflects the uncertainty due to generalizing from student samples to the entire student populations, referred to as sampling variance, and the second reflects uncertainty due to inferring students' performance on the entire assessment from their performance on the subset of items that they took, known as imputation variance. For parameter estimates of variables that are not plausible values, standard errors are based entirely on sampling variance.

Estimating Sampling Variance

TIMSS Advanced makes extensive use of probability sampling to derive achievement results from national samples of students. Because many such samples are possible but only one sample is drawn, some uncertainty about how well the sample represents the population is to be expected. The uncertainty caused by sampling students from a target population, known as sampling variance, can be estimated from the data of the one sample drawn.

Whereas estimating the sampling variance from simple random samples is a relatively easy task, estimating the sampling variance from the complex sample design of TIMSS Advanced is a more challenging endeavor.

A common way to estimate the sampling variance in multi-stage cluster sampling designs is through resampling schemes such as the balanced repeated replication and Jackknife techniques

(Johnson & Rust, 1992; Wolter, 1985). TIMSS Advanced uses one variation of the Jackknife, the Jackknife Repeated Replication (JRR), to estimate sampling variances. JRR was chosen because it is computationally straightforward and provides approximately unbiased estimates of the sampling variances and sampling errors of means, totals, and percentages.

At the core of the JRR technique is the grouping of sampling units into zones based on sample design conditions (e.g., strata) and subsequent repeated draws of subsamples from these zones, i.e., repeated replication. For TIMSS Advanced, the two main features of the TIMSS Advanced sample design that JRR incorporates in its repeated draws of subsamples are the stratification of schools and the clustering of students within schools. This is done by defining Jackknife sampling zones according to the stratification scheme in each zone and by pairing successive schools[1] to model the clustering from each national sample (see Chapter 3 for information on the Sample Design). Since most national samples consist of 150 schools, a total of 75 zones are created. If more than 150 schools are selected, then the additional zones are collapsed into the first 75 zones. The subsampling required by JRR is applied within each sampling zone.

Sampling zones are constructed within explicit strata. When an explicit stratum has an odd number of schools, either by design or because of school non-response, the students in the remaining school are randomly divided to make up two "quasi" schools for the purposes of calculating jackknife standard errors.[2] Each sampling zone then consists of a pair of schools or "quasi" schools.

Exhibit 4.1 lists the number of sampling zones for each TIMSS Advanced 2015 participating country.

Exhibit 4.1: Number of Sampling Zones for Each TIMSS Advanced 2015 Participating Country

Country	TIMSS Advanced 2015 Sampling Zones	
	Advanced Mathematics	Physics
France	73	73
Italy	58	58
Lebanon	75	75
Norway	75	64
Portugal	75	75
Russian Federation	73	50
Russian Federation 6hr+[3]	73	–
Slovenia	75	58
Sweden	71	69
United States	75	75

A dash (-) indicates comparable data are not available.

1 When schools are sampled, schools are ordered within explicit strata by implicit stratification variables and the measure of size. Based on this sorting, successively sampled schools are matched and classified together in each sampling zone. More information can be found in Appendix 3A of Chapter 3.

2 If a remaining school consists of 2 sampled classrooms, each classroom becomes a "quasi" school.

3 For advanced mathematics, the Russian Federation participated in 2015 with an expanded population that included the more specialized students assessed in 1995 and 2008.

The JRR procedure draws two subsamples from each sampling zone: one where the first school in the pair is included and the second school is removed, and another subsample where the second school is included and the first school is removed. In both subsamples, all students in the other sampling zones are included. When a school is removed from the sample, the weights of the remaining school are doubled to make up for the omitted school. With this process applied in each of the 75 sampling zones, the JRR procedure yields a total of 150 replicate subsamples, each one with its own set of replicate sampling weights to account for the successive removal of each school from the pair of schools in any given sampling zone.[4]

The process of creating replicate sampling weights for the replicate subsamples defines replicate factors k_{hj} as follows:

$$k_{hj} = \begin{cases} 2 \text{ for students in school } j \text{ of sampling zone } h \\ 0 \text{ for students in the other school of sampling zone } h \\ 1 \text{ for students in any other sampling zone} \end{cases} \quad (1)$$

These replicate factors are used to compute the 150 sets of replicate sampling weights as follows:

$$W_{hji} = k_{hj} \cdot W_{0i} \quad (2)$$

where W_{0i} is the overall sampling weight of student i and W_{hji} is the resulting replicate sampling weight of student i from sampling zone h when school j is included and the other school in the pair is removed.

Exhibit 4.2 illustrates how the replicate factors, necessary to produce the replicate sampling weights, are derived. Within each sampling zone, each school is assigned randomly an indicator u_{hj}, coded either 0 or 1, such that one school has a value of 0 and the other a value of 1. This indicator serves to identify which schools within each zone will be successively included or removed. When a school is removed from a zone, the replicate factor is set to zero and the sampling weights of all students in that school are set to zero; when a school is included, the replicate factor is set to two and the sampling weights of all students in that school are doubled. The sampling weights of students in all other sampling zones remain unchanged.

4 Prior to 2015, TIMSS Advanced used 75 subsamples and sets of replicate weights to calculate the JRR sampling variances. To provide more accurate estimates, starting in 2015 TIMSS Advanced uses 150 subsamples and sets of replicate weights to calculate the JRR sampling variances. Two subsamples are drawn from each sampling zone rather than one randomly selected subsample.

Exhibit 4.2: Construction of Replicate Factors Across Sampling Zones

Sample Zone	School Replicate Indicator (u_{hj})	Replicate Factors for Computing JRR Replicate Sampling Weights (k_{hj})													
		Zone 1		Zone 2		Zone 3		…	Zone h		…	Zone 75	
		(1)	(2)	(3)	(4)	(5)	(6)		(2h–1)	(2h)		(149)	(150)
1	0	2	0	1	1	1	1	…	1	1	…	1	1
	1	0	2										
2	0	1	1	2	0	1	1	…	1	1	…	1	1
	1			0	2								
3	0	1	1	1	1	2	0	…	1	1	…	1	1
	1					0	2						
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋮	⋮
h	0	1	1	1	1	1	1	…	2	0	…	1	1
	1								0	2			
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮
75	0	1	1	1	1	1	1	…	1	1	…	2	0
	1											0	2

For example, sampling zone 1 yields two sets of replicate sampling weights. The first set has doubled sampling weights ($k_{11} = 2$) for the students in the first school ($u_{11} = 0$) of zone 1, zeroed sampling weights ($k_{12} = 0$) for the students in the second school ($u_{12} = 1$) of zone 1, and unchanged sampling weights ($k_{hj} = 1$) for all students in the other sampling zones. The second set of replicate sampling weights has zeroed sampling weights ($k_{11} = 0$) for the students in the first school ($u_{11} = 0$) of zone 1, doubled sampling weights ($k_{12} = 2$) for the students in the second school ($u_{12} = 1$) of zone 1, and unchanged sampling weights ($k_{hj} = 1$) for all students in the other sampling zones.

The process is repeated across all 75 possible sampling zones, generating 150 sets of replicate sampling weights. The replicate sampling weights are then used to estimate a statistic of interest 150 times. The variation across these 150 jackknife estimates determines the sampling variance.

Given a statistic t to be computed from a national sample, the formula used to estimate the sampling variance of that statistic, based on the TIMSS Advanced JRR algorithm, is given by the following equation:

$$Var_{jrr}(t_0) = \frac{1}{2}\sum_{h=1}^{75}\sum_{j=1}^{2}(t_{hj} - t_0)^2 \qquad (3)$$

where the term t_0 denotes the statistic of interest estimated with the overall student sampling weights W_{0i} and the term t_{hj} denotes the same statistic computed using the set of replicate sampling weights W_{hji} obtained from sampling zone h (h = 1, ... ,75), where the j^{th} school (1^{st} or 2^{nd}) in the zone is included and the other removed.

The sampling variance estimated with the TIMSS Advanced JRR method properly measures the variation arising from having sampled students using the multi-stage stratified cluster sample design. Its square root is the standard error for any statistic derived from variables other than plausible values. Examples of such statistics include the mean age of students, the mean scale score on the TIMSS Advanced *Students Like Learning Advanced Mathematics* contextual scale, and the percentage of students with at least one parent with a university degree.

Estimating Imputation Variance

For variables other than plausible values, standard errors were the result solely of sampling variation, and were computed using the JRR technique. However, the situation for plausible values was more complicated. As described in Chapter 4 of the *TIMSS Advanced 2015 Assessment Frameworks*, the TIMSS Advanced item pool was far too extensive to be administered in its entirety to any one student, and so a matrix-sampling assessment design was adopted whereby each student was given a single test booklet containing only a part of the entire assessment. The results for all of the booklets were then aggregated using item response theory to provide results for the entire assessment. Multiple imputation was used to derive reliable estimates of student performance (plausible values) on the assessment as a whole, even though each student responded to just a subset of the assessment items. Because every student proficiency estimate incorporates a random element, TIMSS Advanced 2015 followed the customary procedure of generating five estimates for each student and using the variability among them as a measure of the imputation uncertainty, or error.

The general procedure for estimating the imputation variance when analyzing student achievement data follows the basic principle of performing any statistical analysis five times—once for each set of plausible values—and aggregating the five sets of results (Mislevy et al., 1992). Thus, for any given achievement-based statistic t, estimating that statistic from each plausible value yields five estimates t_m, m = 1, ... , 5, all of them computed using the overall student sampling weights W_{0i}. The final estimate of that statistic, t_0, is the average of these five estimates:

$$t_0 = \frac{1}{5}\sum_{m=1}^{5} t_m \qquad (4)$$

The imputation variance of the statistic t_0 is simply the variance of the five results from the plausible values, computed as follows:

$$Var_{imp}(t_0) = \frac{6}{5}\sum_{m=1}^{5}\frac{(t_m - t_0)^2}{4} \qquad (5)$$

where the factor $\frac{6}{5}$ is a correction factor required by the multiple imputation methodology. This imputation variance is then added to the sampling variance to produce the total variance estimate of the statistic t_0, as follows:

$$Var_{tot}(t_0) = Var_{jrr}(t_0) + Var_{imp}(t_0) \tag{6}$$

The sampling variance in this context is the average of the sampling variances from the five plausible values, as follows:

$$Var_{jrr}(t_0) = \frac{1}{5}\sum_{m=1}^{5} Var_{jrr}(t_m) \tag{7}$$

where

$$Var_{jrr}(t_m) = \frac{1}{2}\sum_{h=1}^{75}\sum_{j=1}^{2}(t_{mhj} - t_m)^2 \tag{8}$$

and t_{mhj} is the appropriate JRR estimate based on plausible value m computed using the set of replicate sampling weights from sampling zone h where school j is included. The square root of the total variance is then the proper standard error for any statistic based on plausible values, such as the average TIMSS Advanced mathematics achievement for girls and the percentage of students who reach the TIMSS Advanced intermediate international benchmark of physics achievement.

Appendices 4A and 4B provide details on the jackknife sampling variance, the imputation variance, the total variance, and the standard error for each country's mean proficiency estimates in advanced mathematics and physics, respectively.

Estimating Standard Errors for International Averages

Some exhibits in the TIMSS Advanced 2015 reports include international averages and their standard errors. For example, Exhibit M4.1 reports the international average for the percentages of students in three categories of home educational resources and their advanced mathematics achievement. International averages are computed using the data from the participating countries included in the main table of an exhibit. Data from the benchmarking participants is not included in the estimation of international averages.

For any given statistic t_0, its international average is given by:

$$t_{int} = \frac{1}{N}\sum_{i=1}^{N} t_{0i} \tag{9}$$

where N is the number of countries contributing to the international average and t_{0i} is the estimate of our statistic of interest for the i^{th} country.

The variance of the international average t_{int} is given by:

$$Var(t_{int}) = \frac{1}{N^2} \sum_{i=1}^{N} Var_{tot}(t_{0i}) \tag{10}$$

where $Var_{tot}(t_{0i})$ is the total variance of our statistic of interest for the i^{th} country, as given in equation (6) above. For statistics based on plausible values, the total variance includes the sampling variance and the imputation variance. For statistics not based on plausible values, such as percentages, the total variance is based entirely on the sampling variance, as shown in equation (3) above. The standard error of the international average is the square root of the total variance.

References

Johnson, E.G., & Rust, K.F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics*, *17*(2), 175–190.

Mislevy, R.J., Beaton, A., Kaplan, B.A., & Sheehan, K. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*(2), 133–161.

Wolter, K.M. (1985). *Introduction to variance estimation*. New York: Springer-Verlag.

Appendix 4A: Summary Statistics and Standard Errors for Proficiency in Advanced Mathematics

Summary Statistics and Standard Errors for Proficiency in Overall Advanced Mathematics

Country	Sample Size	Overall Advanced Mathematics				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3967	462.664	7.014	2.338	9.352	3.058
Italy	3318	421.944	24.051	4.466	28.517	5.340
Lebanon	1161	531.874	7.817	1.575	9.392	3.065
Norway	2537	459.209	18.505	2.510	21.015	4.584
Portugal	4068	482.253	5.552	0.687	6.239	2.498
Russian Federation	7558	484.662	31.928	0.773	32.701	5.718
Russian Federation 6hr+[1]	3431	540.095	59.743	1.604	61.347	7.832
Slovenia	2922	459.794	10.915	0.635	11.550	3.398
Sweden	3937	431.082	13.203	3.057	16.260	4.032
United States	2954	484.984	25.501	1.261	26.762	5.173

Summary Statistics and Standard Errors for Proficiency in Algebra

Country	Sample Size	Algebra				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3967	469.180	7.722	0.935	8.657	2.942
Italy	3318	414.401	24.501	1.329	25.830	5.082
Lebanon	1161	525.424	9.492	6.391	15.882	3.985
Norway	2537	446.266	16.147	0.639	16.787	4.097
Portugal	4068	494.572	5.655	1.900	7.555	2.749
Russian Federation	7558	494.682	38.602	0.533	39.135	6.256
Russian Federation 6hr+	3431	556.293	72.822	7.897	80.719	8.984
Slovenia	2922	473.548	10.763	1.331	12.095	3.478
Sweden	3937	422.083	13.076	3.442	16.518	4.064
United States	2954	478.150	23.399	1.929	25.327	5.033

1 For advanced mathematics, the Russian Federation participated in 2015 with an expanded population that included the more specialized students assessed in 1995 and 2008.

Summary Statistics and Standard Errors for Proficiency in Calculus

Country	Sample Size	Calculus				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3967	465.938	7.852	2.286	10.138	3.184
Italy	3318	432.688	23.062	4.488	27.550	5.249
Lebanon	1161	543.753	7.425	7.841	15.266	3.907
Norway	2537	463.304	23.197	4.898	28.095	5.300
Portugal	4068	475.762	6.307	0.657	6.964	2.639
Russian Federation	7558	459.057	33.355	1.103	34.459	5.870
Russian Federation 6hr+	3431	513.019	59.430	4.189	63.619	7.976
Slovenia	2922	436.630	13.657	5.376	19.033	4.363
Sweden	3937	438.151	14.000	0.877	14.877	3.857
United States	2954	504.215	31.116	4.456	35.572	5.964

Summary Statistics and Standard Errors for Proficiency in Geometry

Country	Sample Size	Geometry				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3967	440.750	6.582	6.742	13.324	3.650
Italy	3318	413.154	25.306	7.307	32.614	5.711
Lebanon	1161	525.577	11.947	1.583	13.530	3.678
Norway	2537	472.777	16.784	4.583	21.367	4.622
Portugal	4068	463.763	7.048	3.213	10.261	3.203
Russian Federation	7558	499.940	32.884	1.136	34.020	5.833
Russian Federation 6hr+	3431	559.895	61.753	7.986	69.740	8.351
Slovenia	2922	455.854	14.264	1.495	15.759	3.970
Sweden	3937	429.982	11.180	2.227	13.407	3.662
United States	2954	454.953	27.669	4.330	32.000	5.657

Summary Statistics and Standard Errors for Proficiency in Advanced Mathematics Knowing

Country	Sample Size	Advanced Mathematics Knowing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3967	475.334	6.727	0.521	7.248	2.692
Italy	3318	422.828	23.893	6.503	30.396	5.513
Lebanon	1161	542.963	7.240	12.813	20.054	4.478
Norway	2537	445.282	15.335	1.593	16.927	4.114
Portugal	4068	479.314	4.978	3.929	8.907	2.985
Russian Federation	7558	477.578	41.694	2.790	44.484	6.670
Russian Federation 6hr+	3431	537.612	77.292	0.444	77.736	8.817
Slovenia	2922	466.074	12.012	0.503	12.514	3.538
Sweden	3937	404.981	16.326	5.474	21.800	4.669
United States	2954	487.789	30.660	1.869	32.529	5.703

Summary Statistics and Standard Errors for Advanced Mathematics Applying

Country	Sample Size	Advanced Mathematics Applying				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3967	448.880	8.200	3.488	11.688	3.419
Italy	3318	425.189	23.585	5.751	29.336	5.416
Lebanon	1161	529.261	8.395	6.287	14.682	3.832
Norway	2537	458.975	19.848	6.366	26.214	5.120
Portugal	4068	475.842	5.482	2.863	8.345	2.889
Russian Federation	7558	490.747	35.297	2.024	37.321	6.109
Russian Federation 6hr+	3431	543.948	61.129	4.556	65.685	8.105
Slovenia	2922	464.915	10.022	5.620	15.642	3.955
Sweden	3937	434.041	11.844	1.138	12.981	3.603
United States	2954	479.594	28.578	1.551	30.129	5.489

Summary Statistics and Standard Errors for Proficiency in Advanced Mathematics Reasoning

Country	Sample Size	Advanced Mathematics Reasoning				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3967	462.245	7.175	2.539	9.715	3.117
Italy	3318	411.144	26.722	8.630	35.352	5.946
Lebanon	1161	526.658	9.979	5.610	15.589	3.948
Norway	2537	468.619	17.332	2.028	19.361	4.400
Portugal	4068	487.964	6.896	5.330	12.226	3.497
Russian Federation	7558	484.022	26.849	0.964	27.813	5.274
Russian Federation 6hr+	3431	540.789	46.446	5.724	52.170	7.223
Slovenia	2922	442.450	13.550	2.173	15.723	3.965
Sweden	3937	446.651	11.392	3.524	14.916	3.862
United States	2954	484.454	24.881	3.218	28.099	5.301

Appendix 4B: Summary Statistics and Standard Errors for Proficiency in Physics

Summary Statistics and Standard Errors for Proficiency in Overall Physics

Country	Sample Size	Overall Physics				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3958	373.057	13.761	2.110	15.871	3.984
Italy	3424	373.925	43.465	4.287	47.753	6.910
Lebanon	1156	410.159	13.009	7.231	20.240	4.499
Norway	2472	507.262	19.351	1.523	20.874	4.569
Portugal	1783	466.609	19.187	2.186	21.373	4.623
Russian Federation	3822	507.534	48.865	0.923	49.788	7.056
Slovenia	1106	531.033	5.624	0.741	6.365	2.523
Sweden	3727	454.667	33.430	1.327	34.756	5.895
United States	2932	437.338	91.101	2.411	93.512	9.670

Summary Statistics and Standard Errors for Proficiency in Mechanics and Thermodynamics

Country	Sample Size	Mechanics and Thermodynamics				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3958	327.429	21.783	10.630	32.413	5.693
Italy	3424	375.967	40.310	0.553	40.863	6.392
Lebanon	1156	395.429	12.283	7.401	19.685	4.437
Norway	2472	502.581	14.897	1.834	16.731	4.090
Portugal	1783	489.030	19.817	3.639	23.456	4.843
Russian Federation	3822	514.150	44.622	0.524	45.146	6.719
Slovenia	1106	541.431	5.662	1.363	7.025	2.651
Sweden	3727	455.121	35.172	2.015	37.188	6.098
United States	2932	462.238	88.973	3.677	92.650	9.625

Summary Statistics and Standard Errors for Proficiency in Electricity and Magnetism

Country	Sample Size	Electricity and Magnetism				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3958	339.441	12.889	8.845	21.734	4.662
Italy	3424	425.180	40.728	2.913	43.642	6.606
Lebanon	1156	399.127	16.801	9.928	26.729	5.170
Norway	2472	514.368	25.508	4.344	29.852	5.464
Portugal	1783	431.308	21.707	12.508	34.215	5.849
Russian Federation	3822	515.395	58.447	6.236	64.684	8.043
Slovenia	1106	530.263	7.431	10.767	18.198	4.266
Sweden	3727	455.336	34.133	1.319	35.452	5.954
United States	2932	379.529	144.254	4.934	149.187	12.214

Summary Statistics and Standard Errors for Proficiency in Wave Phenomena and Atomic/Nuclear Physics

Country	Sample Size	Wave Phenomena and Atomic/Nuclear Physics				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3958	417.815	14.584	5.345	19.928	4.464
Italy	3424	329.071	54.069	8.577	62.646	7.915
Lebanon	1156	430.516	16.397	30.229	46.626	6.828
Norway	2472	507.447	25.546	1.899	27.445	5.239
Portugal	1783	455.513	20.954	16.939	37.893	6.156
Russian Federation	3822	490.105	54.418	1.300	55.718	7.464
Slovenia	1106	510.914	8.774	11.558	20.332	4.509
Sweden	3727	450.915	38.095	1.638	39.733	6.303
United States	2932	430.688	75.283	0.952	76.235	8.731

Summary Statistics and Standard Errors for Proficiency in Physics Knowing

Country	Sample Size	Physics Knowing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3958	374.787	14.150	0.928	15.078	3.883
Italy	3424	367.266	38.479	5.396	43.874	6.624
Lebanon	1156	378.083	17.288	4.842	22.130	4.704
Norway	2472	529.308	14.771	2.728	17.499	4.183
Portugal	1783	474.028	19.368	2.277	21.646	4.653
Russian Federation	3822	516.718	50.279	6.026	56.305	7.504
Slovenia	1106	521.009	8.776	8.695	17.472	4.180
Sweden	3727	451.645	33.491	2.964	36.455	6.038
United States	2932	444.055	88.814	6.639	95.453	9.770

Summary Statistics and Standard Errors for Proficiency in Physics Applying

Country	Sample Size	Physics Applying				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3958	358.066	14.742	16.803	31.545	5.617
Italy	3424	371.133	47.889	6.133	54.022	7.350
Lebanon	1156	432.517	14.148	15.471	29.618	5.442
Norway	2472	484.133	24.229	3.701	27.930	5.285
Portugal	1783	452.100	20.317	11.804	32.121	5.668
Russian Federation	3822	508.196	55.652	1.411	57.063	7.554
Slovenia	1106	543.490	6.340	8.035	14.374	3.791
Sweden	3727	454.326	36.800	3.648	40.448	6.360
United States	2932	420.403	100.159	3.241	103.400	10.169

Summary Statistics and Standard Errors for Proficiency in Physics Reasoning

Country	Sample Size	Physics Reasoning				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
France	3958	396.765	15.028	2.986	18.013	4.244
Italy	3424	374.825	50.115	2.628	52.743	7.262
Lebanon	1156	374.762	19.943	18.998	38.941	6.240
Norway	2472	518.935	19.645	12.832	32.477	5.699
Portugal	1783	480.645	14.479	0.525	15.004	3.873
Russian Federation	3822	493.011	42.047	2.365	44.412	6.664
Slovenia	1106	514.024	9.735	23.082	32.818	5.729
Sweden	3727	450.388	34.302	3.564	37.866	6.154
United States	2932	454.761	76.075	1.210	77.285	8.791

CHAPTER 5

Sample Implementation in TIMSS Advanced 2015

Sylvie LaRoche
Pierre Foy

Overview

Rigorous sampling of schools and students was a key component of the TIMSS Advanced 2015 project. Implementing the sampling plan was the responsibility of the National Research Coordinator (NRC) in each participating country. NRCs were supported in this endeavor by the TIMSS Advanced 2015 sampling consultants, Statistics Canada and the sampling unit of the IEA Data Processing and Research Center. Sampling consultants conducted the school sampling for most countries and trained NRCs using the Windows® Within-school Sampling Software (WinW3S) provided by the IEA DPC to implement within-school sampling. As an essential part of their sampling activities, NRCs were responsible for providing detailed documentation describing their national sampling plans (sampling data, school sampling frames, and school sample selections). The documentation for each TIMSS Advanced participant was reviewed and completed by the sampling consultants, including detailed information on coverage and exclusion levels, stratification variables, sample selection, participation rates, and variance estimates. The TIMSS & PIRLS International Study Center and the TIMSS Advanced 2015 Sampling Referee, Dr. Keith Rust of Westat, Inc., used this information to evaluate the quality of the samples.

This chapter gives a summary of the major characteristics of the national samples for TIMSS Advanced 2015. More detailed summaries of the sample design for each country, including details of population coverage and exclusions, stratification variables, and schools' sampling allocations, are provided in Appendix 5A: Characteristics of National Samples.

Target Population

As described in Chapter 3 (Sample Design), the TIMSS Advanced 2015 international target population for the advanced mathematics assessments was defined as students in the final year of secondary schooling who have taken courses in advanced mathematics, and for physics students in the final year of secondary schooling who have taken courses in physics.

National Coverage and Exclusions

All participating countries were able to provide full coverage of their defined target populations of advanced mathematics students and physics students. However, countries were allowed specific types of exclusions of schools and students that would have been either too difficult or too costly to assess. For example, very small or remote schools were sometimes excluded. Within some selected schools, students with special needs or students not fluent in the language of the test were sometimes excluded. Exhibits 5.1 and 5.2 summarize population coverage and exclusions for the TIMSS Advanced 2015 advanced mathematics and physics populations. For every participant, the overall percentage of excluded students (combining school-level and within-sample exclusions) was less than 5 percent. Some TIMSS Advanced 2015 participants had no within-school exclusions. Details on national exclusion categories are presented in Appendix 5A: Characteristics of National Samples.

Exhibit 5.1: Coverage of the TIMSS Advanced 2015 Target Population for Advanced Mathematics

Country	International Target Population Coverage	Exclusions from National Target Population		
		School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
France	100%	4.6%	0.1%	4.7%
Italy	100%	0.5%	0.7%	1.1%
Lebanon	100%	1.3%	0.0%	1.3%
Norway	100%	1.4%	0.0%	1.4%
Portugal	100%	0.0%	0.3%	0.3%
Russian Federation	100%	0.2%	0.1%	0.3%
Russian Federation 6hr+	100%	1.0%	0.1%	1.1%
Slovenia	100%	0.3%	2.2%	2.5%
Sweden	100%	1.6%	0.1%	1.7%
United States	100%	0.0%	0.1%	0.1%

Exhibit 5.2: Coverage of the TIMSS Advanced 2015 Target Population for Physics

Country	International Target Population Coverage	Exclusions from National Target Population		
		School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
France	100%	4.6%	0.1%	4.7%
Italy	100%	0.4%	0.4%	0.8%
Lebanon	100%	1.3%	0.0%	1.3%
Norway	100%	3.3%	0.1%	3.4%
Portugal	100%	0.4%	0.1%	0.6%
Russian Federation	100%	0.2%	0.2%	0.4%
Slovenia	100%	1.1%	0.9%	2.0%
Sweden	100%	1.9%	0.0%	2.0%
United States	100%	0.0%	0.1%	0.1%

Target Population Size

Exhibits 5.3 and 5.4 show the number of schools and students in each participant's target population[1] and sample, as well as an estimate of the student population size based on the sample data. The target population figures are derived from the sampling frame used to select the TIMSS Advanced 2015 samples, while the sample figures are based on the number of sampled schools and students that participated in the assessments. The sample figures were computed using sampling weights, which are explained in more detail in Chapter 3. The student population size based on the sampling frame did not take into account the portion of the population excluded within sampled schools and made no adjustment for changes in the population between the date when the information in the sampling frame was collected and the date of the TIMSS Advanced 2015 data collection—usually a 2-year interval. Nevertheless, a comparison of the two estimates of population size can be seen as a validity check on the sampling procedure. In most cases, the population size estimated from the sample closely matched the population size from the sampling frame.

The minimum school sample size required to meet the TIMSS Advanced sampling standards was 150 schools for each study population. Four out of nine countries met this requirement. The sample size in France was very close with 146 sampled schools. Italy was given permission to select 120 schools for each target populations. In Norway all schools were selected, each school being selected for only one study. Due to the relatively small number of eligible schools, the sample sizes were 136 and 130 for advanced mathematics and physics respectively. Sweden was given permission to select slightly smaller school sample sizes of 143 and 134 schools, for the advanced mathematics and physics target population respectively. In Slovenia, there were only 80 schools with advanced

1 After school-level exclusions.

mathematics students, of which 59 also had physics students; all 80 schools were sampled. Most countries sampled one or two classes per sampled school. Details on the national samples of schools and classes are provided in Appendix 5A: Characteristics of National Samples.

Exhibit 5.3: Population and Sample Sizes – TIMSS Advanced 2015 – Advanced Mathematics

Country	Population		Sample		
	Schools	Students	Schools	Students	Student Population Size Estimated From Population
France	2,106	162,106	144	3,967	172,309
Italy	1,820	149,637	113	3,318	142,350
Lebanon	1,635	62,121	251	1,161	4,457
Norway	266	6,903	133	2,537	6,752
Portugal	520	35,428	221	4,068	31,413
Russian Federation	5,534	141,903	346	7,558	138,733
Russian Federation 6hr+	672	26,134	163	3,431	25,855
Slovenia	80	7,138	69	2,922	6,888
Sweden	413	16,283	139	3,937	15,294
United States	33,411	3,816,235	241	2,954	473,872

Exhibit 5.4: Population and Sample Sizes – TIMSS Advanced 2015 – Physics

Country	Population		Sample		
	Schools	Students	Schools	Students	Student Population Size Estimated From Population
France	2,106	162,106	144	3,958	172,309
Italy	1,277	106,005	114	3,424	105,084
Lebanon	1,635	62,121	250	1,156	4,464
Norway	216	4,394	127	2,472	4,166
Portugal	283	6,084	149	1,783	5,669
Russian Federation	3,364	71,009	193	3,822	66,886
Slovenia	59	1,557	50	1,106	1,505
Sweden	407	15,387	133	3,727	15,430
United States	33,411	3,816,235	165	2,932	200,235

Meeting TIMSS Advanced 2015 Standards for Sampling Participation

TIMSS Advanced 2015 participants understood that the goal for sampling participation was 100 percent for all sampled schools, classrooms, and students. Guidelines for reporting achievement data for participants securing less than full participation were modeled after IEA's previous TIMSS Advanced assessment cycles. As summarized in Exhibit 5.5, countries were assigned to one of three categories on the basis of their sampling participation. Countries in Category 1 were considered to have met all TIMSS Advanced 2015 sampling requirements and to have acceptable participation rates. Countries in Category 2 met the participation requirements only after including replacement schools. Countries that failed to meet the participation requirements even with the use of replacement schools were assigned to Category 3. One of the main goals for data quality in TIMSS Advanced 2015 was to have as many countries as possible achieve Category 1 status.

Exhibit 5.5: Categories of Sampling Participation

Category 1	Acceptable sampling participation rate **without** the use of replacement schools. In order to be placed in this category, a country had to have: • An **unweighted** school response rate **without** replacement of at least 85% (after rounding to nearest whole percent) AND an **unweighted** student response rate (after rounding) of at least 85% OR • A **weighted** school response rate **without** replacement of at least 85% (after rounding to nearest whole percent) AND a **weighted** student response rate (after rounding) of at least 85% OR • The product of the (unrounded) **weighted** school response rate **without** replacement and the (unrounded) **weighted** student response rate of at least 75% (after rounding to the nearest whole percent). Countries in this category would appear in the tables and figures in international reports without annotation, and will be ordered by achievement as appropriate.
Category 2	Acceptable sampling participation rate **only when replacement schools are included**. A country would be placed in this category if: • It failed to meet the requirements for Category 1 but had a weighted school response rate **without** replacement of at least 50% (after rounding to the nearest percent) AND HAD EITHER • A **weighted** school response rate **with** replacement of at least 85% (after rounding to nearest whole percent) AND a **weighted** student response rate (after rounding) of at least 85% OR • The product of the (unrounded) **weighted** school response rate **with** replacement and the (unrounded) **weighted** student response rate of at least 75% (after rounding to the nearest whole percent). Countries in this category would be annotated with a "†" in the tables and figures in international reports, and ordered by achievement as appropriate.
Category 3	Unacceptable sampling response rate even when replacement schools are included. Countries that could provide documentation to show that they complied with TIMSS sampling procedures and requirements but did not meet the requirements for Category 1 or Category 2 would be placed in Category 3. Countries in this category would be annotated with a "‡" if they nearly met the requirements for Category 2. Countries would be annotated with a "ǂ" if they failed to meet the participation requirements but had a school participation rate of at least 50% before the use of replacement schools. At last, if none of these conditions are met, countries would appear in a separate section of the achievement tables, below the other countries, in international reports. These countries would be presented in alphabetical order.

Exhibits 5.6 through 5.9 present the school, classroom, student, and overall weighted and unweighted participation rates for each of the participants in the TIMSS Advanced 2015 study, for advanced mathematics and physics, respectively. All but three participants had excellent participation rates and belonged in Category 1. In advanced mathematics, Portugal met the minimum acceptable participation rate only after including replacement schools, and therefore their results were annotated with a dagger (†) in the achievement exhibits of the international reports (Category 2). Despite efforts to secure full participation, Lebanon and the United States

did not meet the required sampling participation rate for either subject, even with the use of replacement schools and were annotated with a triple-dagger (‡) in the achievement exhibits of the international reports (Category 3).

Exhibit 5.6: Participation Rates (Weighted) – TIMSS Advanced 2015 – Advanced Mathematics

	Country	School Participation		Class Participation	Student Participation	Overall Participation	
		Before Replacement	After Replacement			Before Replacement	After Replacement
	France	99%	99%	100%	96%	95%	95%
	Italy	88%	94%	99%	97%	85%	90%
‡	Lebanon	70%	70%	100%	98%	68%	68%
	Norway	100%	100%	100%	93%	93%	93%
†	Portugal	80%	87%	98%	93%	73%	80%
	Russian Federation	100%	100%	100%	98%	98%	98%
	Russian Federation 6hr+	100%	100%	100%	98%	98%	98%
	Slovenia	89%	89%	96%	87%	75%	75%
	Sweden	99%	99%	99%	90%	88%	88%
‡	United States	72%	76%	100%	87%	63%	66%

TIMSS Advanced guidelines for sampling participation: The minimum acceptable participation rates were 85% of both schools and students, or a combined rate (the product of school and student participation) of 75%. Participants not meeting these guidelines were annotated as follows:

† Met guidelines for sample participation rates only after replacement schools were included.

‡ Nearly satisfied guidelines for sample participation rates after replacement schools were included.

‡ Did not satisfy guidelines for sample participation rates.

Exhibit 5.7: Participation Rates (Weighted) – TIMSS Advanced 2015 – Physics

	Country	School Participation		Class Participation	Student Participation	Overall Participation	
		Before Replacement	After Replacement			Before Replacement	After Replacement
	France	99%	99%	100%	96%	95%	95%
	Italy	89%	95%	99%	97%	85%	91%
‡	Lebanon	70%	70%	100%	98%	68%	68%
	Norway	98%	98%	100%	94%	93%	93%
	Portugal	83%	87%	100%	93%	78%	81%
	Russian Federation	97%	100%	100%	98%	95%	98%
	Slovenia	86%	86%	100%	86%	74%	74%
	Sweden	99%	100%	99%	90%	88%	89%
‡	United States	65%	68%	100%	85%	55%	58%

TIMSS Advanced guidelines for sampling participation: The minimum acceptable participation rates were 85% of both schools and students, or a combined rate (the product of school and student participation) of 75%. Participants not meeting these guidelines were annotated as follows:

† Met guidelines for sample participation rates only after replacement schools were included.

‡ Nearly satisfied guidelines for sample participation rates after replacement schools were included.

‡ Did not satisfy guidelines for sample participation rates.

Exhibit 5.8: Participation Rates (Unweighted) – TIMSS Advanced 2015 – Advanced Mathematics

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
France	99%	99%	100%	93%	92%	92%
Italy	87%	94%	99%	95%	82%	89%
Lebanon	71%	71%	100%	95%	67%	67%
Norway	99%	99%	100%	93%	92%	92%
Portugal	82%	88%	98%	91%	73%	79%
Russian Federation	100%	100%	100%	98%	98%	98%
Russian Federation 6hr+	100%	100%	100%	97%	97%	97%
Slovenia	90%	90%	95%	88%	75%	75%
Sweden	99%	99%	99%	90%	88%	88%
United States	73%	76%	100%	86%	63%	66%

Exhibit 5.9: Participation Rates (Unweighted) – TIMSS Advanced 2015 – Physics

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
France	99%	99%	100%	93%	92%	92%
Italy	88%	95%	99%	95%	83%	89%
Lebanon	70%	70%	100%	95%	67%	67%
Norway	98%	98%	100%	94%	92%	92%
Portugal	82%	86%	100%	92%	75%	79%
Russian Federation	97%	100%	100%	98%	95%	98%
Slovenia	85%	85%	100%	86%	73%	73%
Sweden	99%	99%	100%	89%	88%	88%
United States	66%	70%	100%	86%	56%	60%

Exhibits 5.10 through 5.13 show the achieved sample sizes in terms of schools and students for each of the participants in the TIMSS Advanced 2015, for advanced mathematics and physics, respectively.

Exhibit 5.10: School Sample Sizes – TIMSS Advanced 2015 – Advanced Mathematics

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
France	146	145	144	0	144
Italy	120	120	104	9	113
Lebanon	355	354	251	0	251
Norway	136	134	133	0	133
Portugal	251	251	206	15	221
Russian Federation	346	346	346	0	346
Russian Federation 6hr+	181	163	163	0	163
Slovenia	80	77	69	0	69
Sweden	143	141	139	0	139
United States	348	316	230	11	241

Exhibit 5.11: School Sample Sizes – TIMSS Advanced 2015 – Physics

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
France	146	145	144	0	144
Italy	120	120	106	8	114
Lebanon	356	355	250	0	250
Norway	130	130	127	0	127
Portugal	251	173	142	7	149
Russian Federation	193	193	187	6	193
Slovenia	59	59	50	0	50
Sweden	134	134	132	1	133
United States	348	237	156	9	165

Exhibit 5.12: Student Sample Sizes – TIMSS Advanced 2015 – Advanced Mathematics

Country	Within-School Student Participation (Weighted Percentage)	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/School	Number of Students Excluded	Number of Eligible Students	Number of Students Absent	Number of Students Assessed
France	96%	4,310	41	7	4,262	295	3,967
Italy	97%	3,547	28	30	3,489	171	3,318
Lebanon	98%	1,222	0	0	1,222	61	1,161
Norway	93%	2,756	31	1	2,724	187	2,537
Portugal	93%	4,581	109	15	4,457	389	4,068
Russian Federation	98%	7,758	2	12	7,744	186	7,558
Russian Federation 6hr+	98%	3,530	0	3	3,527	96	3,431
Slovenia	87%	3,360	1	42	3,317	395	2,922
Sweden	90%	4,450	85	2	4,363	426	3,937
United States	87%	3,488	57	2	3,429	475	2,954

Students attending a sampled class at the time the sample was chosen but leaving the class before the assessment was administered were classified as "withdrawn."
Students with a disability or language barrier that prevented them from participating in the assessment were classified as "excluded."
Students not present when the assessment was administered, and not subsequently assessed in a make-up session, were classified as "absent."

Exhibit 5.13: Student Sample Sizes – TIMSS Advanced 2015 – Physics

Country	Within-School Student Participation (Weighted Percentage)	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/School	Number of Students Excluded	Number of Eligible Students	Number of Students Absent	Number of Students Assessed
France	96%	4,297	41	7	4,249	291	3,958
Italy	97%	3,652	25	20	3,607	183	3,424
Lebanon	98%	1,215	0	0	1,215	59	1,156
Norway	94%	2,674	44	2	2,628	156	2,472
Portugal	93%	1,968	21	4	1,943	160	1,783
Russian Federation	98%	3,925	2	8	3,915	93	3,822
Slovenia	86%	1,302	6	12	1,284	178	1,106
Sweden	90%	4,236	65	3	4,168	441	3,727
United States	85%	3,539	114	6	3,419	487	2,932

Students attending a sampled class at the time the sample was chosen but leaving the class before the assessment was administered were classified as "withdrawn."
Students with a disability or language barrier that prevented them from participating in the assessment were classified as "excluded."
Students not present when the assessment was administered, and not subsequently assessed in a make-up session, were classified as "absent."

TIMSS Advanced 2015 Trends in Student Populations

Because an important goal of the TIMSS Advanced 2015 assessment was to measure changes in students' advanced mathematics and physics achievement across assessment cycles, it was important to track any changes over time in population composition and coverage that might be related to student achievement. Exhibits 5.14 and 5.15 present, for each country, trends across the TIMSS Advanced cycles (2015, 2008, and 1995) in five important characteristics of the assessment populations: number of years of formal schooling, average student age, percent of students in the national target population excluded from the assessment, TIMSS Advanced coverage indices, and overall participation rates after using replacements. Most countries were very similar with regard to these characteristics across the three assessment cycles, although there have been changes in some countries in the number of years of schooling, the TIMSS Advanced coverage indices of the assessed populations, and in the exclusion rate.

Eight of the countries that participated in TIMSS Advanced 2015 also participated in the assessments of advanced mathematics and physics students in their final year of schooling in 1995 or 2008. The Russian Federation, Slovenia, and Sweden participated in all three TIMSS Advanced assessments, in 1995, 2008, and 2015. In 1995, Italy participated only in advanced mathematics while Norway participated only in physics. France and the United States participated in both assessments in 1995 and 2015. For advanced mathematics, the Russian Federation trend results are available only for the Intensive stream students (6hr+).

The Russian Federation and Slovenia have undergone structural changes in the age at which children enter schools that are reflected in their samples. With regards to the number of years of schooling, the shift from ten years to eleven years of schooling in the Russian Federation was the result of a change in the structure of the education system from 1995 to 2015. Half of the students were still under the older system in 2008. Slovenia has completed a transition toward having all children begin school at an earlier age. For this reason, in 2015 Slovenian students had thirteen years of schooling in their final year of general and technical gymnasia, whereas the 2008 students had twelve years of schooling in their final year.

In 1995, exclusion rates for Advanced Mathematics and Physics were computed based on exclusion rates among all students in the final year of schooling. In the case of the Russian Federation, the figure presented in the 1995 International Report (43.0%) greatly overstated the level of exclusions in the advanced mathematics and physics populations.

The 1995 Advanced Mathematics Coverage Index for Italy was recomputed and is different than from the percentage reported in the 1995 International Report. The 1995 sample for the United States was adjusted to correspond with the course-taking definitions used in 2015, and the 1995 results were recomputed for trend purposes.

Exhibit 5.14: Trends in Student Populations – TIMSS Advanced 2015 – Advanced Mathematics

Country	Years of Formal Schooling*			Average Age at Time of Testing			Overall Exclusion Rates**			Advanced Mathematics Coverage Index			Overall Participation Rates		
	2015	2008	1995	2015	2008	1995	2015	2008	1995	2015	2008	1995	2015	2008	1995
France	12		12	18.0		18.2	4.7%		1.0%	21.5%		19.9%	95%		77%
Italy	13	13	13	18.9	19.0	19.1	1.1%	0.5%	3.8%	24.5%	19.7%	14.1%	90%	95%	68%
Lebanon	12	12		17.8	17.9		1.3%	1.3%		3.9%	5.9%		68%	83%	
Norway	13	13		18.7	18.8		1.4%	1.0%		10.6%	10.9%		93%	83%	
Russian Federation 6hr+	11	10/11	10	17.7	17.0	16.9	1.1%	0.0%	2.0%	1.9%	1.4%	2.%	98%	98%	96%
Slovenia	13	12	12	18.8	18.8	18.9	2.5%	1.3%	6.0%	34.4%	40.5%	75.4%	75%	81%	42%
Sweden	12	12	12	18.8	18.8	18.9	1.7%	1.7%	0.2%	14.1%	12.8%	16.2%	88%	84%	89%
United States	12		12	18.1		18.0	0.1%		3.7%	11.4%		6.4%	66%		71%

* Represents years of schooling counting from the first year of ISCED Level 1.

** In 1995 exclusion rates for Advanced Mathematics were computed based on exclusion rates among all students in the final year of schooling. In the case of the Russian Federation, the figure presented in the 1995 International Report (43.0%) greatly overestimates the level of exclusions in the advanced mathematics population. The figure presented above (2.0%) includes two regions, North Ossetia and Chechen Republic, as well as non-Russian speaking students.

Russian Federation trend results are available only for the Intensive stream students (6hr+). The United States adjusted the 1995 sample to correspond with the course-taking definitions used in 2015, and the 1995 results were recomputed.

An empty cell indicates a country did not participate in that year's assessment.

Exhibit 5.15: Trends in Student Populations – TIMSS Advanced 2015 – Physics

Country	Years of Formal Schooling*			Average Age at Time of Testing			Overall Exclusion Rates**			Physics Coverage Index			Overall Participation Rates		
	2015	2008	1995	2015	2008	1995	2015	2008	1995	2015	2008	1995	2015	2008	1995
France	12		12	18.0		18.2	4.7%		1.0%	21.5%		19.9%	95%		77%
Italy	13	13		18.9	18.9		0.8%	0.9%		18.2%	3.8%		91%	97%	
Lebanon	12	12		17.8	17.9		1.3%	1.3%		3.9%	5.9%		68%	82%	
Norway	13	12	12	18.8	18.8	19.0	3.4%	0.5%	3.8%	6.5%	6.8%	8.4%	93%	73%	83%
Russian Federation	11	10/11	10	17.7	17.1	16.9	0.4%	0.0%	2.0%	4.9%	2.6%	1.5%	98%	97%	95%
Slovenia	13	12	12	18.8	18.7	18.8	2.0%	0.5%	6.0%	7.6%	7.5%	38.6%	74%	67%	43%
Sweden	12	12	12	18.8	18.8	18.9	2.0%	2.3%	0.2%	14.3%	11.0%	16.3%	89%	89%	89%
United States	12		12	18.1		18.0	0.1%		3.7%	4.8%		2.7%	58%		64%

* Represents years of schooling counting from the first year of ISCED Level 1.

** In 1995 exclusion rates for Physics were computed based on exclusion rates among all students in the final year of schooling. In the case of the Russian Federation, the figure presented in the 1995 International Report (43.0%) greatly overestimates the level of exclusions in the advanced mathematics population. The figure presented above (2.0%) includes two regions, North Ossetia and Chechen Republic, as well as non-Russian speaking students.

The United States adjusted the 1995 sample to correspond with the course-taking definitions used in 2015, and the 1995 results were recomputed.

An empty cell indicates a country did not participate in that year's assessment.

Appendix 5A: Characteristics of National Samples

France

A single school sample was used for both advanced mathematics and physics.

Coverage and Exclusions

- Coverage of the national desired target population was 100 percent
- School-level exclusions consisted of overseas territories and private schools without contract
- Within-school exclusions consisted of students with intellectual disabilities and students with functional disabilities

Sample Design

- Explicit stratification by school type (Lycée general et technologique—upper secondary schools with general and technologic streams, Lycée polyvalent—the remaining upper secondary schools) and success rate in the scientific baccalaureate during the 2012 session (3 or 5 success rate levels depending on the school type)
- No implicit stratification
- Schools sampled using probability proportional to (school) size systematic sampling
- Two classes in the selected schools were sampled whenever possible
- Half of the students in the selected classes were randomly assigned an advanced mathematics booklet, and the other half were assigned a physics booklet

Field Test Sample

- 32 schools were sampled for the field test at the same time as the data collection sample, thus no schools were selected for both activities

Allocation of Advanced Mathematics and Physics School Sample in France

			Participating Schools				
Explicit Stratum	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacement	2nd Replacement	Refusal Schools	Excluded Schools
Upper secondary schools with general and technologic streams, with success rate level I	22	0	22	0	0	0	0
Upper secondary schools with general and technologic streams, with success rate level II	22	0	22	0	0	0	0
Upper secondary schools with general and technologic streams, with success rate level III	22	0	22	0	0	0	0
Upper secondary schools with general and technologic streams, with success rate level IV	22	0	22	0	0	0	0
Upper secondary schools with general and technologic streams, with success rate level V	22	0	21	0	0	1	0
Other upper secondary schools, with success rate level I	12	1	11	0	0	0	0
Other upper secondary schools, with success rate level II	12	0	12	0	0	0	0
Other upper secondary schools, with success rate level III	12	0	12	0	0	0	0
Total	**146**	**1**	**144**	**0**	**0**	**1**	**0**

Italy

The sample design for Italy consisted of a mix of the two main designs. Some schools were sampled for advanced mathematics only, some were sampled for physics only, and some were sampled for both subjects.

Coverage and Exclusions

- Coverage of the national desired target population was 100 percent
- School-level exclusions consisted of very small schools (less than 5 eligible students), Slovenian language schools, and German language schools
- Within-school exclusions consisted of students with functional disabilities

Sample Design

- Explicit stratification by school type (technical institute, scientific lyceum) and region (center, southern peninsulas and islands, northeast, northwest, south)
- No implicit stratification
- Courses of interest were found in two types of schools: technical institutes with advanced mathematics classes only and scientific lyceum in which both subjects are compulsory
- A total of 32 schools were sampled from the technical-institute strata for the advanced mathematics sample. From the scientific-lyceum strata, a total of 120 schools were first sampled, from which a subsample of 88 schools was randomly selected for advanced mathematics and physics. The other 32 schools were sampled only for the physics assessment.
- In schools with advanced mathematics classes only, one class was sampled per school, and two classes were sampled in larger schools of 80 or more students
- In schools selected only for the physics target population, one class was sampled per school, and two classes were sampled in larger schools (80 or more students)
- In schools selected for both advanced mathematics and physics target populations, three classes were sampled whenever possible and advanced mathematics and physics booklets were rotated within sampled classes

Field Test Sample

- 34 schools were sampled for the field test. The data collection sample was selected after the field test sample, without controlling for overlap.

Allocation of Advanced Mathematics School Sample in Italy

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacement	Participating Schools: 2nd Replacement	Refusal Schools	Excluded Schools
Technical institute with advanced mathematics program only - Center	6	0	4	2	0	0	0
Technical institute with advanced mathematics program only - Southern Peninsulas and Islands	6	0	5	0	0	1	0
Technical institute with advanced mathematics program only - Northeast	6	0	5	1	0	0	0
Technical institute with advanced mathematics program only - Northwest	8	0	8	0	0	0	0
Technical institute with advanced mathematics program only - South	6	0	6	0	0	0	0
Scientific lyceum with advanced mathematics and physics program - Center	18	0	17	1	0	0	0
Scientific lyceum with advanced mathematics and physics program - Southern Peninsulas and Islands	16	0	13	2	0	1	0
Scientific lyceum with advanced mathematics and physics program - Northeast	14	0	14	0	0	0	0
Scientific lyceum with advanced mathematics and physics program - Northwest	18	0	15	0	1	2	0
Scientific lyceum with advanced mathematics and physics program - South	22	0	17	1	1	3	0
Total	**120**	**0**	**104**	**7**	**2**	**7**	**0**

Allocation of Physics School Sample in Italy

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacement	Participating Schools: 2nd Replacement	Refusal Schools	Excluded Schools
Scientific lyceum with advanced mathematics and physics program - Center	22	0	21	1	0	0	0
Scientific lyceum with advanced mathematics and physics program - Southern Peninsulas and Islands	20	0	17	2	0	1	0
Scientific lyceum with advanced mathematics and physics program - Northeast	20	0	20	0	0	0	0
Scientific lyceum with advanced mathematics and physics program - Northwest	26	0	23	0	1	2	0
Scientific lyceum with advanced mathematics and physics program - South	32	0	25	2	2	3	0
Total	**120**	**0**	**106**	**5**	**3**	**6**	**0**

Lebanon

A single school sample was selected for both advanced mathematics and physics.

Coverage and Exclusions

- Coverage of the national desired target population was 100 percent
- School-level exclusions consisted of very small schools (less than 8 eligible students)
- No within-school exclusions

Sample Design

- Explicit stratification by school type (private, public)
- Implicit stratification by region (7)
- All schools with eligible students were selected for the data collection
- All eligible classes in a school were expected to be sampled. However, a problem during data collection resulted in only one class being sampled per school, leading to a reduced sample size for both target populations.
- Half of the students in the selected classes were randomly assigned an advanced mathematics booklet, and the remaining half were assigned a physics booklet
- In the public school stratum, classes were used as variance estimation strata and half classes were used to build jackknife replicates

Field Test Sample

- A sample of 48 schools was selected for the field test and used for both populations. All schools selected for the field test were also included in the data collection sample.

Allocation of Advanced Mathematics and Physics School Sample in Lebanon

			Participating Schools				
Explicit Stratum	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacement	2nd Replacement	Refusal Schools	Excluded Schools
Private	211	1	121	0	0	89	0
Public	144	0	130	0	0	14	0
Total	**355**	**1**	**251**	**0**	**0**	**103**	**0**

Norway

All eligible schools in Norway were selected for TIMSS Advanced 2015, but each school was selected for only one population, resulting in two separate school samples for advanced mathematics and physics.

Coverage and Exclusions

- Coverage of the national desired target population was 100 percent
- School-level exclusions consisted of very small schools (less than 6 eligible students)
- Within-school exclusions consisted of students with intellectual disabilities

Sample Design

- Explicit stratification by classes taught for each subject (advanced mathematics classes only, both advanced mathematics and physics classes) and the number of eligible physics students (between 6 and 9, between 10 and 19, between 20 and 34, 35 or more). Schools with less than 6 physics students were sampled for advanced mathematics only.
- No implicit stratification
- All schools from the advanced mathematics explicit stratum were selected
- From the remaining advanced mathematics and physics strata, all schools were randomly sampled either for advanced mathematics or for physics through a disproportional allocation procedure. This procedure ensured that sample size requirements for each study population were reached by sampling a higher proportion of schools for physics from the strata of schools with larger numbers of physics students and sampling a higher proportion of schools for advanced mathematics in the strata of schools with less physics students.
- In schools selected for advanced mathematics, all eligible advanced mathematics classes were sampled
- In schools selected for physics, all eligible physics classes were sampled
- In the advanced mathematics school stratum, classes were used as variance estimation strata and half classes were used to build jackknife replicates

Field Test Sample

- A convenience sample of 21 international baccalaureate schools was selected for the field test

Allocation of Advanced Mathematics School Sample in Norway

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacement	Participating Schools: 2nd Replacement	Refusal Schools	Excluded Schools
Schools with advanced mathematics classes	50	2	48	0	0	0	0
Schools with advanced mathematics and physics classes and 35 or more physics students	12	0	12	0	0	0	0
Schools with advanced mathematics and physics classes and 20 to 34 physics students	14	0	14	0	0	0	0
Schools with advanced mathematics and physics classes and 10 to 19 physics students	36	0	36	0	0	0	0
Schools with advanced mathematics and physics classes and less than 10 physics students	24	0	23	0	0	1	0
Total	**136**	**2**	**133**	**0**	**0**	**1**	**0**

Allocation of Physics School Sample in Norway

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacement	Participating Schools: 2nd Replacement	Refusal Schools	Excluded Schools
Schools with advanced mathematics and physics classes and 35 or more physics students	24	0	24	0	0	0	0
Schools with advanced mathematics and physics classes and 20 to 34 physics students	32	0	31	0	0	1	0
Schools with advanced mathematics and physics classes and 10 to 19 physics students	58	0	58	0	0	0	0
Schools with advanced mathematics and physics classes and less than 10 physics students	16	0	14	0	0	2	0
Total	**130**	**0**	**127**	**0**	**0**	**3**	**0**

Portugal

The sample design for Portugal differed from the standard TIMSS Advanced design. It consisted of a mix of the two primary TIMSS Advanced designs described in Chapter 3. Some schools were sampled for advanced mathematics only, some for physics only, and some for both subjects.

Coverage and Exclusions

- Coverage of the national desired target population was 100 percent
- For advanced mathematics, there was no school-level exclusions. For physics, school-level exclusions consisted of very small schools (less than 5 eligible students).
- Within-school exclusions consisted of students with intellectual disabilities and students with functional disabilities

Sample Design

- Explicit stratification by the presence of eligible students from the two study populations based on the sample frame information (schools with only advanced mathematics students, schools with advanced mathematics and physics students), school type (private, public), and region within the public schools strata (7)
- No implicit stratification
- Schools were sampled using probability proportional to school size systematic sampling
- Some sampled schools were composed of classes of advanced mathematics students only. Other sampled schools were composed of some classes that had advanced mathematics students only and other classes that had both advanced mathematics and physics students.
- Classes within the sampled schools were split into 2 class groups where appropriate: classes with students eligible for advanced mathematics only and classes with students eligible for both subjects
- For advanced mathematics, classes were sampled from both class groups
- From the advanced mathematics only class group, one class was sampled and advanced mathematics booklets were distributed
- From the advanced mathematics and physics class group, all classes were expected to be selected. However, a problem during data collection resulted in only one class being selected per school. In each selected class, one out of every six students was randomly assigned an advanced mathematics booklet and the remaining students were assigned a physics booklet.
- During data collection, physics students were found in a number of schools assigned to the advanced mathematics only strata. Also, some schools assigned to the advanced

mathematics and physics strata did not have any physics students. Therefore, all sampled schools were considered eligible for both populations regardless of their classification to one group or the other.

- Out of the 73 schools originally sampled as advanced mathematics only schools, 14 schools ended up having physics students. Out of the 178 schools selected for advanced mathematics and physics, 25 schools ended up having only advanced mathematics students.
- For physics, adjacent advanced mathematics only strata were collapsed to create the jackknife zones and replicates

Field Test Sample

- For the field test, a sample of 40 schools was selected for advanced mathematics and a sample of 24 schools was selected for physics. Only one subject was tested in the sampled schools and 2 classes were sampled whenever possible.

Allocation of Advanced Mathematics School Sample in Portugal

			Participating Schools				
Explicit Stratum	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacement	2nd Replacement	Refusal Schools	Excluded Schools
Private schools with advanced mathematics program - All regions	2	0	2	0	0	0	0
Public schools with advanced mathematics classes - Alentejo	18	0	15	0	0	3	0
Public schools with advanced mathematics classes - Algarve	2	0	2	0	0	0	0
Public schools with advanced mathematics classes - Centro	18	0	18	0	0	0	0
Public schools with advanced mathematics classes - Lisboa	8	0	6	0	0	2	0
Public schools with advanced mathematics classes - Norte	21	0	15	4	0	2	0
Public schools with advanced mathematics classes - R. A. Açores	2	0	0	2	0	0	0
Public schools with advanced mathematics classes - R. A. Madeira	2	0	1	1	0	0	0

Allocation of Advanced Mathematics School Sample in Portugal (Continued)

			Participating Schools				
Explicit Stratum	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacement	2nd Replacement	Refusal Schools	Excluded Schools
Private schools with advanced mathematics classes and advanced mathematics and physics classes - All regions	10	0	10	0	0	0	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - Alentejo	13	0	11	0	0	2	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - Algarve	6	0	5	1	0	0	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - Centro	48	0	42	2	0	4	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - Lisboa	32	0	24	1	0	7	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - Norte	57	0	45	3	0	9	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - R. A. Açores	6	0	5	1	0	0	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - R. A. Madeira	6	0	5	0	0	1	0
Total	**251**	**0**	**206**	**15**	**0**	**30**	**0**

Allocation of Physics School Sample in Portugal

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacement	Participating Schools: 2nd Replacement	Refusal Schools	Excluded Schools
Private schools with advanced mathematics program - All regions	2	1	1	0	0	0	0
Public schools with advanced mathematics classes - Alentejo	18	11	5	0	0	2	0
Public schools with advanced mathematics classes - Algarve	2	1	1	0	0	0	0
Public schools with advanced mathematics classes - Centro	18	14	4	0	0	0	0
Public schools with advanced mathematics classes - Lisboa	8	6	0	0	0	2	0
Public schools with advanced mathematics classes - Norte	21	16	1	2	0	2	0
Public schools with advanced mathematics classes - R. A. Açores	2	2	0	0	0	0	0
Public schools with advanced mathematics classes - R. A. Madeira	2	2	0	0	0	0	0
Private schools with advanced mathematics classes and advanced mathematics and physics classes - All regions	10	2	8	0	0	0	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - Alentejo	13	3	8	0	0	2	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - Algarve	6	1	4	0	0	1	0

Allocation of Physics School Sample in Portugal (Continued)

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacement	Participating Schools: 2nd Replacement	Refusal Schools	Excluded Schools
Public schools with advanced mathematics classes and advanced mathematics and physics classes - Centro	48	8	37	1	0	2	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - Lisboa	32	1	26	0	0	5	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - Norte	57	10	37	3	0	7	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - R. A. Açores	6	0	5	1	0	0	0
Public schools with advanced mathematics classes and advanced mathematics and physics classes - R. A. Madeira	6	0	5	0	0	1	0
Total	**251**	**78**	**142**	**7**	**0**	**24**	**0**

Russian Federation

Two separate school samples were selected for advanced mathematics and physics.

For advanced mathematics, the Russian Federation assessed students from two programs: the Profile and Intensive streams. The summary below describes the samples from both programs.

Advanced Mathematics results were published separately for the students in the Intensive stream. Sampling characteristics for this group of students are described in the Russian Federation 6hr+ section on page 5.29.

Coverage and Exclusions

- Coverage of the national desired target population was 100 percent
- School-level exclusions consisted of very small schools (less than 4 eligible students)
- Within-school exclusions consisted of students with intellectual disabilities and students with functional disabilities

Sample Design

- In a preliminary sampling stage, a sample of 42 regions out of 83 was selected with probabilities proportional to school size. The 14 largest regions were selected with certainty.
- A school sample for advanced mathematics was selected first from the certainty regions and the sampled regions. More details on the sample design for advanced mathematics are given below.
- Following the selection of the advanced mathematics sample, a school sample for physics was selected from the same certainty and sampled regions, minimizing the overlap with the selected advanced mathematics sample using the Chowdhury overlap control method. Additional details on the sample design for physics are provided below.
- School weights were adjusted to take into account the sampling of regions
- Within the strata of certainty regions, schools were paired for variance calculation purposes. Otherwise, selected regions were paired for variance purposes.

Additional details for advanced mathematics

- Schools from the 14 certainty regions were grouped together and were explicitly stratified by the presence of classes from the two advanced mathematics streams: schools with classes from the Intensive stream only, schools with classes from the Profile stream only, and schools with classes from both streams. They also were stratified implicitly by the 14 regions and 9 levels of urbanization.

- Each of the remaining 28 sampled regions became explicit strata and were further stratified by the presence of classes from the two advanced mathematics streams, as was the group of certainty regions
- Classes within schools were split into two class groups: Intensive advanced mathematics classes and Profile advanced mathematics classes, and one class was sampled from each group

Additional details for physics

- The 14 certainty regions were grouped into one large stratum and schools were stratified implicitly by region and 9 levels of urbanization
- Each of the remaining 28 sampled regions became explicit strata
- One eligible physics class was sampled from each selected school

Field Test Sample

- A convenience sample of 45 schools was selected for the field test. Of these, 29 schools were selected for physics only, and 16 schools were selected for advanced mathematics and physics.

Allocation of Advanced Mathematics School Sample in Russian Federation

			Participating Schools				
Explicit Stratum	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacement	2nd Replacement	Refusal Schools	Excluded Schools
Schools with Intensive stream classes in certainty regions	54	0	54	0	0	0	0
Schools with Intensive stream classes in sampled regions	89	0	89	0	0	0	0
Schools with Profile stream classes in certainty regions	56	0	56	0	0	0	0
Schools with Profile stream classes in sampled regions	119	0	119	0	0	0	0
Schools with Intensive and Profile streams classes in certainty regions	7	0	7	0	0	0	0
Schools with Intensive and Profile streams classes in sampled regions	21	0	21	0	0	0	0
Total	**346**	**0**	**346**	**0**	**0**	**0**	**0**

Allocation of Physics School Sample in Russian Federation

			Participating Schools				
Explicit Stratum	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacement	2nd Replacement	Refusal Schools	Excluded Schools
Schools in certainty regions	70	0	70	0	0	0	0
Schools in sampled regions	123	0	117	6	0	0	0
Total	**193**	**0**	**187**	**6**	**0**	**0**	**0**

Russian Federation 6hr+

For advanced mathematics, the Russian Federation assessed students from two programs: the Profile and Intensive streams. Results are provided separately for the students in the Intensive stream since this group corresponds to the group of students assessed in TIMSS Advanced 1995 and 2008. The following summary is for the Intensive stream, designated as Russian Federation 6hr+.

Coverage and Exclusions

- Coverage of the national desired target population was 100 percent
- School-level exclusions consisted of very small schools (less than 4 eligible students)
- Within-school exclusions consisted of students with intellectual disabilities and students with functional disabilities

Sample Design

- In a preliminary sampling stage, a sample of 42 regions out of 83 was selected with probabilities proportional to (school) size. The 14 largest regions were selected with certainty.
- Schools from the 14 certainty regions were grouped together and were explicitly stratified by the presence of classes from the two advanced mathematics streams: schools with classes from the Intensive stream only, schools with classes from both the Profile and Intensive streams. They also were stratified implicitly by the 14 regions and 9 levels of urbanization.
- Each of the remaining 28 sampled regions became explicit strata and were further stratified by the presence of classes from the two advanced mathematics streams
- One class sampled among the Intensive stream classes within each school
- School weights were adjusted to take into account the sampling of regions
- Within the strata of certainty regions, schools were paired for variance calculation purposes. Otherwise, selected regions were paired for variance purposes.

Field Test Sample

- A convenience sample of 34 schools was selected for the field test. Of these, 18 schools were selected for advanced mathematics only and 16 schools were selected for advanced mathematics and physics.

Allocation of Advanced Mathematics School Sample in Russian Federation 6hr+

			Participating Schools				
Explicit Stratum	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacement	2nd Replacement	Refusal Schools	Excluded Schools
Schools with Intensive stream classes in certainty regions	54	1	53	0	0	0	0
Schools with Intensive stream classes in sampled regions	89	11	78	0	0	0	0
Schools with Intensive and Profile streams classes in certainty regions	7	1	6	0	0	0	0
Schools with Intensive and Profile streams classes in sampled regions	21	5	16	0	0	0	0
Schools with Profile stream classes in certainty regions	6	0	6	0	0	0	0
Schools with Profile stream classes in sampled regions	4	0	4	0	0	0	0
Total	**181**	**18**	**163**	**0**	**0**	**0**	**0**

Slovenia

All schools in Slovenia with eligible students were selected for participation in TIMSS Advanced 2015. A total of 80 schools had eligible advanced mathematics students and 59 of these schools had eligible physics students. All eligible physics students took part in the physics assessment.

Since all 59 physics schools also were eligible for the advanced mathematics assessment, some students were selected for both assessments. The two assessments were scheduled on different days to accommodate this and a random mechanism determined which assessment was administered first in each school.

Coverage and Exclusions

- Coverage of the national desired target population was 100 percent
- School-level exclusions consisted of Italian schools and Waldorf schools. For physics, very small schools with less than 7 physics students were also excluded.
- Within-school exclusions consisted of students with intellectual disabilities, and students with functional disabilities

Sample Design

- All schools with eligible students were selected
- Explicit stratification by presence of students from the two study populations (advanced mathematics classes, advanced mathematics classes and physics classes) and the expected number of mathematics "experts" in the final year of secondary school (few, some, many). Mathematics experts were defined as students enrolled in the intensive mathematics classes, and the number of expected experts in each school was estimated based on data from the prior academic year. Schools with "many experts" were expected to have 35 percent or more of the students in the intensive classes; schools with "some experts" were expected to have between 20 and 35 percent of the students in these classes; and schools with "few experts" were those expected to have less than 20 percent of the students in these classes.
- No implicit stratification
- For advanced mathematics, eligible advanced mathematics classes were grouped by type (regular or intense mathematics classes) and one or two eligible classes from each group were sampled
- For physics, all eligible physics classes were sampled
- In strata where at least ninety percent of all schools participated, variance estimation strata and replicates were calculated in one of two ways: (1) when all classes were selected within schools, classes were used as variance strata and half-classes as replicates or (2) when classes were sampled within schools, schools were used as variance strata and classes as replicates. In all other strata, schools were paired within explicit strata for variance calculation purposes.

Field Test Sample

- A sample of 29 schools with advanced mathematics and physics classes was selected and used for both populations. Schools participating in the field test also were part of the data collection sample.

Allocation of Advanced Mathematics School Sample in Slovenia

			Participating Schools				
Explicit Stratum	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacement	2nd Replacement	Refusal Schools	Excluded Schools
Schools with advanced mathematics classes	18	0	17	0	0	1	3
Schools with advanced mathematics classes and physics classes, with few mathematics experts	20	0	17	0	0	3	0
Schools with advanced mathematics classes and physics classes, with some mathematics experts	18	0	16	0	0	2	0
Schools with advanced mathematics classes and physics classes, with many mathematics experts	21	0	19	0	0	2	0
Total	**77**	**0**	**69**	**0**	**0**	**8**	**3**

Allocation of Physics School Sample in Slovenia

			Participating Schools				
Explicit Stratum	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacement	2nd Replacement	Refusal Schools	Excluded Schools
Schools with advanced mathematics classes and physics classes, with few mathematics experts	20	0	15	0	0	5	0
Schools with advanced mathematics classes and physics classes, with some mathematics experts	18	0	16	0	0	2	0
Schools with advanced mathematics classes and physics classes, with many mathematics experts	21	0	19	0	0	2	0
Total	**59**	**0**	**50**	**0**	**0**	**9**	**0**

Sweden

Two separate school samples were selected for advanced mathematics and physics.

Coverage and Exclusions

- Coverage of the national desired target population was 100 percent
- School-level exclusions consisted of very small schools (less than 7 eligible students)
- Within-school exclusions consisted of students with intellectual disabilities and students with functional disabilities

Sample Design

- Eligible schools were initially ordered by the presence of programs offered (natural science, technological, both), by school funding (public, private), and by size. Schools were then split into two homogeneous partitions, each partition being representative of both target populations. The advanced mathematics sample was selected from one partition while the physics sample was selected from the other partition, resulting in two separate school samples for advanced mathematics and physics.
- Explicit stratification by the programs offered in school (natural science, technological, both) and by size (small, medium, large), in both the advanced mathematics partition and physics partition
- Implicit stratification by school type (public, private) in both partitions
- From each partition, a sample of schools was selected with equal probabilities, minimizing the overlap with the field test sample using the Chowdhury overlap control method
- For advanced mathematics, classes within the sampled schools were grouped by program and one or two eligible advanced mathematics classes were sampled from each program
- For physics, classes within the sampled schools were grouped by program and one or two eligible physics classes were sampled from each program

Field Test Sample

- A sample of 48 schools was selected for both populations

Allocation of Advanced Mathematics School Sample in Sweden

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacement	Participating Schools: 2nd Replacement	Refusal Schools	Excluded Schools
Larger schools with natural science and technical programs	28	0	28	0	0	0	0
Medium-size schools with natural science and technical programs	21	0	21	0	0	0	0
Smaller schools with natural science and technical programs	6	0	6	0	0	0	0
Larger schools with natural science program	30	1	28	0	0	1	0
Medium-size schools with natural science program	23	0	22	0	0	1	0
Smaller schools with natural science program	12	0	12	0	0	0	0
Larger schools with technical program	4	0	4	0	0	0	0
Medium-size schools with technical program	11	0	11	0	0	0	0
Smaller schools with technical program	8	1	7	0	0	0	0
Total	**143**	**2**	**139**	**0**	**0**	**2**	**0**

Allocation of Physics School Sample in Sweden

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacement	Participating Schools: 2nd Replacement	Refusal Schools	Excluded Schools
Larger schools with natural science and technical programs	20	0	20	0	0	0	0
Medium-size schools with natural science and technical programs	21	0	21	0	0	0	0
Smaller schools with natural science and technical programs	10	0	10	0	0	0	0
Larger schools with natural science program	31	0	31	0	0	0	0
Medium-size schools with natural science program	16	0	15	1	0	0	0
Smaller schools with natural science program	14	0	13	0	0	1	0
Larger schools with technical program	7	0	7	0	0	0	0
Medium-size schools with technical program	6	0	6	0	0	0	0
Smaller schools with technical program	9	0	9	0	0	0	0
Total	**134**	**0**	**132**	**1**	**0**	**1**	**0**

United States

A single school sample was used for both advanced mathematics and physics.

Coverage and Exclusions

- Coverage of the national desired target population was 100 percent
- No school-level exclusions
- Within-school exclusions consisted of students with intellectual disabilities and students with functional disabilities

Sample Design

- Explicit stratification by the presence of advanced program in school (yes, no), school type (public, private) and census region (4) within public schools
- Implicit stratification by urbanization (4) and ethnicity status (above 15% non-White students in a school, below 15% non-White students in a school)
- The structure of advanced mathematics and physics education required direct student sampling. Within sampled schools, students were assigned to one of three groups: advanced mathematics only, physics only, or advanced mathematics and physics. The advanced mathematics sample was composed of students sampled from the first and third group while the physics sample was composed of students sampled from the second and third group.
- Students selected from the advanced mathematics group were randomly assigned an advanced mathematics booklet
- Students selected from the physics group were randomly assigned a physics booklet
- Students selected from the advanced mathematics and physics group were randomly assigned an advanced mathematics booklet or a physics booklet. Consequently, about half of the students from this group received an advanced mathematics booklet and the other half received a physics booklet. During data collection, 32 advanced mathematics schools and 111 physics schools were found to be ineligible for assessment administration as they did not have eligible advanced mathematics and/or physics students.
- During data collection, 32 advanced mathematics schools and 111 physics schools were found to be ineligible for assessment administration as they did not have eligible advanced mathematics and/or physics students

Field Test Sample

- A sample of 72 schools from 7 states was selected and used for both populations

Allocation of Advanced Mathematics School Sample in the United States

			Participating Schools				
Explicit Stratum	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacement	2nd Replacement	Refusal Schools	Excluded Schools
Private schools with no advanced program	10	8	0	0	0	2	0
Public schools with no advanced program in census regions 1 and 2	14	5	7	0	0	2	0
Public schools with no advanced program in census region 3	10	5	4	0	0	1	0
Public schools with no advanced program in census region 4	10	7	2	0	0	1	0
Private schools with an advanced program	19	0	12	1	0	6	0
Public schools with an advanced program in census region 1	49	2	28	0	0	19	0
Public schools with an advanced program in census region 2	58	1	41	5	0	11	0
Public schools with an advanced program in census region 3	104	0	88	3	0	13	0
Public schools with an advanced program in census region 4	74	4	48	2	0	20	0
Total	**348**	**32**	**230**	**11**	**0**	**75**	**0**

Allocation of Physics School Sample in the United States

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	1st Replacement	2nd Replacement	Refusal Schools	Excluded Schools
Private schools with no advanced program	10	8	0	0	0	2	0
Public schools with no advanced program in census regions 1 and 2	14	11	2	0	0	1	0
Public schools with no advanced program in census region 3	10	8	1	0	0	1	0
Public schools with no advanced program in census region 4	10	9	0	0	0	1	0
Private schools with an advanced program	19	5	9	0	0	5	0
Public schools with an advanced program in census region 1	49	9	20	0	0	20	0
Public schools with an advanced program in census region 2	58	14	26	4	0	14	0
Public schools with an advanced program in census region 3	104	28	64	3	0	9	0
Public schools with an advanced program in census region 4	74	19	34	2	0	19	0
Total	**348**	**111**	**156**	**9**	**0**	**72**	**0**

CHAPTER 6

Survey Operations Procedures in TIMSS Advanced 2015

Ieva Johansone

Overview

As data-based indicators of countries' student achievement profiles and learning contexts, TIMSS and TIMSS Advanced assessments are crucially dependent on the quality of the data collected by each participant. Whereas the development of the assessments is an intensely collaborative process involving all of the partners in the enterprise, the process of administering the assessments and collecting the data is uniquely the responsibility of each individual country.

To ensure the consistency and uniformity of approach necessary for high-quality, internationally comparable data, all participants are expected to follow a set of standardized operations procedures. These procedures have been developed over successive cycles of TIMSS and TIMSS Advanced through a partnership involving the TIMSS & PIRLS International Study Center, the IEA Data Processing and Research Center (IEA DPC), the IEA Secretariat, Statistics Canada, and National Research Coordinators (NRCs) of participating countries. With each new assessment cycle, the operations procedures are updated to enhance efficiency and accuracy and reduce burden, making use of developments in information technology to automate routine activities wherever possible.

In each country, the National Research Coordinator was responsible for the implementation of TIMSS Advanced 2015. Internationally, National Research Coordinators provided the country's perspective in all international discussions, represented the country at international meetings, and were the responsible contact persons for all project activities. Locally, National Research Coordinators were responsible for implementing all the internationally agreed-upon procedures and facilitating all of the national decisions regarding TIMSS Advanced, including any adaptations for the national context.

The daily tasks of the NRCs varied over the course of the TIMSS Advanced 2015 cycle. In the initial phases, National Research Coordinators participated in the TIMSS Advanced 2015 framework review and assessment development process (see Developing the TIMSS Advanced 2015 Achievement Items), and collaborated with Statistics Canada to develop a plan to implement the TIMSS Advanced 2015 sampling design in their countries (see Sample Implementation).

Following the development of the draft achievement items and context questionnaires, all countries conducted a full-scale field test of all instruments and operational procedures in March-April 2014 in preparation for the TIMSS Advanced 2015 data collection, which took place in March-May 2015. The field test allowed the National Research Coordinators and their staff to become acquainted with the operational activities, and the feedback they provided was used to improve the procedures for the data collection. As expected, the field test resulted in some enhancements to survey operations procedures and most definitely contributed to ensuring the successful execution of TIMSS Advanced 2015.

As part of ongoing efforts to improve operations, the TIMSS Advanced National Research Coordinators were asked to complete a Survey Activities Questionnaire (SAQ), which sought feedback on all aspects of their experience conducting TIMSS Advanced 2015. The feedback solicited in the SAQ included an evaluation of the quality of the assessment materials and the effectiveness of the operations procedures and documentation. The results of the TIMSS Advanced 2015 Survey Activities Questionnaire are presented in the final section of this chapter.

TIMSS Advanced 2015 Survey Operations Units, Manuals, and Software

To support the National Research Coordinators in conducting TIMSS Advanced 2015, the TIMSS & PIRLS International Study Center provided step-by-step documentation of all operational activities. Organized into a series of units, the *Survey Operations Procedures* were made available at critical junctures of the project to ensure that NRCs had all the tools and information necessary to discharge their responsibilities.

The *Procedures Units* were accompanied by a series of manuals for use by School Coordinators and Test Administrators that National Research Coordinators could translate and adapt to their local situations. Consistent with the goal of automating and streamlining procedures wherever possible, the IEA DPC provided NRCs with a range of custom-built software products to support activities, including sampling and tracking classes and students, administering school and teacher questionnaires, documenting scoring reliability, and creating and checking data files. The TIMSS & PIRLS International Study Center and the IEA DPC also provided NRCs and their staff with intensive training in constructed response item scoring and data management.

The *Survey Operations Procedures* units were crucial resources for the National Research Coordinators as the units described in detail the tasks the NRCs were responsible for conducting. In the event that some of these tasks were contracted to external agencies or organizations, the units ensured that the NRCs had sufficient knowledge of these matters to supervise the activities of the people who helped conduct the assessment in their countries.

The following units, manuals, and software systems were provided for administering TIMSS Advanced 2015:

- *TIMSS Advanced 2015 Survey Operations Procedures Unit 1: Sampling Schools and Obtaining their Cooperation*
- *TIMSS Advanced 2015 Survey Operations Procedures Unit 2: Preparing for and Conducting the TIMSS Advanced 2015 Field Test.* Unit 2 consisted of the following four sections: Sampling Classes and Field Test Administration, Preparing Achievement Booklets and Background Questionnaires, Scoring the Constructed Response Items, and Creating the Databases. Unit 2 was accompanied by field test versions of the School Coordinator Manual, Test Administrator Manual, National Quality Control Monitor Manual, and three software systems (WinW3S, IEA DME, and IEA OSS – described below).
- *TIMSS Advanced 2015 Survey Operations Procedures Unit 3: Contacting Schools and Sampling Classes for the Data Collection.* Unit 3 was accompanied by the School Coordinator Manual and the Windows® Within-school Sampling Software (WinW3S) and its manual. The WinW3S software enabled TIMSS Advanced 2015 participants to randomly select classes in each sampled school and document in detail the class selection process. The software also was used to track school, teacher, student, and student-teacher linkage information; prepare the survey tracking forms (described later in this chapter); and assign test instruments to students, including printing labels for all the test booklets and questionnaires.
- *TIMSS Advanced 2015 Survey Operations Procedures Unit 4: Preparing Achievement Booklets and Context Questionnaires.* Unit 4 was accompanied by the IEA Online SurveySystem (OSS) and its manual. The IEA Online SurveySystem supported the online administration of the school and teacher questionnaires.
- *TIMSS Advanced 2015 Survey Operations Procedures Unit 5: Conducting the Data Collection.* Unit 5 was accompanied by the Test Administrator Manual, National Quality Control Monitor Manual, and the International Quality Control Monitor Manual.
- *TIMSS Advanced 2015 Survey Operations Procedures Unit 6: Scoring the Constructed Response Items.* Unit 6 was accompanied by the TIMSS Advanced 2015 Scoring Guides, the IEA Coding Expert Software, and the Trend Reliability Scoring Manual. The IEA Coding Expert Software was used to facilitate the trend reliability scoring task.
- *TIMSS Advanced 2015 Survey Operations Procedures Unit 7: Creating the Databases.* Unit 7 was accompanied by the IEA Data Management Expert (DME) software, its manual, and codebooks that specified information on the IEA DME data fields in each of the data files. The IEA DME software is used for data entry and data verification.

TIMSS Advanced 2015 Survey Tracking Forms

TIMSS Advanced uses a series of tracking forms to document class sampling procedures, assign assessment instruments, and track school, teacher, and student information, including the participation status of the respondents. The tracking forms also facilitate the data collection and data verification process.

Four different tracking forms were used for TIMSS Advanced 2015:

- Class Listing Form: This form was completed for each sampled school, listing the eligible classes and providing details about the classes, such as the number of students, and the names of teachers.
- Student-Teacher Linkage Form: This form was completed for each class sampled, listing the names of the students and their teachers, student birth dates, gender, exclusion codes, and linking the students to their teachers.
- Student Tracking Form: This form was created for each class assessed and was completed by the Test Administrators during test administration. The Test Administrators used this form to verify the assignment of assessment instruments to students and to indicate student participation.
- Teacher Tracking Form: This form was completed for each sampled school to indicate the completion of the Teacher Questionnaires.

Operations for Data Collection

The following sections describe the major operational activities coordinated by the National Research Coordinators.

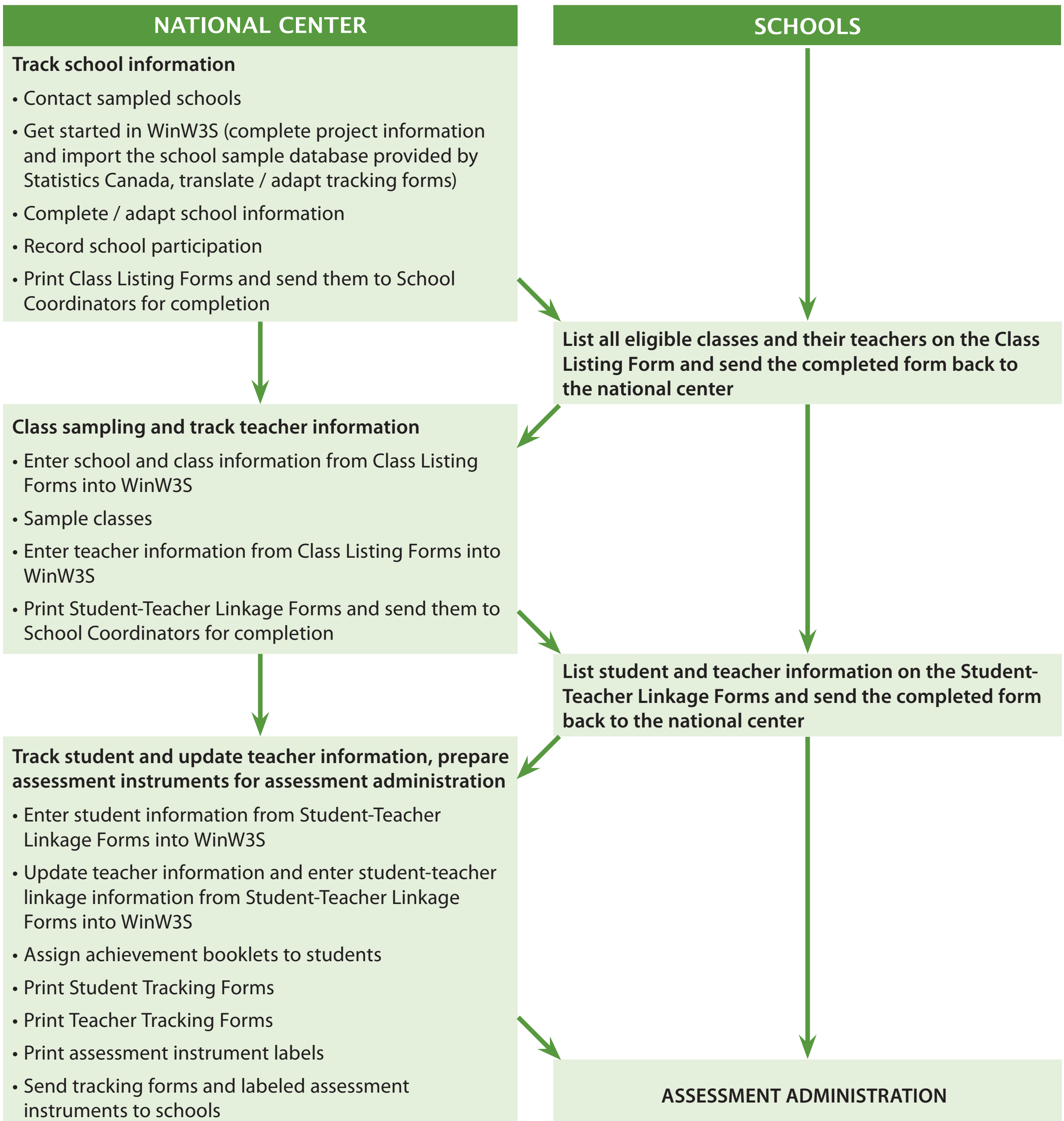
- Contacting schools and sampling classes
- Overseeing translation and preparing assessment instruments
- Managing the administration of the TIMSS Advanced 2015 assessments
- Scoring of the constructed response items
- Creating the TIMSS Advanced 2015 data files

Three other major TIMSS Advanced 2015 operational activities—sampling schools, translation and translation verification of the assessment instruments, and layout verification of the assessment instruments—are described in separate sections of the *Methods and Procedures in TIMSS Advanced 2015* publication (see the Sample Design, Translation and Translation Verification, and Layout Verification chapters).

Contacting Schools and Sampling Classes

Exhibit 6.1 illustrates the major steps in working with schools to sample classes and prepare for the TIMSS Advanced 2015 assessment administration. Once the school samples were drawn, National Research Coordinators were tasked with contacting the sampled schools and encouraging them to take part in the assessment. Depending on the national context, this could involve obtaining support from national or regional educational authorities. *Survey Operations Procedures Unit 1* outlines suggestions on ways to encourage schools to participate in the assessment.

Exhibit 6.1: Diagram of the Sampling Procedures and Preparations for the Assessment Administration Implemented by National Centers and Schools



In cooperation with school principals, National Research Coordinators were responsible for identifying and training School Coordinators for all participating schools. A School Coordinator could be a teacher or guidance counselor in the school, or NRCs could appoint a member of the national center to fill this role. In some countries, a School Coordinator from the national center was responsible for several schools in an area. Each School Coordinator was provided with a School Coordinator Manual, which describes their responsibilities. The School Coordinator Manual was prepared by the TIMSS & PIRLS International Study Center and translated/adapted by National Research Coordinator staff, as necessary.

The responsibilities of the School Coordinator included providing the national center with information on the school; coordinating the date, time, and place for testing; identifying and training a Test Administrator to administer the assessment; coordinating the completion of the TIMSS Advanced 2015 tracking forms; distributing questionnaires; and obtaining parental permission (if necessary). School Coordinators also confirmed receipt of all assessment materials, oversaw the security of the assessment materials, and ensured the return of the assessment materials to the national center following the administration of the assessment.

In addition, School Coordinators played a critical role in providing information for the sampling process, providing the national center with data on eligible classes in the school. With this information, the national centers used the Within-school Sampling Software to sample class(es) within the school. WinW3S tracked school, teacher, and student information, and the software generated the necessary tracking forms and instrument labels facilitating the assessment administration process as well as data checking during the data cleaning process.

Overseeing Translation and Preparing Assessment Instruments

National Research Coordinators also were responsible for preparing the assessment instruments (achievement booklets and context questionnaires) for their countries—a process that included overseeing the translation of the assessment instruments. The overarching goal of assessment instrument preparation is to create internationally comparable achievement booklets and context questionnaires that are appropriately adapted for the national context.

Each student was assigned one of the six advanced mathematics booklets or one of the six physics booklets (see Chapter 4 of the *TIMSS Advanced 2015 Assessment Frameworks* for more information on the matrix sampling design). The achievement booklets are composed of blocks of assessment items, with each block appearing in two booklets. From an operational perspective, each block needed to be translated only once, even though it was included in two different booklets. Adobe®InDesign® software is used by countries to link the translated and adapted assessment blocks to the appropriate booklets. Automating this process through Adobe®InDesign® decreased the chances of human error in the production process.

Six new assessment blocks at each subject were developed for TIMSS Advanced 2015. The new assessment blocks replaced the ones released at the end of the previous assessment cycle. The new assessment items were tried out through the field test in order to investigate the psychometric characteristics of the achievement items and make well-informed decisions about the best items. Similarly, the context questionnaires were evaluated following the field test to gauge the validity and reliability of the various questionnaire scales.

Twice the number of items needed to fill the new assessment blocks were field tested. All participating countries translated and/or adapted the newly developed items into the test administration language(s) and did the same for the questionnaires. After the field test, the best assessment items were chosen for the main data collection and edits were applied to both items and the questionnaires.

National Research Coordinators were responsible for applying these changes to the translated assessment items and questionnaires. Countries that did not participate in TIMSS Advanced 2008 or TIMSS Advanced 1995 had to translate and/or adapt the assessment blocks used in previous assessments (trend blocks) into their language(s) in preparation for the 2015 assessment administration. Countries that had participated in either of the two previous cycles of TIMSS Advanced were required to use the same translations they used in those cycles.

For both the field test and main data collection, the participating countries received the international version (English) of the achievement booklets and context questionnaires with all the necessary instrument production files, including fonts and graphics files. Instructions on how to use the materials to produce high-quality, standardized instruments were included in the corresponding *Survey Operations Procedures* unit.

Once translated and/or adapted, first for the field test and then again for the main data collection, the achievement items and context questionnaires were submitted to the IEA Secretariat for translation verification. The IEA Secretariat worked with independent translators to evaluate each country's translations and when, deemed necessary, suggested changes to the text.

After the translations and adaptations had been verified by the IEA Secretariat, National Research Coordinators assembled the achievement booklets and context questionnaires using Adobe® InDesign® software, and print-ready copies of the instruments were sent to the TIMSS & PIRLS International Study Center for layout verification and a final review of national adaptations. This review checked that each booklet and questionnaire conformed to the international format and that any adaptations made to the instruments did not unduly influence their international comparability.

National Adaptations Forms (NAFs)

While preparing national achievement booklets and context questionnaires, countries sometimes by necessity made adaptations to the international versions. All national adaptations to the international assessment instruments, other than direct translation, were documented using the National Adaptations Forms. There is a separate set of NAFs for the achievement booklets and for the context questionnaires. During the translation verification and layout review, the verifiers checked whether the national adaptations were likely to influence the ability to produce internationally comparable data for the items involved. Any questions raised were directed to the NRC for consideration via the NAFs.

The NAFs were completed and reviewed at various stages of preparing national assessment instruments. Version I of the forms was completed during the internal translation and review process and sent along with the rest of the materials for international translation verification. After translation verification, the forms (Version II) were updated in response to the translation verifier's comments and reflecting any changes resulting from the verification, and sent along with the national assessment instruments for layout verification. Following layout verification, the national instruments and NAFs were finalized (Version III) and submitted to the IEA Secretariat, the TIMSS & PIRLS International Study Center, and the IEA DPC as the final documentation of the national adaptations.

Managing the Administration of the TIMSS Advanced 2015 Assessments

Printing assessment materials and distributing them to the participating schools required careful organization and planning on the part of the National Research Coordinator. Each student was assigned one of the six advanced mathematics booklets or one of the six physics booklets according to a systematic distribution plan implemented by the WinW3S sampling software. This process is facilitated by the tracking forms and labels generated by WinW3S.

Each student also was assigned a Student Questionnaire, which was labeled so that it could be linked to the achievement booklet. In addition, an individually labeled Teacher Questionnaire was sent to each teacher listed on the Teacher Tracking Form and a School Questionnaire was sent to the principal. These materials were packaged and sent to the School Coordinators prior to the testing date, giving ample time for the School Coordinators to confirm the receipt and correctness of the materials. The School Questionnaire and Teacher Questionnaires were then distributed while the other instruments were kept in a secure room until the testing date.

Each sampled class was assigned a Test Administrator who followed procedures described in the Test Administrator Manual to administer the achievement booklets and Student Questionnaire. This person was chosen and trained by the School Coordinator. In many cases, the School Coordinator doubled as the Test Administrator. The Test Administrator was responsible for distributing materials to the appropriate students, reading to the students the instructions provided in the Test Administrator's manual, and timing the sessions.

The Test Administrator documented the timing of the testing session on the Test Administration Form. The Test Administration Form also solicited information about anything out of the ordinary that took place during assessment administration.

The time allotted for the achievement test was standardized to 90 minutes. If a student completed the test before the allotted time, the student was not allowed to leave the testing room. Students who completed the assessment early were asked to review their answers or read quietly. Following the assessment administration, students were given at least 30 minutes to complete the Student Questionnaire; extra time was given when necessary.

The Test Administrator was required to use the Student Tracking Form and labels to distribute the booklets to the correct students and to document student participation. If the participation rate was below 90 percent in any class, it was the School Coordinator's responsibility to hold a makeup session for the absent students before returning all of the testing materials to the national center.

Linking Students to their Teachers and Classes

Exhibit 6.2 illustrates the hierarchical identification system codes that are used to link the data among schools, classes, students, and teachers. The school, class, and student IDs are strictly hierarchical, with classes nested within schools and students nested within classes.

Exhibit 6.2: Hierarchical Identification System Codes Used to Link Schools, Classes, Students, and Teachers

Participant	ID Components	ID Structure	Numeric Example
School	School	CCCC	0001
Class	School + Class within the school	CCCCKK	000101 000102
Student	School + Class within the school + Student within the class	CCCCKKSS	00010101 00010201
Teacher	School + Teacher within the school + Linkage number to the sampled class	CCCCTTLL	00010101 00010201

Each teacher is assigned a teacher identification number consisting of the four-digit school number followed by a two-digit teacher number. Since a teacher could be teaching advanced mathematics and/or physics to some or all of the students in a class, it is necessary to have a unique identification number for each teacher linked to a class and to certain students within the class. This is achieved by adding a two-digit link number to the six digits of the teacher identification number to create a unique eight-digit identification number.

Online Administration of the School and Teacher Questionnaires

Countries could choose to administer the school and/or teacher questionnaires online. The benefits of administering the questionnaires online included saving money and time in printing, and improving the efficiency of questionnaire distribution, data entry, and data cleaning. For the online administration of the questionnaires, the IEA DPC provided its IEA Online SurveySystem software that incorporates design, presentation, and monitoring components.

The design component, known as the Designer, supports the preparation of the online surveys, data management, and data output to the IEA DPC. Through the IEA Online SurveySystem Designer component, national centers could tailor the online questionnaires to their national language. To facilitate translation and adaptation, the Designer concurrently stored the original English question text and the translations and/or national adaptations. It also stored the variable names and data validation rules. If a national center decided not to administer a particular international question or option, it could be disabled in the Designer and would not be administered during the online questionnaire administration. The Designer also included an integrated preview function to allow for a visual side-by-side comparison of the paper/PDF and online versions of the questionnaires, facilitating the layout verification process.

For the online presentation, the Web Component presents the questionnaires to the respondents. The navigation capabilities of the Web Component are designed to allow respondents to pick and choose their order of response. Buttons marked "next" and "previous" facilitated navigation between adjacent pages, so users could browse through the questionnaire in the same way that they flip through the pages of the paper questionnaire. A hyperlinked interactive "table of contents" allowed the respondents to readily navigate to specific questions. Overall, these two functions permitted the respondents to answer questions in the order of their choosing, and skip questions just as they could if they were responding to the paper questionnaire. The online questionnaires could be accessed through any standard Internet browser on all standard operating systems without the user needing any additional software.

Finally, the Web-based Monitor component allows for monitoring the survey responses in real time. Many national centers made extensive use of the Web-based Monitor to follow-up with non-respondents.

The IEA Data Processing and Research Center followed a stringent set of procedures in order to safeguard the confidentiality of the respondents and maintain the integrity of the data. Each respondent received a statement of confidentiality and information on how to access the online questionnaire. For most countries, the online questionnaire administration was hosted on the IEA DPC's customized high-performance server. The IEA DPC server allowed for the 24-hour availability of the questionnaires during the data-collection period, and it also ensured backup and recovery provisions for the data.

Scoring the Constructed Response Items

Constructed response items represent a substantial portion of the TIMSS Advanced assessments, and because reliable and valid scoring of these items is critical to the assessment results, the TIMSS & PIRLS International Study Center provided explicit scoring guides for each item and extensive training in their use. Also, the *Survey Operations Procedures* units specified a procedure for efficiently organizing and implementing the scoring activity.

International scoring training sessions (one for the field test and one for the main data collection) were conducted where all National Research Coordinators (or country representatives appointed by the NRCs) were trained to score the constructed response items. At these training sessions, the scoring guide for each item was reviewed and applied to a set of example student responses that had already been scored. These example papers were chosen to represent a range of response types and to demonstrate the guides as clearly as possible. Following the example papers, the training participants applied the scoring guides to a different set of student responses that had not yet been scored. The scores to these practice papers were then shared with the group and any discrepancies were discussed.

Following the international scoring training, national centers trained their scoring staff on how to apply the scoring guides for the constructed response items. National Research Coordinators were encouraged to create additional example papers and practice papers from student responses collected in their country.

Documenting Scoring Reliability

Because reliable scoring of the constructed response items is essential for high quality data, it is important to document the reliability of the scoring process. A high degree of scorer agreement is evidence that scorers have applied the scoring guides in the same way. The procedure for scoring the TIMSS Advanced 2015 constructed response items provided for documenting scoring reliability within each country (within-country reliability scoring) and over time (trend reliability scoring).

The method for establishing the reliability of the scoring within each country was for two scorers to independently score a random sample of 200 responses for each constructed response item. The degree of agreement between the scores assigned by the two scorers is a measure of the reliability of the scoring process. In collecting the within-country reliability data, it was vital that the scorers independently scored the items assigned to them, and each scorer did not have prior knowledge of the scores assigned by the other scorer. This was achieved by the first scorer marking his/her scores on Reliability Scoring Sheets, leaving the second scorer to mark his/her scores directly into the booklets. The within-country reliability scoring was integrated within the main scoring procedure and ongoing throughout the scoring process.

The purpose of the trend reliability scoring was to measure the reliability of the scoring from one assessment cycle to the next (i.e., from TIMSS Advanced 2008 to TIMSS Advanced 2015). The trend reliability scoring required scorers of the current assessment to score student responses collected in the previous cycle. The scores of the current cycle were then compared with the scores awarded in the previous assessment cycle. Trend reliability scoring was conducted using the IEA Coding Expert Software provided by the IEA DPC.

Trend reliability scoring for TIMSS Advanced 2015 involved four secured item blocks. Student responses included in the trend reliability scoring (150-200 responses per item) were actual student responses collected during the previous assessment cycle in each country. These responses were scanned and provided for each participating country and benchmarking entity along with the IEA Coding Expert Software. All scorers who scored the trend assessment blocks in 2015 were required to participate in the trend reliability scoring. If all scorers were trained to score all trend items, the software divided the student responses equally among the scorers. If scorers were trained to score specific item blocks, National Research Coordinators were able to specify within the software which scorers would score particular item blocks, and the software allocated the student responses accordingly. Similar to the within-country reliability scoring, the trend reliability scoring had to be integrated within the main scoring procedure.

Creating the TIMSS Advanced 2015 Databases

The data entry process took place March-May 2014 for the field test and June-September 2015 following the main data collection. The procedure for creating the TIMSS Advanced 2015 databases included entering sampling and assessment administration information into the WinW3S database and adding responses from the context questionnaires and achievement booklets using the IEA Data Management Expert (DME) software.

The IEA DPC provided DME software to accommodate keyboard data entry and data verification. The DME software also offers data and file management capabilities, a convenient checking and editing mechanism, interactive error detection, and quality-control procedures. For the TIMSS Advanced 2015 context questionnaires administered online on the IEA DPC's server, the data were directly accessible by the IEA DPC and no further data entry was required.

Along with the DME software, the IEA DPC provided international codebooks describing all variables and their characteristics, thus ensuring that the data files met the internationally defined rules and standards for data entry. The files within the DME database for entering the TIMSS Advanced 2015 data were based on these codebooks. However, the codebooks had to match exactly the national assessment instruments so that the answers of the respondents could be entered properly. Therefore, any adaptations to the international instruments also required adaptations to the international codebooks. The adapted national codebooks then were used to create the TIMSS Advanced 2015 data files in each country, with the responses to the context questionnaires, achievement booklets, and Reliability Scoring Sheets keyed into the DME database.

Quality control throughout the data entry process was essential to maintain accurate data. Therefore, National Research Coordinators were responsible for performing periodic reliability checks during data entry and for applying a series of data verification checks provided by both WinW3S and DME software prior to submitting the databases to the IEA DPC. To ensure the reliability of the data entry process, the data-entry staff was required to double enter at least 5 percent of each instrument type. An error rate of 1 percent or less was acceptable for the background questionnaire files. An error rate of 0.1 percent or less was required for the student achievement files and the reliability scoring files. If the required agreement was not reached, retraining of the data entry staff was required.

Additionally, the data verification module of WinW3S and DME could identify a range of problems, such as inconsistencies of identification codes and out-of-range or otherwise invalid codes. The data quality control procedures also verified the integrity of the linkage between the students, teachers, and schools entered into the DME database and tracking of information for those specified in WinW3S.

When all data files had passed the quality control checks, they were submitted to the IEA DPC, along with data documentation, for further checking and processing. For information on data processing at the IEA DPC, please refer to the Creating the International Database chapter of this publication.

TIMSS Advanced 2015 Survey Activities Questionnaire

The Survey Activities Questionnaire was designed to elicit information about NRCs' experiences in preparing for and conducting the TIMSS Advanced 2015 data collection. The questionnaire was composed of six sections and focused on the following:

- Sampling schools and classes
- Preparing assessment instruments
- Administering the assessments
- Implementing the National Quality Control Program
- Preparing for and scoring the constructed response items
- Creating the databases

All items in the Survey Activities Questionnaire included accompanying comment fields, in which NRC respondents were encouraged to explain their responses, provide additional information, and suggest improvements in the process.

The *TIMSS Advanced 2015 Survey Activities Questionnaire* was administered online via the IEA's Online SurveySystem and was completed by NRCs. The following sections summarize information gathered from the Survey Activities Questionnaire, reflecting the quality of the TIMSS Advanced 2015 survey materials and procedures in the participating countries.

Sampling Schools and Classes

The first section of the Survey Activities Questionnaire asked NRCs about the *Survey Operations Procedures* for sampling both schools and classes within the sampled schools. As shown in Exhibit 6.3, all but one of the countries considered that *Survey Operations Procedures Unit 1* was clear and sufficient, and all countries considered that *Survey Operations Procedures Unit 3* was clear and sufficient. Five countries reported deviating from the basic TIMSS Advanced sampling design. Their reasons for these modifications to the sampling procedures included allowing for census participation, oversampling certain regions, and that all or some classes and/or students of the target population were eligible to be sampled for both advanced mathematics and physics assessments. Statistics Canada sampling experts selected the school samples for almost all countries. Only one country reported selecting their TIMSS Advanced 2015 school sample at the national center working in collaboration with Statistics Canada.

Exhibit 6.3: Survey Activities Questionnaire, Section One—Sampling (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "TIMSS Advanced 2015 Survey Operations Procedures Unit 1 – Sampling Schools and Obtaining their Cooperation" clear and sufficient?	8	1	0
Were there any conditions or organizational constraints that necessitated deviations from the basic TIMSS Advanced sampling design described in the "Survey Operations Procedures Unit 1"?	5	4	0
Did you use the Within-school Sampling Software (WinW3S) to sample classes?	8	1	0
Did you experience any problems or inconveniences when using the WinW3S software?	3	5	1
Was the information provided in the "TIMSS Advanced 2015 Survey Operations Procedures Unit 3 – Contacting Schools and Sampling Classes for the Data Collection" clear and sufficient?	9	0	0
Did you follow the procedures outlined in "Survey Operations Procedures Unit 3" for working with the schools to sample classes (e.g., using the appropriate tracking forms in the proposed order to obtain information from School Coordinators)?	8	1	0

All countries but one selected classes within the sampled schools using the Windows® Within-school Sampling Software (WinW3S) provided by the IEA Data Processing and Research Center. Three countries experienced some problems using the WinW3S software, such as issues importing tracking form information.

One NRC reported having to modify the procedures outlined in the *Survey Operations Procedures Unit 3*. In this particular case, all classes were selected in the sampled schools and, as such, no class sampling was necessary.

Translating, Adapting, and Producing Assessment Instruments

The second section of the Survey Activities Questionnaire asked NRCs about translating, adapting, assembling, and printing the test materials, as well as issues related to checking the materials and securely storing them.

As reported in Exhibit 6.4, all of the NRCs answered that they were able to assemble the test booklets and questionnaires according to the instructions provided. All NRCs also reported applying corrections to their survey instruments as suggested by the external translation verifier and/or the layout verifier. However, five countries reported experiencing some problems using the assessment instrument production materials. These problems included the following: issues with equations and formulas in MathType, CopyFlow Gold overwriting the preset table styles, issues with fonts for equations and formulas and fonts for special characters (e.g., for Cyrillic alphabet), and difficulty fitting longer national text in the context questionnaires. All of the identified problems were resolved either by specialists at the national center or with assistance from the TIMSS & PIRLS International Study Center.

Exhibit 6.4: Survey Activities Questionnaire, Section Two—Translating, Adapting, and Producing Assessment Instruments (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "TIMSS Advanced 2015 Survey Operations Procedures Unit 4 – Preparing Achievement Booklets and Context Questionnaires" clear and sufficient?	9	0	0
Did you encounter any major problems using the assessment instrument production materials (e.g., instrument production files, fonts, support materials) provided by the TIMSS & PIRLS International Study Center?	5	4	0
After the translation verification, did you correct your translations/adaptations as suggested by the verifier in the majority of cases?			
Advanced Mathematics booklets	9	0	0 *(Not Answered)* 0 *(Not Applicable)*
Physics booklets	9	0	0 *(Not Answered)* 0 *(Not Applicable)*
Context questionnaires	9	0	0 *(Not Answered)* 0 *(Not Applicable)*
After the layout verification, did you correct your assessment instruments as noted by the verifier in the majority of cases?			
Advanced Mathematics booklets	9	0	0 *(Not Answered)* 0 *(Not Applicable)*
Physics booklets	9	0	0 *(Not Answered)* 0 *(Not Applicable)*
Context questionnaires	9	0	0 *(Not Answered)* 0 *(Not Applicable)*
Did you apply any quality control measures to check the achievement booklets and context questionnaires during the printing process (e.g., checking for missing pages, upside down pages, text too bright or too dark)?	9	0	0
Did you take measures to protect the security of the assessment instruments during the translation, assembly, and printing process?	9	0	0
Did you detect any potential breaches in security of the assessment instruments?	0	9	0

Question	Yes	No	Not Answered
Did you encounter any problems preparing the Online SurveySystem files for administering the school and/or teacher questionnaires online?	2	2	0 *(Not Answered)* 5 *(Not Applicable)*

All countries conducted the recommended quality control checks during the process of printing the testing materials, and no country expressed concerns about a breach of security of the assessment instruments. Two countries reported that they experienced problems with the Online SurveySystem related to the date and time fields, which appeared as plain text fields.

Assessment Administration

The third section of the Survey Activities Questionnaire addressed the extent to which NRCs detected errors in the testing materials during packaging for shipment to schools. As shown in Exhibit 6.5, a small number of errors were found in the materials. Errors found after distribution usually were very minor, and either were fixed by School Coordinators or replacement materials were provided.

Exhibit 6.5: Survey Activities Questionnaire, Section Three—Assessment Administration (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "TIMSS Advanced 2015 Survey Operations Procedures Unit 5 – Conducting the Data Collection" clear and sufficient?	9	0	0
Were any errors detected in any of the following assessment materials after they were sent to schools?			
Achievement booklets	3	6	0 *(Not Answered)* 0 *(Not Applicable)*
Achievement booklet ID labels	0	9	0 *(Not Answered)* 0 *(Not Applicable)*
Student Questionnaires	1	8	0 *(Not Answered)* 0 *(Not Applicable)*
Student Questionnaire ID labels	0	8	0 *(Not Answered)* 1 *(Not Applicable)*
Student Tracking Forms	2	7	0 *(Not Answered)* 0 *(Not Applicable)*
Teacher Questionnaires	1	8	0 *(Not Answered)* 0 *(Not Applicable)*
Teacher Tracking Forms	0	9	0 *(Not Answered)* 0 *(Not Applicable)*
School Questionnaires	0	9	0 *(Not Answered)* 0 *(Not Applicable)*

Exhibit 6.5: Survey Activities Questionnaire, Section Three—Assessment Administration (Numbers of NRC Responses) (Continued)

Question	Yes	No	Not Answered
School Coordinator Manuals	1	7	0 *(Not Answered)* 1 *(Not Applicable)*
Test Administrator Manuals	0	9	0 *(Not Answered)* 0 *(Not Applicable)*
If any errors were detected, did you correct the error(s) before the testing began?	1	6	0 *(Not Answered)* 2 *(Not Applicable)*
Does your country have a confidentiality policy that restricts putting student names on tracking forms and survey instrument covers?	3	6	0
Did you encounter any problems translating and/or adapting the School Coordinator Manual?	0	9	0
Did you encounter any problems translating and/or adapting the Test Administrator Manual?	0	9	0
Were School Coordinators appointed from within the participating schools?	9	0	0
Did you hold formal training session(s) for School Coordinators?	4	5	0
Were Test Administrators trained by School Coordinators within the participating schools?	5	4	0
Did Test Administrators document any problems or special circumstances that occurred frequently during the assessment administration (please refer to the completed Test Administration Forms)?	2	7	0
If you administered school and/or teacher questionnaires online, did any of the respondents in your country encounter any problems responding to the online questionnaires?	1	3	0 *(Not Answered)* 5 *(Not Applicable)*

In all countries, School Coordinators were appointed within the participating schools. In some, mostly larger countries, School Coordinator training was conducted either online or in a written format via extended manuals. In five of the countries, the School Coordinators appointed the Test Administrators from within the participating school and in these countries the School Coordinators trained the Test Administrators. In the remaining countries, the Test Administrators were from the national center or contracted externally.

Two NRCs reported problems documented by Test Administrators during assessment administration. The documented problems included numerous absent students, which resulted in make-up sessions, and difficulties creating the student-teacher linkage for new students.

National Quality Control Program

The fourth section of the Survey Activities Questionnaire addressed the national quality control program that each country implemented during data collection. As part of the national quality assurance activities, NRCs were instructed to send National Quality Control Observers to 10 percent of the participating schools in order to observe test administration and document compliance with prescribed procedures. This was in addition to the program of school visits by International Quality Control Monitors conducted by the TIMSS & PIRLS International Study Center.

As shown in Exhibit 6.6, due to budgetary constraints, one country did not send national monitors to the testing sessions, and one country sent monitors to less than ten percent of participating schools. Seven countries conducted their quality assurance program using the National Quality Control Monitor Manual provided by the TIMSS & PIRLS International Study Center. One country reported that their NQCMs documented a frequent problem occurring during the testing sessions linked to the language of the test.

Exhibit 6.6: Survey Activities Questionnaire, Section Four—National Quality Control Program (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Did you conduct a national quality control program that observed the data collection in the participating schools?	7	2	0
Did you use the National Quality Control Monitor (NQCM) Manual and the Classroom Observation Record provided by the TIMSS & PIRLS International Study Center to conduct your national quality control program?	7	1	0 *(Not Answered)* 1 *(Not Applicable)*
Did your national quality control monitors (NQCMs) document any major problems or special circumstances that occurred frequently during the assessment administration?	1	7	0 *(Not Answered)* 1 *(Not Applicable)*

Preparing for and Scoring the Constructed Response Items

Exhibit 6.7 provides data on responses to questions about NRCs experiences preparing for and scoring the constructed response items. All NRCs found the scoring procedures outlined in *Survey Operations Procedures Unit 6—Scoring the Constructed Response Items* to be clear and sufficient. Two countries reported that they would have liked to have scoring training practice materials for all items instead of a select group of items and that they found it difficult to translate the scoring guides and the training materials. Five NRCs reported creating their own national examples and practice papers for training their scorers, as suggested by the TIMSS & PIRLS International Study Center. Six countries scanned their achievement booklets and scored student responses electronically, and no problems were encountered during the Trend Reliability Scoring. Most countries had all scorers participate in the scoring of the trend items, with two countries having only a subset score these items.

Exhibit 6.7: Survey Activities Questionnaire, Section Five—Preparing for and Scoring the Constructed Response Items (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "TIMSS Advanced 2015 Survey Operations Procedures Unit 6 – Scoring the Constructed Response Items" clear and sufficient?	9	0	0
Did you encounter any problems using the scoring training materials, provided by the TIMSS & PIRLS International Study Center?	2	7	0
Did you create national scoring training materials in addition to the international scoring training materials?	5	4	0
Did you scan the achievement booklets for electronic image scoring?			
Advanced Mathematics booklets	6	2	0 *(Not Answered)* 1 *(Not Applicable)*
Physics booklets	6	2	0 *(Not Answered)* 1 *(Not Applicable)*
Did you encounter any problems during the Trend Reliability Scoring?			
Procedural problems	0	6	0 *(Not Answered)* 3 *(Not Applicable)*
Technical, software related problems	0	6	0 *(Not Answered)* 3 *(Not Applicable)*
Did all your scorers participate in scoring student responses of the trend items?	4	2	0 *(Not Answered)* 3 *(Not Applicable)*

Creating the Databases

The last section of the Survey Activities Questionnaire addressed data entry and quality control activities. As shown in Exhibit 6.8, all of the NRCs found the instructions in the *Survey Operations Procedures Unit 7* to be clear and sufficient. Two NRCs expressed a wish for a more automated data entry process in WinW3S and would like the software to allow multiple users to work at the same time via a server. Five countries used optical scanning of the assessment instruments instead of manual data entry. All countries reported applying all required data quality checks and storing their assessment instruments in a secure storage area.

Exhibit 6.8: Survey Activities Questionnaire, Section Six—Creating Databases (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "TIMSS Advanced 2015 Survey Operations Procedures Unit 7 – Creating the Databases" clear and sufficient?	9	0	0
Did you encounter any problems entering test administration information and exporting your WinW3S database(s)?	2	7	0
Who primarily entered the data for your country?			
National center staff	4	-	0
Temporarily hired data entry staff	3	-	0
An external data entry firm	1	-	0
Combination of the above	1	-	0
Other	0	-	0
Did you use manual (key) data entry to create the data files for your country?			
Achievement booklets	4	5 *(Optical Scanning)*	0 *(Not Answered)* 0 *(Not Applicable)*
Context questionnaires	4	5 *(Optical Scanning)*	0 *(Not Answered)* 0 *(Not Applicable)*
Did you encounter any problems using the IEA's Data Manager Expert (DME) software?	1	8	0
If you entered data manually, did you enter 5% of each survey instrument twice as a quality control measure?	3	1	0 *(Not Answered)* 5 *(Not Applicable)*
Did you apply all the data quality checks described in the "TIMSS Advanced 2015 Survey Operations Procedures Unit 7 – Creating the Databases" before submitting your data to the IEA Data Processing and Research Center?	9	0	0
Have you stored all achievement booklets and context questionnaires in a secure storage area until the original documents can be discarded?	9	0	0

CHAPTER 7

Translation and Translation Verification for TIMSS Advanced 2015

David Ebbs
Michelle Djekić

Introduction

This chapter describes the activities and procedures related to the preparation of the national versions of the TIMSS Advanced 2015 instruments, focusing on two major activities:

- Translation and adaptation of the international version of the TIMSS Advanced 2015 instruments into national survey languages
- International translation verification of the national translations/adaptations

The TIMSS & PIRLS International Study Center is responsible for the development of the international version of the TIMSS Advanced 2015 instruments in English. After the release of the international source version, all participating countries are required to translate and/or adapt the international version into their language(s) of instruction. To ensure that the translated national instruments are equivalent to the international version, linguistic and assessment experts conduct a series of reviews based on the international source version in English.

The translation and translation verification processes aim to ensure that the national versions consist of high quality translations and are internationally comparable and adapted appropriately to every country's context and education system. The international source version comprises the achievement items and questionnaires. At the national level, all participating countries are required to translate and/or adapt the international source version according to the international guidelines for TIMSS Advanced 2015, conduct a review of their translation's quality and appropriateness, and document all national adaptations for reference at later stages in the process. It is also necessary for all participating countries to follow standard, internationally agreed-upon procedures from the initial translation through to the final printing of their national instruments. For countries whose survey language is English, national adaptations to the materials are also required to accommodate the variations used in different English-speaking countries.

At the international level, the IEA Secretariat is responsible for ensuring that every country's translated and adapted materials undergo international translation verification. As part of the international quality assurance program, the international translation verification process requires that all national instruments go through a formal external review of the translations and adaptations prior to the commencement of the assessment. The translation verifiers provide detailed feedback to improve the accuracy of the national instruments for every country, in comparison to the international instruments. When the verified materials are returned, the National Research Coordinators (NRCs) are tasked with reviewing the translation verification feedback, revising their materials as needed, and updating their documentation for use during data processing and analysis. The translation and translation verification processes occur twice—first before the field test and then before the commencement of data collection. The IEA Secretariat manages these processes, which entails the careful documentation of outcomes at the various stages of translation, adaptation, translation verification, and revision.

Prior to the field test and before the commencement of data collection, the same general international translation verification procedures apply for all items except those designed to measure trends from previous cycles. Trend items undergo a separate international translation verification procedure to ensure consistency across assessment cycles.

The TIMSS Advanced 2015 materials required to undergo translation verification are:

- Student achievement items (assembled in blocks of advanced mathematics items and blocks of physics items)
- Background questionnaires for school principals, teachers, and students
- Covers and directions (for each achievement booklet and background questionnaire)
- Online covers and directions (for countries administering the teacher and school questionnaires online)

The TIMSS Advanced procedural manuals and scoring guides for the constructed response items typically are translated but not subject to the international verification procedure.

The Translation Process

The TIMSS & PIRLS International Study Center provides directions for translating the achievement items and questionnaires, and requests that a skilled and experienced translator translate the international instruments. To ensure that national versions of the TIMSS Advanced 2015 instruments are consistent with the international version, the assessment translation guidelines allow for national adaptations where necessary. Following translation of the international instruments, one or more qualified reviewers independently review the completed translations to ensure the nationally translated instruments are of the highest quality and student-level appropriate. Certain countries employ multiple translators and reviewers, either working together to complete the tasks on schedule, or working independently to provide two or more reviews. When countries use more than one translator, they must reconcile the translation differences to ensure the production of a single consistently translated set of materials. Similarly, when using more than one reviewer, countries are responsible for ensuring consistency of the reviews across the translated materials. When countries prepare translations in more than one language, professionals proficient in both languages should be involved to ensure equivalency across the national translations.

Guidelines for Translation and Adaptation

The general purpose of translation and adaptation is to maintain the same meaning and level of difficulty as the international version while following the rules of the target language(s) and considering the participant's cultural context. English-speaking countries are also required to adapt the international version to suit the varieties of English used in their national contexts. This also applies to countries using and adapting translations developed by other country participants.

In particular, translators and reviewers are asked to ensure that:

- The translation is at an appropriate level for the target population
- No information is omitted, added, or clarified in the translated text
- The translated text has the same meaning and uses equivalent terminology as in the international version
- The translated text has the same register (language level and degree of formality) and level of difficulty as the international version
- Idiomatic expressions are translated appropriately, not necessarily word for word
- The translated text uses correct grammar, punctuation, qualifiers, and modifiers, as appropriate for the target language

After the field test, the TIMSS & PIRLS International Study Center provides NRCs with a list of changes to the international version that they can refer to while preparing their national instruments. This information minimizes the translation burden while highlighting the necessary changes to the translations prior to commencement of the assessment.

The Target Language

Identifying the language of the assessment (the "target" language) for most countries is relatively straightforward because there is a dominant language used in both the public and private sectors of society. However, some countries use more than one language of instruction in their national educational systems. In such cases, countries translate the student instruments into several target languages to ensure that they can be administered in the language of instruction used for teaching in schools.

Scope of Translation and Verification in TIMSS Advanced 2015

For the TIMSS Advanced 2015 cycle, nine countries prepared ten national sets of assessment materials in eight languages. Exhibit 7.1 lists the TIMSS Advanced 2015 countries, the target languages identified for each country, and the instruments to be translated.

Exhibit 7.1: Languages Used for the TIMSS Advanced 2015 Assessment Instruments

Country	Language	Instruments				
		Achievement Test	Student Questionnaire	Teacher Questionnaire – Advanced Mathematics	Teacher Questionnaire – Physics	School Questionnaire
France	French	●	●	●	●	●
Italy	Italian	●	●	●	●	●
Lebanon	English	●	●	●	●	●
	French	●	●	●	●	●
Norway	Norwegian	●	●	●	●	●
Portugal	Portuguese	●	●	●	●	●
Russian Federation	Russian	●	●	●	●	●
Slovenia	Slovene	●	●	●	●	●
Sweden	Swedish	●	●	●	●	●
United States	English	●	●	●	●	●

Providing the Instruments for Translation and Adaptation

The TIMSS & PIRLS International Study Center provides NRCs with electronic files consisting of all materials to be translated, including special forms for documenting each step of the adaptation, translation, and translation verification processes. According to the TIMSS Advanced 2015 assessment design, all of the achievement item blocks appear in more than one booklet, and therefore the component parts of the booklets (blocks, covers, and directions) are prepared as separate files to facilitate translation. This approach allows countries to translate each component only once before assembling the booklets. In addition to the international instruments, NRCs receive detailed manuals and instructional videos that provide information on how to work with the electronic files, including guidelines for translation and adaptation, and instructions for booklet assembly.

Translators and Reviewers

Countries are strongly advised to hire highly qualified translators and reviewers who are well suited to the task of working with the TIMSS Advanced 2015 materials.

Essential qualifications for translators and reviewers include:

- Excellent knowledge of English
- Excellent knowledge of the target language
- Experience in the country's cultural context
- Experience in translating texts in the subject areas related to the TIMSS Advanced 2015 assessment (advanced mathematics and physics, respectively)

In addition to the above, reviewers are also expected to have experience, preferably as a school teacher, with students in the target grade. Reviewers are primarily responsible for evaluating the readability and accuracy of the translation for the target population.

Translation and Adaptation of the Achievement Test

While translating the TIMSS Advanced achievement test, it can be challenging to select appropriate terms and expressions in the target language(s) of each country that convey the same meaning and style of the text used in the international source version. When adapting and translating expressions with more contextually appropriate terms, translators must ensure that the meaning or difficulty of the item remains the same as in the international source version. For example, it is important that adaptation/translation of an item does not simplify or clarify the text in such a way as to provide a hint or definition of the meaning of a question. Translators must also ensure the consistency of adaptations and translations from item to item and across the materials. Similarly, for multiple-choice items, translators are instructed to pay particular attention to the literal and synonymous matches of the text in both the question stem and answer options, maintaining matches between international version and the translated national version.

NRCs are strongly encouraged to keep adaptations to a minimum, but some adaptations are necessary in order to prevent students from facing unfamiliar vocabulary or contexts that could hinder their ability to read and understand the item. At times, changes to the instruments may be required in order to follow national conventions of measurement, mathematical notation (e.g., decimal separator, multiplication sign), punctuation, and expressions of date and time. For example, a word such as "flashlight" in American English would be adapted to "torch" in British English. In addition, names of fictional characters and places may be modified to similar names applicable in the target language. When the names of fictional cities or towns are adapted, translators are advised not to use real place names to prevent students' responses from being influenced by their perception and knowledge of the names.

Some terms in the text are not to be changed or adapted beyond translation, such as the proper names of actual people and places, as well as the fictional currency "zed" (which is used in the TIMSS Advanced 2015 items pertaining to money). To aid in the standardization of the most common adaptations across countries, the TIMSS & PIRLS International Study Center provides a list of specific examples of acceptable and unacceptable adaptations, including a list of measurement conversions.

Blocks of Achievement Items Designated to Measure Trends

According to a carefully specified design, a number of blocks (approximately 33%) are carried over to the next cycle (see Chapter 1: Developing the TIMSS Advanced 2015 Achievement Items) for the purpose of measuring changes in student achievement over time. To ensure the quality of the trend measurement, these "trend blocks" must be administered in exactly the same way during every cycle. For countries that previously participated in TIMSS Advanced 2008, the translations of the trend blocks used in the previous assessment(s) were compared against the 2015 assessment translations.

If a country determines that changes to the trend blocks are unequivocally required (e.g., in order to correct a mistranslation discovered in a previous translation), the changes are carefully documented and reviewed. Items with changes may not be included in the trend analyses for that country.

The preparation of the trend blocks for countries not participating in the trend comparison follows the same general procedure for preparation as the newly developed assessment blocks for the current cycle.

Translation and Adaptation of the Questionnaires

The translation of the questionnaires differ from the assessment items in that participating countries are required to adapt some terms, and to ensure that questions are appropriate within the national context and education system. The terms requiring adaptation are listed in angle brackets in the international version with their country-specific information. For instance, <language of test> and <twelfth grade> would be adapted to the name of the actual language and grade in which the assessment is being administered. In Lebanon, these terms would be replaced by equivalents "English" and "Third secondary year". Some terms related to specific aspects of teaching and learning also are designated for adaptation—<in-service/professional development> should be adapted to the local term that denotes the supplemental training provided to teachers during their professional careers. Items assessing levels of education use the current version of the International Standard Classification of Education (ISCED) system, ISCED 2011 (UNESCO Institute for Statistics, 2012), and require adaptation to the nationally equivalent educational terms for each participating country.

The guidelines for translation and adaptation provide countries with detailed descriptions of the intent of each required adaptation to clarify the meaning of the terms used and to enable the translators to select the appropriate national term or expression to convey the intended meaning. Countries are permitted to add a limited number of national interest questions to the questionnaires. To avoid influencing the responses to international questions, NRCs are advised to place any national interest questions at the end of the corresponding module or questionnaire, and to ensure these adopt the same format as the rest of the questionnaire. Please note that all national interest questions must be documented and approved by the TIMSS & PIRLS International Study Center before their inclusion in the questionnaires.

The National Adaptation Form

The National Adaptation Form (NAF) is an Excel document formatted to contain the complete translation, adaptation, and translation verification history of each set of national instruments. During various stages of the instrument preparation process the form is completed and reviewed. All national adaptations should be documented in the NAF, and NRCs must prepare one NAF for each language and set of instruments. During the translation and adaptation processes, the first version of the NAF is filled out in collaboration with the translator(s), reviewer(s), and NRC. The translator and reviewer document the initial adaptations made to the instruments, to be reviewed and consolidated by the NRC.

When documenting an adaptation, the following information is recorded in the NAF: 1) identifying information (location and/or question number), 2) an English back translation of the adaptation, and 3) recoding instructions (if applicable). All locations of required adaptations are listed in the NAF with yellow cell backgrounds and with the terms in a different font color and in angle brackets to enable easy identification and review. For ease of use and documentation of the different stages of translation and translation verification, the NAF includes designated areas for each item, respondent, and instrument. After each round of international translation verification, the NAF is updated and revised with commentary from the international translation verifier and the NRC. The NAF is an important record of each country's final instruments, as it contains information used throughout the different stages of translation, adaptation, and translation verification.

In addition, the International Quality Control Monitors (IQCMs) also use the NAF after data collection to review the implementation of translation verification feedback (see the Survey Operations Procedures and Quality Assurance chapters). The NAF is referenced when adding national data to the international database and during data analysis.

International Translation Verification

After the international instruments are translated and internally reviewed by countries, the national translations of the instruments are then submitted for international translation verification. The IEA Secretariat manages the international translation verification process in coordination with an external translation verification company, cApStAn Linguistic Quality Control (based in Brussels, Belgium).

Translation Verifiers

For TIMSS Advanced 2015, the international translation verifiers are responsible for reviewing and documenting the quality of the national instruments and their comparability to the international instruments. The required qualifications for verifiers include:

- Fluency in English
- Mother tongue proficiency in the target language
- Formal credentials as translators working in English
- University-level education and (if possible) familiarity with the subject area
- Residency in the target country, or close contact with the country and its culture

The IEA Secretariat trains all international translation verifiers, supplying them with a comprehensive set of instructional materials to support their work. For TIMSS Advanced 2015, international translation verifiers were trained through web-based seminars and provided with information about TIMSS Advanced 2015 and the assessment instruments. Each international translation verifier received a document containing 1) the description of the adaptation and translation guidelines, 2) the relevant manuals and instruments, and 3) a document with the directions and instructions for reviewing the national instruments and registering deviations from the international version. During the verification of the final assessment instruments, international translation verifiers were given a list of changes to the international instruments made after the field test, and given access to the relevant national field test NAFs.

The International Translation Verification Process

The instructions and training given to the international translation verifiers emphasize the importance of maintaining the same meaning and difficulty level in the translations and adaptations as in the international versions, and ensuring that translations and adaptations are adequate and consistent within and across national instruments. The translation verification process involves:

- Checking the accuracy, linguistic correctness, and comparability of the translation and adaptations of the achievement items and questionnaires

- Documenting any deviations between the national and international versions, including additions, deletions, and mistranslations
- Suggesting an alternative translation/adaptation to improve the accuracy and comparability of the national instruments

Verifiers provided feedback from translation verification in both the instruments and the NAFs. Verifiers are asked to correct the text of the assessment items and questionnaires and/or to add notes specifying errors using either "Sticky Notes" in Adobe PDFs or "Track Changes" and "New Comment" functions in Microsoft Word. Some of the typical errors identified by verifiers during translation verification include mistranslations, inconsistent translations (mathematical symbols, adaptation of ISCED levels, literal versus synonymous matches), omissions/additions of text, adaptations of names (fictional versus real), gender agreement, and grammar. After reviewing the documented comments and suggestions from the verifiers, NRCs revise and improve their national versions.

Translation verifiers record all comments viewed as major deviations or deviations in adaptations in the NAF. All verifier comments contain a code to help NRCs understand the severity and the type of deviation of the translated text with regard to the international version. In addition, verifiers review and comment on all adaptations reported in the NAF.

Codes Used in Verification Feedback

To help establish the quality and comparability of the translated/adapted instruments, the international translation verifiers aim to provide meaningful feedback to the NRCs, TIMSS & PIRLS International Study Center staff, and other members of the study consortium. To standardize the verification feedback across countries, verifiers are asked to assign a code to each intervention, indicating the nature and severity of the issue identified. These codes are accompanied by explanatory information, along with corrections or suggestions for improvement, if applicable. The criteria for coding are as follows:

CODE 1 indicates a major change or error. Examples include the omission or addition of a question or answer option; incorrect translation that changes the meaning or difficulty of the item or question; and incorrect order of questions or answer options in a multiple-choice question.

If in any doubt, verifiers are instructed to use **CODE 1?** so that the error can be referred to the TIMSS & PIRLS International Study Center for further consultation.

CODE 2 indicates a minor change or error, such as a spelling or grammar error that does not affect comprehension.

CODE 3 indicates that while the translation is adequate, the verifier has a suggestion for an alternative wording.

CODE 4 indicates that an adaptation is acceptable and appropriate. For example, a reference to winter for a country in the Southern Hemisphere is changed from January to July.

Verification of the Trend Assessment Blocks

For all countries assessing trends, the international verification procedure includes a "trend check" for the achievement instruments to ensure that the trend items have not been changed. This involves:

- Checking that each of the trend items for the current cycle are equivalent to the trend items administered in the previous cycle
- Documenting any differences in content

The translation verifiers are instructed to record any discrepancies found in the trend items within the NAF. NRCs are instructed to carefully review all discrepancies listed by the translation verifiers and discuss any proposed changes with the TIMSS & PIRLS International Study Center.

Outcomes and Summary for TIMSS Advanced 2015

In accordance with previous cycles of TIMSS Advanced, stringent procedures for translation, adaptation, and translation verification were implemented to ensure the production of high quality translations and internationally comparable TIMSS Advanced 2015 instruments. In addition, the TIMSS & PIRLS International Study Center provided the NRCs with comprehensive guidelines containing information about the NRC's responsibilities, including the importance of employing highly skilled and experienced translators and reviewers for instrument production.

After the completion of the international translation verification processes (field test and data collection), NRCs reviewed the feedback from the translation verifiers. The feedback contained commentary and suggestions on errors in the texts, ranging from grammar and typographical errors to additions/deletions of text and mistranslations. Based on this important feedback and in agreement with the translation guidelines for TIMSS Advanced 2015, NRCs revised and improved the quality of their national versions.

The outcomes and feedback from translation verification confirm that countries followed the guidelines and procedures (including submission of materials and review of post-verification materials) to produce high quality and internationally comparable instruments for the TIMSS Advanced 2015 cycle.

References

UNESCO Institute for Statistics. (2012). *ISCED: International Standard Classification of Education*. Retrieved January 20, 2016, from http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx

CHAPTER 8

Layout Verification for TIMSS Advanced 2015

Erin Wry
Ieva Johansone

Layout verification is the final external review and ratification of each participating country's national assessment instruments (achievement booklets and questionnaires) and their corresponding National Adaptations Forms. To ensure that the instruments are of the highest quality and are comparable across all of the participating countries, countries follow standard internationally agreed-upon procedures in preparing national versions of the instruments (see Chapter 6 on TIMSS Advanced Survey Operations Procedures). TIMSS Advanced translation guidelines allow for national adaptations to instruments as long as international comparability is maintained. Countries are required to document any national adaptations applied to the international assessment instruments within the TIMSS Advanced 2015 National Adaptations Forms. This documentation is verified nationally and internationally throughout all stages of preparing each country's national instruments.

Prior to both the field test and main data collection, all national instruments undergo independent translation verification coordinated by the IEA Secretariat, and after the contents of the achievement booklets and context questionnaires have completed translation verification, the national instruments are sent to the TIMSS & PIRLS International Study Center for layout verification. During the layout verification process, the TIMSS & PIRLS International Study Center checks to ensure that all national assessment instruments conform to the international format and that any national adaptations made to the TIMSS Advanced 2015 international instruments do not unduly influence their international comparability. In particular, layout verification focuses on the following:

- Reviewing the national assessment instruments for acceptable layout structure including pagination, page breaks, item sequence, response options, text formats, and graphics
- Reviewing the national adaptations applied to both the international achievement booklets and context questionnaires with respect to how they may influence the international comparability of the data

Scope of Layout Verification for TIMSS Advanced 2015

Participating countries prepare national versions of the instruments for the field test and then again for the main data collection. This includes translating and/or adapting the newly developed items and questionnaires in preparation for the field test. Then, changes resulting from the field test are applied to the achievement items selected for the main data collection and similar modifications are applied to the context questionnaires. Accordingly, in preparation for TIMSS Advanced 2015 assessment administration, layout verification was conducted twice for each participating country—once for the field test and again for the main data collection.

To complete layout verification, each country submits the following documentation for each language in which they are administering the assessment:

- A set of all achievement booklets assembled in complete, ready-to-print booklet form
- Context questionnaires for school principals, teachers, and students in complete, ready-to-print booklet form
- National Adaptations Forms for both the achievement booklets and context questionnaires, including documentation of national adaptations and the feedback received from translation verification

For the TIMSS Advanced 2015 main data collection, layout verification was completed for each of the 9 participating countries. With one country administering the assessment in two languages, the TIMSS & PIRLS International Study Center reviewed a total of 10 sets of national TIMSS Advanced 2015 assessment instruments (each set including achievement booklets and context questionnaires). A list of assessment instruments and the languages in which they were administered in each of the participating countries can be found in Exhibit 7.1 of the TIMSS Advanced Translation and Translation Verification chapter of this volume.

Layout Verification of Achievement Booklets

The primary goal of layout verification is to ensure that students in different countries experience the assessment instruments in the same way. Thus, the national achievement booklets were checked against the international versions to identify any deviations from the international format.

Due to differences in languages, the national assessment instruments varied slightly in length and format. The international versions, however, were designed with this in mind, and extra space was provided in the margins of the pages to facilitate the use of longer text and different paper sizes (letter versus A4) without necessitating extensive changes to the layout of each page.

National Research Coordinators (NRCs) were directed to document all national adaptations (apart from direct translations) made to the achievement booklets within the achievement booklet National Adaptations Forms. During layout verification, the verifiers also checked the achievement items for international comparability while taking into consideration the national adaptations documented by the NRCs. Any layout deviations or errors, as well as any concerns of international incomparability of assessment items, were documented by the verifiers in the National Adaptations Forms.

Per the TIMSS Advanced assessment design, the TIMSS Advanced 2015 achievement instruments include blocks of items from TIMSS Advanced 2008 and TIMSS Advanced 1995. These "trend blocks" provide the foundation for the measurement of change in student achievement over time and therefore must be administered in the same way across subsequent cycles. As such, for countries that previously participated in TIMSS Advanced 2008 and/or TIMSS Advanced 1995, the TIMSS Advanced 2015 trend blocks were reviewed during the layout verification against those from the last cycle in which the country participated. Any deviations from the previous cycle were documented by the verifiers within the National Adaptations Forms.

Following layout verification, the National Adaptations Forms containing the verifiers' comments were sent back to the National Research Coordinators for consideration. The National Research Coordinators were asked to confirm that each of the suggested changes was implemented or provide an explanation for not implementing the suggested change.

Layout Verification of Context Questionnaires

As with the achievement booklets, the context questionnaires were also checked against the international versions to identify any potential layout issues as well as to ensure the international comparability of the questionnaire data.

In an effort to make the questionnaires general enough for international analyses but appropriate for each intended audience, participating countries were required to adapt certain phrases and designations in the text of the questionnaires. The text that requires country-specific adaptations is enclosed in brackets (e.g., <language of test>) in the international instruments. To assist the NRCs in finding comparable and appropriate substitutions for the bracketed text, the TIMSS & PIRLS International Study Center supplied documentation in one of the Survey Operations Units, providing explanations of the intended meaning of each bracketed text, and where applicable, offered examples to guide the National Research Coordinators in selecting appropriate replacements.

National Research Coordinators were directed to document all national adaptations made to the context questionnaires within the National Adaptations Forms. During the layout verification, the verifiers checked the instruments for international comparability, taking into consideration the national adaptations documented by the National Research Coordinators. Any adaptations or errors that were not internationally comparable were documented by the verifiers in the National Adaptations Forms along with recommendations for recoding or rewording.

Additionally, the verifiers ensured that all bracketed text, requiring country-specific adaptations, was properly documented with English back translations. The documentation for these universally adapted questionnaire items is intended for later use in the National Adaptations Database. The database is a compilation of each country's intended adaptations, to be used during data processing by the IEA Data Processing and Research Center, and the information included in the database is reported as a supplement to the user guide for the TIMSS Advanced 2015 International Database.

Similar to the layout verification process for the achievement items, layout verifiers provided the NRCs with feedback through the National Adaptations Forms, and the NRCs were asked to respond to the feedback by either confirming the implementation of the suggested modifications or providing an explanation as to why the changes were not applied.

CHAPTER 9

Quality Assurance Program for TIMSS Advanced 2015

Ieva Johansone
Erin Wry

Considerable effort has been made to develop standardized materials and survey operations procedures so that the TIMSS Advanced 2015 data meet the highest standards. To document data collection activities and verify that the standardized TIMSS Advanced procedures were followed, the TIMSS & PIRLS International Study Center, working in coalition with the IEA Secretariat, developed and implemented an ambitious International Quality Assurance Program. The purpose of this chapter is to provide an overview of the International Quality Assurance Program and report on the data collected through the program.

Overview

The International Quality Assurance Program was implemented by independent International Quality Control Monitors (IQCMs) appointed by the IEA Secretariat. The major task of the IQCMs was to conduct site visits during the data collection process and report on their observations. In each country, the IQCM visited a sample of 24 participating schools during the testing sessions, 12 for advanced mathematics and 12 for physics.

For each school visit, IQCMs observed the testing session and recorded their observations, noting any deviations from the standardized administration script, timing, and procedures. In addition, IQCMs interviewed the School Coordinators about their experiences coordinating the TIMSS Advanced assessment. IQCMs also checked whether the suggestions made by the international translation and layout verifiers had been integrated into the final assessment instruments, as documented in the National Adaptations Forms.

Prior to beginning their assignments, the IQCMs were mandated to attend a training session conducted by the TIMSS & PIRLS International Study Center. During the training, IQCMs were introduced to the TIMSS Advanced survey operations procedures and the design of the TIMSS Advanced 2015 achievement booklets and context questionnaires. IQCMs were also supplied with a manual detailing their role and responsibilities as well as the necessary materials for completing the quality control tasks.

An important aspect of the International Quality Assurance Program is the independence of the IQCMs from the national centers. The Quality Control Monitor could not be a member of the national center, or a family member or personal friend of the NRC. Often, the ICQM was a school inspector, ministry official, or retired school teacher. The IQCM was required to be fluent in both English and the language(s) spoken in the country. In most participating countries, the IEA Secretariat recruited IQCMs who had served in the same role in previous IEA assessments. For the remaining countries, National Research Coordinators assisted the IEA Secretariat in nominating an International Quality Control Monitor.

When necessary, the IQCMs were permitted to recruit assistants in order to effectively cover the territory and testing timetable. One TIMSS Advanced IQCM was trained for each of the participating countries. In addition, the IQCMs trained 23 assistant monitors to assist them. Altogether, Quality Control Monitors observed 108 advanced mathematics testing sessions and 110 physics testing sessions. The results of the TIMSS Advanced 2015 IQCM observations are reported in the following sections of this chapter.

Quality Control Observations of the TIMSS Advanced 2015 Data Collection

International Quality Control Monitors (IQCMs) conducted site visits during TIMSS Advanced test administration to a sample of 24 participating schools (12 per subject) in each country. For each school visit, the IQCMs completed the TIMSS Advanced 2015 Classroom Observation Record. For purposes of reporting, the advanced mathematics records were combined with the physics records.

The observation records were organized into four sections:

- Section A—Documentation of the TIMSS Advanced Testing Session
- Section B—Summary Observations of the TIMSS Advanced Administration
- Section C—Student Questionnaire Administration
- Section D—Interview with the School Coordinator

Documentation and Summary Observations of the TIMSS Advanced 2015 Testing Sessions

Sections A and B of the Classroom Observation Record addressed activities that took place during the actual testing sessions. The achievement test was administered in 90 minutes. During test administration, IQCMs were asked to observe the activities of the Test Administrator, specifically the following:

- Distributing, collecting, and securing the test booklets
- Following the assessment administration script
- Making time announcements during the testing sessions

As shown in Exhibit 9.1, the IQCMs reported that the TIMSS Advanced assessments were conducted in accordance with the international procedures, particularly, in regard to booklet distribution and security. In 99% of the observations, the test booklets were distributed according to the booklet assignment on the Student Tracking Forms and secured immediately following the testing session. Furthermore, in 87% of the observations the total testing time was equal to the time allowed. In the cases where the testing time was not equal to the time allowed, this was usually due to students finishing early. IQCMs reported that in nearly 80% of the observed testing administrations the students finished the assessment in less than the 90 minutes allotted.

Exhibit 9.1: Observations of TIMSS Advanced 2015 Assessment Administration Sessions—218 Advanced Mathematics and Physics Sessions (Percent of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered (%)
Did the Test Administrator distribute the test booklets according to the booklet assignment on the *Student Tracking Form* and booklet labels?	99	1	0
Did the total testing time of the testing session equal the time allowed?	87	13	0
Did the Test Administrator announce "you have 10 minutes left" prior to the end of the testing session?	91	9	0
Were there any other time remaining announcements made during the testing session?	20	79	1
Did any students finish the TIMSS Advanced assessment early (before the time allowed was up)?	79	21	0
Did the test administrator have a watch with a seconds hand, a stopwatch, or a timer for accurately timing the testing session?	86	13	1
Were the booklets collected and secured after the testing session?	99	1	0

If Test Administrators observed students working faster than expected, a remaining-time announcement was made prior to the planned 10 minute warning to inform students that they still had ample time to complete their work.

Exhibit 9.2 reports on the activities conducted during the assessment sessions. One of the most important methods of standardizing the assessment administration was to have all test administrators follow the script provided in the TIMSS Advanced Test Administrator Manual. IQCMs reported that in 66% of the observations, the Test Administrators followed the script exactly. In the circumstances in which the Test Administrator deviated from the script, nearly all modifications were described as "minor." In addition, the IQCMs reported that the test administrators appropriately addressed student questions, according to the guidelines outlined in the test administrator manual, in 98% of the observed sessions.

Exhibit 9.2: Following the Test Administration Script—218 Advanced Mathematics and Physics Sessions (Percent of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered (%)
Had the test administrator familiarized himself or herself with the test administration script prior to the testing?	88	10	2 *(I Cannot Answer)* 0 *(Not Answered)*
Did the test administrator follow the test administration script in the Test Administrator Manual?	66	24 *(Minor changes)* 10 *(Major changes)*	0
If the Test Administrator made changes to the script, how would you describe them?			
Additions	19	20	0 *(Not Answered)* 61 *(Not Applicable)*
Revisions	23	15	0 *(Not Answered)* 62 *(Not Applicable)*
Deletions	15	24	0 *(Not Answered)* 61 *(Not Applicable)*
Did the test administrator address student questions appropriately?	98	2	0

Exhibit 9.3 presents observations on student compliance with instructions and overall cooperation during the assessment administration. According to the IQCMs' observations, in almost all of the sessions, students complied well or very well with the instruction to stop work at the end of testing session. In addition, IQCMs described the students as mostly orderly and cooperative during the testing sessions, with 70% of the observations cited as having students who were extremely orderly and cooperative.

Exhibit 9.3: Student Cooperation During Assessment Administration—218 Advanced Mathematics and Physics Sessions (Percent of IQCM Responses)

Question	Very Well (%)	Fairly Well (%)	Not well at all (%)	Not Answered (%)	
When the Test Administrator ended the testing session, how well did the student comply with the instruction to stop work?	85	12	2	1	
	Extremely (%)	**Moderately (%)**	**Somewhat (%)**	**Hardly (%)**	**Not answered (%)**
To what extent would you describe the students as orderly and cooperative?	70	24	6	0	0

Summary Observations of the TIMSS Advanced 2015 Testing Sessions

Exhibit 9.4 reports on the IQCMs' general observations of the TIMSS Advanced assessment administration. Overall, IQCMs reported that the quality of testing sessions was good, very good or excellent in 93% of their observations. In most of the testing sessions the IQCMs observed, no participation problems occurred; however, in 4% of the cases the IQCMs recorded several students refusing to take the test. These cases were mostly credited to time conflicts with other classes or other tests occurring within or near the same timeframe.

The IQCMs reported few problems with the instruments themselves. In only 1% of the observed testing sessions was a defective booklet detected after the testing began, and in 100% of the sessions the student identification information on the booklets corresponded exactly with the Student Tracking Form. Furthermore, 92% of the observed TIMSS Advanced test administrations took place under favorable room conditions that were suitable for students to work without distraction. For students requiring special accommodations, accommodations were implemented in 6% of the observed testing sessions. The most common special accommodations provided included booklets with larger print, extended time, use of a magnifying glass, and a separate room for increased focusing ability.

Exhibit 9.4: General Observations of the Testing Session—218 Advanced Mathematics and Physics Sessions (Percent of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered (%)
Did the student identification information on the booklets correspond with the *Student Tracking Form*?	100	0	0
Were any defective test booklets detected and replaced?	0 *(BEFORE the testing began)* 1 *(AFTER the testing began)*	100 *(BEFORE the testing began)* 97 *(AFTER the testing began)*	0 *(BEFORE the testing began)* 2 *(AFTER the testing began)*
If any defective test booklets were replaced, did the Test Administrator replace them appropriately?	2	2	0 *(Not Answered)* 96 *(Not Applicable)*
Did any students refuse to take the test?	4	96	0
If a student refused, did the Test Administrator accurately follow the instructions for excusing the student?	4	0	0 *(Not Answered)* 96 *(Not Applicable)*
Were any late students admitted to the testing room?	8 *(BEFORE the testing began)* 5 *(AFTER the testing began)*	82 *(There were no late students)* 5 *(Late students were not admitted)*	0
Did any students leave the room for an "emergency" during the testing?	28	72	0
If a student left the room for an emergency during the testing, did the Test Administrator address the situation appropriately (collect the test booklet, and if re-admitted, return the test booklet)?	17	11	0 *(Not Answered)* 72 *(Not Applicable)*

Exhibit 9.4: General Observations of the Testing Session—218 Advanced Mathematics and Physics Sessions (Percent of IQCM Responses) (Continued)

Question	Yes (%)	No (%)	Not Answered (%)
Were there any students requiring special accommodations (e.g., students with visual or hearing impairment, Dyslexia)?	6	94	0
Did students store away everything, including all electronic devices, having only a pen or a pencil and the test booklet for the duration of the test administration?	92	8	0
Were the conditions in the testing room suitable (lighting, temperature, noise, etc.) for the students to work without distractions?	92	8	0
Did the seating arrangement provide adequate space for students to work and not be distracted by each other?	94	6	0

Question	Excellent (%)	Very Good (%)	Good (%)	Fair (%)	Poor (%)	Not Answered (%)
In general, how would you describe the overall quality of the testing session?	48	30	15	4	3	0

Student Questionnaire Administration

Exhibit 9.5 summarizes the IQCMs' observations of the Student Questionnaire administration. In order to link the achievement scores to Student Questionnaire data, it is essential that the students receive the correct student questionnaire. IQCMs reported that the Student Questionnaires were distributed according to the Student Tracking Forms and questionnaire labels in 94% of the observed sessions. In some cases, the Student Questionnaire was attached to the student's booklet and therefore the Student Tracking Form was not needed. In 65% of the observed cases, Test Administrators followed the Student Questionnaire administration script exactly. If the Test Administrator deviated from the script, most frequently the modifications were "minor." It should be noted that some schools chose to administer the questionnaire on a different date than the TIMSS Advanced achievement booklets, and in these cases, IQCMs were not required to observe the Student Questionnaire administration.

Exhibit 9.5: Student Questionnaire Administration—218 Advanced Mathematics and Physics Sessions (Percent of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered (%)
When the test administrator read the script to end the assessment session followed by the Student Questionnaire administration, did the test administrator announce a break?	76	22	2
Did the Test Administrator distribute the Student Questionnaires according to the *Student Tracking Form* and questionnaire labels?	94	2	4 *(Not Applicable)*
Did the test administrator follow the questionnaire administration script in the Test Administrator Manual?	65	21 *(Minor changes)* 10 *(Major changes)*	0 *(Not Answered)* 4 *(Not Applicable)*
If the Test Administrator made changes to the script, how would you describe them?			
Additions	14	24	0 *(Not Answered)* 62 *(Not Applicable)*
Revisions	21	17	0 *(Not Answered)* 62 *(Not Applicable)*
Deletions	12	25	0 *(Not Answered)* 63 *(Not Applicable)*

Interview with the School Coordinator

Section D was the final component of the Classroom Observation Record and involved the IQCM conducting an interview with the School Coordinator. The interview addressed issues such as the following:

- Shipment of assessment materials
- Arrangements for test administration
- Responsiveness of the national center to queries
- Necessity for make-up sessions
- Organization of classes in the school (to validate within-school sampling procedure)

As shown in Exhibit 9.6, 90% of the School Coordinators considered that the TIMSS Advanced 2015 administration in their school went very well overall. Additionally, 89% of the School Coordinators felt that the provided School Coordinator Manual worked well and did not need improvement. The overall attitude of the other school staff members was regarded as mostly positive (73%) or neutral (24%). Several School Coordinators cited some resistance towards testing in general but not specific to TIMSS Advanced.

Exhibit 9.6: Interview with the School Coordinator, Overview—218 Advanced Mathematics and Physics Sessions (Percent of School Coordinator Responses)

Question	Very well, no problems (%)	Satisfactorily, few problems (%)	Unsatisfactorily, many problems (%)	Not Answered (%)
Overall, how would you say the testing went?	90	10	0	0
	Positive (%)	Neutral (%)	Negative (%)	Not Answered (%)
Overall, how would you rate the attitude of the other school staff members towards the TIMSS Advanced testing?	73	24	3	0
	Worked well (%)	Needs improvement (%)	Not Answered (%)	
Overall, do you feel the School Coordinator Manual worked well for you or does it need improvement?	89	11	0	

In 92% of the interviews, the School Coordinators felt the national center was responsive to their questions and concerns, as show in Exhibit 9.7. Additionally, nearly half of the School Coordinators reported that the School Questionnaire and Teacher Questionnaire were administered online with very few problems.

Because the sampling of classes requires a complete list of all classes in the school at the target grade, IQCMs were also asked to verify that the class list did indeed include all classes. Almost all of the School Coordinators confirmed that the complete list of classes had been documented and that all students appeared in one and only one of these classes. In 32% of the schools, the School Coordinator anticipated the need for a makeup session, and almost all of these coordinators intended to conduct one.

As a reflection of the successful planning and implementation of TIMSS Advanced 2015, 91% of respondents affirmed that they would be willing to serve as a School Coordinator in future international assessments. Finally, it is notable that the response rate for the Classroom Observation Records was considerably high on all questions, with only a handful of questions going unanswered.

Exhibit 9.7: Interview with the School Coordinator, Details—218 Advanced Mathematics and Physics Sessions (Percent of School Coordinator Responses)

Question	Yes (%)	No (%)	Not Answered (%)
Prior to the testing day, did you have time to check your shipment of materials from the national center?	73	24	3
Did you receive the correct shipment of the materials as listed in your School Coordinator Manual and according to the tracking forms?	74	19	7
If no, did the national center provide the missing materials in time for the testing?	1	13	0 *(Not Answered)* 86 *(Not Applicable)*
Was the national center responsive to your questions or concerns?	92	5	3
Was the Teacher Questionnaire administered online?	42	56	2
If the Teacher Questionnaire was administered online, did the teacher(s) encounter any problems?	3	50	0 *(Not Answered)* 47 *(Not Applicable)*
Was the School Questionnaire administered online?	43	56	1
If the School Questionnaire was administered online, did the person completing it encounter any problems?	3	49	0 *(Not Answered)* 48 *(Not Applicable)*
Do you anticipate that a makeup session will be required at your school?	32	67	1
If yes, do you intend to conduct one?	30	3	0 *(Not Answered)* 67 *(Not Applicable)*
Did the students receive any special instructions, motivational talk, or incentives to prepare them for the assessment?	65	35	0
Is this a complete list of the classes in this grade in this school?	97	3	0
To the best of your knowledge, are there any students in the target population who are not in any of these classes?	6	93	1
To the best of your knowledge, are there any students in the target grade in more than one of these classes?	9	90	1
If there was another international assessment, would you be willing to serve as a School Coordinator?	91	9	0

CHAPTER 10

Creating the TIMSS Advanced 2015 International Database

Sebastian Meyer
Mark Cockle
Milena Taneva

This chapter describes the procedures implemented by the IEA Data Processing and Research Center (IEA DPC) for checking the TIMSS Advanced 2015 data and creating the TIMSS Advanced 2015 International Database (IDB).

Preparing the TIMSS Advanced 2015 International Database and ensuring its integrity was a complex endeavor requiring extensive collaboration between the IEA Data Processing and Research Center, the TIMSS & PIRLS International Study Center, Statistics Canada, and the national centers of participating countries. Once the countries had created their data files and submitted them to the IEA DPC, an exhaustive process of checking and editing known as "data cleaning" began.

Data cleaning is the process of checking data for inconsistencies and formatting the data to create a standardized output. The overriding concerns of the data cleaning process were to ensure:

- All information in the database conformed to the internationally defined data structure
- The content of all codebooks and documentation appropriately reflected national adaptations to questionnaires
- All variables used for international comparisons were in fact comparable across countries (after harmonization, where necessary)
- All institutions involved in this process applied quality control measures throughout in order to assure the quality and accuracy of the TIMSS Advanced 2015 data

The IEA DPC was responsible for checking the data files from each country, applying standardized data cleaning rules to verify the accuracy and consistency of the data and documenting any deviations from the international file structure. Data files were created at each country's national center and reviewed prior to submission to the IEA DPC. The National Research Coordinators (NRCs) collaborated with the IEA DPC to resolve any queries which emerged during the data cleaning process, and the NRCs checked interim versions of the national participant database(s) produced by the IEA DPC. The TIMSS & PIRLS International Study Center provided the NRCs with univariate data almanacs containing summary statistics on each variable so that the national centers could evaluate their data from an international perspective.

The TIMSS & PIRLS International Study Center also scaled the achievement and background data, as documented in Chapter 13: Scaling the TIMSS Advanced 2015 Achievement Data, and produced achievement scores (plausible values) and scores on the background scales. Using the Within-School Sampling Software (WinW3S)[1] database and response data provided by the IEA DPC, Statistics Canada in collaboration with the IEA DPC calculated the sampling weights, population coverage, and school and student participation rates—as documented in Chapter 3 and Chapter 5.

Data Sources

Data Entry and Verification of Paper Questionnaires

Each national center was responsible for inputting the information collected in test booklets and paper-based questionnaires into computer data files using the IEA Data Management Expert (DME) software. The DME is a software system developed by the IEA DPC that facilitates data entry and includes validation checks to identify inconsistencies. As a general rule of thumb, national centers were instructed to enter data for any questionnaire that contained at least one valid response, discarding unused or empty questionnaires.

National centers entered responses from the paper instruments into data files using a predefined international codebook. The codebook contained information about the names, lengths, labels, valid ranges for continuous measures or counts or valid values for nominal or ordinal questions, and missing codes for each variable.

As documented in Chapter 7: Translation and Translation Verification, countries participating in TIMSS Advanced are expected to make national adaptations to certain questions in the international questionnaires (e.g., the questions about parents' education must be adapted to the national context). Countries making such adaptations were required to adapt the codebook structure to reflect the adaptations made to the national questionnaire versions before beginning the data entry process.

To ensure consistency across participating countries, the basic rule for data entry in the DME required national staff to enter data "as is" without any interpretation, correction, truncation, imputation, or cleaning.

The rules for data entry included the following:

- Responses to closed response items coded as "1" if the first option was used, "2" if the second option is marked, and so on

1 WinW3S is a software developed by the IEA DPC that stores participation information at school, teacher, class, and student levels in a relational database while maintaining a hierarchical ID system. The software allows users to perform all necessary within-school sampling according to the TIMSS Advanced standards, and also provides some data validation in and across these levels.

- Responses to open response questions, for example number of students in the TIMSS Advanced class, entered "as is" even if the value is outside the originally expected range
- Responses to filter questions and filter-dependent questions entered exactly as filled in by the respondent, even if the information provided is logically inconsistent
- Non-response, ambiguous responses, responses given outside of the expected format, or conflicting responses (e.g., selection of two options in a multiple-choice question), coded as "omitted or invalid"

As each respondent ID number was entered it was checked by the DME software for alignment with a five-digit checksum generated by WinW3S. A mistype in either the ID or the checksum resulted in an error message prompting the data-entry person to check the entered values. The data-verification module of DME also checked for a range of other issues such as inconsistencies in identification codes and out-of-range or otherwise invalid codes. When such issues were flagged by the software, the individuals entering the data were prompted to resolve the inconsistency or confirm that an issue existed before resuming data entry.

Double-Data Entry

To check data entry reliability in participating countries, national centers were required to enter a 5% sample of each survey instrument (achievement booklet or questionnaire) twice by two different data entry persons (punchers). The IEA DPC recommended that countries begin the double-data entry process as early as possible during the data capture period in order to identify possible systematic misunderstandings or mishandlings of data-entry rules and to initiate appropriate remedial actions—for example, retraining national center staff. Those entering the data were required to resolve discrepancies between the first and second data entries by consulting the original questionnaire and applying the international rules in a uniform way.

While it was desirable that each and every discrepancy be resolved before submission of the complete dataset, the acceptable level of disagreement between the originally entered and double-entered data was established at 1 percent or less for questionnaire data and at the 0.1 percent or less level for achievement data. Values above this level required a complete re-entry of data.

The level of disagreement between the originally entered and double-entered data was evaluated by the IEA DPC, and it was found that in general the margin of error observed for processed data was well below the required threshold.

Data Verification at the National Centers

Before sending the data to the IEA DPC for further processing, national centers carried out mandatory validation and verification steps on all entered data and undertook corrections as necessary.

While the questionnaire data were being entered, the data manager or other staff at each national center used the information from the Teacher Tracking Forms to verify the completeness of the materials. Student participation information (e.g., whether a student participated in the assessment or was absent) was entered via WinW3S.

The validation process was supported by an option in WinW3S to generate an inconsistency report. This report listed all of the types of discrepancies between variables recorded during the within-school sampling and test administration process and made it possible to cross-check these data against data entered in the DME, the database for online respondents, and the uploaded student data on the central international server.

Data managers were requested to resolve such issues before final data submission to the IEA DPC. If inconsistencies remained or the national center could not solve them, the DPC asked the center to provide documentation on these problems.

As well as submitting the validated data to the IEA DPC, NRCs also provided extensive documentation. In addition to documentation on inconsistencies, national centers submitted hard copies or electronic scans of all original student and Teacher Tracking Forms, Student-Teacher Linkage Forms, and when applicable a report on data-capture activities collected as part of the online Survey Activities Questionnaire.

Data from Online Questionnaire Administration

As documented in Chapter 6: Survey Operations Procedures, national centers had the option of administering the principal and teacher questionnaires online instead of, or in addition to, using paper-based questionnaires.

To ensure confidentiality, national centers provided every respondent with a letter that contained individual login information along with information on how to access the online questionnaire. This login information corresponded to the ID and checksum provided from WinW3S, meaning that the identity validation step occurring at the national centers for paper-based questionnaires occurred when the respondents' logged-in to the survey. Also, since responses were collected in digital format and stored directly on the IEA DPC server, there was no need for data entry, reducing the workload for national centers.

As a further advantage of online administration, the data tended to have fewer inconsistencies when compared with the data collected through the paper-based questionnaires, mitigating the number of issues needed to be resolved by the IEA DPC and the national centers. This is partly because, to some extent, the online system does not allow inconsistent response patterns. For example, if the directions ask the respondent to "Check one circle for each line," the system does not allow the respondent to check more than one response category on each line.

The TIMSS Advanced 2015 online questionnaires also include skip logic, which minimized response burden and improved data consistency. The TIMSS Advanced questionnaires have a number of questions that filter out respondents—meaning the subsequent questions are not applicable given the response to the filter question. For example, Question 8a of the school questionnaire reads "Does your school have a school library? If yes, go to 8b, and if no, go to 9." If a respondent chooses "No", the online survey skips directly to Question 9, omitting Questions 8b and 8c. Not only does the skip logic save the respondents' time, it also results in fewer inconsistencies in the data received by the IEA DPC.

Cleaning the International and National Databases

Overview

In order to ensure the integrity of the international database, a uniform data cleaning process was followed, involving regular consultation between the IEA Data Processing and Research Center and the NRCs. After each country had submitted its data, codebooks, and documentation, the DPC, in collaboration with the NRCs, conducted a four-step cleaning procedure upon the submitted data and documentation:

1. A structural check
2. A check of the identification (ID) variables
3. Linkage cleaning
4. Background cleaning

Data cleaning was an iterative process. Numerous iterations of the four-step cleaning procedure were completed on each national data set. This repetition ensured that all data were properly cleaned and that any new errors that could have been inadvertently introduced during the data cleaning were rectified. The cleaning process was repeated as many times as necessary until all data were made consistent and comparable. Any inconsistencies detected during the cleaning process were resolved in collaboration with national centers, and all corrections made during the cleaning process were documented in a cleaning report, produced for each country.

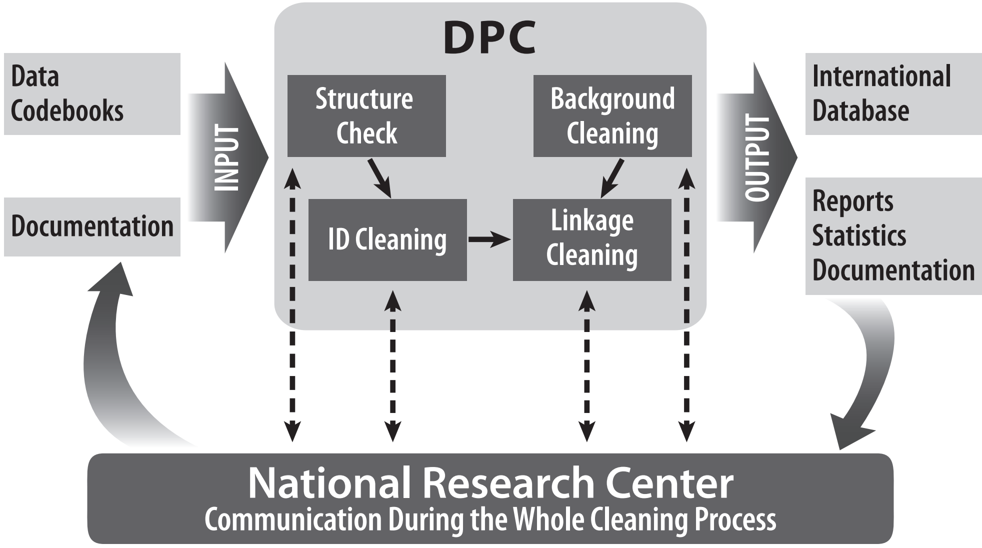
After the final cleaning iteration, each country's data were sent to Statistics Canada for the calculation of sampling weights, and then the data, including sampling weights, were sent to the TIMSS & PIRLS International Study Center so that scaling could be performed. The NRCs were provided with interim data products to review at two different points in the process.

Preparing National Data Files for Analysis

The main objectives of the data cleaning process were to ensure that the data adhered to international formats, that school, teacher, and student information could be linked across various data files, and that the data reflected the information collected within each country in an accurate and consistent manner.

As illustrated in Exhibit 10.1, the program-based data cleaning consisted of a set of activities explained in the following subsections. The IEA DPC carried out all of these activities in close communication with the national centers.

Exhibit 10.1: Overview of Data Processing at the IEA Data Processing and Research Center



Checking Documentation, Import, and Structure

For each country, data cleaning began with an exploratory review of its data-file structures and its data documentation, including a review of National Adaptation Forms, Student Tracking Forms, Teacher Tracking Forms, Student-Teacher Linkage Forms, and the Survey Activities Questionnaire.

The IEA DPC first merged the tracking information and sampling information captured in the WinW3S database with the student-level database containing the corresponding student survey instrument data. During this step, IEA DPC staff also merged the data from the school and teacher questionnaires for both the online and paper modes of administration. At this stage, data from the different sources were transformed and imported into one structured query language (SQL) database so that this information would be available during all further data-processing stages.

The first checks identified differences between the international and the national file structures. Some countries made adaptations (such as adding national variables or omitting or modifying international variables) to their questionnaires. The extent and nature of such changes differed across countries: some countries administered the questionnaires without any modifications (apart from translations and necessary adaptations relating to cultural or language-specific terms), whereas other countries inserted response categories within existing international variables or added national variables.

To keep track of adaptations, staff at the TIMSS & PIRLS International Study Center asked the national centers to complete National Adaptation Forms while they were translating the international version of the survey instruments. Where necessary, the IEA DPC modified the structure and values of the national data files to ensure that the resulting data remained comparable across countries. Details about country-specific adaptations to the international instruments can be found in Supplement 2 of the *TIMSS Advanced 2015 User Guide for the International Database*.

The IEA DPC then discarded variables created purely for verification purposes during data entry, and made provision for adding new variables necessary for analysis and reporting, including reporting variables, derived variables, sampling weights, and scale scores.

Once IEA DPC staff had ensured that each data file matched the international format, they applied a series of standard data cleaning rules for further processing. Processing during this step employed software developed by the IEA DPC that could identify and correct inconsistencies in the data. Each potential problem flagged at this stage was identified by a unique problem number, and then described and recorded in a database. The action taken by the cleaning program or IEA DPC staff with respect to each problem was also recorded.

The IEA DPC referred problems that could not be rectified automatically through the program to the responsible NRC so that national center staff could check the original data-collection instruments and tracking forms to trace the source of these errors. Wherever possible, staff at the IEA DPC suggested a remedy and asked the national centers to either accept it or propose an alternative. If a national center could not solve an issue through verification of the instruments or forms, the IEA DPC applied a general cleaning rule to the files to rectify the error. When all automatic updates had been applied, IEA DPC staff used SQL recoding scripts to directly apply any remaining corrections to the data files.

Cleaning Identification Variables

Each record in a data file needs to have a unique identification number. The existence of records with duplicate ID numbers in a file implies an error of some kind. Some countries administered the school and teacher questionnaires online, in addition to the paper mode. This could yield the theoretical possibility that a respondent completed both the paper and the online versions of the questionnaire. If two records in a TIMSS Advanced 2015 database shared the same ID number and contained exactly the same data, the IEA DPC deleted one of the records and kept the other one in the database. In the rare case that both records contained different data and IEA DPC staff found it impossible to identify which record contained the "true data," national centers were asked which record to keep.

Although the ID cleaning covered all data from all instruments, it focused mainly on the student file. In addition to checking the unique student ID number, it was crucial to check variables pertaining to student participation and exclusion status, as well as students' dates of birth and dates of testing in order to calculate student age at the time of testing. The Student Tracking Forms provided an important tool for resolving anomalies in the database.

As mentioned previously, the IEA DPC conducted all cleaning procedures in close cooperation with the national centers. After national center staff had cleaned the identification variables, they passed the clean databases with information about student participation and exclusion on to Statistics Canada, which used this information to calculate students' participation rates, exclusion rates, and student sampling weights.

Checking Linkages

As data on students, parents, teachers, and schools appeared in a number of different data files, a process of linkage cleaning was implemented to ensure that the data files would correctly link together. The linking of the data files followed a hierarchical system of identification codes that included school, class, and student components. These codes linked the students with their class and/or school membership. Further information on linkage codes can be found in Chapter 6: Survey Operations Procedures.

Linkage cleaning consisted of a number of checks to verify that student entries matched between achievement files, student background files, and scoring reliability files. In addition, at this stage, checks were conducted to ensure that teacher and student records linked correctly with their corresponding schools. The Student Tracking Forms, Teacher Tracking Forms, and Student-Teacher Linkage Forms were crucial in resolving any anomalies. The IEA DPC also liaised with NRCs about any problematic cases, and the national centers were provided with standardized reports listing all inconsistencies identified within the data.

Resolving Inconsistencies in Questionnaire Data

The amount of inconsistent and implausible responses in questionnaire data files varied considerably across countries. The IEA DPC determined the treatment of inconsistent responses on a question-by-question basis, using all available documentation to make an informed decision. IEA DPC staff also checked all questionnaire data for consistency across the responses given. For example, Question 1 in the school questionnaire asked for the total school enrollment in all grades, while Question 2 asked for the enrollment in the target grade only. Logically, the number given as a response to Question 2 could not exceed the number provided by school principals in Question 1. Similarly, it is not possible that the amount of years a teacher has been teaching altogether (Question 1 in the teacher questionnaires) exceeds his/her age (Question 3 in the teacher questionnaires). The IEA DPC flagged inconsistencies of this kind and then asked the national centers to review these issues. IEA DPC staff recoded as "invalid" those cases that could not be corrected.

Filter questions, which appeared in some questionnaires, directed respondents to a particular subquestion. The IEA DPC applied the following cleaning rule to these filter questions and the dependent questions that followed: If a respondent answered "No" to Question 8a in the school questionnaire "Does your school have a school library?" IEA DPC recoded any responses to the dependent questions as "logically not applicable." Also, following the same example, if the filter question was omitted but at least one valid response was found in the dependent questions then the IEA DPC recoded the filter question to "Yes." This of course is only possible for dichotomous filter questions (e.g., with response options such "Yes/No").

The IEA DPC also applied what are known as split variable checks to questions where the answer was coded into several variables. For example, Question 6 in the student questionnaire asked students: "Do you have any of these things?" Student responses were captured in a set of ten variables, each one coded as "Yes" if the corresponding "Yes" option was filled in and "No" if the "No" option was filled in. Occasionally, students checked the "Yes" boxes but left the "No" boxes unchecked. Because, in these cases, it was clear that the unchecked boxes actually meant "No," these responses were recoded accordingly.

Resolving Inconsistencies Between Tracking Information and Questionnaire Data

Two different sets of TIMSS Advanced 2015 data indicated age and gender for students. The first set was the tracking information provided by the school coordinator or test administrator throughout the within-school sampling and test/questionnaire administration process. The second set comprised the actual responses given by students in the student questionnaires. In some cases, data across these two sets did not match and resolution was needed.

If the information on gender or birth year and month was missing in the student questionnaire but the student participated, this information, when available, was copied over from the tracking data to the questionnaire. If discrepancies were found between existing tracking and questionnaire gender and age data, the IEA DPC queried the case with the national center, and the national center investigated which source of information was correct.

Handling of Missing Data

Two types of entries were possible during the TIMSS Advanced 2015 data capture: valid data values and missing data values. Missing data can be assigned a value of omitted/invalid, or not administered during data capture. The IEA DPC applied additional missing codes to the data to facilitate further analyses. This process led to four distinct types of missing data in the international database:

- **Omitted or invalid:** The respondent had a chance to answer the question but did not do so, leaving the corresponding item or question blank. This code was also used if the response was uninterpretable or out-of-range.
- **Not administered:** This signified that the item or question was not administered to the respondent, which meant that the respondent could not read and answer the question. The not administered missing code was used for those student test items that were not in the set of assessment blocks administered to a student either deliberately (due to the rotation of assessment blocks) or, in a very few cases, due to technical failure or incorrect translations. This missing code was also used for those records that were included in the international database but did not contain a single response to one of the assigned questionnaires. In addition, the not administered code was used for individual

questionnaire items that a national center decided not to include in the country-specific version of the questionnaire.

- **Logically not applicable:** The respondent answered a preceding filter question in a way that made the following dependent questions not relevant to him or her.
- **Not reached:** This applied only to the individual items of the student achievement test and indicated those items that students did not attempt due to a lack of time. "Not reached" codes were derived as follows: First, the last answer given by a student in a session is identified. This could be either a valid or invalid response to an item. The first omitted response after this last answer is coded as "omitted", but all following responses to these items in the session are then coded as "Not reached". For example, the response pattern "1 9 4 2 9 9 9 9 9 9" (where "9" represents "omitted") is recoded to "1 9 4 2 9 R R R R R" (where "R" represents "Not reached").

Data Cleaning Quality Control

Because TIMSS Advanced 2015 was a large and highly complex study with very high standards for data quality, maintaining these standards required an extensive set of interrelated data-checking and data cleaning procedures. To ensure that all procedures were conducted in the correct sequence, that no special requirements were overlooked, and that the cleaning process was implemented independently of the persons in charge, the data quality control process included the following steps:

- **Thorough testing of all data cleaning programs:** Before applying the programs to real datasets, the IEA DPC applied them to simulation datasets containing all possible problems and inconsistencies
- **Registering all incoming data and documents in a specific database:** The IEA DPC recorded the date of arrival as well as specific issues requiring attention
- **Carrying out data cleaning according to strict rules:** Deviations from the cleaning sequence were not possible, and the scope for involuntary changes to the cleaning procedures was minimal
- **Documenting all systematic data recodings that applied to all countries:** The IEA DPC recorded all changes to data in the comprehensive cleaning documentation provided to national centers
- **Logging every "manual" correction to a country's data files in a recoding script:** Logging these changes, which occurred only occasionally, allowed IEA DPC staff to undo changes or to redo the whole manual-cleaning process at any later stage of the data cleaning process

- **Repeating, on completion of data cleaning for a country, all cleaning steps from the beginning:** This step allowed the IEA DPC to detect any problems that might have been inadvertently introduced during the data cleaning process
- **Working closely with national centers at various steps of the cleaning process:** The IEA DPC provided national centers with the processed data files and accompanying documentation so that center staff could thoroughly review and correct any identified inconsistencies

The IEA DPC compared national adaptations recorded in the documentation for the national datasets with the structure of the submitted national data files. IEA DPC staff then recorded any identified deviations from the international data structure in the national adaptation database and for the supplementary materials provided with the *TIMSS Advanced 2015 User Guide for the International Database*. Whenever possible, the IEA DPC recoded national deviations to ensure consistency with the international data structure.

Interim Data Products

Before the TIMSS Advanced International Databases were finalized, two major interim versions of the data files were sent to each country—each country receiving only its own data. The first version was sent as soon as the data could be considered "clean" as regards identification codes and linkage issues. Documentation, with a list of the cleaning checks and corrections made in the data, was included to enable the NRC to review the cleaning process before the 7th NRC meeting in Lisbon in December 2015. A second version of the data files was sent to countries when the weights and international achievement scores were available and had been merged with the data files. This version, containing only records that satisfied the sampling standards, allowed the NRCs to replicate the results presented in the international reports.

Interim data products were accompanied by detailed data processing and national adaptation documentation, codebooks, and summary statistics. The summary statistics, preliminary versions of the TIMSS Advanced 2015 Almanacs, were created by the TIMSS & PIRLS International Study Center and included weighted univariate statistics for all questionnaire variables for each country. For categorical variables, representing the majority of variables, the percentages of respondents choosing each of the response options were displayed. For continuous numeric variables, various descriptive statistics were reported, including the minimum, maximum, mean, standard deviation, median, mode, and percentiles. For both types of variables, the percentages of missing data were reported. Additionally, for the achievement items, the TIMSS & PIRLS International Study Center provided item analysis and reliability statistics listing information regarding the number of valid cases, percentages, percentage correct, Rasch item difficulty, scoring reliability, and so forth. These statistics were used for a more in-depth review of the data at the international and national levels in terms of plausibility, unexpected response patterns, etc.

Final Product—the TIMSS Advanced 2015 International Databases

The data cleaning effort implemented at the IEA DPC ensured that the TIMSS Advanced 2015 international databases contained high-quality data. More specifically, the process ensured that:

- Information coded in each variable was internationally comparable
- National adaptations were reflected appropriately in all variables
- All entries in the database could be successfully linked within and across levels
- Sampling weights and student achievement scores were available for international comparisons

Supplements to the *TIMSS Advanced 2015 International Database and User Guide* document all national adaptations made to questionnaires by individual countries and how they were handled in the data. The meaning of country-specific items also can be found in this supplement, as well as recoding requirements by the TIMSS & PIRLS International Study Center.

CHAPTER 11

Reviewing the TIMSS Advanced 2015 Achievement Item Statistics

Pierre Foy
Michael O. Martin
Ina V.S. Mullis
Liqun Yin
Kerry Cotter
Jenny Liu

The TIMSS & PIRLS International Study Center conducted a review of a range of diagnostic statistics to examine and evaluate the psychometric characteristics of each achievement item across the countries that participated in the TIMSS Advanced 2015 assessments. This review of item statistics is essential to the successful application of item response theory (IRT) scaling to derive student achievement scores for analysis and reporting. This review played a crucial role in the quality assurance of the TIMSS Advanced 2015 achievement data prior to scaling, making it possible to detect unusual item properties that could signal a problem or error for a particular country. For example, an item that was uncharacteristically easy or difficult, or had an unusually low discriminating power, could indicate a potential problem with either translation or printing. Similarly, a constructed response item with unusually low scoring reliability could indicate a problem with a scoring guide in a particular country. In the rare instances where such items were found, the country's translation verification documents and printed booklets were examined for flaws or inaccuracies and, if necessary, the item was removed from the international database for that country.

Statistics for Item Review

The TIMSS & PIRLS International Study Center computed item statistics for all achievement items in the 2015 assessments, including advanced mathematics (102 items) and physics (103 items). The item statistics for each of the participating countries were then carefully reviewed. Exhibits 11.1 and 11.2 show actual samples of the statistics calculated for a multiple-choice and a constructed response item, respectively.

Exhibit 11.1: Example International Item Statistics for a TIMSS Advanced 2015 Multiple-Choice Item

Trends in International Mathematics and Science Study - TIMSS Advanced 2015 Assessment Results - Physics
International Item Review Statistics (Unweighted)

Physics: Mechanics & Thermodynamics / Knowing (P4_02 - PA33088) Type: MC Key: A
Label: Force of Sun on two planets

				Percentages							Point Biserials								
Country	Cases	DIFF	DISC	P_A	P_B	P_C	P_D	P_E	P_OM	P_NR	PB_A	PB_B	PB_C	PB_D	PB_E	PB_OM	PB_NR	RDIFF	Flags
France	1323	28.7	0.49	28.7	49.9	9.1	12.0	.	0.3	0.0	0.49	-0.25	-0.13	-0.18	.	-0.00	.	0.21	__H_F_
Italy	1148	55.2	0.55	55.2	33.4	2.6	7.1	.	1.7	0.2	0.55	-0.39	-0.14	-0.20	.	-0.10	-0.04	-1.02	__E_F_
Lebanon	386	25.6	0.53	25.6	41.5	11.5	15.7	.	5.7	0.8	0.53	-0.16	-0.17	-0.21	.	-0.08	-0.09	0.51	__H___
Norway	831	65.8	0.36	65.8	25.2	4.7	3.0	.	1.3	0.0	0.36	-0.25	-0.13	-0.14	.	-0.11	.	-0.57	____F_
Portugal	590	56.2	0.50	56.2	29.9	4.2	9.5	.	0.2	0.2	0.50	-0.31	-0.15	-0.26	.	0.00	-0.03	-0.58	____F_
Russian Federation	1269	77.7	0.52	77.7	15.1	3.6	3.3	.	0.2	0.2	0.52	-0.40	-0.17	-0.23	.	-0.04	0.00	-1.30	__E_F_
Slovenia	369	75.6	0.43	75.6	16.3	2.4	5.7	.	0.0	0.0	0.43	-0.30	-0.12	-0.24	.	.	.	-0.98	__E_F_
Sweden	1232	54.8	0.48	54.8	28.4	5.4	4.0	.	7.4	0.1	0.48	-0.33	-0.14	-0.14	.	-0.13	-0.03	-0.44	__H_F_
United States	985	64.2	0.58	64.2	26.2	4.9	4.0	.	0.8	0.0	0.58	-0.42	-0.19	-0.22	.	-0.09	.	-0.87	__E_F_
International Avg (9)	8133	56.0	0.49	56.0	29.5	5.4	7.1	.	2.0	0.2	0.49	-0.31	-0.15	-0.20	.	-0.07	-0.03	-0.56	____F_

Keys: DIFF= Percent correct score; DISC= Item discrimination; P_A...P_D= Percentage choosing each option; P_OM, P_NR= Percentage Omitted, Not Reached;
PB_A...PB_D= Point Biserial for each option; PB_OM, PB_NR= Point Biserial for Omitted, Not Reached; RDIFF= Rasch difficulty.
Flags: A= Attractive distractor; C= Difficulty less than chance; D= Negative/low discrimination; E= Easier than average;
F= Distractor chosen by less than 10%; H= Harder than average; V= Difficulty greater than 95%.

Exhibit 11.2: Example International Item Statistics for a TIMSS Advanced 2015 Constructed Response Item

Trends in International Mathematics and Science Study - TIMSS Advanced 2015 Assessment Results - Adv. Mathematics
International Item Review Statistics (Unweighted)

Mathematics: Algebra / Knowing (M2_02 - MA33225) Type: CR 2 Points
Label: Polynomials satisfying conditions (DERIVED)

Country	Cases	DIFF	DISC	Percentages P_0	P_1	P_2	P_OM	P_NR	Point Biserials PB_0	PB_1	PB_2	PB_OM	PB_NR	RDIFF	Reliability N	Score	Code	Flags
France	1328	52.0	0.46	37.8	15.1	44.4	2.7	0.0	-0.40	-0.04	0.44	-0.06	.	-0.64	.	.	.	______
Italy	1103	39.8	0.50	44.2	18.9	30.3	6.5	0.0	-0.34	-0.12	0.52	-0.10	.	-0.52	.	.	.	______
Lebanon	385	69.7	0.40	21.6	12.7	63.4	2.3	0.0	-0.29	-0.10	0.39	-0.21	.	-0.64	.	.	.	______
Norway	853	33.2	0.32	47.1	32.1	17.1	3.6	0.0	-0.23	0.01	0.32	-0.05	.	0.08	.	.	.	__H___
Portugal	1365	52.1	0.47	35.8	20.1	42.1	2.1	0.0	-0.37	-0.11	0.48	-0.06	.	-0.63	.	.	.	______
Russian Federation	2522	54.5	0.56	38.5	7.7	50.7	3.1	0.0	-0.47	-0.13	0.57	-0.11	.	-0.33	.	.	.	__H___
Russian Federation 6hr+	1151	65.6	0.53	29.3	6.2	62.6	2.0	0.0	-0.46	-0.12	0.53	-0.16	.	-0.29	.	.	.	__H___
Slovenia	981	74.5	0.37	19.8	11.1	68.9	0.2	0.0	-0.33	-0.10	0.36	-0.03	.	-1.42	.	.	.	__E___
Sweden	1321	41.0	0.51	48.2	17.9	32.0	1.9	0.1	-0.44	-0.01	0.49	-0.03	0.04	-0.45	.	.	.	__H___
United States	977	60.8	0.49	31.3	14.7	53.5	0.5	0.3	-0.41	-0.14	0.50	-0.06	-0.05	-0.74	.	.	.	__E___
International Avg (9)	10835	53.1	0.45	36.0	16.7	44.7	2.5	0.0	-0.37	-0.08	0.45	-0.08	-0.01	-0.59	.	.	.	______

Keys: DIFF= Percent correct score; DISC= Item discrimination; P_0...P_2= Percentage obtaining score level; P_OM, P_NR= Percentage Omitted, Not Reached;
PB_0...PB_2= Point Biserial for score level; PB_OM, PB_NR= Point Biserial for Omitted, Not Reached; RDIFF= Rasch difficulty;
Reliability: N= Responses double scored; Score= Percentage agreement on score; Code= Percentage agreement on code.
Flags: A= Point Biserials not ordered; C= Difficulty less than chance; D= Negative/low discrimination; E= Easier than average;
F= Score obtained by less than 10%; H= Harder than average; R= Scoring reliability less than 85%; V= Difficulty greater than 95%.

For all items, regardless of format (i.e., multiple-choice or constructed response), statistics included the number of students that responded in each country, the difficulty level (the percentage of students that answered the item correctly), and the discrimination index (the point-biserial correlation between success on the item and total score).[1] Also provided was an estimate of the difficulty of the item using a Rasch one-parameter IRT model. Statistics for each item were displayed alphabetically by country, together with an international average—i.e., based on all participating countries listed above the international average for each statistic. The international averages of the item difficulties and item discriminations served as guides to the overall statistical properties of the items.

Statistics displayed for multiple-choice items included the percentage of students that chose each response option—as well as the percentage of students that omitted or did not reach the item—and the point-biserial correlations for each response option. Statistics displayed for constructed response items (which could have 1 or 2 score points) included the percent correct and point-biserial of each score level. Constructed response item tables also provided information about the reliability with which each item was scored in each country, showing the total number of double-scored responses, the percentage of score agreement between the scorers, and—because TIMSS Advanced has a 2-digit scoring scheme—the percentage of code agreement between scorers.

During item review, “not-reached” responses (i.e., items toward the end of the booklet that the student did not attempt)[2] were treated as “not administered” and thus did not contribute to the calculation of the item statistics. However, the percentage of students not reaching each item was reported. Omitted responses, although treated as incorrect, were tabulated separately from incorrect responses for the sake of distinguishing students who provided no form of response from students who attempted a response.

The definitions and detailed descriptions of the statistics that were calculated are given below. The statistics were calculated separately by subject, and within each table are listed in order of their appearance in the item review outputs:

CASES: This is the number of students to whom the item was administered. Not-reached responses were not included in this count.

DIFF: The item difficulty is the average percent correct on an item. For a 1-point item, including all multiple-choice items, it is the percentage of students providing a fully correct response to the item. For 2-point items, it is the average percentage of points. For example, if 25 percent of students scored 2 points, 50 percent scored 1 point on a 2-point item, and the other 25 percent score 0 points, then the average percent correct for such an item would be 50 percent. For this statistic, not-reached responses were not included.

1 For computing point-biserial correlations, the total score is the percentage of points a student has scored on the items (s)he was administered. Not-reached responses are not included in the total score.

2 An item was considered “not-reached” if the item itself and the item immediately preceding it were not answered and no subsequent items had been attempted.

DISC: The item discrimination is computed as the correlation between the response to an item and the total score on all items administered to a student. Items exhibiting good measurement properties should have a moderately positive correlation, indicating that the more able students get the item right, the less able get it wrong. For this statistic, not-reached items were not included.

PCT_A, PCT_B, PCT_C, PCT_D, and PCT_E: Available for multiple-choice items. Each column indicates the percentage of students choosing the particular response option for the item (A, B, C, D, or E). Not-reached responses were excluded from the denominator.

PCT_0, PCT_1, and PCT_2: Available for constructed response items. Each column indicates the percentage of students responding at that particular score level, up to and including the maximum score level for the item. Not-reached items were excluded from the denominator.

PCT_OM: Percentage of students who, having reached the item, did not provide a response. Not reached responses were excluded from the denominator.

PCT_NR: Percentage of students who did not reach the item. This statistic is the number of students who did not reach an item as a percentage of all students who were administered that item, including those who omitted or did not reach that item.

PB_A, PB_B, PB_C, PB_D, or PB_E: Available for multiple-choice items. These columns show the point-biserial correlations between choosing each of the response options (A, B, C, D, or E) and the total score on all of the items administered to a student. Items with good psychometric properties have moderately positive correlations for the correct option and negative correlations for the distracters (the incorrect options). Not-reached responses were not included in these calculations.

PB_0, PB_1, and PB_2: Available for constructed response items. These columns present the point-biserial correlations between the score levels on the item (0, 1, or 2) and the overall score on all of the items the student was administered. For items with good measurement properties, the correlation coefficients should monotonically increase from negative to positive as the score on the item increases. Not-reached responses were not included in these calculations.

PB_OM: The point-biserial correlation between a binary variable indicating an omitted response to the item, and the total score on all items administered to a student. This correlation should be negative or near zero. Not-reached responses were not included in this statistic.

PB_NR: The point-biserial correlation between a binary variable indicating a not-reached response to the item, and the total score on all items administered to a student. This correlation should be negative or near zero.

RDIFF: An estimate of the difficulty of an item based on a Rasch one-parameter IRT model applied to the achievement data of a given country. The difficulty estimate is expressed in the logit metric (with a positive logit indicating a difficult item) and was scaled so that the average Rasch item difficulty across all items within each country was zero.

Reliability (N): To provide a measure of the reliability of the scoring of the constructed response items, items in approximately 25 percent of the test booklets in each country were independently scored by two scorers. This column indicates the number of responses that were double-scored for a given item in a country.

Reliability (Score): This column contains the percentage of agreement on the score value of the two-digit diagnostic codes assigned by the two independent TIMSS Advanced scorers.

Reliability (Code): This column contains the percentage of agreement on the two-digit diagnostic codes assigned by the two independent TIMSS Advanced scorers.

As an aid to the reviewers, the item-review displays included a series of flags signaling the presence of one or more conditions that might indicate a problem with an item. The following conditions were flagged:

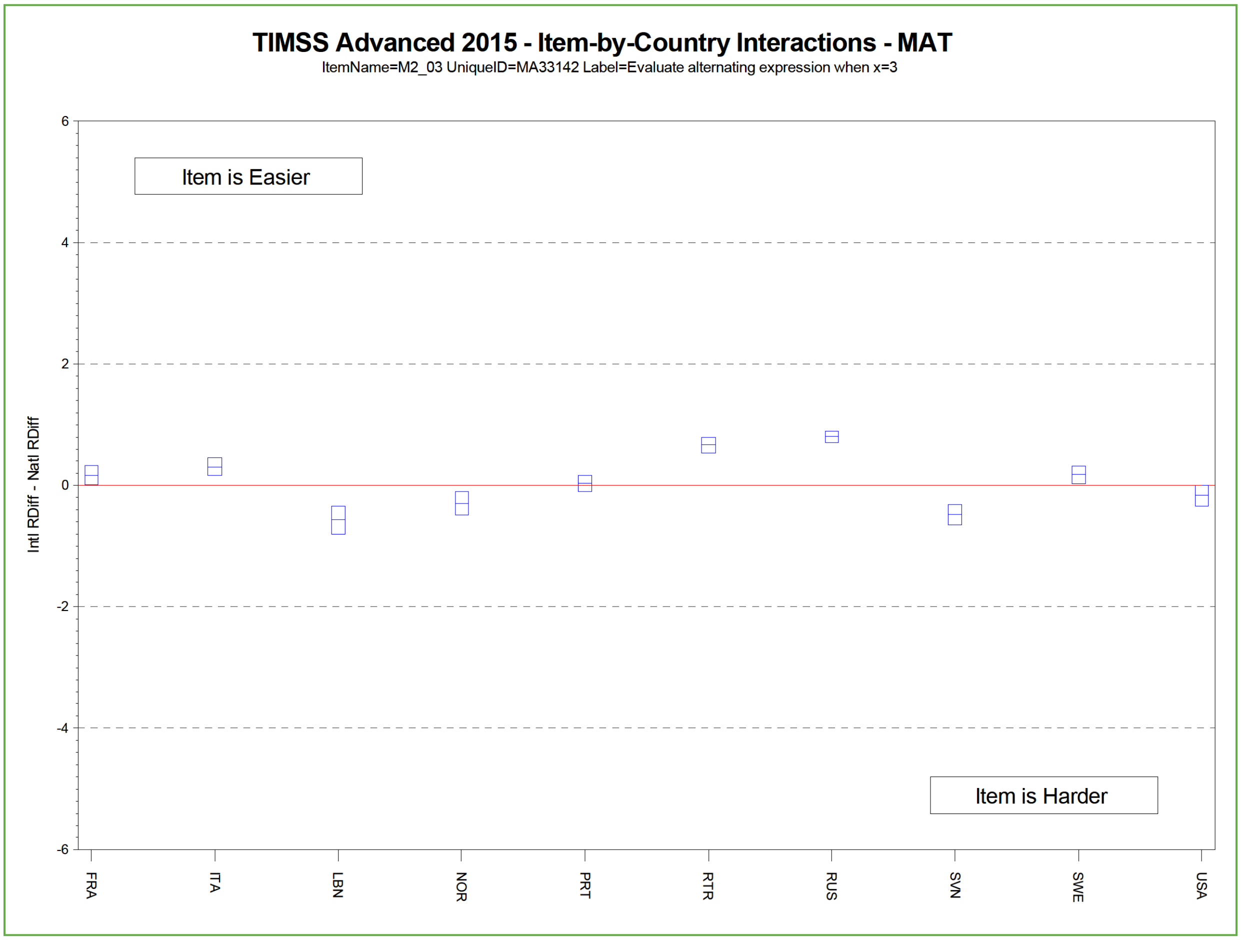
- The item discrimination (DISC) was less than 0.10 (flag D)
- The item difficulty (DIFF) was less than .25 for multiple-choice items (flag C)
- The item difficulty (DIFF) exceeded .95 (flag V)
- The Rasch difficulty estimate (RDIFF) for a given country made the item either easier (flag E) or more difficult (flag H) relative to the international average for that item
- The point-biserial correlation for at least one distracter in a multiple-choice item was positive, or the point-biserial correlations across the score levels of a constructed response item were not ordered (flag A)
- The percentage of students selecting one of the response options for a multiple-choice item, or one of the score values for a constructed response item, was less than 10 percent (flag F)
- Scoring reliability for agreement on the score value of a constructed response item was less than 85 percent (flag R)

Although not all of these conditions necessarily indicated a problem, the flags were a useful tool to draw attention to potential sources of concern.

Item–by–Country Interaction

Although countries are expected to exhibit some variation in performance across items, in general countries with high average performance on the assessment should perform relatively well on each of the items, and low-scoring countries should do less well on each of the items. When this does not occur (e.g., when a high-performing country has low performance on an item on which other countries are doing well), there is said to be an item-by-country interaction. When large, such item-by-country interactions may be a sign that an item is flawed in some way and that steps should be taken to address the problem. To assist in detecting sizeable item-by-country interactions, the TIMSS & PIRLS International Study Center produced a graphical display for each item showing the difference between each country's Rasch item difficulty and the international average Rasch item difficulty across all countries. An example of the graphical displays is provided in Exhibit 11.3.

Exhibit 11.3: Example Plot of Item-by-Country Interaction for a TIMSS Advanced 2015 Item



In each of these item-by-country interaction displays, the difference in Rasch item difficulty for each country is presented as a 95 percent confidence interval, which includes a built-in Bonferroni correction for multiple comparisons across the participating countries. The limits for this confidence interval were computed as follows:

$$\text{Upper Limit} = RDIFF_{i.} - RDIFF_{ik} + SE(RDIFFik) \cdot Z_b$$
$$\text{Lower Limit} = RDIFF_{i.} - RDIFF_{ik} - SE(RDIFFik) \cdot Z_b$$

where $RDIFF_{\mathrm{ik}}$ is the Rasch difficulty of item i in country k, $RDIFF_{i.}$ is the international average Rasch difficulty of item i, $SE(RDIFF_{ik})$ is the standard error of the Rasch difficulty of item i in country k, and Z_b is the 95% critical value from the Z distribution corrected for multiple comparisons using the Bonferroni procedure.

Trend Item Review

In order to measure trends, TIMSS Advanced 2015 included achievement items from previous assessments as well as items developed for use for the first time in 2015. Accordingly, the TIMSS Advanced 2015 assessments included items from 1995, 2008, and 2015. An important review step, therefore, was to check that these “trend items” had statistical properties in 2015 similar to those they had in the previous assessments (e.g., a TIMSS Advanced item that was relatively easy in 2008 should still be relatively easy in 2015).

As can be seen in the example in Exhibit 11.4, the trend item review focused on statistics for trend items from the current and previous assessments (2015 and 2008) for countries that participated in both. For each country, trend item statistics included the percentage of students in each score category (or response option for multiple-choice items) for each assessment, as well as the difficulty of the item and the percent correct by gender. In reviewing these item statistics, the aim was to detect any unusual changes in item difficulties between administrations, which might indicate a problem in using the item to measure trends.

Exhibit 11.4: Example Item Statistics for a TIMSS Advanced 2015 Trend Item

Trends in International Mathematics and Science Study - TIMSS Advanced 2015 Assessment Results - Mathematics
Trend Achievement Data Almanac for Advanced Mathematics Items (Weighted)

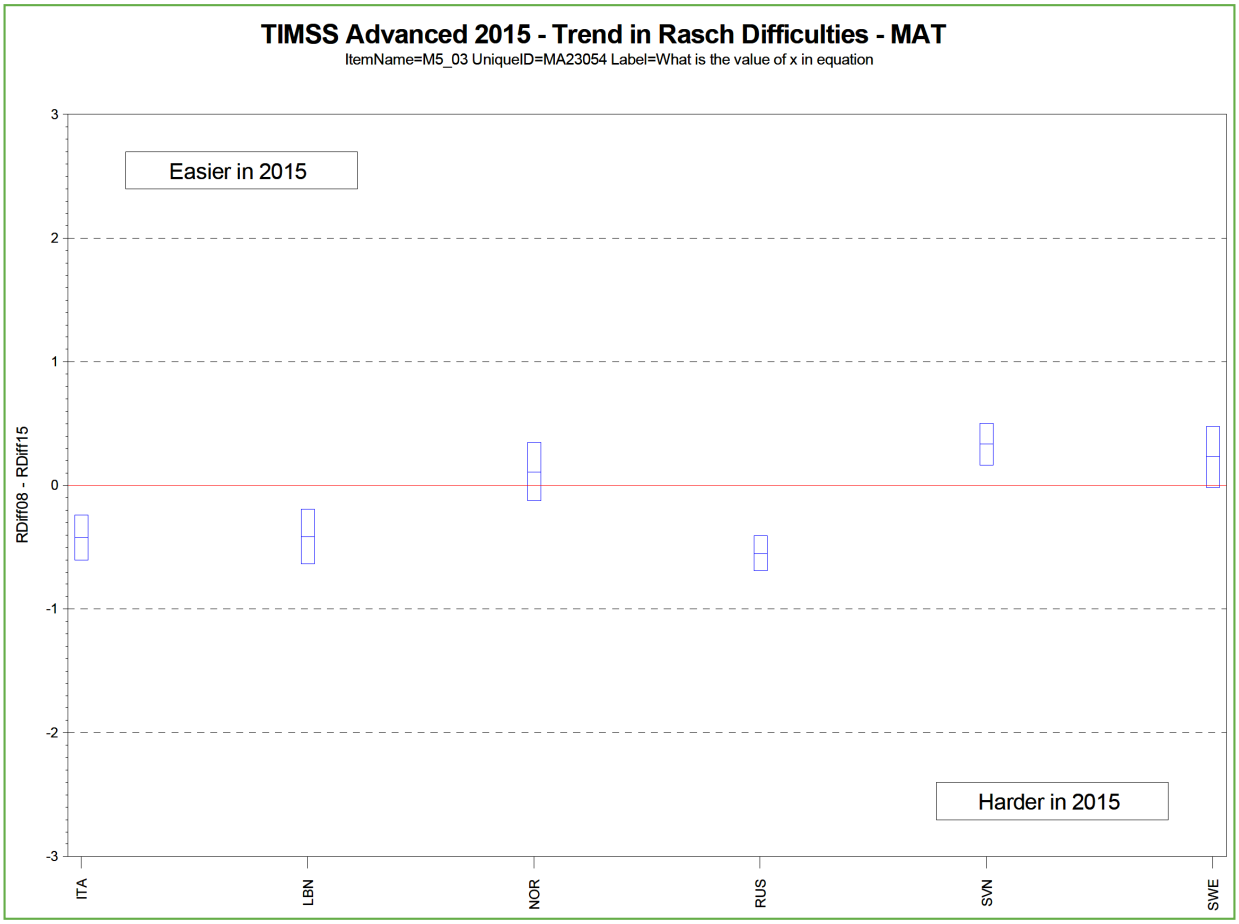
M3_03 (MA23187): Algebra / Applying   Type: CR   2 Points
Label: New diameter of soup can

COUNTRY	Year	N	20 %	21 %	10 %	11 %	12 %	70 %	79 %	OMITTED %	NOT REACHED %	V1 %	V2 %	GIRL PCT RIGHT	BOY PCT RIGHT
Italy	2008	1069	32.1	0.1	5.3	0.0	0.0	0.2	39.3	22.9	0.1	37.5	32.2	27.4	34.6
	2015	1117	31.6	0.0	3.3	6.0	0.0	0.1	35.2	23.6	0.1	41.0	31.6	30.3	32.4
Lebanon	2008	802	42.2	0.1	7.4	0.0	0.0	0.2	34.5	15.6	0.0	49.7	42.4	44.7	41.4
	2015	383	36.5	0.0	10.7	2.1	0.0	0.0	32.0	18.7	0.0	49.3	36.5	31.3	39.5
Norway	2008	966	55.6	0.7	2.7	15.3	0.2	0.5	17.6	7.5	0.0	74.4	56.3	57.7	55.4
	2015	839	50.9	0.3	8.9	14.5	0.0	0.0	17.0	8.3	0.1	74.6	51.2	58.3	47.2
Russian Federation 6hr+	2008	1588	55.8	0.1	8.1	2.0	0.2	0.0	21.5	12.4	0.0	66.0	55.9	54.9	56.6
	2015	1134	54.4	0.6	12.3	0.4	0.1	0.1	20.0	12.1	0.0	67.8	55.0	52.8	57.2
Slovenia	2008	1090	45.2	0.0	16.3	1.6	0.0	0.0	27.5	9.4	0.0	63.1	45.2	45.6	44.8
	2015	983	32.2	0.0	16.4	17.4	0.2	0.0	26.9	7.0	0.0	66.1	32.2	28.9	37.4
Sweden	2008	1144	59.5	0.2	5.2	9.8	0.0	0.2	19.3	5.8	0.1	74.6	59.7	62.1	58.0
	2015	1318	57.4	0.1	8.3	0.7	0.0	0.0	25.1	8.3	0.1	66.5	57.5	63.0	54.1
International Avg (n=5)	2008	5071	46.9	0.2	7.4	5.3	0.0	0.2	27.7	12.2	0.0	59.9	47.1	47.5	46.9
	2015	4640	41.7	0.1	9.5	8.1	0.0	0.0	27.2	13.2	0.1	59.5	41.8	42.4	42.1

V1 = Percent scoring 1 pt or better;   V2 = Percent scoring 2 pts;   Percent right for boys and girls corresponds to percent obtaining full credit.
Because of missing gender information, some totals may appear inconsistent.

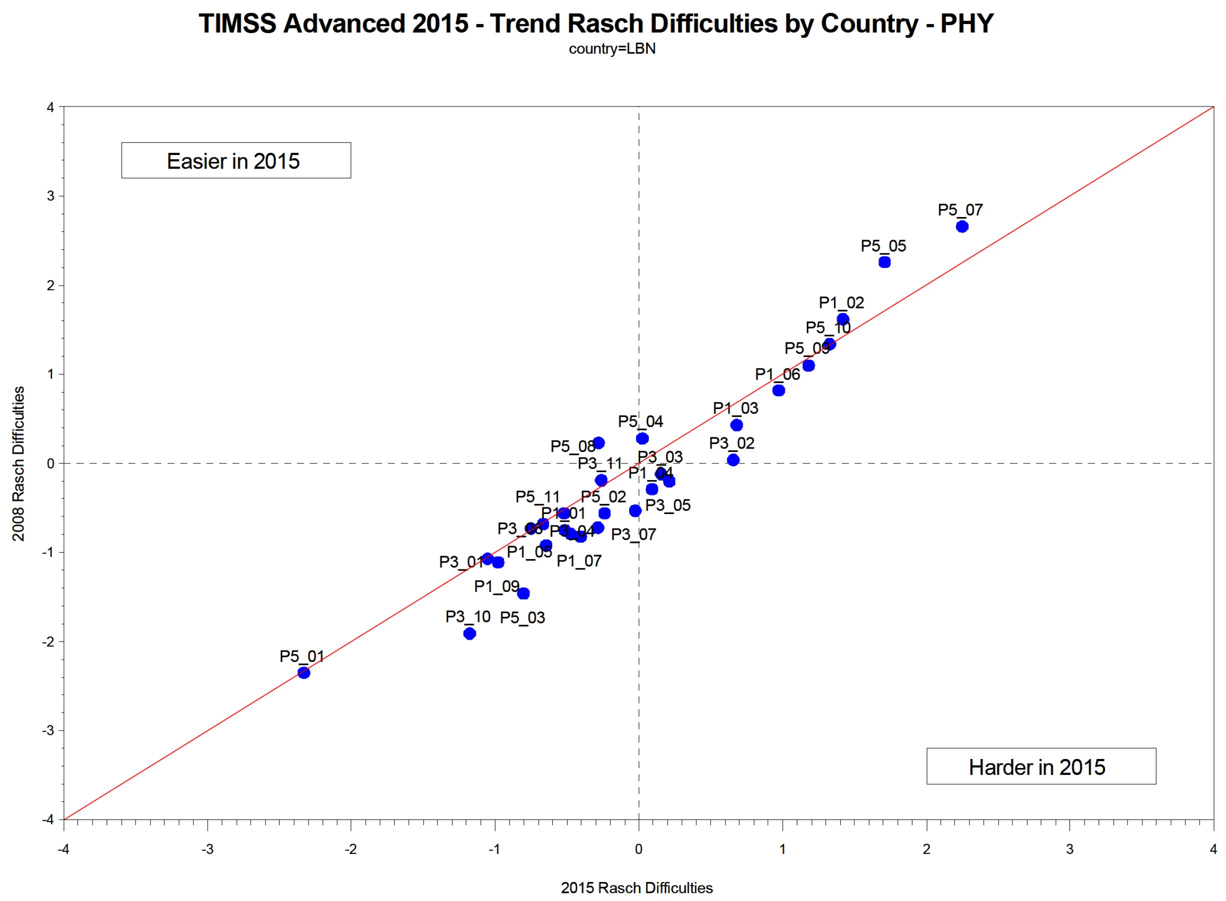
While some changes in item difficulties were anticipated as countries' overall achievement may have improved or declined, items were noted if the difference between the Rasch difficulties across the two assessments for a particular country was greater than 2 logits. The TIMSS & PIRLS International Study Center used two different graphical displays to examine the differences in item difficulties. The first of these, shown for an example item in Exhibit 11.5, displays the difference in Rasch item difficulty of the item between 2015 and 2008 for each country. A positive difference for a country indicates that the item was relatively easier in 2015, and a negative difference indicates that the item was relatively more difficult.

Exhibit 11.5: Example Plot of Differences in Rasch Item Difficulties Between 2015 and 2008 for a TIMSS Advanced Trend Item



The second graphical display, presented in Exhibit 11.6, shows the performance of a given country on all trend items simultaneously. For each country, the graph plots the 2015 Rasch difficulty of every trend item against its Rasch difficulty in 2008. Where there were no differences between the difficulties in the two successive administrations, the data points aligned on or near the diagonal.

Exhibit 11.6: Example Plot of Rasch Item Difficulties Across TIMSS Advanced Trend Items by Country



Reliability

Documenting the reliability of the TIMSS Advanced 2015 assessments was a critical quality control step in reviewing the items. As one indicator of reliability, the review considered Cronbach's Alpha coefficient of reliability calculated at the assessment booklet level. Secondly, the scoring of the constructed response items had to meet specific reliability criteria in terms of consistent within-country scoring and across assessment or trend–scoring.

Test Reliability

Exhibit 11.7 displays the TIMSS Advanced 2015 advanced mathematics and physics test reliability coefficients for every country, respectively. These coefficients are the median Cronbach's alpha reliability across all TIMSS Advanced 2015 assessment booklets. In general, reliabilities were relatively high, with international median reliabilities (the median of the reliability coefficients for all countries) of 0.88 for advanced mathematics and 0.84 for physics.

Exhibit 11.7: Cronbach's Alpha Reliability Coefficient—TIMSS Advanced 2015

Country	Reliability Coefficient	
	Advanced Mathematics	Physics
France	0.86	0.70
Italy	0.90	0.82
Lebanon	0.88	0.77
Norway	0.84	0.84
Portugal	0.86	0.78
Russian Federation	0.93	0.88
Russian Federation 6hr+[3]	0.93	—
Slovenia	0.88	0.85
Sweden	0.89	0.85
United States	0.91	0.84
International Median	**0.88**	**0.84**

Scoring Reliability for Constructed Response Items

A sizeable proportion of the items in the TIMSS Advanced 2015 assessments were constructed response items, comprising about half of the assessment score points. An essential requirement for use of such items is that they be reliably scored by all participants. That is, a particular student response should receive the same score, regardless of the scorer. In conducting TIMSS Advanced 2015, measures taken to ensure that the constructed response items were scored reliably in all countries included developing scoring guides for each constructed response question (that provided descriptions of acceptable responses for each score point value) and providing extensive training in the application of the scoring guides. See Chapter 1: Developing the TIMSS Advanced 2015 Achievement Items for more information on the scoring guides and see Chapter 6: Survey Operations Procedures for information on the scoring process.

Within-Country Scoring Reliability

To gather and document information about the within-country agreement among scorers for TIMSS Advanced 2015, a random sample of approximately 25 percent of the assessment booklets was selected to be scored independently by two scorers. The inter-scorer agreement for each item in each country was examined as part of the item review process. Exact percent agreement across items was high on average across countries, with 97 percent agreement in advanced mathematics and 96 percent agreement in physics, on average internationally. In TIMSS Advanced 2015 there also was high agreement at the diagnostic score level, with 96 percent agreement in advanced mathematics and 94 percent agreement in physics, on average internationally. See Appendix 11A

3 For advanced mathematics, the Russian Federation participated in 2015 with an expanded population that included the more specialized students assessed in 1995 and 2008.

for the average and range of the within-country percentage of correctness score agreement across all items. The TIMSS Advanced Within-Country Scoring Reliability documents also provide the average and range of the within-country percentage of diagnostic score agreement.

Trend Item Scoring Reliability

The TIMSS & PIRLS International Study Center also took steps to show that the 2015 constructed response items used in TIMSS Advanced 2008 were scored in the same way in both assessments. In anticipation of this, countries that participated in TIMSS Advanced 2008 sent samples of scored student booklets from the 2008 data collections to the IEA Data Processing and Research Center (IEA DPC), where they were digitally scanned and stored for later use. As a check on scoring consistency from one administration to the next, staff members working in each country on scoring the 2015 data were asked also to score these 2008 responses using the Trend Reliability Scoring Software developed by the IEA DPC. Each country scored 200 responses for each of 11 advanced mathematics and 11 physics items.

There was a very high degree of scoring consistency in TIMSS Advanced 2015. The exact agreement between the scores awarded in 2008 and those given by the 2015 scorers was 93 percent in advanced mathematics and 92 percent in physics, on average internationally. There also was high agreement in TIMSS Advanced at the diagnostic score level, although somewhat less in physics than in advanced mathematics, on average. The average and range of scoring consistency over time can be found in Appendix 11B.

Item Review Procedures

Using the information from the comprehensive collection of item analyses and reliability data that were computed and summarized for TIMSS Advanced 2015, the TIMSS & PIRLS International Study Center thoroughly reviewed all item statistics for every participating country to ensure that the items were performing comparably across countries. In particular, items with the following problems were considered for possible deletion from the international database:

- An error was detected during translation verification but was not corrected before test administration
- Data checking revealed a multiple-choice item with more or fewer options than in the international version
- The item analysis showed the item to have a negative biserial, or, for an item with more than 1 score point, point biserials that did not increase with each score level
- The item-by-country interaction results showed a very large negative interaction for a particular country

- For constructed response items, the within-country scoring reliability data showed an agreement of less than 70 percent
- For trend items, an item performed substantially differently in 2015 compared to the TIMSS Advanced 2008 administration, or an item was not included in the previous assessment for a particular country

When the item statistics indicated a problem with an item, the documentation from the translation verification was used as an aid in checking the test booklets. If a question remained about potential translation or cultural issues, however, then the National Research Coordinator was consulted before deciding how the item should be treated.

The checking of the TIMSS Advanced 2015 achievement data involved review of more than 200 items and resulted in the detection of very few items that were inappropriate for international comparisons. Among the few items singled out in the review process were mostly items with differences attributable to either translation or printing problems. See Appendix 11C: Country Adaptations to Items and Item Scoring for a list of deleted items, as well as a list of recodes made to constructed response item codes. There also were a number of items in each study that were combined, or derived, for scoring purposes. See Appendix 11D for details about how score points were awarded for each derived item.

Appendix 11A: TIMSS Advanced 2015 Within-Country Scoring Reliability for the Constructed Response Items

TIMSS Advanced 2015 Within-Country Scoring Reliability for the Advanced Mathematics Constructed Response Items

Country	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Minimum	Maximum		Minimum	Maximum
France	97	84	100	96	81	100
Italy	97	90	100	96	86	100
Lebanon	92	78	99	88	77	98
Norway	98	87	100	97	83	100
Portugal	100	99	100	100	99	100
Russian Federation	99	96	100	99	96	100
Russian Federation 6hr+[1]	99	96	100	99	96	100
Slovenia	99	92	100	99	92	100
Sweden	96	88	100	95	82	100
United States	96	83	100	95	83	100
International Avg.	**97**	**89**	**100**	**96**	**87**	**100**

1 For advanced mathematics, the Russian Federation participated in 2015 with an expanded population that included the more specialized students assessed in 1995 and 2008.

TIMSS Advanced 2015 Within-Country Scoring Reliability for the Physics Constructed Response Items

Country	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Minimum	Maximum		Minimum	Maximum
France	96	83	100	94	83	100
Italy	95	85	100	94	85	100
Lebanon	90	59	100	83	50	99
Norway	97	91	100	96	90	100
Portugal	100	98	100	99	97	100
Russian Federation	97	88	100	97	86	100
Slovenia	96	86	100	95	77	100
Sweden	96	90	100	94	85	100
United States	96	84	100	94	79	100
International Avg.	**96**	**85**	**100**	**94**	**81**	**100**

Appendix 11B: Trend Scoring Reliability for the Constructed Response Items

TIMSS Advanced 2015 Trend Scoring Reliability for the Advanced Mathematics Constructed Response Items

Country	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Minimum	Maximum		Minimum	Maximum
Italy	93	73	100	92	73	99
Norway	96	91	99	94	85	99
Russian Federation 6hr+[1]	92	72	100	91	72	100
Slovenia	92	77	99	90	76	98
Sweden	94	87	100	90	82	97
International Avg.	**93**	**80**	**100**	**91**	**78**	**98**

1 For advanced mathematics, the Russian Federation participated in 2015 with an expanded population that included the more specialized students assessed in 1995 and 2008.

TIMSS Advanced 2015 Trend Scoring Reliability for the Physics Constructed Response Items

Country	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Minimum	Maximum		Minimum	Maximum
Italy	93	88	97	89	79	92
Norway	91	74	99	87	73	96
Russian Federation	92	83	99	88	72	96
Slovenia	90	81	98	82	69	92
Sweden	93	76	98	87	75	93
International Avg.	**92**	**80**	**98**	**87**	**73**	**94**

Appendix 11C: Country Adaptations to Items and Item Scoring

TIMSS Advanced Advanced Mathematics

Deleted Items

ALL COUNTRIES

MA13014, M1_04 (attractive distracter)

ITALY

MA23185, M5_02 (negative discrimination)

PORTUGAL

MA23185, M5_02 (negative discrimination)

SLOVENIA

MA33140, M2_07 (translation error)

Constructed Response Items with Category Recodes

ALL COUNTRIES

MA33039, M2_11 (recode 20 to 10, 21 to 11, 10 to 79)

TIMSS Advanced Physics

Deleted Items

ALL COUNTRIES

PA13020, P1_10 (low discrimination)

PA23131, P5_06 (outside scope of TIMSS Advanced 2015 Assessment Framework)

PA33056, P9_01 (poor discrimination)

FRANCE

PA33120, P4_07 (translation error)

PA33119, P6_07 (translation error)

PA33111B, P9_05B (translation error)

Constructed Response Items with Category Recodes

ALL COUNTRIES

PA23088, P3_10 (recode 11 to 71)

PA33073, P6_02 (recode 70 to 12)

Appendix 11D: Score Points for Derived Items in TIMSS Advanced 2015

TIMSS Advanced Advanced Mathematics
MA33225, M2_02 – Item parts A, B, C, D, and E are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 4 parts are correct
MA33118, M8_06 – Item parts A, B, and C are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 2 parts are correct
MA33236, M9_08 – Item parts A, B, C, and D are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 3 parts are correct

TIMSS Advanced Physics
PA33102A, P2_05 – Item parts A, B, and C are combined to create a 1-piont item, where 1 score point is awarded if all parts are correct
PA33008, P8_10 – Item parts A, B, and C are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 2 parts are correct

CHAPTER 12

TIMSS Advanced 2015 Achievement Scaling Methodology[1]

The TIMSS and TIMSS Advanced approach to scaling the achievement data, based on item response theory (IRT) scaling with marginal estimation, was developed originally by Educational Testing Service for use in the U.S. National Assessment of Educational Progress (NAEP). It is based on psychometric models that were first used in the field of educational measurement in the 1950s and have become popular since the 1970s for use in large-scale surveys, test construction, and computer adaptive testing.[2]

Three distinct IRT models, depending on item type and scoring procedure, were used in the analysis of the TIMSS Advanced 2015 assessment data. Each is a "latent variable" model that describes the probability that a student will respond in a specific way to an item in terms of the student's proficiency, which is an unobserved or "latent" trait, and various characteristics (or "parameters") of the item. A three-parameter model was used with multiple-choice items, which were scored as correct or incorrect, and a two-parameter model for constructed response items with just two response options, which also were scored as correct or incorrect. Since each of these item types has just two response categories, they are known as dichotomous items. A partial credit model was used with polytomous constructed response items, i.e., those with more than two response options.

Two- and Three-Parameter IRT Models for Dichotomous Items

The fundamental equation of the three-parameter (3PL) model gives the probability that a student whose proficiency on a scale *k* is characterized by the unobservable variable will respond correctly to item *i* as:

$$P\left(x_i = 1 \middle| \theta_k, a_i, b_i, c_i\right) = c_i + \frac{1 - c_i}{1 + \exp\left(-1.7 \cdot a_i \cdot (\theta_k - b_i)\right)} \equiv P_{i,1}\left(\theta_k\right) \qquad (1)$$

1 This description of the TIMSS and TIMSS Advanced achievement scaling methodology has been adapted with permission from the TIMSS 1999 Technical Report (Yamamoto and Kulick, 2000).

2 For a description of IRT scaling see Birnbaum (1968); Lord and Novick (1968); Lord (1980); Van Der Linden and Hambleton (1996). The theoretical underpinning of the multiple imputation methodology was developed by Rubin (1987), applied to large-scale assessment by Mislevy (1991), and studied further by Mislevy, Johnson and Muraki (1992) and Beaton and Johnson (1992). For a recent overview, see von Davier and Sinharay (2014) and von Davier (2014). The procedures used in TIMSS and TIMSS Advanced have been used in several other large-scale surveys, including the U.S. National Assessment of Educational Progress (NAEP), the U.S. National Adult Literacy Survey (NALS), the International Adult Literacy Survey (IALS), and the International Adult Literacy and Life Skills Survey (IALLS).

where

x_i is the response to item i, 1 if correct and 0 if incorrect;

θ_k is the proficiency of a student on a scale k (note that a student with higher proficiency has a greater probability of responding correctly);

a_i is the slope parameter of item i, characterizing its discriminating power;

b_i is the location parameter of item i, characterizing its difficulty;

c_i is the lower asymptote parameter of item i, reflecting the chances of students with very low proficiency selecting the correct answer.

The probability of an incorrect response to the item is defined as:

$$P_{i,0} = P\left(x_i = 0 \middle| \theta_k, a_i, b_i, c_i\right) = 1 - P_{i,1}\left(\theta_k\right) \tag{2}$$

The two-parameter (2PL) model was used for the constructed response items that were scored as either correct or incorrect. The form of the 2PL model is the same as Equations (1) and (2) with the c_i parameter fixed at zero.

IRT Model for Polytomous Items

In TIMSS Advanced, constructed response items requiring an extended response were scored for partial credit, with 0, 1, and 2 as the possible score levels. These polytomous items were scaled using a generalized partial credit model (Muraki, 1992). The fundamental equation of this model gives the probability that a student with proficiency θ_k on scale k will have, for the i^{th} item, a response x_i that is scored in the l^{th} of m_i ordered score categories as:

$$P\left(x_i = l \middle| \theta_k, a_i, b_i, d_{i,1}, \cdots, d_{i,m_i-1}\right) = \frac{\exp\left(\sum_{v=0}^{l} 1.7 \cdot a_i \cdot \left(\theta_k - b_i + d_{i,v}\right)\right)}{\sum_{g=0}^{m_i-1} \exp\left(\sum_{v=0}^{g} 1.7 \cdot a_i \cdot \left(\theta_k - b_i + d_{i,v}\right)\right)} \equiv P_{i,l}\left(\theta_k\right) \tag{3}$$

where

m_i is the number of response categories for item i, usually 3;

x_i is the response to item i, ranging between 0 and $m_i - 1$;

θ_k is the proficiency of a student on a scale k;

a_i is the slope parameter of item i;

b_i is its location parameter, characterizing its difficulty;

$d_{i,l}$ is the category l threshold parameter.

The indeterminacy of model parameters in the polytomous model is resolved by setting $d_{i,0} = 0$ and $\sum_{j=1}^{m_i-1} d_{i,j} = 0$.

For all of the IRT models there is a linear indeterminacy between the values of item parameters and proficiency parameters, i.e., mathematically equivalent but different values of item parameters can be estimated on an arbitrarily linearly transformed proficiency scale. This linear indeterminacy can be resolved by setting the origin and unit size of the proficiency scale to arbitrary constants, such as a mean of 500 and a standard deviation of 100, as was done originally for TIMSS and TIMSS Advanced in 1995. The indeterminacy is most apparent when the scale is set for the first time.

IRT modeling relies on a number of assumptions, the most important being conditional independence. Under this assumption, item response probabilities depend only on θ_k (a measure of a student's proficiency) and the specified parameters of the item, and are unaffected by the demographic characteristics or unique experiences of the students, the data collection conditions, or the other items presented in the test. Under this assumption, the joint probability of a particular response pattern x across a set of n items is given by:

$$P\left(x \middle| \theta_k, \textit{item parameters}\right) = \prod_{i=1}^{n} \prod_{l=0}^{m_i - 1} P_{i,l}\left(\theta_k\right)^{u_{i,l}} \tag{4}$$

where $P_{i,l}(\theta_k)$ is of the form appropriate to the type of item (dichotomous or polytomous), m_i is equal to 2 for dichotomously scored items, and $u_{i,l}$ is an indicator variable defined as:

$$u_{i,l} = \begin{cases} 1 \textit{ if response is } x_i \textit{ is in category } l; \\ 0 \textit{ otherwise} \end{cases} \tag{5}$$

Replacing the hypothetical response pattern with the real scored data, the above function can be viewed as a likelihood function to be maximized by a given set of item parameters. In TIMSS Advanced, the item parameters for each scale are estimated independently of the parameters of other scales. Once items were calibrated in this manner, a likelihood function for the proficiency θ_k was induced from student responses to the calibrated items. This likelihood function for the proficiency θ_k is called the posterior distribution of the θ's for each student.

Proficiency Estimation Using Plausible Values

Most cognitive skills testing is concerned with accurately assessing the performance of individual students for the purposes of diagnosis, selection, or placement. Regardless of the measurement model used, whether classical test theory or item response theory, the accuracy of these measurements can be improved—that is, the amount of measurement error can be reduced—by increasing the number of items given to the individual. Thus, it is common to see achievement tests designed to provide information on individual students that contain more than 70 items. Since the uncertainty associated with each θ in such tests is negligible, the distribution of θ, or the joint distribution of θ with other variables, can be approximated using each individual's estimated θ.

For the distribution of proficiencies in large populations, however, more efficient estimates can be obtained from a matrix-sampling design like that used in TIMSS Advanced. This design solicits relatively few responses from each sampled student while maintaining a wide range of content representation when responses are aggregated across all students. With this approach, however, the advantage of estimating population characteristics more efficiently is offset by the inability to make precise statements about individuals. Indeed, the uncertainty associated with individual θ estimates becomes too large to be ignored. In this situation, aggregations of individual student scores can lead to seriously biased estimates of population characteristics (Wingersky, Kaplan, & Beaton, 1987).

Plausible values methodology was developed as a way to address this issue. Instead of first computing estimates of individual θ's and then aggregating these to estimate population parameters, the plausible values approach uses all available data, students' responses to the items they were administered together with all background data, to estimate directly the characteristics of student populations and subpopulations. Although these directly estimated population characteristics could be used for reporting purposes, instead the usual plausible values approach is to generate multiple imputed scores, called plausible values, from the estimated ability distributions and to use these in analyses and reporting, making use of standard statistical software. By including all available background data in the model, a process known as "conditioning", relationships between these background variables and the estimated proficiencies will be appropriately accounted for in the plausible values. Because of this, analyses conducted using plausible values will provide an accurate representation of these underlying relationships. A detailed review of the plausible values methodology is given in Mislevy (1991).[3]

The following is a brief overview of the plausible values approach. Let y represent the responses of all sampled students to background questions or background data of sampled students collected from other sources, and let θ represent the proficiency of interest. If θ were known for all sampled students, it would be possible to compute a statistic $t(\theta,y)$, such as a sample mean or sample percentile point, to estimate a corresponding population quantity T.

Because of the latent nature of the proficiency, however, θ values are not known even for sampled students. The solution to this problem is to follow Rubin (1987) by considering θ as "missing data" and approximate $t(\theta,y)$ by its expectation given (x,y), the data that actually were observed, as follows:

$$\begin{aligned} t^{*}(x, y) &= E\left| t(\underline{\theta}, \underline{y}) \mid \underline{x}, \underline{y} \right| \\ &= \int t(\underline{\theta}, \underline{y})\, p(\underline{\theta} \mid \underline{x}, \underline{y})\, d\underline{\theta} \end{aligned} \tag{6}$$

3 Along with theoretical justifications, Mislevy presents comparisons with standard procedures; discusses biases that arise in some secondary analyses; and offers numerical examples.

It is possible to approximate t^* using random draws from the conditional distribution of the scale proficiencies given the student's item responses x_j, the student's background variables y_j, and model parameters for the items. These values are referred to as imputations in the sampling literature, and as plausible values in large-scale surveys such as PIRLS, TIMSS, NAEP, NALS, and IALLS. The value of θ for any student that would enter into the computation of t is thus replaced by a randomly selected value from his or her conditional distribution. Rubin (1987) proposed repeating this process several times so that the uncertainly associated with imputation can be quantified. For example, the average of multiple estimates of t, each computed from a different set of plausible values, is a numerical approximation of t^* of the above equation; the variance among them reflects the uncertainty due to not observing $\underline{\theta}$. It should be noted that this variance does not include the variability of sampling from the population. That variability is estimated separately by a jackknife variance estimation procedure.

Plausible values are not intended to be estimates of individual student scores, but rather are imputed scores for like students—students with similar response patterns and background characteristics in the sampled population—that may be used to estimate population characteristics correctly. When the underlying model is correctly specified, plausible values will provide consistent estimates of population characteristics, even though they are generally biased estimates of the proficiencies of the individuals with whom they are associated. Taking the average of the plausible values still will not yield suitable estimates of individual student scores.[4]

Plausible values for each student j are drawn from the conditional distribution $P\left(\theta_j \middle| x_j, y_j, \Gamma, \Sigma\right)$, where Γ is a matrix of regression coefficients for the background variables, and Σ is a common variance matrix of residuals. Using standard rules of probability, the conditional probability of proficiency can be represented as:

$$P\left(\theta_j \middle| x_j, y_j, \Gamma, \Sigma\right) \propto P\left(x_j \middle| \theta_j, y_j, \Gamma, \Sigma\right) P\left(\theta_j \middle| y_j, \Gamma, \Sigma\right) = P\left(x_j \middle| \theta_j\right) P\left(\theta_j \middle| y_j, \Gamma, \Sigma\right) \qquad (7)$$

where θ_j is a vector of scale values, $P(x_j|\theta_j)$ is the product over the scales of the independent likelihoods induced by responses to items within each scale, and $P(\theta_j|y_j, \Gamma, \Sigma)$ is the multivariate joint density of proficiencies for the scales, conditional on the observed values y_j of background responses and parameters Γ and Σ. Item parameter estimates are fixed and regarded as population values in the computations described in this section.

4 For further discussion, see Mislevy, Beaton, Kaplan, and Sheehan (1992).

Conditioning

A multivariate normal distribution was assumed for $P(\theta_j|y_j, \Gamma, \Sigma)$, with a common variance Σ, and with a mean given by a linear model with regression parameters Γ. Since in large-scale studies like TIMSS and TIMSS Advanced there are many hundreds of background variables, it is customary to conduct a principal components analysis to reduce the number of variables to be used in Γ. Typically, components accounting for 90 percent of the variance in the data are selected. These principal components are referred to as the conditioning variables and denoted as y^c. The following model is then fit to the data:

$$\theta = \Gamma' y^c + \varepsilon \tag{8}$$

where ε is normally distributed with mean zero and variance Σ. As in a regression analysis, Γ is a matrix each of whose columns is the effects for each scale and Σ is the matrix of residual variance between scales.

Note that in order to be strictly correct for all functions Γ of θ, it is necessary that $P(\theta|y)$ be correctly specified for all background variables in the survey. Estimates of functions Γ involving background variables not conditioned in this manner are subject to estimation error due to misspecification. The nature of these errors is discussed in detail in Mislevy (1991). In TIMSS Advanced, however, the principal components account for almost all of the variance in the student background variables, so that the computation of marginal means and percentile points of θ for these variables is nearly optimal.

The basic method for estimating Γ and Σ with the Expectation and Maximization (EM) procedure is described in Mislevy (1985) for a single scale case. The EM algorithm requires the computation of the mean θ, and variance Σ, of the posterior distribution in Equation (7).

Generating Proficiency Scores

After completing the EM algorithm, plausible values for all sampled students are drawn from the joint distribution of the values of Γ in a three-step process. First, a value of Γ is drawn from a normal approximation to $P(\Gamma,\Sigma|x_j, y_j)$ that fixes Σ at the value $\hat{\Sigma}$ (Thomas, 1993). Second, conditional on the generated value of Γ (and the fixed value of $\Sigma=\hat{\Sigma}$), the mean θ_j and variance Σ_j^p of the posterior distribution in Equation (7), where p is the number of scales, are computed using the methods applied in the EM algorithm. In the third step, the proficiency values are drawn independently from a multivariate normal distribution with mean θ_j and variance Σ_j^p. These three steps are repeated five times, producing five imputations of θ_j for each sampled student.

For students with an insufficient number of responses, the Γ's and Σ's described in the previous paragraph are fixed. Hence, all students—regardless of the number of items attempted—are assigned a set of plausible values.

The plausible values can then be employed to evaluate Equation (6) for an arbitrary function T as follows:

- Using the first vector of plausible values for each student, evaluate T as if the plausible values were the true values of θ. Denote the result as T_1
- Evaluate the sampling variance of T_1, or Var_1, with respect to students' first vector of plausible values
- Carry out steps 1 and 2 for the second through fifth vectors of plausible values, thus obtaining T_u and Var_u, for $u = 2, \ldots, 5$
- The best estimate of T obtainable from the plausible values is the average of the five values obtained from the different sets of plausible values:

$$\hat{T} = \frac{\sum_u T_u}{5} \tag{9}$$

- An estimate of the variance of $\hat{T}$ is the sum of two components: an estimate of Var_u obtained by averaging as in the previous step, and the variance among the T_u's

  Let $\overline{U} = \frac{\sum_u Var_u}{M}$, and let $B_M = \frac{\sum_u \left(T_u - \hat{T}\right)^2}{M-1}$ be the variance among the M plausible values.

  Then the estimate of the total variance of $\hat{T}$ is:

$$Var\left(\hat{T}\right) = \overline{U} + \left(1 + M^{-1}\right) B_M \tag{10}$$

The first component in $Var\left(\hat{T}\right)$ reflects the uncertainty due to sampling students from the population; the second reflects the uncertainty due to the fact that sampled students' θ's are not known precisely, but only indirectly through x and y.

Working with Plausible Values

The plausible values methodology is used in TIMSS Advanced to ensure the accuracy of estimates of the proficiency distributions for the TIMSS Advanced populations as a whole and particularly for comparisons between subpopulations. A further advantage of this method is that the variation between the five plausible values generated for each student reflects the uncertainty associated with proficiency estimates for individual students. However, retaining this component of uncertainty requires that additional analytical procedures be used to estimate students' proficiencies.

If the θ values were observed for all sampled students, the statistic $(t-T)\big/U^{\frac{1}{2}}$ would follow a t-distribution with d degrees of freedom. Then the incomplete-data statistic $(T-\hat{T})\big/\left[Var(\hat{T})\right]^{\frac{1}{2}}$ is approximately t-distributed, with degrees of freedom (Johnson & Rust, 1992) given by:

$$\nu = \frac{1}{\dfrac{f_M^{\ 2}}{M-1} + \dfrac{(1-f_M)^2}{d}} \tag{11}$$

where d is the degrees of freedom for the complete-data statistic, and f_M is the proportion of total variance due to not observing the values:

$$f_M = \frac{(1+M^{-1})\, B_M}{Var(\hat{T})} \tag{12}$$

When B_M is small relative to $\overline{U}$, the reference distribution for the incomplete-data statistic differs little from the reference distribution for the corresponding complete-data statistic. If, in addition, d is large, the normal approximation can be used instead of the t-distribution.

For a k-dimensional function T, such as the k coefficients in a multiple regression analysis, each U and $\overline{U}$ is a covariance matrix, and B_M is an average of squares and cross-products rather than simply an average of squares. In this case, the quantity $(\underline{T}-\hat{\underline{T}})\ Var^{-1}(\hat{\underline{T}})\ (\underline{T}-\hat{\underline{T}})'$ is approximately F-distributed with degrees of freedom equal to k and ν, with ν defined as above but with a matrix generalization of f_M:

$$f_M = (1+M^{-1})\ \text{Trace}\left[B_M\, Var^{-1}(\hat{T})\right]\big/k \tag{13}$$

For the same reason that the normal distribution can approximate the t-distribution, a chi-square distribution with k degrees of freedom can be used in place of the F-distribution for evaluating the significance of the above quantity $(\underline{T}-\hat{\underline{T}})\ Var^{-1}(\hat{\underline{T}})\ (\underline{T}-\hat{\underline{T}})'$.

Statistics $\hat{T}$, the estimates of proficiency conditional on responses to cognitive items and background variables, are consistent estimates of the corresponding population values T, as long as background variables are included in the conditioning variables. The consequences of violating this restriction are described by Beaton and Johnson (1992), Mislevy (1991), and Mislevy and Sheehan (1987). To avoid such biases, the TIMSS Advanced analyses include nearly all student background variables, in the form of principal components, as well as the class means to preserve between-class differences—the between-classroom and within-classroom variance structure essential for hierarchical modeling.

References

Beaton, A.E., & Johnson, E.G. (1992). Overview of the scaling methodology used in the National Assessment. *Journal of Educational Measurement*, *26*(2), 163–175.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick, *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing.

Johnson, E.G., & Rust, K.F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics.*

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Redding, MA: Addison-Wesley.

Lord, F.M. (1980). *Applications of items response theory to practical testing problems*. Hillsdales, NJ: Lawrence Erlbaum Associates.

Mislevy, R.J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association*, *80*, 993–997.

Mislevy, R.J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, *56*, 177- 196.

Mislevy, R.J., Beaton, A.E., Kaplan, B., & Sheehan, K.M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*(2), 133–161.

Mislevy, R.J., Johnson, E.G. & Muraki, E. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics*, *17*(2), 131–154.

Mislevy, R.J., & Sheehan, K.M. (1987). Marginal estimation procedures. In A.E. Beaton (Ed.), *Implementing the new design: The NAEP 1983–84 technical report* (pp. 293–360) (no. 15-TR-20). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*(2), 159–176.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Thomas, N. (1993). Asymptotic corrections for multivariate posterior moments with factored likelihood functions. *Journal of Computational and Graphical Statistics*, *2*, 309-322.

Van Der Linden, W.J., & Hambleton, R. (1996). *Handbook of Modern Item Response Theory*. New York. Springer-Verlag.

von Davier, M. (2014). Imputing proficiency data under planned missingness in population models. In L. Rutkowski, M. von Davier, & D. Rutkowski, *Handbook of international large-scale assessment: Background, technical issues and methods of data analysis* (pp. 175-201). Boca Raton: Chapman & Hall/CRC.

von Davier, M., & Sinharay, S. (2014). Analytics in international large-scale assessments: item response theory and population models. In L. Rutkowski, M. von Davier, & D. Rutkowski, *Handbook of international large-scale assessment: Background, technical issues and methods of data analysis* (pp. 155–174). Boca Raton: Chapman & Hall/CRC.

Wingersky, M., Kaplan, B., & Beaton, A.E. (1987). Joint estimation procedures. In A.E. Beaton (Ed.), *Implementing the new design: The NAEP 1983–84 technical report* (pp.285–92) (No. 15-TR-20). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Yamamoto, K., & Kulick, E. (2000). Scaling methodology and procedures for the TIMSS mathematics and science scales. In M.O. Martin, K.D. Gregory, & S.E. Stemler (Eds.), *TIMSS 1999 technical report*. Chestnut Hill, MA: Boston College.

CHAPTER 13

Scaling the TIMSS Advanced 2015 Achievement Data

Pierre Foy
Liqun Yin

Overview

The TIMSS Advanced assessments cover a wide range of topics in advanced mathematics and physics. Given this broad coverage, a matrix-sampling booklet design is used such that each student is administered only a subset of the entire TIMSS Advanced mathematics and physics item pools (see Chapter 4 of *TIMSS Advanced 2015 Assessment Frameworks*). Given the complexities of the data collection and the need to have student scores on the entirety of each assessment for analysis and reporting purposes, TIMSS Advanced relies on item response theory (IRT) scaling to describe student achievement and to provide accurate measures of trends. As each student responded to only a part of the assessment item pool, the TIMSS Advanced scaling approach uses multiple imputation—or plausible values—methodology to obtain proficiency scores in advanced mathematics and physics for all students. To enhance the reliability of the student scores, the TIMSS Advanced scaling approach uses conditioning, a process in which student responses to the items are combined with information about students' backgrounds.

This scaling chapter begins with a general description of the scaling approach and its use of plausible values. It then describes the concurrent calibration method used specifically to measure trends. Next, it explains how the proficiency scores are generated through the use of conditioning and describes the process of transforming the proficiency scores to place them on the metrics used to measure trends. A description of the technical details involved in the scaling can be found in Chapter 12: TIMSS Advanced 2015 Achievement Scaling Methodology.

Implementing the TIMSS Advanced Scaling Procedures

The application of IRT scaling and plausible values methodology to the data from the TIMSS Advanced assessments involves four major tasks: calibrating the achievement items (estimating model parameters for each item), creating principal components from the student questionnaire data for use in conditioning, generating proficiency scores for advanced mathematics and physics, and placing these proficiency scores on the metrics used to report trend results from previous assessments. TIMSS Advanced has separate scales for advanced mathematics and physics. The scaling procedures also generate proficiency scores for the domains of the overall subjects: the content and cognitive domains of advanced mathematics and physics.

Linking Assessments Cycles with Concurrent Calibration

The metric of the TIMSS Advanced reporting scales for overall advanced mathematics and physics were originally established in TIMSS Advanced 1995 by setting the mean of the national average scores for all countries that participated in TIMSS Advanced 1995 to 500 and the standard deviation to 100. To enable measurement of trends over time, achievement data from successive TIMSS Advanced assessments were transformed to these same metrics. This is done by concurrently scaling the data from each successive assessment with the data from the previous assessment—a process known as concurrent calibration—and applying linear transformations to place the results from each successive assessment on the same scale as the results from the previous assessment. This procedure enables TIMSS Advanced to measure trends across all three assessment cycles: 1995, 2008, and 2015.[1]

The first step in linking the assessments for trend scaling is to estimate (calibrate) the item parameters for the items in the current assessment through a concurrent calibration of the data from the current assessment and from the previous assessment. In 2015, the TIMSS Advanced concurrent calibration consisted of combining achievement data from the 2015 and 2008 assessments.

In linking successive assessments, concurrent calibration relies on having a substantial number of trend items, items that are retained from one assessment to the next. The TIMSS Advanced assessment consists of 9 advanced mathematics item blocks and 9 physics item blocks. In TIMSS Advanced 2015, 6 of the advanced mathematics blocks and 6 of the physics blocks consisted of newly developed items. The remaining 3 advanced mathematics blocks and 3 physics blocks were carried forward from the TIMSS Advanced 2008 assessment and are the basis for linking TIMSS Advanced 2015 to the TIMSS Advanced achievement scale and maintaining trends over time. Exhibits 13.1 through 13.2 list the number of items present for TIMSS Advanced concurrent calibration by item type and content and cognitive domain for both subjects.

1 See Mazzeo and von Davier (2014) for a discussion of the linking procedure used by TIMSS.

Exhibit 13.1: TIMSS Advanced 2015 Advanced Mathematics Items for Concurrent Calibration

Item Type	Points	Items Released in 2008		Items Common in 2008 and 2015		Items Introduced in 2015		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple-Choice	1	25	25	20	20	35	35	80	80
Constructed Response	1	13	13	5	5	22	22	40	40
	2	2	4	6	12	13	26	21	42
Total		**40**	**42**	**31**	**37**	**70**	**83**	**141**	**162**

TIMSS Advanced 2015 Advanced Mathematics Items for Concurrent Calibration by Content and Cognitive Domains

Advanced Mathematics Content Domains	Items Released in 2008		Items Common in 2008 and 2015		Items Introduced in 2015		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Algebra	13	14	12	14	25	29	50	57
Calculus	16	16	9	12	25	31	50	59
Geometry	11	12	10	11	20	23	41	46
Advanced Mathematics Cognitive Domains	**Items Released in 2008**		**Items Common in 2008 and 2015**		**Items Introduced in 2015**		**Total**	
	Items	**Points**	**Items**	**Points**	**Items**	**Points**	**Items**	**Points**
Knowing	16	16	11	12	21	23	48	51
Applying	14	14	17	21	23	27	54	62
Reasoning	10	12	3	4	26	33	39	49
Total	**40**	**42**	**31**	**37**	**70**	**83**	**141**	**162**

Exhibit 13.2: TIMSS Advanced 2015 Physics Items for Concurrent Calibration

Item Type	Points	Items Released in 2008		Items Common in 2008 and 2015		Items Introduced in 2015		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple-Choice	1	20	20	20	20	39	39	79	79
Constructed Response	1	11	11	8	8	21	21	40	40
	2	6	12	3	6	11	22	20	40
Total		**37**	**43**	**31**	**34**	**71**	**82**	**139**	**159**

TIMSS Advanced 2015 Physics Items for Concurrent Calibration by Content and Cognitive Domains

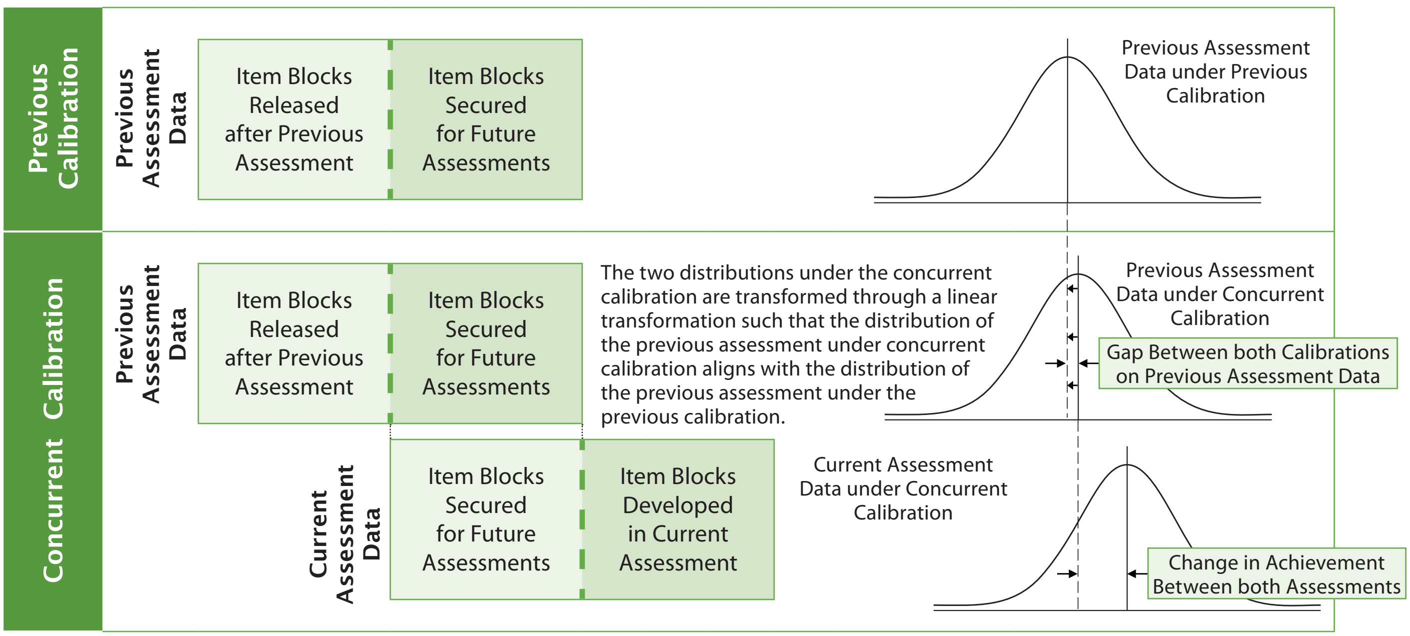
Physics Content Domains	Items Released in 2008		Items Common in 2008 and 2015		Items Introduced in 2015		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Electricity & Magnetism	12	13	8	9	20	23	40	45
Mechanics & Thermodynamics	18	22	13	15	26	32	57	69
Wave Phenomena & Atomic/ Nuclear Physics	7	8	10	10	25	27	42	45
Physics Cognitive Domains	**Items Released in 2008**		**Items Common in 2008 and 2015**		**Items Introduced in 2015**		**Total**	
	Items	**Points**	**Items**	**Points**	**Items**	**Points**	**Items**	**Points**
Knowing	9	9	10	11	21	21	40	41
Applying	18	20	14	15	27	35	59	70
Reasoning	10	14	7	8	23	26	40	48
Total	**37**	**43**	**31**	**34**	**71**	**82**	**139**	**159**

In concurrent calibration, item parameters for the current assessment are estimated based on the data from both the current and previous assessments, recognizing that some items (the trend items) are common to both. It is then possible to estimate the latent ability distributions of students in both assessments using the item parameters from the concurrent calibration. The difference between these two distributions is the change in achievement between the previous and current assessments.

After the calibration, the next step is to find a linear transformation that transforms the distribution of the previous assessment data under the concurrent calibration to match the distribution of these same data under the calibration that was done in the previous assessment. The final step entails applying this linear transformation to the current assessment data scaled using the concurrent calibration. This places the current assessment data on the trend scale.

Exhibit 13.3 illustrates how the concurrent calibration approach is applied in the context of TIMSS Advanced trend scaling. The gap between the distributions of the previous assessment data (in this case TIMSS Advanced 2008) under the previous calibration and under the concurrent calibration is typically small and is the result of slight differences in the item parameter estimates from the two calibrations (Exhibit 13.3, second panel). The linear transformation removes this gap by shifting the two distributions from the concurrent calibration such that the distribution of the previous assessment data from the concurrent calibration aligns with the distribution of the previous assessment data from the previous calibration,[2] while preserving the gap between the previous and current assessment data under the concurrent calibration. This latter gap is the change in achievement between the previous and current assessments that TIMSS Advanced sets out to measure as trend.

Exhibit 13.3: Concurrent Calibration Model Used for TIMSS Advanced



Calibrating the TIMSS Advanced 2015 Assessment Data

Item calibration was conducted by the TIMSS & PIRLS International Study Center using the commercially-available Parscale software (Muraki & Bock, 1991) and included data from the previous assessment (TIMSS Advanced 2008) and data from the 2015 assessment for countries that participated in both assessment cycles. The calibration used all available item response data from each country's student samples and from both current and previous assessments. All student samples were weighted so that each country contributed equally to the item calibration.

2 The difference between the ability distributions of the previous assessment data under the two calibrations is a measure of the linkage error in the trend scaling procedure.

Exhibits 13.4 and 13.5 show the sample sizes for scaling the TIMSS Advanced 2015 data. A total of 9 countries from TIMSS Advanced 2015 contributed to the concurrent calibration at each subject.

Exhibit 13.4: TIMSS Advanced 2015 Sample Sizes for Scaling the Advanced Mathematics Data

Country	Item Calibration		Proficiency Estimation	
	2015	2008	2015	2008
Armenia	—	858	—	858
France	3,967	—	3,967	—
Iran, Islamic Rep. of	—	2,425	—	2,425
Italy	3,318	2,143	3,318	2,143
Lebanon	1,161	1,612	1,161	1,612
Netherlands	—	1,537	—	1,537
Norway	2,537	1,932	2,537	1,932
Portugal	4,068	—	4,068	—
Russian Federation	—	3,185	7,558	3,185
Russian Federation 6hr+	3,431	—	3,431	—
Slovenia	2,922	2,156	2,922	2,156
Sweden	3,937	2,303	3,937	2,303
United States	2,954	—	2,954	—
Total	**28,295**	**18,151**	**35,853**	**18,151**

Exhibit 13.5: TIMSS Advanced 2015 Sample Sizes for Scaling the Physics Data

Country	Item Calibration		Proficiency Estimation	
	2015	2008	2015	2008
Armenia	—	894	—	894
France	3,958	—	3,958	—
Iran, Islamic Rep. of	—	2,434	—	2,434
Italy	3,424	1,861	3,424	1,861
Lebanon	1,156	1,595	1,156	1,595
Netherlands	—	1,511	—	1,511
Norway	2,472	1,640	2,472	1,640
Portugal	1,783	—	1,783	—
Russian Federation	3,822	3,166	3,822	3,166
Slovenia	1,106	1,097	1,106	1,097
Sweden	3,727	2,291	3,727	2,291
United States	2,932	—	2,932	—
Total	**24,380**	**16,489**	**24,380**	**16,489**

The item parameters estimated from these concurrent calibrations, based on the countries that have participated in both the previous and current assessments, were used to estimate student proficiency for all countries participating in the TIMSS Advanced 2015 assessments. These item parameters were also used to estimate student proficiency in the advanced mathematics and physics content and cognitive domains. The item parameters estimated from the TIMSS Advanced concurrent calibrations for advanced mathematics and physics are presented in Appendix 13A-13B.

Treatment of Omitted and Not-Reached Responses

Given the matrix-sampling design used by TIMSS Advanced, whereby a student is administered only a sample of the assessment blocks (three advanced mathematics or three physics blocks) most items are missing by design for each student. However, missing data can also result from a student not answering an item, which can occur when the student does not know the answer, omits the item by mistake, or does not have sufficient time to attempt the item. An item is considered "not reached" when the item itself and the item immediately preceding it are not answered, and there are no other items completed in the remainder of the booklet.

Not-reached items are treated differently in estimating item parameters and in generating student proficiency scores. In estimating the values of the item parameters, items in the assessment booklets that are considered not to have been reached by students are treated as if they have not been administered. This approach is considered optimal for parameter estimation. However, not-reached items are always considered as incorrect responses when student proficiency scores are generated.

Evaluating Fit of IRT Models to the TIMSS Advanced Assessment Data

After the item calibrations were completed, checks were performed to verify that the item parameters obtained from Parscale adequately reproduce the observed distribution of student responses across the proficiency continuum. The fit of the IRT models to the TIMSS Advanced assessment data is examined by comparing the item response function curves generated using the item parameters estimated from the data with the empirical item response functions calculated from the latent abilities estimated for each student that responded to the item. When the empirical results for an item fall near the fitted curves, the IRT model fits the data well and provides an accurate and reliable measurement of the underlying proficiency scale. Graphical plots of these response function curves are called item characteristic curves (ICC).

The plots in the Exhibits 13.6 and 13.7 show examples of the empirical and fitted item response functions for dichotomously scored (right/wrong) multiple-choice and constructed response items, respectively. In each plot, the horizontal axis represents the proficiency scale, and the vertical axis represents the probability of a correct response. The fitted curve based on the estimated item parameters is shown as a solid line. Empirical results are represented by circles. The empirical results are obtained by first dividing the proficiency scale into intervals of equal size and then counting the number of students responding to the item whose estimated latent abilities (EAP scores) from Parscale fall in each interval. Then the proportion of students in each interval that responded correctly to the item is calculated. In the exhibits, the center of each circle represents this empirical proportion of correct responses. The size of each circle is proportional to the number of students contributing to the estimation of the empirical proportion correct.

Exhibit 13.6: Example of Item Response Function for a Dichotomous Multiple-Choice Item from the TIMSS Advanced 2015 Advanced Mathematics Assessment

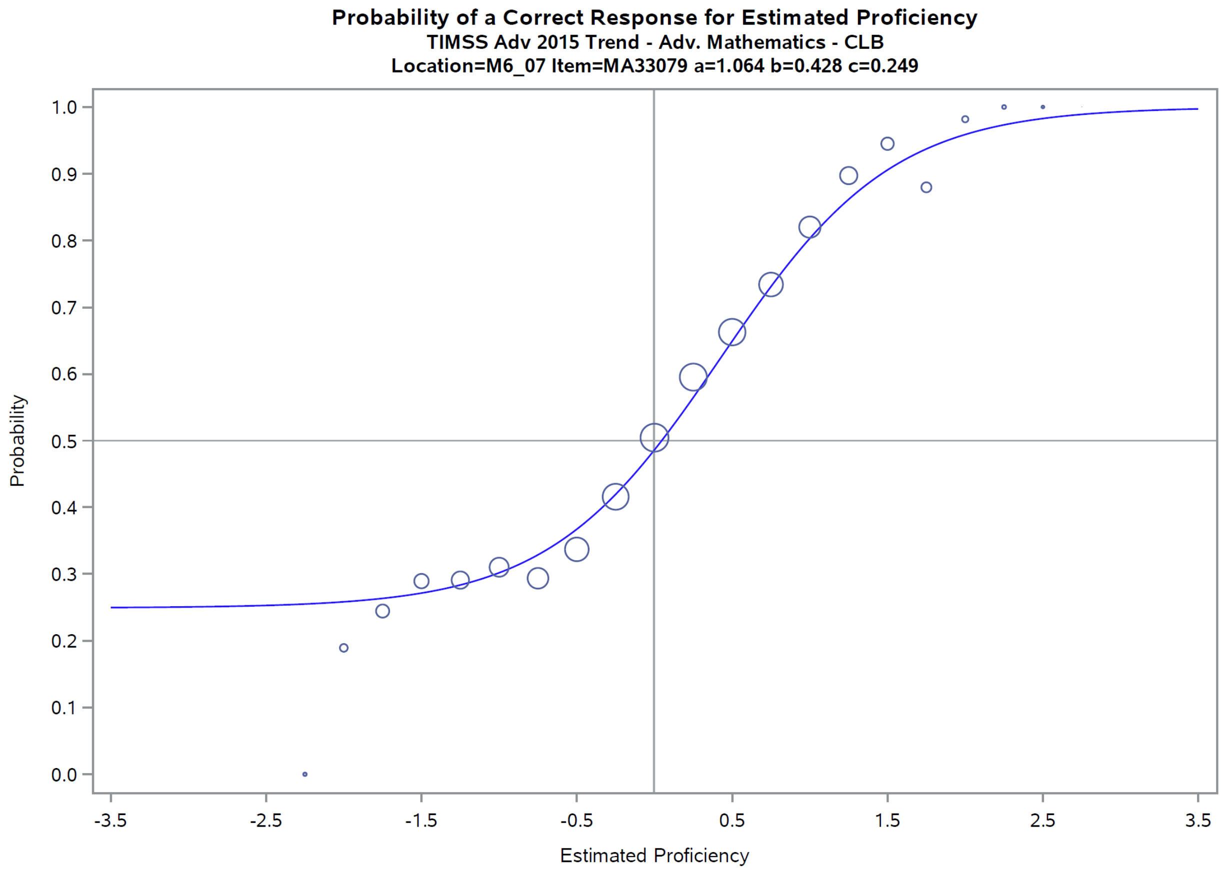
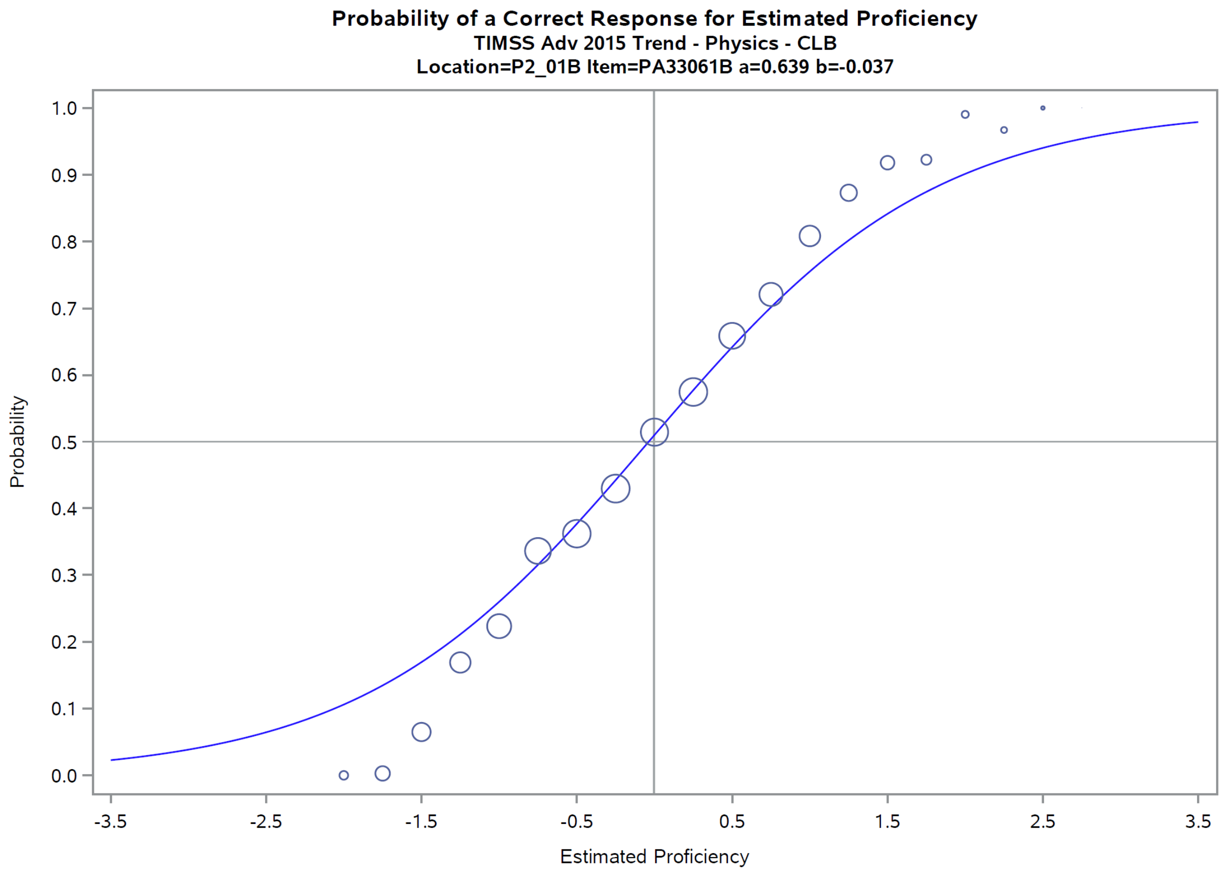
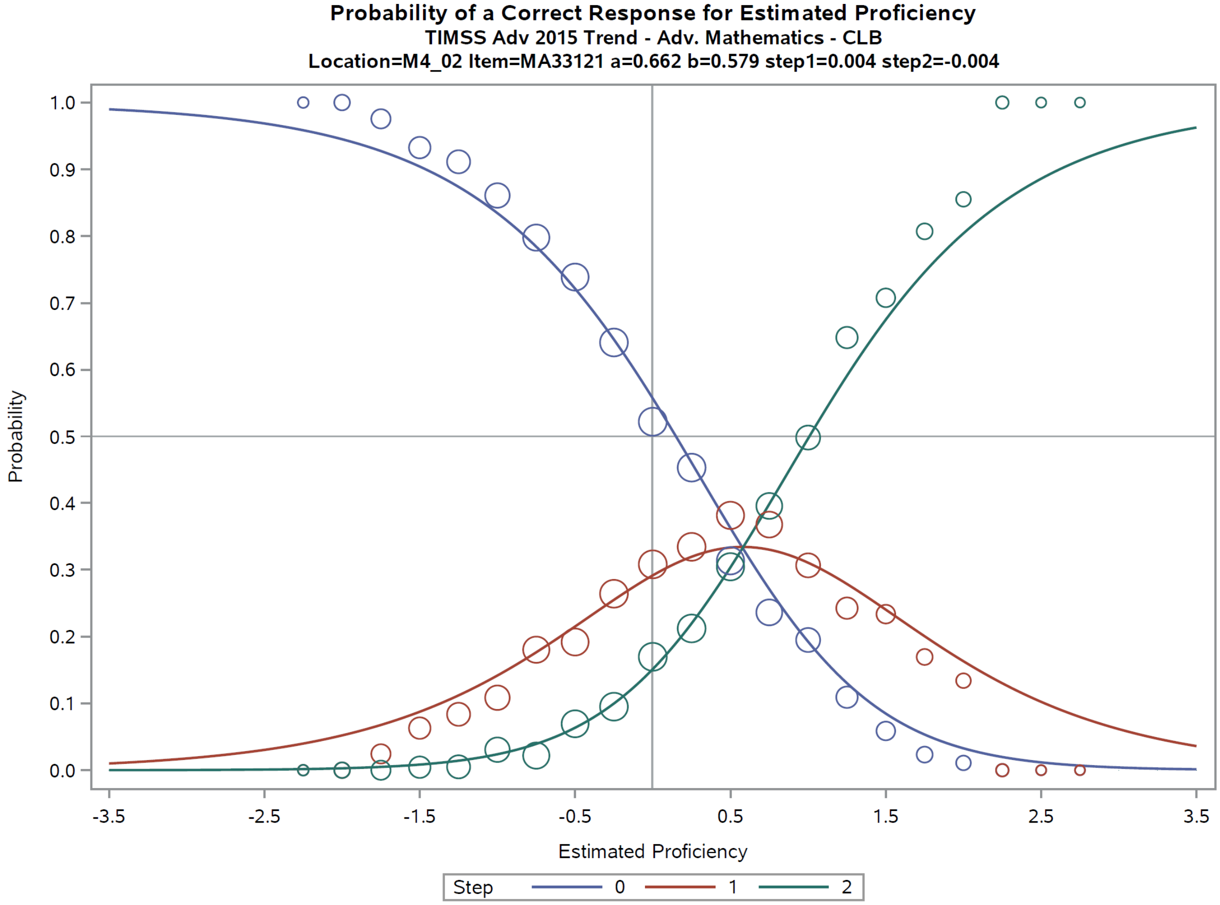


Exhibit 13.7: Example of Item Response Function for a Dichotomous Constructed Response Item from the TIMSS Advanced 2015 Physics Assessment



The plot in Exhibit 13.8 shows the empirical and fitted item response functions for a polytomous item (scored 0, 1, or 2). As for the dichotomous item plots, the horizontal axis represents the proficiency scale, but in this example the vertical axis represents the probability of having a response in a given response category. The fitted curves based on the estimated item parameters are shown as solid lines and again the empirical results are represented by circles. The interpretation of the circles is the same as in Exhibits 13.6 and 13.7. The curve starting at the top left of the chart plots the probability of a score of zero on the item. This probability decreases as proficiency increases. The bell-shaped curve shows the probability of a score of one point—partial credit, starting low for low-ability students, reaching a maximum for medium-ability students, and decreasing for high-ability students. The curve ending at the top right corner of the chart shows the probability of a score of two points—full credit, starting low for low-ability students and increasing as proficiency increases.

Exhibit 13.8: Example of Item Response Function for a Polytomous Constructed Response Item from the TIMSS Advanced 2015 Advanced Mathematics Assessment



Variables for Conditioning the TIMSS Advanced Assessment Data

Conditioning is the practice of using all available students' background information to improve the reliability of the estimated student proficiency scores. Ideally all background data would be included in the conditioning model, but because TIMSS Advanced has so many student background variables that could be used in conditioning, the TIMSS & PIRLS International Study Center follows the practice established by NAEP and followed by other large-scale studies of using principal components analysis to reduce the number of variables while explaining most of their common variance. Principal components for the TIMSS Advanced student background variables were constructed as follows:

- For categorical variables (questions with a small number of fixed response options), a dummy coded variable was created for each response option, with a value of one if the option is chosen and zero otherwise. If a student omitted or was not administered a particular question, all dummy coded variables associated with that question were assigned the value zero.

- Background variables with numerous response options (such as year of birth) were recoded using criterion scaling.[3] This was done by replacing the response option with the mean interim achievement score of all students choosing that option. Criterion scaling maximizes the correlation between the scaled variable and achievement. For TIMSS Advanced, the interim achievement score was the average of the advanced mathematics and physics EAP scores produced from the item calibrations.
- Separately for each country, all the dummy-coded and criterion-scaled variables were included in a principal components analysis. Those principal components accounting for 90 percent of the variance of the background variables were retained for use as conditioning variables.[4] Because the principal components analysis was performed separately for each country, different numbers of principal components were required to account for 90% of the common variance in each country's background variables.

In addition to the principal components, student gender (dummy coded), the language of the test (dummy coded), an indicator of the classroom in the school to which a student belongs (criterion scaled), and an optional country-specific variable (dummy coded) were included as primary conditioning variables, thereby accounting for most of the variance between students and preserving the between-classroom and within-classroom variance structure in the scaling model. Exhibits 13.9 and 13.10 provide details on the conditioning models used for proficiency estimation in advanced mathematics and physics, respectively.

3 The process of generating criterion-scaled variables is described in Beaton (1969).

4 The number of principal components retained is limited to no more than 5% of a country's student sample size, thereby possibly reducing the percentage of variance accounted for, to avoid over-specification of the conditioning model.

Exhibit 13.9: TIMSS Advanced 2015 Conditioning Models for Advanced Mathematics Proficiency Estimation

Country	2015				2008			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
France	2	349	180	90	—	—	—	—
Armenia	—	—	—	—	2	271	42	52
Iran, Islamic Rep. of	—	—	—	—	2	279	121	80
Italy	2	355	165	86	2	270	107	78
Lebanon	3	354	58	56	3	277	80	63
Netherlands	—	—	—	—	2	269	76	65
Norway	2	355	126	77	2	270	96	74
Portugal	2	354	178	90	—	—	—	—
Russian Federation	3	354	182	90	2	277	157	90
Russian Federation 6hr+	2	354	171	89	—	—	—	—
Slovenia	2	355	146	83	2	270	107	78
Sweden	2	351	182	90	2	268	115	81
United States	10	350	147	86	—	—	—	—

Exhibit 13.10: TIMSS Advanced 2015 Conditioning Models for Physics Proficiency Estimation

Country	2015				2008			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Armenia	—	—	—	—	2	275	44	53
France	2	347	177	90	—	—	—	—
Iran, Islamic Rep. of	—	—	—	—	2	282	121	79
Italy	2	355	171	87	2	206	93	78
Lebanon	3	355	57	56	3	281	79	63
Netherlands	—	—	—	—	2	272	75	64
Norway	2	354	123	76	2	272	82	68
Portugal	2	353	89	69	—	—	—	—
Russian Federation	2	354	178	90	2	283	158	90
Slovenia	2	355	55	56	2	272	54	53
Sweden	2	351	180	90	2	268	114	79
United States	10	350	146	86	—	—	—	—

Generating IRT Proficiency Scores for the TIMSS Advanced Assessment Data

Educational Testing Service's MGROUP program (Sheehan, 1985) was used to generate the IRT proficiency scores. This program takes as input the students' responses to the items they were given, the item parameters estimated at the calibration stage, and the conditioning variables, and generates as output the plausible values that represent student proficiency. TIMSS Advanced estimates overall advanced mathematics and physics using two separate MGROUP runs.

A useful feature of MGROUP is its ability to perform multi-dimensional scaling using the responses to all items across the proficiency scales and the correlations among the scales to improve the reliability of each individual scale. This feature of MGROUP was used to generate proficiency scores for the TIMSS Advanced 2015 content and cognitive domains using the item parameters estimated for the overall advanced mathematics and overall physics scales as well the same conditioning variables. The content domain scaling used two three-dimensional models, one to estimate proficiency scores for the three content domains in advanced mathematics and a second for the three physics content domains. The cognitive domain scaling relied on two three-dimensional models to estimate the three cognitive domains in advanced mathematics and physics.

In addition to generating plausible values on the overall advanced mathematics and physics scales for the 2015 assessment data, the item parameters estimated at the calibration stage also were used to generate plausible values for the TIMSS Advanced 2008 assessment for the countries included in the concurrent calibration. These additional plausible values were used to establish the linear transformation necessary to place the 2015 assessment data on the appropriate trend scales.

Transforming the Overall Scores to Measure Trends

To provide results for the TIMSS Advanced 2015 assessments on the existing TIMSS Advanced achievement scales, the 2015 proficiency scores (plausible values) for overall advanced mathematics and overall physics had to be transformed to the TIMSS Advanced reporting metric. This was accomplished through a set of linear transformations as part of the concurrent calibration approach. These linear transformations were given by:

$$PV^{*}_{k,i} = A_{k,i} + B_{k,i} \times PV_{k,i}$$

where

$PV_{k,i}$ is the TIMSS Advanced 2015 plausible value i of scale k prior to transformation;

$PV^{*}_{k,i}$ is the TIMSS Advanced 2015 plausible value i of scale k after transformation; and

$A_{k,i}$ and $B_{k,i}$ are the linear transformation constants.

The linear transformation constants were obtained by first computing the international means and standard deviations of the proficiency scores for the overall advanced mathematics and physics scales using the plausible values produced in 2008 based on the 2008 item calibrations for the trend countries. These were the plausible values published in 2008. Next, the same calculations were done using the plausible values from the re-scaled TIMSS 2008 assessment data based on the 2015 concurrent item calibrations for the same set of countries. From these calculations, the linear transformation constants were defined as:

$$B_{k,i} = \sigma_{k,i} / \sigma^*_{k,i}$$

$$A_{k,i} = \mu_{k,i} - B_{k,i} \cdot \mu^*_{k,i}$$

where

$\mu_{k,i}$ is the international mean of scale *k* based on plausible value *i* published in 2008;

$\mu^*_{k,i}$ is the international mean of scale *k* based on plausible value *i* from the 2008 assessment based on the 2015 concurrent calibration;

$\sigma_{k,i}$ is the international standard deviation of scale *k* based on plausible value *i* published in 2008;

$\sigma^*_{k,i}$ is the international standard deviation of scale *k* based on plausible value *i* from the 2008 assessment based on the 2015 concurrent calibration.

There are five sets of transformation constants for each scale, one for each plausible value. The trend countries contributed equally in the calculation of these transformation constants. Exhibits 13.11 and 13.12 show the TIMSS Advanced 2015 transformation constants for each subject, respectively.

Exhibit 13.11: TIMSS Advanced 2015 Linear Transformation Constants for Advanced Mathematics Achievement Scores

Overall Mathematics	TIMSS Advanced 2008 Published Scores		TIMSS Advanced 2008 Re-Scaled Scores		$A_{k,i}$	$B_{k,i}$
	Mean	Standard Deviation	Mean	Standard Deviation		
PV1	482.08004	102.02112	0.02399	0.96695	479.54919	105.50867
PV2	484.06485	101.87143	0.01989	0.97315	481.98240	104.68254
PV3	481.55003	101.94905	0.01522	0.98122	479.96851	103.90001
PV4	483.32440	101.66813	0.02280	0.97990	480.95897	103.75383
PV5	483.13244	102.27646	0.01689	0.98610	481.38015	103.71804

Exhibit 13.12: TIMSS Advanced 2015 Linear Transformation Constants for Physics Achievement Scores

Overall Physics	TIMSS Advanced 2008 Published Scores		TIMSS Advanced 2008 Re-Scaled Scores		$A_{k,i}$	$B_{k,i}$
	Mean	Standard Deviation	Mean	Standard Deviation		
PV1	498.80619	105.32288	0.20174	0.87253	474.45477	120.70987
PV2	499.21094	104.17338	0.19521	0.87781	476.04490	118.67351
PV3	498.78611	105.03022	0.19684	0.87638	475.19542	119.84490
PV4	498.93406	104.72493	0.19855	0.87162	475.07806	120.14980
PV5	498.86035	105.38060	0.20000	0.88298	474.99048	119.34691

These linear transformation constants were applied to the overall proficiency scores—advanced mathematics and physics—for all participating countries. This provided student achievement scores for the TIMSS Advanced 2015 assessments that are directly comparable to the scores from all previous assessments.

The linear transformation constants for the overall scales also were applied to the scales for the content and cognitive domains. The transformation constants for advanced mathematics were applied to the proficiency scores of the advanced mathematics content domains and cognitive domains, and the transformation constants for physics were applied to the proficiency scores of the physics content domains and cognitive domains. In this approach to measuring trends in content and cognitive domains, achievement changes over time are established in the context of achievement in each subject overall. Trends are not established separately for each content or cognitive domain; rather differential changes in performance in the domains are considered in the light of trends in the subject overall.

References

Beaton, A.E. (1969). Criterion scaling of questionnaire items. *Socio-Economic Planning Sciences*, 2, 355–362.

Mazzeo, J., & von Davier, M. (2014). Linking scales in international large-scale assessments. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues and methods of data analysis* (pp. 229–258). Boca Raton: Chapman & Hall/CRC.

Muraki, E., & Bock, R.D. (1991). PARSCALE: Parameter scaling of rating data [Computer Program]. Chicago, IL: Scientific Software, Inc.

Sheehan, K.M. (1985). M-Group: Estimation of group effects in multivariate models [Software Version 3.2]. Princeton, NJ: Educational Testing Service.

Appendix 13A: TIMSS Advanced 2015 Advanced Mathematics Item Parameters from Concurrent Calibration

TIMSS Advanced 2015 Advanced Mathematics Item Parameters from Concurrent Calibration

Item		Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})
Items Released in 2008:						
M1_01	MA13001	1.070 (0.069)	-0.322 (0.059)	0.111 (0.027)		
M1_02	MA13002	1.325 (0.099)	0.118 (0.051)	0.204 (0.025)		
M1_03	MA13003	0.580 (0.073)	0.532 (0.139)	0.200 (0.043)		
M1_04	MA13004	0.941 (0.090)	0.176 (0.087)	0.275 (0.034)		
M1_06	MA13006	0.895 (0.068)	0.390 (0.057)	0.095 (0.023)		
M1_07	MA13007	0.913 (0.083)	0.398 (0.073)	0.192 (0.029)		
M1_08	MA13008	1.064 (0.080)	-0.293 (0.074)	0.205 (0.034)		
M1_09	MA13009	0.732 (0.082)	0.538 (0.101)	0.221 (0.035)		
M3_01	MA13021	0.453 (0.069)	0.031 (0.292)	0.278 (0.070)		
M3_04	MA13024	1.243 (0.116)	0.572 (0.055)	0.255 (0.023)		
M3_05A	MA13025A	0.567 (0.034)	0.152 (0.048)			
M3_05B	MA13025B	0.877 (0.062)	1.922 (0.093)			
M3_06A	MA13026A	0.600 (0.035)	-0.340 (0.052)			
M3_06B	MA13026B	0.847 (0.052)	1.449 (0.064)			
M3_07	MA13027	0.468 (0.016)	0.770 (0.037)		-2.253 (0.119)	2.253 (0.124)
M3_08	MA13028	0.469 (0.044)	2.422 (0.192)			
M3_09	MA13029	0.426 (0.016)	0.548 (0.040)		-1.736 (0.109)	1.736 (0.112)
M6_01	MA23069	1.036 (0.124)	0.460 (0.077)	0.170 (0.030)		
M6_02	MA23135	0.573 (0.046)	0.013 (0.067)			
M6_03	MA23208	0.896 (0.127)	0.799 (0.089)	0.166 (0.030)		
M6_04	MA23165	0.645 (0.051)	0.460 (0.066)			
M6_05	MA23039	0.718 (0.080)	-0.306 (0.131)	0.151 (0.045)		
M6_06	MA23159	0.640 (0.048)	-0.420 (0.066)			
M6_07	MA23198	1.237 (0.084)	0.681 (0.042)			
M6_08	MA23042	1.046 (0.166)	0.571 (0.102)	0.333 (0.036)		
M6_09	MA23055	0.964 (0.112)	0.500 (0.076)	0.134 (0.029)		
M6_10	MA23080	1.182 (0.119)	0.272 (0.063)	0.135 (0.027)		
M6_11	MA23021	1.931 (0.286)	1.191 (0.055)	0.160 (0.016)		
M7_01	MA23004	0.797 (0.152)	0.904 (0.136)	0.300 (0.044)		
M7_02	MA23063	1.242 (0.207)	1.477 (0.087)	0.144 (0.020)		
M7_03	MA23141	1.030 (0.076)	0.953 (0.053)			

TIMSS Advanced 2015 Advanced Mathematics Item Parameters from Concurrent Calibration (Continued)

M7_04	MA23133	0.937 (0.135)	1.207 (0.085)	0.108 (0.024)		
M7_05	MA23158	1.486 (0.211)	1.304 (0.064)	0.124 (0.017)		
M7_06	MA23151	0.833 (0.133)	0.749 (0.116)	0.227 (0.041)		
M7_07A	MA23035A	0.995 (0.069)	0.424 (0.045)			
M7_07B	MA23035B	0.897 (0.076)	1.250 (0.076)			
M7_08	MA23050	1.085 (0.226)	1.482 (0.111)	0.237 (0.027)		
M7_09	MA23041	0.882 (0.144)	1.014 (0.101)	0.185 (0.035)		
M7_10	MA23182	1.291 (0.166)	0.660 (0.069)	0.198 (0.030)		
M7_11	MA23170	1.005 (0.077)	0.833 (0.054)			
Items Common in 2008 and 2015:						
M1_01	MA13011	0.925 (0.060)	0.201 (0.056)	0.155 (0.024)		
M1_02	MA13012	1.098 (0.089)	1.030 (0.043)	0.176 (0.015)		
M1_03	MA13013	1.051 (0.067)	0.088 (0.055)	0.207 (0.025)		
M1_05	MA13015	0.682 (0.054)	-0.171 (0.117)	0.214 (0.041)		
M1_06	MA13016	0.579 (0.079)	1.694 (0.105)	0.154 (0.025)		
M1_07	MA13017	0.743 (0.055)	0.790 (0.055)	0.091 (0.020)		
M1_08	MA13018	0.364 (0.040)	-0.530 (0.315)	0.186 (0.071)		
M1_09	MA13019	1.136 (0.138)	1.793 (0.074)	0.185 (0.012)		
M1_10	MA13020	1.580 (0.156)	1.308 (0.042)	0.274 (0.012)		
M3_01	MA23005	0.554 (0.053)	0.270 (0.121)	0.179 (0.038)		
M3_02	MA23145	0.561 (0.027)	0.595 (0.044)			
M3_03	MA23187	0.321 (0.010)	-0.145 (0.036)		-1.660 (0.092)	1.660 (0.090)
M3_04	MA23201	0.813 (0.031)	0.031 (0.027)			
M3_05	MA23154	0.449 (0.012)	-0.164 (0.027)		-1.246 (0.069)	1.246 (0.067)
M3_06	MA23206	1.232 (0.074)	-0.048 (0.046)	0.227 (0.022)		
M3_07	MA23166	1.158 (0.057)	1.542 (0.048)			
M3_08	MA23043	0.560 (0.016)	0.969 (0.031)		-1.102 (0.060)	1.102 (0.069)
M3_09	MA23076	0.744 (0.056)	0.177 (0.075)	0.171 (0.028)		
M3_10	MA23176	0.734 (0.074)	0.417 (0.096)	0.308 (0.031)		
M3_11	MA23098	1.044 (0.109)	1.096 (0.056)	0.275 (0.017)		
M5_01	MA23144	1.045 (0.086)	0.965 (0.045)	0.182 (0.016)		
M5_02	MA23185	0.733 (0.072)	0.787 (0.077)	0.195 (0.027)		
M5_03	MA23054	0.593 (0.018)	0.851 (0.027)		-0.546 (0.048)	0.546 (0.056)
M5_04	MA23064	0.753 (0.074)	1.207 (0.064)	0.138 (0.020)		
M5_05A	MA23131A	0.927 (0.036)	0.434 (0.026)			
M5_05B	MA23131B	0.889 (0.043)	1.490 (0.052)			
M5_06	MA23157	0.675 (0.025)	0.798 (0.027)		0.489 (0.035)	-0.489 (0.046)

TIMSS Advanced 2015 Advanced Mathematics Item Parameters from Concurrent Calibration (Continued)

M5_07	MA23045	1.133 (0.121)	1.531 (0.059)	0.175 (0.012)		
M5_08	MA23082	0.739 (0.074)	0.356 (0.104)	0.300 (0.034)		
M5_09	MA23020	0.892 (0.107)	1.157 (0.071)	0.316 (0.021)		
M5_10	MA23094	0.491 (0.019)	1.547 (0.054)		-0.479 (0.059)	0.479 (0.082)
Items Introduced in 2015:						
M2_01	MA33086	0.989 (0.097)	-0.371 (0.109)	0.160 (0.054)		
M2_02	MA33225	0.517 (0.023)	-0.186 (0.039)		-0.786 (0.090)	0.786 (0.087)
M2_03	MA33142	0.693 (0.049)	0.816 (0.065)			
M2_04	MA33044	0.792 (0.142)	1.109 (0.118)	0.251 (0.037)		
M2_05	MA33179	1.134 (0.064)	0.485 (0.036)			
M2_06	MA33076	0.414 (0.131)	1.128 (0.389)	0.245 (0.100)		
M2_07	MA33140	0.809 (0.102)	0.335 (0.115)	0.145 (0.047)		
M2_08	MA33007	0.878 (0.096)	0.105 (0.105)	0.149 (0.046)		
M2_09	MA33214	1.100 (0.068)	0.934 (0.048)			
M2_10	MA33171	0.898 (0.105)	0.456 (0.089)	0.147 (0.037)		
M2_11	MA33039	0.805 (0.049)	0.025 (0.043)			
M2_12	MA33180	0.541 (0.043)	0.892 (0.084)			
M4_01	MA33008	0.780 (0.050)	0.356 (0.047)			
M4_02	MA33121	0.662 (0.034)	0.579 (0.039)		0.004 (0.060)	-0.004 (0.072)
M4_03	MA33240	0.439 (0.022)	0.318 (0.047)		-0.627 (0.095)	0.627 (0.101)
M4_04	MA33050	0.945 (0.095)	0.502 (0.065)	0.070 (0.027)		
M4_05	MA33046	0.766 (0.090)	-0.608 (0.187)	0.177 (0.078)		
M4_06	MA33162	1.094 (0.143)	0.629 (0.077)	0.243 (0.031)		
M4_07	MA33163	1.214 (0.122)	0.223 (0.065)	0.175 (0.031)		
M4_08	MA33066	0.804 (0.096)	0.507 (0.089)	0.098 (0.037)		
M4_09	MA33182	1.014 (0.127)	0.606 (0.078)	0.191 (0.032)		
M4_10	MA33232	0.726 (0.032)	0.338 (0.032)		-0.378 (0.063)	0.378 (0.068)
M4_11	MA33178	0.709 (0.050)	0.202 (0.052)			
M6_01	MA33201	0.992 (0.169)	0.869 (0.100)	0.340 (0.034)		
M6_02	MA33016	1.171 (0.196)	0.758 (0.093)	0.419 (0.031)		
M6_03	MA33083	0.846 (0.053)	0.286 (0.042)			
M6_04	MA33143	1.055 (0.066)	0.801 (0.045)			
M6_05	MA33198	0.572 (0.101)	-0.223 (0.323)	0.243 (0.100)		
M6_06	MA33227	0.875 (0.061)	0.983 (0.059)			
M6_07	MA33079	1.064 (0.132)	0.428 (0.085)	0.249 (0.036)		
M6_08	MA33220	0.543 (0.030)	0.145 (0.039)		-0.039 (0.075)	0.039 (0.077)
M6_09	MA33150	0.545 (0.080)	-0.294 (0.268)	0.072 (0.095)		

TIMSS Advanced 2015 Advanced Mathematics Item Parameters from Concurrent Calibration (Continued)

M6_10	MA33233	0.459 (0.019)	0.530 (0.045)		-1.741 (0.122)	1.741 (0.128)
M6_11	MA33157	0.657 (0.114)	1.047 (0.122)	0.124 (0.044)		
M6_12	MA33155	1.093 (0.090)	1.584 (0.085)			
M7_01	MA33202	0.902 (0.100)	0.493 (0.077)	0.103 (0.034)		
M7_02	MA33042	0.978 (0.057)	-0.323 (0.040)			
M7_03	MA33094	0.587 (0.044)	0.079 (0.057)			
M7_04	MA33123	1.153 (0.147)	0.844 (0.065)	0.180 (0.026)		
M7_05	MA33137	0.894 (0.124)	0.655 (0.095)	0.201 (0.038)		
M7_06	MA33067	1.347 (0.075)	0.392 (0.031)			
M7_07	MA33012	0.831 (0.039)	1.005 (0.039)		-0.430 (0.062)	0.430 (0.077)
M7_08	MA33075	1.250 (0.157)	0.926 (0.060)	0.161 (0.022)		
M7_09	MA33212	1.496 (0.209)	1.000 (0.059)	0.234 (0.021)		
M7_10	MA33186	1.113 (0.186)	0.763 (0.096)	0.379 (0.033)		
M7_11	MA33239	0.467 (0.026)	0.874 (0.058)		-0.629 (0.094)	0.629 (0.111)
M7_12	MA33038	0.847 (0.060)	0.695 (0.054)			
M8_01	MA33027	1.007 (0.106)	-0.035 (0.092)	0.183 (0.043)		
M8_02	MA33091	0.518 (0.042)	0.430 (0.070)			
M8_03	MA33106	1.461 (0.164)	0.601 (0.053)	0.211 (0.023)		
M8_04	MA33090	0.527 (0.027)	0.280 (0.040)		-0.327 (0.077)	0.327 (0.083)
M8_05	MA33126	0.999 (0.131)	0.225 (0.106)	0.301 (0.042)		
M8_06	MA33118	0.311 (0.021)	0.458 (0.069)		-0.477 (0.123)	0.477 (0.135)
M8_07	MA33243	0.657 (0.027)	0.887 (0.040)		-1.967 (0.133)	1.967 (0.141)
M8_08	MA33229	1.220 (0.091)	1.390 (0.069)			
M8_09	MA33011	0.851 (0.084)	0.314 (0.073)	0.050 (0.032)		
M8_10	MA33159	0.921 (0.058)	0.600 (0.046)			
M8_11	MA33054	0.891 (0.124)	0.962 (0.082)	0.129 (0.028)		
M9_01	MA33085	1.228 (0.138)	0.319 (0.073)	0.253 (0.033)		
M9_02	MA33190	0.449 (0.039)	-0.664 (0.088)			
M9_03	MA33115	1.059 (0.166)	1.097 (0.085)	0.249 (0.027)		
M9_04	MA33237	0.933 (0.056)	0.407 (0.041)			
M9_05	MA33077	0.769 (0.118)	0.678 (0.122)	0.210 (0.045)		
M9_06	MA33132	0.611 (0.130)	0.915 (0.188)	0.263 (0.059)		
M9_07	MA33218	0.636 (0.026)	0.737 (0.038)		-1.373 (0.099)	1.373 (0.107)
M9_08	MA33236	0.621 (0.029)	0.809 (0.043)		-0.519 (0.074)	0.519 (0.087)
M9_09	MA33181	0.697 (0.097)	-0.240 (0.210)	0.202 (0.079)		
M9_10	MA33002	0.822 (0.053)	0.430 (0.047)			
M9_11	MA33169	1.228 (0.252)	1.341 (0.103)	0.373 (0.024)		
M9_12	MA33235	1.248 (0.082)	0.970 (0.047)			

Appendix 13B: TIMSS Advanced 2015 Physics Item Parameters from Concurrent Calibration

TIMSS Advanced 2015 Physics Item Parameters from Concurrent Calibration

Item		Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})
Items Released in 2008:						
P1_01	PA13001	0.670 (0.099)	0.478 (0.173)	0.337 (0.052)		
P1_02	PA13002	1.089 (0.110)	0.451 (0.076)	0.304 (0.032)		
P1_03	PA13003	1.059 (0.094)	0.114 (0.082)	0.253 (0.037)		
P1_04	PA13004	0.578 (0.056)	-1.001 (0.231)	0.199 (0.079)		
P1_05	PA13005	0.720 (0.088)	-0.266 (0.205)	0.352 (0.066)		
P1_06	PA13006	1.047 (0.183)	1.676 (0.099)	0.283 (0.020)		
P1_09	PA13009	0.877 (0.143)	1.571 (0.096)	0.234 (0.024)		
P3_01	PA13021	0.812 (0.090)	0.644 (0.087)	0.190 (0.035)		
P3_02	PA13022	0.693 (0.026)	1.214 (0.036)		-0.890 (0.065)	0.890 (0.077)
P3_03	PA13023	0.635 (0.038)	0.030 (0.045)			
P3_04	PA13024	0.671 (0.027)	1.093 (0.036)		-0.466 (0.056)	0.466 (0.068)
P3_05	PA13025	0.589 (0.024)	1.267 (0.043)		-0.815 (0.070)	0.815 (0.085)
P3_06	PA13026	1.051 (0.076)	1.817 (0.080)			
P3_07A	PA13027A	0.834 (0.053)	1.234 (0.057)			
P6_01	PA23050	1.010 (0.201)	1.148 (0.109)	0.330 (0.035)		
P6_02	PA23056	0.441 (0.085)	-0.032 (0.385)	0.273 (0.091)		
P6_03	PA23142	1.231 (0.169)	0.727 (0.078)	0.261 (0.033)		
P6_04	PA23072	0.677 (0.053)	0.352 (0.057)			
P6_05	PA23022	0.473 (0.027)	1.702 (0.087)		-1.701 (0.150)	1.701 (0.179)
P6_06	PA23030	1.727 (0.253)	1.395 (0.058)	0.111 (0.015)		
P6_07	PA23078	0.320 (0.043)	0.870 (0.139)			
P6_08	PA23113	0.880 (0.184)	1.268 (0.122)	0.284 (0.038)		
P6_09	PA23128	0.454 (0.047)	0.874 (0.100)			
P6_10	PA23058	1.163 (0.167)	0.767 (0.084)	0.267 (0.034)		
P6_11	PA23115	1.365 (0.196)	1.171 (0.063)	0.165 (0.022)		
P7_01	PA23110	0.651 (0.124)	0.897 (0.169)	0.233 (0.055)		
P7_02	PA23014	0.570 (0.049)	0.190 (0.067)			
P7_03	PA23025	0.779 (0.040)	1.323 (0.047)		-0.831 (0.088)	0.831 (0.104)
P7_04	PA23028	1.024 (0.142)	0.626 (0.098)	0.247 (0.041)		
P7_05	PA23034	0.448 (0.050)	1.293 (0.128)			
P7_06	PA23044	0.795 (0.067)	1.328 (0.080)			

TIMSS Advanced 2015 Physics Item Parameters from Concurrent Calibration (Continued)

P7_07	PA23082	0.719 (0.055)	0.169 (0.055)			
P7_08	PA23140	0.687 (0.138)	1.328 (0.140)	0.194 (0.044)		
P7_09	PA23084	0.752 (0.038)	1.297 (0.045)		-1.635 (0.136)	1.635 (0.146)
P7_10	PA23059	0.587 (0.083)	-0.666 (0.286)	0.258 (0.087)		
P7_11	PA23138	1.436 (0.198)	0.302 (0.096)	0.429 (0.040)		
P7_12	PA23137	0.591 (0.052)	0.674 (0.070)			
Items Common in 2008 and 2015:						
P1_01	PA13011	0.301 (0.057)	0.750 (0.411)	0.230 (0.077)		
P1_02	PA13012	1.392 (0.110)	1.357 (0.037)	0.099 (0.009)		
P1_03	PA13013	0.914 (0.112)	1.525 (0.072)	0.240 (0.017)		
P1_04	PA13014	0.896 (0.069)	0.512 (0.059)	0.201 (0.024)		
P1_05	PA13015	0.903 (0.116)	0.616 (0.105)	0.555 (0.025)		
P1_06	PA13016	1.132 (0.097)	1.341 (0.045)	0.124 (0.012)		
P1_07	PA13017	0.301 (0.043)	0.276 (0.360)	0.189 (0.069)		
P1_08	PA13018	0.633 (0.082)	1.147 (0.097)	0.254 (0.030)		
P1_09	PA13019	0.521 (0.080)	1.626 (0.117)	0.178 (0.031)		
P3_01	PA23071	0.674 (0.140)	1.595 (0.139)	0.470 (0.025)		
P3_02	PA23146	0.665 (0.028)	-0.170 (0.033)			
P3_03	PA23029	1.108 (0.066)	0.424 (0.038)	0.127 (0.017)		
P3_04	PA23104	0.725 (0.082)	0.723 (0.094)	0.301 (0.030)		
P3_05	PA23038	0.788 (0.061)	0.628 (0.058)	0.138 (0.022)		
P3_06	PA23041	0.584 (0.026)	-0.584 (0.043)			
P3_07	PA23053	0.509 (0.016)	0.589 (0.028)		-0.403 (0.051)	0.403 (0.057)
P3_08	PA23148	0.446 (0.055)	-0.054 (0.254)	0.254 (0.062)		
P3_09	PA23119	0.484 (0.014)	1.508 (0.044)		-1.893 (0.088)	1.893 (0.101)
P3_10	PA23088	0.726 (0.029)	-0.081 (0.030)			
P3_11	PA23066	0.699 (0.030)	0.636 (0.035)			
P5_01	PA23048	0.616 (0.064)	-0.638 (0.226)	0.390 (0.062)		
P5_02	PA23039	0.797 (0.091)	1.266 (0.070)	0.220 (0.021)		
P5_03	PA23035	0.830 (0.032)	0.104 (0.027)			
P5_04	PA23042	0.857 (0.108)	1.127 (0.077)	0.349 (0.023)		
P5_05	PA23012	0.841 (0.039)	1.308 (0.047)			
P5_06 *	PA23131	0.984 (0.160)	1.606 (0.093)	0.253 (0.022)		
P5_07	PA23051	0.982 (0.037)	0.504 (0.025)			
P5_08	PA23085	0.882 (0.117)	1.391 (0.077)	0.309 (0.020)		
P5_09	PA23130	0.502 (0.014)	1.102 (0.033)		-1.489 (0.071)	1.489 (0.080)
P5_10	PA23086	0.535 (0.030)	1.284 (0.067)			
P5_11	PA23064	0.784 (0.078)	0.643 (0.081)	0.264 (0.029)		

* Item P5_06 is not included in the TIMSS Advanced 2015 Physics Framework. All responses to the item in the 2015 assessment were set to Not Administered.

TIMSS Advanced 2015 Physics Item Parameters from Concurrent Calibration (Continued)

Items Introduced in 2015:						
P2_01A	PA33061A	1.223 (0.118)	0.664 (0.051)	0.096 (0.019)		
P2_01B	PA33061B	0.639 (0.042)	-0.037 (0.053)			
P2_02	PA33004	0.749 (0.049)	0.679 (0.058)			
P2_03	PA33044	0.487 (0.131)	1.771 (0.218)	0.161 (0.048)		
P2_04	PA33075	0.623 (0.042)	0.248 (0.057)			
P2_05A	PA33102A	0.284 (0.032)	-1.777 (0.203)			
P2_05B	PA33102B	0.366 (0.034)	-0.168 (0.087)			
P2_06	PA33121	0.990 (0.172)	0.958 (0.102)	0.358 (0.029)		
P2_07	PA33115	1.070 (0.126)	0.765 (0.067)	0.165 (0.024)		
P2_08	PA33005	0.631 (0.041)	-0.509 (0.058)			
P2_09A	PA33101A	0.647 (0.090)	0.181 (0.158)	0.185 (0.054)		
P2_09B	PA33101B	0.492 (0.053)	2.319 (0.229)			
P4_01	PA33078	0.701 (0.078)	0.310 (0.098)	0.082 (0.036)		
P4_02	PA33088	1.149 (0.099)	-0.122 (0.064)	0.140 (0.029)		
P4_03	PA33058	0.462 (0.019)	0.200 (0.044)		-1.190 (0.104)	1.190 (0.109)
P4_04	PA33057	0.870 (0.122)	0.835 (0.089)	0.192 (0.029)		
P4_05	PA33047	0.567 (0.067)	2.686 (0.281)			
P4_06	PA33012	0.892 (0.127)	1.050 (0.087)	0.159 (0.025)		
P4_07	PA33120	0.862 (0.153)	1.326 (0.110)	0.178 (0.027)		
P4_08A	PA33079A	0.802 (0.124)	0.894 (0.103)	0.217 (0.032)		
P4_08B	PA33079B	0.828 (0.051)	0.524 (0.051)			
P4_09	PA33116	0.986 (0.114)	0.497 (0.076)	0.185 (0.028)		
P4_10	PA33070	0.591 (0.039)	-0.716 (0.064)			
P4_11	PA33011	0.466 (0.037)	0.041 (0.072)			
P6_01	PA33059	0.581 (0.106)	-0.300 (0.315)	0.351 (0.083)		
P6_02	PA33073	0.481 (0.025)	0.027 (0.043)		-0.177 (0.084)	0.177 (0.087)
P6_03	PA33019	1.282 (0.244)	1.429 (0.097)	0.250 (0.018)		
P6_04	PA33015	0.650 (0.113)	0.752 (0.141)	0.216 (0.045)		
P6_05	PA33086	0.782 (0.129)	0.904 (0.110)	0.226 (0.035)		
P6_06	PA33035	0.315 (0.033)	-0.232 (0.100)			
P6_07	PA33119	0.558 (0.035)	2.005 (0.112)		-1.533 (0.155)	1.533 (0.201)
P6_08	PA33046	0.765 (0.048)	0.229 (0.048)			
P6_09	PA33083	0.715 (0.092)	0.061 (0.139)	0.186 (0.050)		
P6_10	PA33069	0.728 (0.046)	-0.994 (0.063)			
P6_11	PA33114	0.559 (0.106)	-0.218 (0.326)	0.338 (0.085)		
P6_12	PA33080	0.832 (0.056)	0.871 (0.062)			

TIMSS Advanced 2015 Physics Item Parameters from Concurrent Calibration (Continued)

P7_01	PA33065	0.936	(0.174)	0.876	(0.114)	0.382	(0.032)				
P7_02A	PA33009A	0.693	(0.032)	1.099	(0.050)			-0.811	(0.083)	0.811	(0.102)
P7_02B	PA33009B	0.895	(0.116)	0.527	(0.090)	0.200	(0.034)				
P7_03	PA33002	0.725	(0.113)	-0.498	(0.248)	0.411	(0.071)				
P7_04	PA33098	0.769	(0.135)	0.991	(0.116)	0.236	(0.035)				
P7_05A	PA33028A	0.545	(0.023)	0.811	(0.047)			-1.434	(0.108)	1.434	(0.120)
P7_05B	PA33028B	0.730	(0.097)	0.051	(0.146)	0.217	(0.052)				
P7_06A	PA33054A	1.312	(0.290)	1.551	(0.114)	0.304	(0.018)				
P7_06B	PA33054B	0.866	(0.065)	1.351	(0.086)						
P7_07	PA33040	0.475	(0.113)	0.316	(0.359)	0.295	(0.089)				
P7_08A	PA33095A	0.720	(0.047)	-1.107	(0.067)						
P7_08B	PA33095B	0.671	(0.105)	0.466	(0.146)	0.207	(0.050)				
P8_01A	PA33066A	0.686	(0.198)	1.422	(0.197)	0.410	(0.038)				
P8_01B	PA33066B	0.823	(0.131)	-0.175	(0.190)	0.466	(0.053)				
P8_02	PA33090	0.383	(0.022)	0.796	(0.070)			-0.603	(0.105)	0.603	(0.127)
P8_03A	PA33064A	0.323	(0.034)	0.233	(0.106)						
P8_03B	PA33064B	0.552	(0.027)	1.355	(0.070)			-1.079	(0.105)	1.079	(0.133)
P8_04	PA33110	0.335	(0.119)	-0.607	(1.247)	0.352	(0.213)				
P8_05	PA33118	0.465	(0.153)	1.582	(0.264)	0.281	(0.061)				
P8_06	PA33109	0.755	(0.090)	0.202	(0.110)	0.149	(0.041)				
P8_07	PA33029	0.488	(0.077)	0.512	(0.171)	0.050	(0.057)				
P8_08	PA33097	0.595	(0.041)	-0.575	(0.061)						
P8_09	PA33099	0.684	(0.124)	0.666	(0.152)	0.292	(0.046)				
P8_10	PA33008	0.209	(0.017)	-0.795	(0.101)			-0.694	(0.197)	0.694	(0.181)
P9_02	PA33072	0.440	(0.036)	-0.932	(0.094)						
P9_03	PA33063	1.187	(0.198)	0.792	(0.091)	0.414	(0.027)				
P9_04	PA33077	0.623	(0.025)	0.501	(0.037)			-1.007	(0.085)	1.007	(0.092)
P9_05A	PA33111A	0.642	(0.026)	0.509	(0.036)			-0.936	(0.082)	0.936	(0.089)
P9_05B	PA33111B	0.900	(0.153)	0.910	(0.106)	0.277	(0.034)				
P9_06	PA33003	0.813	(0.109)	-0.831	(0.224)	0.384	(0.075)				
P9_07	PA33081	0.963	(0.192)	1.259	(0.114)	0.324	(0.027)				
P9_08	PA33045	0.512	(0.128)	0.566	(0.310)	0.358	(0.075)				
P9_09A	PA33094A	0.419	(0.039)	0.837	(0.108)						
P9_09B	PA33094B	0.628	(0.146)	1.208	(0.169)	0.282	(0.046)				
P9_09C	PA33094C	0.848	(0.048)	1.263	(0.052)			-0.098	(0.060)	0.098	(0.088)

CHAPTER 14

Using Scale Anchoring to Interpret the TIMSS Advanced 2015 Achievement Scales

Ina V.S. Mullis
Kerry E. Cotter
Victoria A.S. Centurino
Bethany G. Fishbein
Katherine A. Reynolds

Introduction

As described in Chapter 13: Scaling the TIMSS Advanced 2015 Achievement Data, the TIMSS Advanced 2015 achievement results are summarized using item response theory (IRT) scaling and reported on 0 to 1,000 achievement scales, with most achievement scores ranging from 300 to 700. Countries' average scores provide users of the data with information about how achievement compares among countries and whether scores are improving or declining over time.

To provide important information for policy and curriculum reform, however, it is important to understand the advanced mathematics and physics competencies associated with different locations within the range of scores on the achievement scales. For example, in terms of levels of student understanding, what does it mean for a country to have average achievement of 513 or 426, and how are these scores different?

The TIMSS Advanced 2015 International Benchmarks provide information about what students know and can do at different points along the achievement scales. More specifically, TIMSS Advanced has identified three points along the achievement scales to use as international benchmarks of achievement—Advanced International Benchmark (625), High International Benchmark (550), and Intermediate International Benchmark (475). For each assessment, the TIMSS & PIRLS International Study Center works with the expert international committee, Science and Mathematics Item Review Committee (SMIRC), to conduct a scale anchoring analysis to describe student competencies at the benchmarks.

This chapter describes the scale anchoring procedures that were applied to describe student performance at the international benchmarks for TIMSS Advanced 2015. The analysis was conducted separately for advanced mathematics and for physics. In brief, scale anchoring involved identifying items that students scoring at the international benchmarks answered correctly, and then having experts examine the content of each item to determine the kind of knowledge, skill, or reasoning demonstrated by students who responded correctly to the item. The experts then summarized the detailed list of item competencies in a brief description of achievement at each international benchmark. Thus, the scale anchoring procedure yielded a content-referenced interpretation of the achievement results that can be considered in light of the TIMSS Advanced 2015 frameworks for assessing advanced mathematics and physics.

Classifying the Items

As the first step, students scoring within 40 scale-score points of each benchmark (i.e., the benchmark plus or minus 20) were identified for the benchmark analysis. The range of 40 points provided an adequate sample of students scoring at the benchmark, yet was small enough so that performance at one international benchmark was still distinguishable from the next. The score ranges around each international benchmark and the number of students scoring in each range are shown in Exhibit 14.1.

Exhibit 14.1: Range Around Each International Benchmark and Number of Students Within Each Range

	Intermediate (475)	High (550)	Advanced (625)
Range of Scale Scores	*455–495*	*530–570*	*605–645*
TIMSS Advanced Advanced Mathematics	5,887	4,369	1,687
TIMSS Advanced Physics	3,083	2,318	1,071

The second step involved computing the percentage of those students scoring in the range around each international benchmark that answered each item correctly. To compute these percentages, students in each country were weighted proportionally to the size of the student population in the country. For multiple-choice items and constructed response items worth 1 point, it was a straightforward matter of computing the percentage of students at each benchmark who answered each item correctly. For constructed response items scored for partial and full credit, percentages were computed for students receiving partial credit (1-point) as well as for students receiving full credit (2-points).

Third, the criteria described below were applied to identify the items that anchored at each benchmark. An important feature of the scale anchoring method is that it yields descriptions of the performance demonstrated by students reaching each of the international benchmarks on the scales, and that the descriptions reflect demonstrably different accomplishments by students reaching each successively higher benchmark. Because the process entails the delineation of sets of items that students at each international benchmark are likely to answer correctly and that discriminate between one benchmark and the next, the criteria for identifying the items that anchor considers performance at more than one benchmark.

For multiple-choice items, 65 percent was used as the criterion for anchoring at each benchmark being analyzed, since students would be likely (about two thirds of the time) to answer the item correctly. A somewhat less strict criterion was used for the constructed response items, because students had much less scope for guessing. For constructed response items, the criterion of 50 percent was used for the benchmark without any discrimination criterion for the next lower benchmark. In addition, a criterion of less than 50 percent was used for the next lower benchmark, because with this response probability, students were more likely to have answered the item incorrectly than correctly.

Using a multiple-choice item as an example, the criteria for each benchmark are outlined below:

- A multiple-choice item anchored at the Intermediate International Benchmark (475) if at least 65 percent of students scoring in the range answered the item correctly. Because this was the lowest benchmark described, there were no further criteria.
- A multiple-choice item anchored at the High International Benchmark (550) if at least 65 percent of students scoring in the range answered the item correctly, and less than 50 percent of students at the Intermediate International Benchmark answered the item correctly.
- A multiple-choice item anchored at the Advanced International Benchmark (625) if at least 65 percent of students scoring in the range answered the item correctly, and less than 50 percent of students at the High International Benchmark answered the item correctly.

To include all of the multiple-choice items in the anchoring process and provide information about content domains and cognitive processes that might not otherwise have had many anchor items, the concept of items that "almost anchored" was introduced. These were items that met slightly less stringent criteria for being answered correctly. The criteria to identify multiple-choice items that "almost anchored" were that 60 to 65 percent of students scoring in the range answered the item correctly and less than 50 percent of students at the next lowest benchmark answered the item correctly. To be completely inclusive for all items, items that met only the criterion that

60 to 65 percent of the students answered correctly (regardless of the performance of students at the next lower point) were also identified. The categories of items were mutually exclusive, and ensured that all of the items were available to inform the descriptions of student achievement at the anchor levels. A multiple-choice item was considered to be "too difficult" to anchor if less than 60 percent of students at the advanced benchmark answered the item correctly. A constructed response item was considered to be "too difficult" to anchor if less than 50 percent of students at the advanced benchmark answered the item correctly.

Exhibit 14.2 presents the number of TIMSS Advanced 2015 advanced mathematics and physics items that anchored at each international benchmark. A description of the items for advanced mathematics and physics can be found in Appendix 14A and 14B, respectively. It should be noted that a partial credit item can anchor twice, typically at a higher benchmark for full credit, and a lower benchmark for partial credit (but sometimes both anchored at the same level). Scale anchoring for the physics items considered partial credit and full credit responses separately. Scale anchoring for advanced mathematics used only the full credit anchoring results.

Exhibit 14.2: Number of Items Anchoring and Almost Anchoring at Each International Benchmark

	Intermediate (475)	High (550)	Advanced (625)	Above Advanced	Total
TIMSS Advanced – Advanced Mathematics					
Algebra	6	20	11	0	37
Calculus	5	12	13	4	34
Geometry	6	9	12	3	30
Advanced Mathematics Total	**17**	**41**	**36**	**7**	**101**
TIMSS Advanced – Physics					
Mechanics & Thermodynamics	12	19	7	8	46
Electricity & Magnetism	10	6	9	6	31
Wave Phenomena & Atomic/Nuclear Physics	14	8	11	5	38
Physics Total	**36**	**33**	**27**	**19**	**115**

Writing the Scale Anchoring Descriptions

The scale anchoring for TIMSS Advanced 2015 was conducted in the spring of 2016 at a four-day meeting in Seoul, South Korea. In preparation for review by SMIRC, staff at the TIMSS & PIRLS International Study Center used examples from previous assessments to draft short descriptions of the student competencies demonstrated by a correct (or partially correct) response to each advanced mathematics and physics item. Then, the advanced mathematics and physics items were separately grouped by international benchmark, and within each benchmark the items were sorted by content area. The final categorization was by the anchoring criteria the items met—items that anchored, followed by items that almost anchored, followed by items that met only the 60 to 65 percent criteria. Also, in addition to the short draft descriptions, the following information was included for each item: framework classification, answer key or scoring guide, secure status, percent correct at each benchmark, and overall international percent correct.

At the scale anchoring meetings, the expert committees 1) worked through each item to finalize the description of the student competencies demonstrated by a correct (or a partially correct) response, 2) summarized the proficiency demonstrated by students reaching each international benchmark for publication in reports, and 3) selected example items that supported and illustrated the benchmark descriptions to publish together with the descriptions.

Following the scale anchoring meeting, the descriptions and example items published in the TIMSS Advanced 2015 report were reviewed by National Research Coordinators at their 8th meeting in Quebec City, Canada.

Appendix 14A: TIMSS Advanced 2015 Advanced Mathematics Item Descriptions Developed During the TIMSS Advanced 2015 Benchmarking

Items at Intermediate International Benchmark (475)	
Algebra	
M2_01	Recognizes the graph of the absolute value of a function given the graph of the function
M3_01	Determines which term has a given value in a geometric sequence
M3_04	Analyzes steps in a given solution of a simple logarithmic equation and identifies an error
M7_02	Computes the value of a composite function at a given value
M8_01	Identifies the expression that results from the composite of a function with itself
M9_02	Evaluates an exponential expression with three unknowns given three possible values for each unknown
Calculus	
M1_05	Differentiates an exponential function where the exponent is a simple polynomial
M3_06	Analyzes the graph of a function to determine the sign of its derivative
M4_05	Computes the limit of an exponential function
M6_05	Integrates the sum of an exponential function and a monomial
M8_05	Determines the limit of a rational function in terms of an unknown constant
Geometry	
M1_08	Calculates the difference between vectors in coordinate form
M2_11	Finds the length of a diagonal of a regular hexagon of given side length
M3_09	Evaluates the shortest path between opposite vertices on the surface of a cube
M3_10	Solves a word problem about height given the distance and angle of elevation
M6_09	Recognizes a diagram of the sum of three vectors
M9_09	Identifies the length of a side of an isosceles triangle using properties of a similar triangle

Items at High International Benchmark (550)	
Algebra	
M1_01	Rationalizes the denominator of an expression
M1_03	Determines when a rational function with numerator and denominator in factored form is negative
M2_02	Indicates whether factored polynomials satisfy two given conditions (2 of 2 points)
M2_05	Determines the values of two constants in a rational expression given its graph with two specified points
M3_02	Solves a word problem about the number of permutations
M3_03	Solves a word problem involving dimensions of two cylindrical containers given their volumes (2 of 2 points)
M4_01	Simplifies an expression with log base 10 in the exponent
M4_04	Determines the values of two constants in a rational function given its asymptotes
M5_02	Identifies two constants in a rational function given two points on its graph
M5_05A	Solves a word problem by finding the distance between the points at which a parabola intersects the x-axis
M6_01	Recognizes the graph that could represent a function and its inverse
M6_02	Identifies the solution of a quadratic inequality
M6_03	Solves an exponential decay equation for the time at which a specified amount of substance remains
M7_01	Determines the interval on which a given rational function is greater than the square of that function
M7_03	Multiplies complex numbers
M7_04	Determines the domain of a logarithm of a rational function
M8_02	Finds the value of a particular term of an arithmetic sequence
M8_03	Uses the initial value of a fractional expression with three unknowns to evaluate the expression after the unknowns are divided by multiples of 2
M9_01	Identifies an increasing function defined for all real numbers
M9_04	Determines the value of an unknown in a logarithmic equation given its two solutions
Calculus	
M2_07	Identifies the graph of a function that satisfies given conditions for the first and second derivatives
M2_08	Determines the limit of a rational function in terms of an unknown constant

M3_05	Finds the second derivative of a rational function (2 of 2 points)
M4_06	Identifies the derivative of a composite trigonometric function
M4_07	Identifies the graph of a function given the graph of its first derivative
M4_08	Identifies the values of a definite integral with an unknown upper bound
M6_07	Identifies the local maximum of a function given intervals on which its first and second derivatives are positive, negative, and zero
M6_08	Sketches the graph of a function on a specified interval with three given properties (2 of 2 points)
M7_05	Uses the additivity of intervals to identify the value of a definite integral
M7_06	Determines the derivative of the product of a monomial and an unspecified function
M9_05	Identifies a true statement about discontinuity and non-differentiability for a graph of a piecewise function
M9_06	Recognizes the graph of the derivative of a curvilinear, discontinuous function
Geometry	
M2_10	Determines the value of a trigonometric function given the value of a related function
M4_09	Determines the ratio of the squares of two sides of a scalene triangle given two of its angles
M4_10	Finds the maximum value of a trigonometric function and a value of the independent variable at which it occurs (2 of 2 points)
M4_11	Proves that a quadrilateral with given coordinates of its vertices is a parallelogram
M5_08	Identifies coordinates of the fourth vertex of a parallelogram when three vertices are given
M7_10	Recognizes the description of a sine graph transformation
M8_09	Identifies a vector that is perpendicular to a given vector in a coordinate system
M8_10	Determines the lengths of two sides of a triangle given its area, the sum of the lengths of the two sides, and the angle included between them
M9_10	Determines the coordinates of line segment endpoints given the midpoint

Items at Advanced Benchmark (625)	
Algebra	
M1_02	Calculates the cube of a complex number given in trigonometric form
M2_03	Finds the sum of the first 100 terms of an alternating series at a given value of x
M2_04	Determines the sum of an infinite alternating geometric series

M4_02	Determines the intersection of two functions in terms of an unknown, non-zero coefficient (2 of 2 points)
M4_03	Determines the value when one cost becomes less than another and explains whether increasing the initial costs will change the value (2 of 2 points)
M5_01	Given the first three terms, calculates the sum of an infinite geometric series
M5_03	Solves a logarithmic equation (2 of 2 points)
M5_04	Given one imaginary root, identifies the constant term of a third-degree polynomial with known coefficients
M6_04	Determines the coefficient of the linear term and the constant of a quadratic equation given its solution
M8_04	Determines the amount of time that a ball is at or above a specified height given the quadratic function for its height (2 of 2 points)
M9_03	Determines the values of an unknown coefficient for which the graph of a parabola lies above the x-axis
Calculus	
M1_06	Maximizes the volume of a cylinder given a relationship between its height and diameter
M2_06	Identifies the value of a definite integral from areas shown on a graph
M2_09	Explains whether a given piecewise function is continuous at a given value
M3_08	Calculates the area between the graphs of a linear and a quadratic function (2 of 2 points)
M5_06	Given the graph of the derivative of a function, determines the x-values of the maximum point and the point of inflection of the function (2 of 2 points)
M6_06	Explains whether a right-hand limit and a left-hand limit of a function are equal
M7_07	Maximizes the area of a rectangle with constraint on the sum of three sides and explains why the solution gives the maximum area (2 of 2 points)
M7_08	Identifies specific properties of the first and second derivatives of a function given its graph
M7_09	Determines the limit of a rational function in terms of an unknown constant
M8_06	Indicates whether statements about the continuity or differentiability of a function with given conditions are true (2 of 2 points)
M8_07	Solves a multi-step word problem by maximizing the profit given a quadratic cost function and the linear income function (2 of 2 points)
M9_08	Determines the equation of a line parallel to a tangent line of a given function at a specified point (2 of 2 points)
M9_08	Indicates whether statements about a function are true given a graph of the derivative (2 of 2 points)

Geometry	
M1_07	Identifies the equation of a line through a given point and perpendicular to a given line
M1_10	Uses vector sums and differences to express a relationship among three vectors shown in a figure
M2_12	Determines the length of a line segment in a problem involving similar right triangles
M3_11	Uses properties of vectors to analyze equivalence of conditions involving the sum and difference of two vectors
M5_09	Compares amplitudes and periods of sine functions
M6_10	Justifies a statement regarding the length of the radius of a circle drawn on a square grid (2 of 2 points)
M6_11	Identifies the parameter of a sine function given the graphs of a function and its transformation
M7_11	Solves a word problem involving concentric circles and areas of sectors (2 of 2 points)
M7_12	Explains why the sum of a sine and a cosine function does not exceed a specified value
M8_11	Determines a diagonal length of a rhombus in terms of the length of a side given the ratio between the obtuse and acute angles
M9_11	Identifies the parameters of a cosine function used to model data presented in a graph
M9_12	Proves that a trigonometric relation holds for a triangle with specified angle and side measures

Items Above the Advanced International Benchmark (625)	
Calculus	
M3_07	Solves a multi-step word problem by maximizing the profit given a quadratic cost function and the unit selling price
M5_05B	Solves a multi-step word problem by calculating the area between two intersecting parabolas
M5_07	Determines the vertical line that divides a specified area between a parabola and the x-axis into equal parts
M8_08	Shows a process for integrating the product of a linear and a trigonometric function
Geometry	
M1_09	Given two points, identifies an equation that represents the set of all points twice as far from one of the given points as from the other
M5_10	Calculates the two possible lengths of a side of a triangle given an angle and the lengths of two sides that do not include the angle (2 of 2 points)
M6_12	Proves the equality of sines of supplementary angles

Appendix 14B: TIMSS Advanced 2015 Physics Item Descriptions Developed During the TIMSS Advanced 2015 Benchmarking

Items at Intermediate International Benchmark (475)	
Mechanics & Thermodynamics	
P1_05	Recognizes the process of energy transfer from the Sun to the Earth
P2_01B	Compares the amount of time it takes an object to reach the apex of its motion from a given point and the time it takes to fall from the apex back to the given point
P3_01	Selects the graph that best represents the potential energy of a ball rolling up and down an inclined plane
P3_02	Describes the direction of the acceleration of a body moving in a circular path with constant speed
P3_08	Identifies the best explanation for the temperature change in a rising air mass
P4_02	Recognizes how the force exerted by the Sun on Planet X compares with the force exerted by the Sun on the Earth, given the masses of the planets and the relationship between their distances to the Sun
P4_03	Calculates the final velocity of two skiers after they collide inelastically (1 of 2 points)
P6_01	Recognizes the relationship between the change in internal energy and the change in temperature of a gas when work is done on it by the environment
P6_02	Calculates the initial height from which a body began moving vertically down (1 of 2 points)
P7_03	Identifies an energy transformation that occurs when a meteor enters Earth's atmosphere and is incinerated
P8_01B	Identifies the diagram that best represents the path of a ball attached to a string after it has been released from circular motion at a constant speed
P9_02	Calculates work done by friction to stop an object sliding along a rough surface
Electricity & Magnetism	
P2_05A	Identifies the direction of the force on a point charge in various positions in an electric field
P2_05B	Orders three points in an electric field by increasing field strength
P3_04	Identifies the direction of the electric force on a charged object in an electric field
P3_06	States the meaning of the symbols in a formula for a charged particle moving in a magnetic field
P6_06	Completes a diagram to indicate the direction of current induced in a coil that is moving towards a stationary current-carrying coil
P7_05B	Identifies the path of a negatively charged particle as it passes between two charged plates

P7_07	Identifies the best explanation for why a fluorescent tube lights up when it is positioned close to a charged balloon
P8_04	Recognizes the changes in magnitude of the magnetic flux through a conducting coil as a magnet enters, moves inside, and leaves the coil
P8_06	Evaluates descriptions of the resistance in an unknown electrical component based on its current-voltage graph
P9_06	Evaluates descriptions of processes by which a flashlight containing a coil of wire and a magnet that can slide through the coil produces light
Wave Phenomena & Atomic/Nuclear Physics	
P1_01	Recognizes a correct statement about black lines in the continuous spectrum of sunlight
P1_07	Evaluates reasons for the difference between the input and output energies associated with the photoelectric effect
P2_08	Calculates the wavelength of a sound wave above water
P2_09A	Evaluates descriptions of the result of increasing the temperature of a black body on the radiation it emits
P3_10	Completes a table to indicate the number of protons and neutrons in given isotopes
P4_10	States what happens to the wavelengths of water waves that decrease in speed as they approach the shore
P5_01	Recognizes a range of wavelengths associated with visible light
P6_09	Evaluates experimental set-ups to compare the effect of changing the apex angle of a prism on the angle between incident and refracted rays of light and chooses the best pair
P6_10	Calculates the wavelength of a musical note
P6_11	Recognizes the best explanation for why electromagnetic radiation is characterized by photon energy, radiation frequency, and radiation wavelength
P7_08A	Calculates the speed of a wave moving down an oscillating rope
P8_08	Determines the wavelength of a wave presented as a graphical trace
P8_10	Determines the atomic numbers and mass numbers of 2 isotopes involved in nuclear reactions (1 of 2 points)
P9_08	Evaluates possible factors that account for the differences in interference patterns produced by two different subatomic particles with equal kinetic energies

Items at High International Benchmark (550)

Mechanics & Thermodynamics	
P1_04	Derives an expression for the speed at the top of the trajectory of an object moving in a vertical circular path
P2_01A	Recognizes the acceleration at the apex of an object thrown vertically upward

P2_04	Shows the steps in a calculation of the amount of energy required to increase the temperature of water in a given context
P3_03	Recognizes the best explanation for a ball rebounding to a height that is less than the initial height of release
P3_07	Calculates the new volume of an ideal gas when pressure and temperature change (1 of 2 points)
P4_01	Identifies the ratio of the maximum temperature to the minimum temperature of a sample of a gas during a closed cycle represented in a volume-pressure graph
P4_03	Shows the steps in a calculation of the final velocity of two skiers after they collide inelastically (2 of 2 points)
P4_04	Identifies the direction two balls will travel after they collide inelastically
P5_07	Calculates the energy released when a container of water cools
P6_02	Shows the steps in a calculation of the initial height from which a body began moving vertically down (2 of 2 points)
P6_04	Recognizes the equation describing the force of friction acting on an object sliding down an inclined plane
P6_05	Recognizes the information required to calculate the speed of a satellite in orbit around the Earth
P7_01	Identifies the best estimate for the coefficient of friction between an object and the surface along which it is being dragged
P7_02B	Identifies a reason that the height of a spacecraft above the surface of a planet varies during its orbit
P8_01A	Recognizes the diagram that best represents the direction of the net force acting on a ball attached to a string and moving in a circle at constant speed
P8_02	Describes one step in a sequence for checking the calibration of a thermometer, given a list of available equipment (1 of 2 points)
P8_03A	Calculates the magnitude of the normal force on a body sliding on the inside of a smooth cylindrical surface at a specified angle
P9_03	Identifies the best estimate for the acceleration of an elevator
P9_04	Calculates the temperature of a gas after compression (1 of 2 points)
Electricity & Magnetism	
P2_06	Evaluates explanations for the increase in temperature of an iron plate positioned near a coil of wire connected to an alternating voltage source
P3_05	Ranks the equivalent resistance for four different combinations of resistors
P5_03	Calculates the resistance in a flashlight bulb using Ohm's Law and Joule's law
P5_04	Recognizes paths of a neutral particle and a positively charged particle in a magnetic field shown in a diagram
P6_08	Completes a diagram to indicate the direction of net force on a point charge influenced by two other point charges
P9_05A	Calculates the magnitude of the magnetic field acting on a proton (1 of 2 points)

Wave Phenomena & Atomic/Nuclear Physics	
P2_07	Recognizes the source of energy used to generate electricity in a nuclear power plant
P4_08B	Compares two types of electromagnetic radiation and explains which type is more harmful to humans in terms of the frequency and energy of the photons, given a diagram of the electromagnetic spectrum
P4_09	Recognizes what accounts for the difference in the mass of an atom before and after a nuclear reaction
P4_11	Explains which semiconductor is appropriate to use in a solar panel, given a graph of the performance of each semiconductor across a range of wavelengths of light
P5_11	Identifies an estimate of the age of an organic specimen, given the concentration of carbon-14 in it
P7_08B	Identifies the relationship between the initial and final frequencies and wavelengths of a wave with a final speed less than its initial speed
P8_07	Identifies the index of refraction of a piece of glass, given a diagram showing the glass, the angle of incidence, and angle of refraction
P8_09	Orders examples of electromagnetic radiation in terms of increasing photon energies

Items at Advanced International Benchmark (625)

Mechanics & Thermodynamics	
P2_02	Evaluates a mechanical system run by an electric motor and predicts the difference between the theoretical and actual final temperatures of the system
P5_05	Deduces the tension in the string connecting two unequal masses in freefall
P5_08	Identifies the temperature at which two rods of different metals will have the same length
P7_02A	Calculates an estimate of the mass of a planet given information about the speed of a spacecraft in orbit around it and the radius of the orbit (1 of 2 points)
P8_02	Describes a sequence of steps for checking the calibration of a thermometer, given a list of available equipment (2 of 2 points)
P8_03B	Calculates the speed of a body at the lowest point of its trajectory (1 of 2 points)
P9_04	Shows the steps in a calculation for the temperature of a gas after compression (2 of 2 points)
Electricity & Magnetism	
P1_08	Analyzes a complex circuit diagram to determine the power consumption of light bulbs
P4_06	Identifies the distance at which an electric field is four times less than it is at a given distance from the source
P4_07	Identifies the prediction for the change in the path of a horizontal electron beam as a result of the presence of a magnetic field
P5_02	Interprets a current vs. resistance graph to identify the internal resistance of a battery

P7_05A	Calculates the electric force on a negatively charged particle when it is in between two charged plates (1 of 2 points)
P7_05A	Shows the steps in a calculation of the electric force on a negatively charged particle when it is in between two charged plates (2 of 2 points)
P8_05	Identifies the diagram of an electromagnet that depicts the direction of the current and the polarity, given the orientation of the battery
P9_05A	Shows the steps in a calculation of the magnitude of the magnetic field acting on a proton (2 of 2 points)
P9_05B	Identifies the direction of the magnetic field acting on a proton
Wave Phenomena & Atomic/Nuclear Physics	
P3_11	Completes the equation for a nuclear fission reaction
P4_08A	Recognizes the type of radiation associated with a given range of wavelengths in the electromagnetic spectrum
P5_09	States that a red object absorbs light of all wavelengths from a green light source (1 of 2 points)
P5_10	Justifies an argument that it might be more appropriate to indicate that an object is hot by associating it with the color blue
P6_12	Explains which of two heated bars is hotter in terms of the relationship between the color of emitted light and temperature
P7_04	Recognizes a correct statement about an oscillating electric field in a transmitting antenna generating a magnetic field
P8_10	Determines the atomic numbers and mass numbers of 3 isotopes involved in nuclear reactions (2 of 2 points)
P9_07	Recognizes whether the frequency and wavelength of light change as the light passes from air to water
P9_09A	Explains which pair of atomic reactants can most likely be used in a fusion reaction to produce usable energy for humans, given a temperature-reaction rate graph
P9_09B	Recognizes the information needed to calculate the energy production of three pairs of atomic reactants in a fusion reaction
P9_09C	Calculates the mass lost in a fusion reaction (1 of 2 points)

Items Above the Advanced International Benchmark (625)

Mechanics & Thermodynamics	
P1_02	Identifies the force recorded on one of two spring balances, given the force recorded on the other spring balance and the relationship between the two spring constants
P2_03	Recognizes the path of motion of the center of mass of a curved bar as it flies through the air
P3_07	Shows the steps in a calculation of the new volume of an ideal gas when pressure and temperature change (2 of 2 points)
P3_09	Interprets the context of a solar cooker and states that food in the cooker comes to equilibrium with its surroundings (1 of 2 points)

P3_09	Interprets the context of a solar cooker to explain why, when the cooker is placed in the sun, food heats steadily and then stays at a constant temperature (2 of 2 points)
P6_03	Identifies the relationship between two forces exerted on an object at the apex of its curved motion
P7_02A	Shows the steps in a calculation for an estimate of the mass of a planet given information about the speed of a spacecraft in orbit around it and the radius of the orbit (2 of 2 points)
P8_03B	Shows the steps in a calculation for the speed of a body at the lowest point of its trajectory (2 of 2 points)
Electricity & Magnetism	
P1_06	Identifies the direction of the force on a current-carrying conductor in a given magnetic field
P4_05	Explains how a charged balloon sticks to a wall
P6_07	Predicts the direction of movement of a foil strip as a permanent magnet approaches it (1 of 2 points)
P6_07	Predicts the direction of movement of a foil strip as a permanent magnet approaches it and explains the prediction (2 of 2 points)
P7_06A	Identifies an explanation for the change in polarity of the induced emf as a magnet passes through a coil of conducting wire
P7_06B	Explains the difference in magnitude of the induced emf at its extrema as a magnet passes through a coil of conducting wire
Wave Phenomena & Atomic/Nuclear Physics	
P1_03	Identifies the component of Rutherford's experimental set-up that should be varied to obtain the appropriate data
P1_09	Identifies the effect of a nuclear reaction on the atomic and mass numbers of an atom
P2_09B	Evaluates the conclusion that Wien's Law holds for an object, based on three observations of temperature and wavelength of emitted radiation
P5_09	States that a red object absorbs light of all wavelengths from a green light source and explains the observation (2 of 2 points)
P9_09C	Shows the steps in a calculation of the mass lost in a fusion reaction (2 of 2 points)

CHAPTER 15

Creating and Interpreting the TIMSS Advanced 2015 Context Questionnaire Scales

Michael O. Martin
Ina V.S. Mullis
Martin Hooper
Liqun Yin
Pierre Foy
Lauren Palazzo

Overview

As described in Chapter 2: Developing the TIMSS Advanced 2015 Context Questionnaires, many of the TIMSS Advanced 2015 context questionnaire items were developed to be combined into scales measuring a single underlying latent construct. For reporting, the scales were constructed using item response theory (IRT) scaling methods, specifically the Rasch partial credit model (Masters and Wright, 1997). As a parallel to the TIMSS Advanced International Benchmarks of achievement, each context scale allowed students to be classified into regions corresponding to high, middle, and low values on the construct. To facilitate interpretation of the regions, the cutpoints delimiting the regions were defined in terms of combinations of response categories.

This chapter describes the procedures for constructing, interpreting, and validating scales based on responses to student, teacher, and school questionnaires.

Reporting TIMSS Advanced 2015 Context Questionnaire Scales

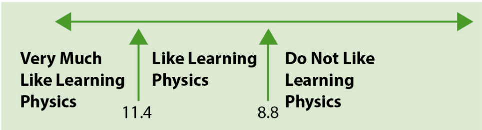
As an example illustrating the TIMSS Advanced approach to reporting context questionnaire data, Exhibit 15.1 presents the TIMSS Advanced 2015 *Students Like Learning Physics* scale. As the name suggests, this scale seeks to measure students' feelings towards learning physics. For each of the twelve statements, students were asked to indicate the degree of their agreement with the statement: agree a lot, agree a little, disagree a little, or disagree a lot. Using IRT partial credit scaling, the data from student responses were placed on a scale constructed so that the scale centerpoint of 10 was located at the mean score across all TIMSS Advanced countries. The units of the scale were chosen so that 2 scale score points corresponded to the standard deviation across all countries. Statements expressing negative sentiment were reverse coded during the scaling (statements 3, 6, and 11). Students who Very Much Like Learning Physics had a scale score greater than or equal to the point on the scale, 11.4 in this case, corresponding to agreeing a lot, on average, with six of the twelve statements and agreeing a little with six of the statements. Students who Do Not Like Learning Physics had a score no higher than the point (8.8) on the scale corresponding to disagreeing a little with six of the statements, on average, and agreeing a little with six of them.

Exhibit 15.1: Items in the TIMSS Advanced 2015 *Students Like Learning Physics* Scale

How much do you agree with these statements about the physics you are studying?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
PSBP20A	1 I enjoy conducting experiments or investigations in physics	○	○	○	○
PSBP20B	2) I get a sense of satisfaction when I solve physics problems	○	○	○	○
PSBP20C*	3) I feel bored when I do my physics schoolwork*	○	○	○	○
PSBP20D	4) I like studying for my physics class outside of school	○	○	○	○
PSBP20E	5) It is interesting to learn physics laws and principles	○	○	○	○
PSBP20F*	6) I dread my physics class*	○	○	○	○
PSBP20G	7) I am studying physics because I like to learn new things	○	○	○	○
PSBP20H	8) I enjoy figuring out challenging physics	○	○	○	○
PSBP20 I	9) Physics is one of my favorite subjects	○	○	○	○
PSBP20J	10) Jobs that require physics skills seem interesting to me	○	○	○	○
PSBP20K*	11) I wish I did not have to study physics*	○	○	○	○
PSBP20L	12) I enjoy thinking about the world in terms of laws of physics	○	○	○	○

* Reverse coded



Scaling Procedure

Partial credit IRT scaling is based on a statistical model that relates the probability that a person will choose a particular response to an item to that person's location on the underlying construct. In the TIMSS Advanced 2015 *Students Like Learning Physics* scale, the underlying construct is students' attitudes towards learning physics, and students who agree in general with the twelve statements are assumed to have more interest in learning physics while students who disagree with the statements are assumed to have less interest.

The partial credit model is shown below:

$$P_{x_i}(\theta_n) = \frac{e^{\sum_{j=0}^{x_i}(\theta_n - \delta_i + \tau_{ij})}}{\sum_{h=0}^{m_i} e^{\sum_{j=0}^{x_i}(\theta_n - \delta_i + \tau_{ij})}} \quad x_i = 0, 1, \ldots, m_i$$

where $P_{x_i}(\theta_n)$ denotes the probability that person n with location θ_n on the latent construct would choose response level x_i to item i out of the m_i possible response levels for the item. The item parameter δ_i gives the location of the item on the latent construct and τ_{ij} denotes step parameters for the response levels. For each scale, the scaling procedure involves first estimating the δ_i and τ_{ij} item parameters, and then using the model with these parameters to estimate θ_n, the score on the latent construct, for each on the n respondents. Depending on the scale, respondents may be students, teachers, or school principals.

The TIMSS Advanced 2015 context questionnaire scaling was conducted using the ConQuest 2.0 software (Wu, Adams, Wilson, & Haldane, 2007).

In preparation for the context questionnaire scaling effort, the TIMSS & PIRLS International Study Center developed a system of production programs that could effectively calibrate the items on each scale using ConQuest and produce scale scores for each scale respondent. Each TIMSS Advanced assessment population (advanced mathematics, physics) consisted of approximately 30,000 students, as well as their teachers and school principals. The estimation of the item parameters, a procedure also known as item calibration, was conducted on the combined data from all countries, with each country contributing equally to the calibration. This was achieved by assigning weights that sum to 500 for each country's student data. Exhibit 15.2 shows the international item parameters for the *Students Like Learning Physics* scale. For each item, the delta parameter δ_i shows the estimated overall location of the item on the scale, and the tau parameters τ_{ij} show the location of the steps, expressed as deviations from delta. Also, included in the right column is the Rasch infit item statistic, which is a measure of how well the data matches the model, with values above 1.3 indicating unexpected response patterns. As can be seen in this exhibit, the data seemed to match the model well for the twelve items of the *Like Learning Physics* scale.

Exhibit 15.2: Item Parameters for the TIMSS Advanced 2015 *Students Like Learning Physics* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
PSBP20A	-0.51423	-1.07025	-0.41990	1.49015	1.15
PSBP20B	-0.90801	-0.86799	-0.43946	1.30745	1.10
PSBP20C*	0.51315	-1.69256	-0.14245	1.83501	1.23
PSBP20D	1.03682	-1.72677	-0.08320	1.80997	1.11
PSBP20E	-0.21380	-1.38699	-0.26425	1.65124	0.90
PSBP20F*	-0.39443	-1.02035	0.04349	0.97686	1.82
PSBP20G	-0.18466	-1.35598	-0.31080	1.66678	0.81
PSBP20H	0.04092	-1.36190	-0.07935	1.44125	0.85
PSBP20I	0.54534	-0.86829	-0.06458	0.93287	0.75
PSBP20J	0.17389	-0.97760	-0.33432	1.31192	0.94
PSBP20K*	-0.07291	-0.60736	-0.22268	0.83004	1.02
PSBP20L	-0.02208	-1.18389	-0.23501	1.41890	1.02

* Reverse Coded

Once the calibration was completed and international item parameters were estimated, individual scores for each respondent (students, teachers, or principals) were generated using weighted maximum likelihood estimation (Warm, 1989). All cases with valid responses to at least two items on a scale were included in the calibration and scoring processes.

The scale scores produced by the weighted likelihood estimation are in the logit metric with measured values ranging from approximately -5 to +5. To convert to a more convenient reporting metric, a linear transformation was applied to the international distribution of logit scores for each scale, so that the resulting distribution across all countries had a mean of 10 and a standard deviation of 2. Exhibit 15.3 presents the scale transformation constants applied to the international distribution of logit scores for the *Students Like Learning Physics* scale to transform them to the (10, 2) reporting metric.

Exhibit 15.3: Scale Transformation Constants for the TIMSS Advanced 2015 *Students Like Learning Physics* Scale

Scale Transformation Constants	
A = 8.81822	Transformed Scale Score = 8.81822 + 1.464843 • Logit Scale Score
B = 1.464843	

To provide an approach to reporting the context questionnaire scales analogous to the TIMSS Advanced International Benchmarks for the TIMSS Advanced achievement scales, a method was developed to divide each scale into high, middle, and low regions and provide a content-referenced interpretation for these regions. For the TIMSS Advanced achievement scales, the Intermediate, High, and Advanced International Benchmarks are specific reference points on the scale that can be used to monitor progress in student achievement. Using a scale anchoring procedure, student performance at each Benchmark is described in terms of the advanced mathematics and physics (depending on the subject) that students reaching that Benchmark know and can do. The percentage of students reaching each of these International Benchmarks can serve as a profile of student achievement in a country.

For the high, middle, and low regions of the context questionnaire scales, the interpretation is content-referenced to the extent that the boundaries of the regions were defined in terms of identifiable combinations of response categories. The particular response combinations that defined the regions boundaries, or cutpoints, were based on a judgment of what constituted a high or low region on each individual scale. For example, based on a consideration of the questions making up the *Students Like Learning Physics* scale, it was determined that in order to be in the high region of the scale and labeled "Very Much Like Learning Physics," a student would have to agree a lot, on average, to at least six of the twelve statements and agree a little to the other six. Similarly, it was determined that a student who, on average, at most agreed a little with six of the statements and disagreed a little with the other six would be labeled "Do Not Like Learning Physics."

The scale region cutpoints were quantified by assigning a numeric value to each response category, such that each respondent's responses to the scale's questions could be expressed as a "raw score." Assigning 0 to "Disagree a lot," 1 to "Disagree a little," 2 to "Agree a little," and 3 to "Agree a lot," results in raw scores on the *Students Like Learning Physics* scale ranging from 0 (disagree a lot with all twelve statements) to 36 (agree a lot to all twelve). A student who agreed a lot with six of the statements and agreed a little with the other six would have a raw score of 30 (6×3 + 6×2). Following this approach, a student with a raw score of 30 or more would be in the "Very Much Like Learning Physics" region of the scale. Similarly, agreeing a little with six statements and disagreeing a little with six statements would result in a raw score of 18 (6×2 + 6×1), so that a student with a raw score less than or equal to 18 would be in the "Do Not Like Learning Physics" region.

A property of a Rasch scale is that each raw score has a unique scale score associated with it. Exhibit 15.4 presents a raw score-scale score equivalence table for the *Students Like Learning Physics* scale. From this table, it can be seen that a raw score of 18 corresponds to a scale score of 8.8 (rounding up) and a raw score of 30 corresponds to a scale score of 11.4 (rounding down).[1] These scale scores were the cutpoints used to divide the scale into the three regions.

1 The reason for rounding was to facilitate reporting, and it was decided that the highest cutpoint would be rounded down to ensure that those with an unrounded scale score (e.g., 11.42601 for the *Like Learning Physics* scale) at the cutpoint were included within the highest region. For a similar reason, the lower cutpoint was rounded up.

Exhibit 15.4: Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 *Students Like Learning Physics* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	2.29834	
1	3.92594	
2	4.70005	
3	5.22580	
4	5.63169	
5	5.97016	
6	6.26389	
7	6.52678	
8	6.76750	
9	6.99121	
10	7.20367	
11	7.40630	
12	7.60136	
13	7.79069	
14	7.97591	
15	8.15835	
16	8.33926	
17	8.51979	
18	8.70102	8.8
19	8.88404	
20	9.06991	
21	9.25980	
22	9.45491	
23	9.65655	
24	9.86615	
25	10.08468	
26	10.31566	
27	10.56116	
28	10.82464	
29	11.11077	
30	11.42601	11.4
31	11.77922	
32	12.18651	
33	12.67364	
34	13.29295	
35	14.17455	
36	15.92718	

Validating the TIMSS Advanced 2015 Context Questionnaire Scales

As evidence that the context questionnaire scales provide comparable measurement across countries, reliability coefficients were computed for each scale for every country and benchmarking participant, and a principal components analysis of the scale items was conducted. Exhibit 15.5 presents the results of this analysis for the *Students Like Learning Physics* scale. The Cronbach's Alpha reliability coefficients generally were at an acceptable level, with all above 0.8 and many above 0.9. The exhibit also shows the percentage of variance among the scale items accounted for by the first principal component in each country. In most cases this was acceptably high, indicating that the items could be adequately represented by a single scale. The component loadings of each questionnaire item from the principal components analysis are positive and substantial, indicating a strong correlation between each item and the scale in every country.

Exhibit 15.5: Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the TIMSS Advanced 2015 *Students Like Learning Physics* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item											
			PSBP20A	PSBP20B	PSBP20C*	PSBP20D	PSBP20E	PSBP20F*	PSBP20G	PSBP20H	PSBP20I	PSBP20J	PSBP20K*	PSBP20L
France	0.89	47	0.59	0.50	0.63	0.70	0.78	0.30	0.78	0.75	0.83	0.69	0.80	0.66
Italy	0.90	49	0.62	0.54	0.60	0.78	0.78	0.33	0.81	0.79	0.82	0.77	0.75	0.64
Lebanon	0.83	39	0.67	0.59	0.34	0.48	0.70	0.16	0.75	0.72	0.81	0.73	0.51	0.67
Norway	0.89	47	0.44	0.35	0.69	0.67	0.78	0.44	0.78	0.79	0.85	0.74	0.77	0.68
Portugal	0.89	47	0.58	0.67	0.61	0.70	0.76	0.40	0.78	0.78	0.82	0.68	0.73	0.66
Russian Federation	0.92	53	0.53	0.78	0.63	0.76	0.82	0.38	0.79	0.81	0.85	0.82	0.67	0.70
Slovenia	0.87	41	0.50	0.59	0.48	0.57	0.70	0.49	0.72	0.77	0.80	0.73	0.59	0.64
Sweden	0.91	51	0.52	0.61	0.69	0.72	0.79	0.51	0.79	0.81	0.84	0.74	0.78	0.69
United States	0.92	52	0.63	0.61	0.55	0.66	0.82	0.63	0.78	0.80	0.85	0.75	0.75	0.78

* Reverse Coded

As indicators of effective environments for learning, a positive relationship with achievement is an important aspect of validity for the TIMSS Advanced context questionnaire scales. For the *Students Like Learning Physics* scale, Exhibit 15.6 presents the Pearson correlation with physics achievement in TIMSS Advanced 2015 for each country, together with *r*-squared—the proportion of variance in achievement attributable to the *Students Like Learning Physics* scale. These figures show a moderate relationship with achievement across participating countries. Also shown is the proportion of variance in achievement attributable to differences between the regions of the *Students Like Learning Physics* scale. This is very similar to the proportion of variance explained by the scale as a whole, indicating that dividing the scale into regions has little effect on its power to account for achievement differences.

Exhibit 15.6: Relationship Between the TIMSS Advanced 2015 *Students Like Learning Physics* Scale and TIMSS Advanced 2015 Physics Achievement

Country	Pearson's Correlation with Physics Achievement		Variance in Physics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.46	0.22	0.18
Italy	0.39	0.15	0.13
Lebanon	0.22	0.05	0.04
Norway	0.50	0.25	0.22
Portugal	0.43	0.18	0.17
Russian Federation	0.38	0.15	0.13
Slovenia	0.44	0.20	0.15
Sweden	0.44	0.20	0.16
United States	0.41	0.17	0.16
International Median	**0.43**	**0.18**	**0.16**

Item parameter estimates and item and scale statistics similar to those above are available in Appendix 15A or each of the TIMSS Advanced 2015 advanced mathematics context questionnaire scales and in Appendix 15B for each of the physics context questionnaire scales.

References

Masters, G.N., & Wright, B.D. (1997). The partial credit model. In M.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory*. Berlin: Springer.

Warm, T.A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427–450.

Wu, M.L., Adams, R.J, Wilson, M.R., & Haldane, S. (2007). Conquest 2.0 [computer software]. Camberwell, Australia: Australian Council for Educational Research.

Appendix 15A: TIMSS Advanced 2015 Context Questionnaire Scales, Advanced Mathematics

Home Educational Resources Scale, Advanced Mathematics

The Home Educational Resources (HER) scale was created based on students' responses concerning the availability of four resources described below.

Items in the TIMSS Advanced 2015 Home Educational Resources Scale, Advanced Mathematics

MSBG04

Number of books in the home:

1) 0-10
2) 11-25
3) 26-100
4) 101-200
5) More than 200

MSDG06S[1]

Number of home study supports:

1) None
2) Study desk/table or own room
3) Both

MSDGEDUP[1]

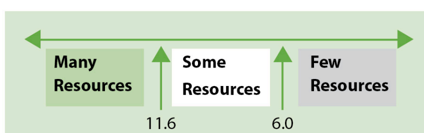
Highest level of education of either parent:

1) Finished some primary or lower secondary or did not go to school
2) Finished lower secondary
3) Finished upper secondary
4) Finished post-secondary education
5) Finished university or higher

MSDGOCCP[1]

Highest level of occupation of either parent:

1) Has never worked outside home for pay, general laborer, or semi-professional (skilled agricultural or fishery worker, craft or trade worker, plant or machine operator)
2) Clerical (clerk or service or sales worker)
3) Small business owner
4) Professional (corporate manager or senior official, professional, or technician or associate professional)



1 Derived variable. For more details, see Supplement 3 of the TIMSS Advanced 2015 User Guide for the International Database.

Item Parameters for the TIMSS Advanced 2015 Home Educational Resources Scale, Advanced Mathematics

Item	delta	tau_1	tau_2	tau_3	tau_4	Infit
MSBG04	0.63886	-0.93281	-0.47613	0.80516	0.60378	1.10
MSDG06S	-0.97302	-0.28994	0.28994			1.27
MSDGEDUP	-0.07938	-0.66004	-0.41662	0.85158	0.22508	0.96
MSDGOCCP	0.41354	-0.67697	1.24010	-0.56313		0.97

*Reverse coded

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 Home Educational Resources Scale, Advanced Mathematics

Scale Transformation Constants	
A = 7.705317	Transformed Scale Score = 7.705317 + 2.192074 • Logit Scale Score
B = 2.192074	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Home Educational Resources Scale, Advanced Mathematics

Raw Score	Transformed Scale Score	Cutpoint
0	1.48945	
1	3.86046	
2	5.08504	
3	5.98800	6.0
4	6.72756	
5	7.36772	
6	7.93623	
7	8.46945	
8	8.98566	
9	9.51161	
10	10.07874	
11	10.73945	
12	11.62325	11.6
13	13.29245	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Home Educational Resources Scale, Advanced Mathematics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item			
			MSBG04	MSDG06S	MSDGEDUP	MSDGOCCP
France	0.59	46	0.66	0.24	0.83	0.81
Italy	0.62	48	0.69	0.27	0.84	0.82
Lebanon	0.53	42	0.62	0.38	0.77	0.74
Norway	0.54	44	0.68	0.33	0.79	0.75
Portugal	0.69	52	0.77	0.18	0.87	0.84
Russian Federation	0.41	38	0.62	0.07	0.76	0.75
Russian Federation 6hr+	0.40	38	0.62	0.25	0.75	0.72
Slovenia	0.53	42	0.67	0.17	0.80	0.76
Sweden	0.61	48	0.72	0.47	0.77	0.76
United States	0.53	43	0.66	0.57	0.79	0.58
International Avg.	0.56	45	0.68	0.30	0.80	0.76

Relationship Between the TIMSS Advanced 2015 Home Educational Resources Scale, Advanced Mathematics, and TIMSS Advanced 2015 Advanced Mathematics Achievement

Country	Pearson's Correlation with Advanced Mathematics Achievement		Variance in Advanced Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.32	0.10	0.07
Italy	0.30	0.09	0.06
Lebanon	0.22	0.05	0.03
Norway	0.28	0.08	0.05
Portugal	0.24	0.06	0.05
Russian Federation	0.18	0.03	0.02
Russian Federation 6hr+	0.23	0.05	0.03
Slovenia	0.21	0.04	0.04
Sweden	0.34	0.11	0.08
United States	0.25	0.06	0.04
International Median	0.25	0.06	0.05

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Safe and Orderly School–Teachers' Reports Scale, Advanced Mathematics

The Safe and Orderly School–Teachers' Reports (SOS) scale was created based on teachers' degree of agreement with the eight statements described below.

Items in the TIMSS Advanced 2015 Safe and Orderly School–Teachers' Reports Scale, Advanced Mathematics[1]

Thinking about your current school, indicate the extent to which you agree or disagree with each of the following statements.

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
MTBG07A	1) This school is located in a safe neighborhood	○	○	○	○
MTBG07B	2) I feel safe at this school	○	○	○	○
MTBG07C	3) This school's security policies and practices are sufficient	○	○	○	○
MTBG07D	4) The students behave in an orderly manner	○	○	○	○
MTBG07E	5) The students are respectful of the teachers	○	○	○	○
MTBG07F	6) The students respect school property	○	○	○	○
MTBG07G	7) This school has clear rules about student conduct	○	○	○	○
MTBG07H	8) This school's rules are enforced in a fair and consistent manner	○	○	○	○

Very Safe and Orderly ← 9.9 → **Safe and Orderly** ← 6.5 → **Less than Safe and Orderly**

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Disagree a little" and "Disagree a lot" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

Item Parameters for the TIMSS Advanced 2015 Safe and Orderly Schools - Teachers' Reports Scale, Advanced Mathematics

Item	delta	tau_1	tau_2	Infit
MTBG07A	-0.82109	-1.51458	1.51458	1.15
MTBG07B	-1.74585	-1.72523	1.72523	0.92
MTBG07C	-0.33658	-1.56042	1.56042	1.04
MTBG07D	0.39769	-1.95319	1.95319	0.83
MTBG07E	0.19117	-2.14541	2.14541	0.85
MTBG07F	1.15478	-2.08336	2.08336	0.94
MTBG07G	0.17538	-1.48879	1.48879	1.10
MTBG07H	0.98450	-1.53824	1.53824	1.05

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 Safe and Orderly Schools - Teachers' Reports Scale, Advanced Mathematics

Scale Transformation Constants	
A = 8.173896	Transformed Scale Score = 8.173896 + 0.95666 • Logit Scale Score
B = 0.95666	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Safe and Orderly Schools - Teachers' Reports Scale, Advanced Mathematics

Raw Score	Transformed Scale Score	Cutpoint
0	3.37625	
1	4.67674	
2	5.40155	
3	5.94826	
4	6.41573	6.5
5	6.84650	
6	7.26412	
7	7.68378	
8	8.11417	
9	8.55968	
10	9.01824	
11	9.48591	
12	9.96410	9.9
13	10.46579	
14	11.02538	
15	11.73354	
16	12.98778	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Safe and Orderly Schools - Teachers' Reports Scale, Advanced Mathematics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			MTBG07A	MTBG07B	MTBG07C	MTBG07D	MTBG07E	MTBG07F	MTBG07G	MTBG07H
France	0.90	59	0.70	0.71	0.78	0.84	0.80	0.71	0.77	0.79
Italy	0.85	49	0.60	0.67	0.69	0.81	0.79	0.75	0.57	0.68
Lebanon	0.84	48	0.50	0.62	0.56	0.71	0.74	0.74	0.80	0.82
Norway	0.84	48	0.50	0.64	0.52	0.74	0.77	0.80	0.76	0.75
Portugal	0.88	54	0.62	0.68	0.81	0.80	0.77	0.75	0.70	0.75
Russian Federation	0.83	47	0.50	0.69	0.74	0.83	0.81	0.68	0.73	0.43
Russian Federation 6hr+	0.78	41	0.70	0.55	0.59	0.76	0.72	0.62	0.70	0.41
Slovenia	0.91	61	0.74	0.75	0.83	0.79	0.82	0.78	0.75	0.80
Sweden	0.80	43	0.50	0.54	0.63	0.83	0.76	0.76	0.51	0.66
United States	0.90	61	0.54	0.70	0.81	0.85	0.85	0.86	0.78	0.83
International Avg.	0.86	52	0.58	0.67	0.71	0.80	0.79	0.76	0.71	0.72

Relationship Between the TIMSS Advanced 2015 Safe and Orderly Schools - Teachers' Reports Scale, and TIMSS Advanced 2015 Advanced Mathematics Achievement

Country	Pearson's Correlation with Advanced Mathematics Achievement		Variance in Advanced Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.16	0.02	0.02
Italy	0.07	0.00	0.03
Lebanon	0.13	0.02	0.00
Norway	0.07	0.01	0.01
Portugal	0.04	0.00	0.00
Russian Federation	0.14	0.02	0.03
Russian Federation 6hr+	0.17	0.03	0.05
Slovenia	0.17	0.03	0.03
Sweden	0.08	0.01	0.01
United States	0.03	0.00	0.02
International Median	0.08	0.01	0.02

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

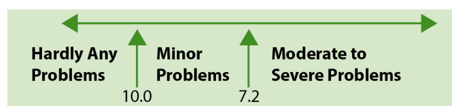
School Discipline Problems–Principals' Reports Scale, Advanced Mathematics

The School Discipline Problems–Principals' Reports (DAS) scale was created based on principals' responses concerning the eleven potential school problems described below.

Items in the TIMSS Advanced 2015 School Discipline Problems–Principals' Reports Scale, Advanced Mathematics

To what degree is each of the following a problem among twelfth grade students in your school?

		Not a problem	Minor problem	Moderate problem	Serious problem
MCBG11A	1) Arriving late at school	◯	◯	◯	◯
MCBG11B	2) Absenteeism (i.e., unjustified absences)	◯	◯	◯	◯
MCBG11C	3) Classroom disturbance	◯	◯	◯	◯
MCBG11D	4) Cheating	◯	◯	◯	◯
MCBG11E	5) Profanity	◯	◯	◯	◯
MCBG11F	6) Vandalism	◯	◯	◯	◯
MCBG11G	7) Theft	◯	◯	◯	◯
MCBG11H	8) Intimidation or verbal abuse among students (including texting, emailing, etc.)	◯	◯	◯	◯
MCBG11 I	9) Physical injury to other students	◯	◯	◯	◯
MCBG11J	10) Intimidation or verbal abuse of teachers or staff (including texting, emailing, etc.)	◯	◯	◯	◯
MCBG11K	11) Physical injury to teachers or staff	◯	◯	◯	◯



Item Parameters for the TIMSS Advanced 2015 School Discipline Problems - Principals' Reports Scale, Advanced Mathematics

Item	delta	tau_1	tau_2	tau_3	Infit
MCBG11A	1.17742	-3.15684	-0.15587	3.31271	1.32
MCBG11B	1.45295	-2.15166	-0.29819	2.44985	1.30
MCBG11C	0.19376	-1.84784	-0.67552	2.52336	0.88
MCBG11D	0.54002	-2.57666	-0.59148	3.16814	1.15
MCBG11E	-0.23367	-1.81279	-0.33749	2.15028	0.87
MCBG11F	-0.49122	0.42442	-1.36426	0.93984	0.69
MCBG11G	-0.42249	0.12894	-1.20999	1.08105	0.69
MCBG11H	0.03725	-0.81699	-1.25629	2.07328	0.78
MCBG11I	-0.61113	0.68322	-1.40198	0.71876	0.61
MCBG11J	-0.63015	0.69434	-1.66531	0.97097	0.74
MCBG11K	-1.01274	1.42144	-0.55467	-0.86677	0.56

Scale Transformation Constants for the TIMSS Advanced 2015 School Discipline Problems - Principals' Reports Scale, Advanced Mathematics

Scale Transformation Constants	
A = 7.71961	Transformed Scale Score = 7.71961 + 0.975134 • Logit Scale Score
B = 0.975134	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 School Discipline Problems - Principals' Reports Scale, Advanced Mathematics

Raw Score	Transformed Scale Score	Cutpoint
0	3.51299	
1	4.63283	
2	5.15247	
3	5.48517	
4	5.72627	
5	5.91593	
6	6.07244	
7	6.20925	
8	6.33188	
9	6.44442	
10	6.55006	
11	6.65149	
12	6.75093	
14	6.95014	
15	7.05580	
16	7.16645	7.2
17	7.28461	
18	7.41311	
19	7.55522	
20	7.71471	
21	7.89534	
22	8.10206	
23	8.33660	
24	8.60078	
25	8.89568	
26	9.22471	
27	9.59087	
28	10.00271	10.0
29	10.47049	
30	11.00615	
31	11.62931	
32	12.40529	
33	13.73050	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 School Discipline Problems - Principals' Reports Scale, Advanced Mathematics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item										
			MCBG11A	MCBG11B	MCBG11C	MCBG11D	MCBG11E	MCBG11F	MCBG11G	MCBG11H	MCBG11I	MCBG11J	MCBG11K
France	0.94	65	0.61	0.65	0.79	0.70	0.88	0.89	0.80	0.88	0.91	0.87	0.85
Italy	0.96	72	0.49	0.77	0.87	0.72	0.74	0.94	0.95	0.91	0.95	0.94	0.92
Lebanon	0.98	84	0.84	0.88	0.90	0.87	0.94	0.95	0.96	0.91	0.94	0.93	0.94
Norway	0.85	46	0.65	0.74	0.67	0.58	0.78	0.72	0.66	0.63	0.74	0.69	0.61
Portugal	0.93	62	0.50	0.75	0.81	0.71	0.75	0.86	0.86	0.86	0.88	0.85	0.72
Russian Federation	0.75	34	0.56	0.67	0.65	0.58	0.73	0.52	0.48	0.56	0.58	0.39	-
Russian Federation 6hr+	0.77	34	0.65	0.66	0.72	0.60	0.70	0.56	0.50	0.54	0.57	0.05	-
Slovenia	0.76	37	0.26	0.27	0.53	0.53	0.66	0.68	0.66	0.58	0.80	0.64	0.83
Sweden	0.78	32	0.52	0.54	0.59	0.41	0.53	0.71	0.68	0.67	0.58	0.62	0.20
United States	0.88	47	0.55	0.56	0.79	0.62	0.78	0.77	0.78	0.74	0.73	0.65	0.47
International Avg.	0.87	53	0.55	0.65	0.73	0.64	0.75	0.78	0.76	0.75	0.79	0.73	0.69

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS Advanced 2015 School Discipline Problems - Principals' Reports Scale, and TIMSS Advanced 2015 Advanced Mathematics Achievement

Country	Pearson's Correlation with Advanced Mathematics Achievement		Variance in Advanced Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.10	0.01	0.00
Italy	0.21	0.04	0.06
Lebanon	0.10	0.01	0.01
Norway	0.18	0.03	0.01
Portugal	0.03	0.00	0.00
Russian Federation	0.15	0.02	0.00
Russian Federation 6hr+	0.02	0.00	0.00
Slovenia	0.25	0.06	0.03
Sweden	0.09	0.01	0.01
United States	0.16	0.03	0.02
International Median	0.15	0.02	0.01

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

School Supports Advanced Mathematics and Physics Education–Principal Version Scale, Advanced Mathematics

The School Supports Advanced Mathematics and Physics Education–Principal Version (SMP) scale was created based on principals' responses characterizing the seven aspects described below.

Items in the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education–Principal Version Scale, Advanced Mathematics[1]

How much do you agree with these statements about advanced mathematics and physics education within your school?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
MCBG10A	1) The school encourages students to study advanced mathematics and physics	○	○	○	○
MCBG10B	2) The school promotes professional development for teachers of advanced mathematics and physics	○	○	○	○
MCBG10C	3) The school provides students with information about career options in advanced mathematics and physics	○	○	○	○
MCBG10D	4) The school has initiatives to promote student interest in advanced mathematics and physics (e.g., student clubs, competitions)	○	○	○	○
MCBG10E	5) The school has partnership initiatives with industry/businesses in advanced mathematics and physics	○	○	○	○
MCBG10F	6) Advanced mathematics and physics teachers are admired by other teachers in the school	○	○	○	○
MCBG10G	7) Students at this school respect students who excel in advanced mathematics and physics	○	○	○	○

Strongly Supportive		Supportive		Less than Supportive
	11.0		6.5	

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Disagree a little" and "Disagree a lot" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

Item Parameters for the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Principal Version Scale, Advanced Mathematics

Item	delta	tau_1	tau_2	Infit
MCBG10A	-0.97501	-1.02632	1.02632	0.90
MCBG10B	-0.49096	-1.21322	1.21322	0.96
MCBG10C	-0.81296	-1.29416	1.29416	0.94
MCBG10D	-0.54344	-1.00501	1.00501	0.95
MCBG10E	1.78310	-0.57696	0.57696	1.06
MCBG10F	1.03554	-0.58860	0.58860	0.94
MCBG10G	0.00373	-0.96389	0.96389	0.97

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Principal Version Scale, Advanced Mathematics

Scale Transformation Constants	
A = 8.749167	Transformed Scale Score = 8.749167 + 1.410095 • Logit Scale Score
B = 1.410095	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Principal Version Scale, Advanced Mathematics

Raw Score	Transformed Scale Score	Cutpoint
0	2.89865	
1	4.69620	
2	5.68613	
3	6.45357	6.5
4	7.12107	
5	7.73084	
6	8.30187	
7	8.84425	
8	9.36982	
9	9.89666	
10	10.43956	
11	11.02451	11.0
12	11.71058	
13	12.61631	
14	14.32251	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Principal Version Scale, Advanced Mathematics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			MCBG10A	MCBG10B	MCBG10C	MCBG10D	MCBG10E	MCBG10F	MCBG10G
France	0.66	34	0.70	0.60	0.62	0.56	0.50	0.60	0.46
Italy	0.73	40	0.81	0.73	0.68	0.60	0.67	0.31	0.46
Lebanon	0.75	41	0.61	0.64	0.73	0.76	0.56	0.60	0.56
Norway	0.65	35	0.34	0.56	0.70	0.49	0.40	0.79	0.73
Portugal	0.73	39	0.55	0.66	0.58	0.69	0.49	0.69	0.66
Russian Federation	0.72	40	0.55	0.58	0.58	0.71	0.55	0.77	0.66
Russian Federation 6hr+	0.69	41	0.71	0.78	0.65	0.65	0.52	0.62	0.53
Slovenia	0.69	36	0.62	0.64	0.63	0.64	0.63	0.51	0.49
Sweden	0.71	37	0.78	0.60	0.62	0.58	0.60	0.52	0.51
United States	0.78	44	0.57	0.63	0.71	0.77	0.60	0.67	0.70
International Avg.	0.71	38	0.62	0.63	0.65	0.64	0.56	0.61	0.58

Relationship Between the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Principal Version Scale, and TIMSS Advanced 2015 Advanced Mathematics Achievement

Country	Pearson's Correlation with Advanced Mathematics Achievement		Variance in Advanced Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.02	0.00	0.00
Italy	0.06	0.00	0.00
Lebanon	0.10	0.01	0.00
Norway	0.17	0.03	0.05
Portugal	0.03	0.00	0.01
Russian Federation	0.16	0.03	0.01
Russian Federation 6hr+	0.21	0.04	0.02
Slovenia	0.29	0.09	0.04
Sweden	0.06	0.00	0.00
United States	0.10	0.01	0.01
International Median	0.10	0.01	0.01

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

School Supports Advanced Mathematics and Physics Education–Teacher Version Scale, Advanced Mathematics

The School Supports Advanced Mathematics and Physics Education–Teacher Version (SMP) scale was created based on teachers' responses characterizing the seven aspects described below.

Items in the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education–Teacher Version Scale, Advanced Mathematics[1]

How much do you agree with these statements about advanced mathematics and physics education within your school?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
MTBG06A	1) The school encourages students to study advanced mathematics and physics	○	○	○	○
MTBG06B	2) The school promotes professional development for teachers of advanced mathematics and physics	○	○	○	○
MTBG06C	3) The school provides students with information about career options in advanced mathematics and physics	○	○	○	○
MTBG06D	4) Advanced mathematics and physics teachers are admired by other teachers in the school	○	○	○	○
MTBG06E	5) Teachers have high expectations for student achievement in advanced mathematics and physics	○	○	○	○
MTBG06F	6) Students at this school respect students who excel in advanced mathematics and physics	○	○	○	○
MTBG06G	7) Parents expect their children to study advanced mathematics and physics	○	○	○	○

Strongly Supportive | 11.6 | Supportive | 7.4 | Less than Supportive

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Disagree a little" and "Disagree a lot" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

Item Parameters for the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Teacher Version Scale, Advanced Mathematics

Item	delta	tau_1	tau_2	Infit
MTBG06A	-0.83359	-1.13874	1.13874	0.90
MTBG06B	0.36769	-0.90518	0.90518	1.09
MTBG06C	0.01503	-1.05692	1.05692	0.98
MTBG06D	0.90364	-0.63126	0.63126	0.93
MTBG06E	-0.28192	-1.24591	1.24591	0.98
MTBG06F	-0.41669	-1.02854	1.02854	1.09
MTBG06G	0.24584	-1.19301	1.19301	1.10

Scale Transformation Constants for the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Teacher Version Scale, Advanced Mathematics

Scale Transformation Constants	
A = 9.498786	Transformed Scale Score = 9.498786 + 1.425402 • Logit Scale Score
B = 1.425402	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Teacher Version Scale, Advanced Mathematics

Raw Score	Transformed Scale Score	Cutpoint
0	3.88749	
1	5.67991	
2	6.64185	
3	7.36800	7.4
4	7.98333	
5	8.54226	
6	9.06318	
7	9.56156	
8	10.04970	
9	10.54426	
10	11.05900	
11	11.62342	11.6
12	12.28865	
13	13.18972	
14	14.91360	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Teacher Version Scale, Advanced Mathematics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			MTBG06A	MTBG06B	MTBG06C	MTBG06D	MTBG06E	MTBG06F	MTBG06G
France	0.58	29	0.47	0.41	0.47	0.52	0.58	0.73	0.52
Italy	0.72	38	0.69	0.59	0.70	0.66	0.62	0.40	0.62
Lebanon	0.72	38	0.74	0.72	0.65	0.52	0.61	0.48	0.56
Norway	0.72	38	0.70	0.52	0.60	0.77	0.67	0.62	0.34
Portugal	0.72	38	0.66	0.71	0.64	0.59	0.62	0.41	0.64
Russian Federation	0.83	50	0.65	0.81	0.74	0.72	0.57	0.76	0.66
Russian Federation 6hr+	0.75	42	0.62	0.76	0.77	0.66	0.59	0.54	0.55
Slovenia	0.66	33	0.64	0.71	0.63	0.56	0.55	0.51	0.34
Sweden	0.66	33	0.63	0.38	0.70	0.48	0.63	0.61	0.53
United States	0.78	44	0.68	0.60	0.74	0.70	0.52	0.70	0.68
International Avg.	0.71	38	0.65	0.60	0.65	0.61	0.60	0.58	0.54

Relationship Between the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Teacher Version Scale, and TIMSS Advanced 2015 Advanced Mathematics Achievement

Country	Pearson's Correlation with Advanced Mathematics Achievement		Variance in Advanced Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.06	0.00	0.00
Italy	0.13	0.02	0.02
Lebanon	0.04	0.00	0.01
Norway	0.08	0.01	0.02
Portugal	0.06	0.00	0.00
Russian Federation	0.22	0.05	0.03
Russian Federation 6hr+	0.15	0.02	0.01
Slovenia	0.24	0.06	0.03
Sweden	0.07	0.00	0.00
United States	0.01	0.00	0.01
International Median	0.07	0.00	0.01

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Students Like Learning Advanced Mathematics Scale

The Students Like Learning Advanced Mathematics (SLM) scale was created based on students' degree of agreement with the twelve statements described below.

Items in the TIMSS Advanced 2015 Students Like Learning Advanced Mathematics Scale

How much do you agree with these statements about the mathematics you are studying?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
MSBM20A	1) When I do mathematics problems, I sometimes get completely absorbed	◯	◯	◯	◯
MSBM20B	2) I get a sense of satisfaction when I solve mathematics problems	◯	◯	◯	◯
MSBM20C*	3) I feel bored when I do my mathematics schoolwork*	◯	◯	◯	◯
MSBM20D	4) I like studying for my mathematics class outside of school	◯	◯	◯	◯
MSBM20E	5) It is interesting to learn mathematics theory	◯	◯	◯	◯
MSBM20F*	6) I dread my mathematics class*	◯	◯	◯	◯
MSBM20G	7) I am studying mathematics because I like to learn new things	◯	◯	◯	◯
MSBM20H	8) I enjoy figuring out challenging mathematics	◯	◯	◯	◯
MSBM20 I	9) Mathematics is one of my favorite subjects	◯	◯	◯	◯
MSBM20J	10) Jobs that require advanced mathematics skills seem interesting to me	◯	◯	◯	◯
MSBM20K*	11) I wish I did not have to study mathematics*	◯	◯	◯	◯
MSBM20L	12) I enjoy thinking about the world in terms of mathematical relationships	◯	◯	◯	◯

* Reverse coded

Very Much Like Learning Advanced Mathematics		Like Learning Advanced Mathematics		Do Not Like Learning Advanced Mathematics
	11.8		9.1	

Item Parameters for the TIMSS Advanced 2015 Students Like Learning Advanced Mathematics Scale

Item	delta	tau_1	tau_2	tau_3	Infit
MSBM20A	-0.59588	-1.40851	-0.22921	1.63772	1.28
MSBM20B	-1.08963	-0.57708	-0.49790	1.07498	1.10
MSBM20C*	0.39039	-1.70328	-0.19453	1.89781	1.30
MSBM20D	0.79348	-1.59115	-0.11154	1.70269	1.13
MSBM20E	0.39789	-1.25999	-0.18804	1.44803	1.03
MSBM20F*	-0.55711	-1.10780	0.07358	1.03422	2.03
MSBM20G	0.00711	-1.53161	-0.11150	1.64311	0.89
MSBM20H	-0.16988	-1.21083	-0.10307	1.31390	0.85
MSBM20I	0.21802	-0.68925	-0.03127	0.72052	0.73
MSBM20J	0.15044	-0.99874	-0.14656	1.14530	0.95
MSBM20K*	-0.28490	-0.61660	-0.08438	0.70098	1.31
MSBM20L	0.74007	-1.30692	0.04757	1.25935	1.11

*Reverse coded

Scale Transformation Constants for the TIMSS Advanced 2015 Students Like Learning Advanced Mathematics Scale

Scale Transformation Constants	
A = 9.128252	Transformed Scale Score = 9.128252 + 1.556567 • Logit Scale Score
B = 1.556567	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Students Like Learning Advanced Mathematics Scale

Raw Score	Transformed Scale Score	Cutpoint
0	2.17529	
1	3.91526	
2	4.74499	
3	5.30942	
4	5.74558	
5	6.11032	
6	6.42755	
7	6.71198	
8	6.97262	
9	7.21536	
10	7.44367	
11	7.66218	
12	7.87192	
13	8.07472	
14	8.27223	
15	8.46589	
16	8.65702	
17	8.84686	
18	9.03665	9.1
19	9.22759	
20	9.42092	
21	9.61793	
22	9.82004	
23	10.02874	
24	10.24570	
25	10.47196	
26	10.71137	
27	10.96623	
28	11.24022	
29	11.53811	
30	11.86658	11.8
31	12.23465	
32	12.65864	
33	13.16494	
34	13.80785	
35	14.72417	
36	16.55930	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Students Like Learning Advanced Mathematics Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item											
			MSBM20A	MSBM20B	MSBM20C*	MSBM20D	MSBM20E	MSBM20F*	MSBM20G	MSBM20H	MSBM20I	MSBM20J	MSBM20K*	MSBM20L
France	0.88	45	0.71	0.46	0.61	0.68	0.76	0.32	0.76	0.76	0.83	0.65	0.76	0.56
Italy	0.85	46	-0.47	0.41	0.64	0.78	0.66	0.39	0.81	0.80	0.85	0.79	0.76	0.63
Lebanon	0.74	30	0.47	0.59	0.33	0.45	0.63	0.08	0.68	0.66	0.72	0.67	0.30	0.62
Norway	0.89	45	0.66	0.42	0.66	0.68	0.75	0.41	0.76	0.79	0.84	0.67	0.69	0.59
Portugal	0.90	47	0.64	0.57	0.64	0.73	0.70	0.48	0.74	0.76	0.83	0.69	0.78	0.61
Russian Federation	0.91	51	0.65	0.75	0.59	0.67	0.74	0.42	0.80	0.84	0.84	0.82	0.62	0.68
Russian Federation 6hr+	0.91	51	0.69	0.76	0.57	0.67	0.73	0.43	0.78	0.85	0.85	0.81	0.60	0.68
Slovenia	0.90	47	0.71	0.44	0.51	0.67	0.52	0.65	0.78	0.82	0.84	0.78	0.76	0.65
Sweden	0.91	49	0.66	0.57	0.69	0.72	0.79	0.43	0.78	0.80	0.85	0.71	0.73	0.60
United States	0.91	50	0.68	0.58	0.58	0.63	0.74	0.66	0.76	0.79	0.82	0.76	0.70	0.71
International Avg.	0.87	46	0.52	0.53	0.58	0.67	0.70	0.43	0.76	0.78	0.82	0.73	0.68	0.63

*Reverse coded

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Relationship Between the TIMSS Advanced 2015 Students Like Learning Advanced Mathematics Scale, and TIMSS Advanced 2015 Advanced Mathematics Achievement

Country	Pearson's Correlation with Advanced Mathematics Achievement		Variance in Advanced Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.51	0.26	0.21
Italy	0.35	0.13	0.10
Lebanon	0.29	0.08	0.06
Norway	0.48	0.23	0.20
Portugal	0.50	0.25	0.21
Russian Federation	0.37	0.13	0.12
Russian Federation 6hr+	0.37	0.14	0.13
Slovenia	0.52	0.27	0.24
Sweden	0.54	0.29	0.24
United States	0.38	0.14	0.12
International Median	0.48	0.23	0.20

Students' Sense of School Belonging Scale, Advanced Mathematics

The Students' Sense of School Belonging (SSB) scale was created based on students' degree of agreement with the nine statements described below.

Items in the TIMSS Advanced 2015 Students' Sense of School Belonging Scale, Advanced Mathematics

What do you think about your school? Tell how much you agree with these statements.

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
MSBM22A	1) I enjoy school	◯	◯	◯	◯
MSBM22B	2) I feel safe when I am at school	◯	◯	◯	◯
MSBM22C	3) I feel like I belong at this school	◯	◯	◯	◯
MSBM22D	4) I like to see my classmates at school	◯	◯	◯	◯
MSBM22E	5) Teachers at my school are fair to me	◯	◯	◯	◯
MSBM22F	6) I am proud to go to this school	◯	◯	◯	◯
MSBM22G	7) I learn a lot in school	◯	◯	◯	◯
MSBM22H	8) My classmates respect students who excel in school subjects	◯	◯	◯	◯
MSBM22 I	9) My classmates respect students who struggle learning school subjects	◯	◯	◯	◯

High Sense of School Belonging | 10.6 | Sense of School Belonging | 7.7 | Little Sense of School Belonging

Item Parameters for the TIMSS Advanced 2015 Students' Sense of School Belonging Scale, Advanced Mathematics

Item	delta	tau_1	tau_2	tau_3	Infit
MSBM22A	0.39975	-1.19908	-0.58547	1.78455	1.12
MSBM22B	-0.15512	-0.80900	-0.70858	1.51758	1.03
MSBM22C	0.28834	-1.11210	-0.31747	1.42957	0.95
MSBM22D	-0.80465	-0.67223	-0.64404	1.31627	1.22
MSBM22E	0.00294	-1.42359	-0.45338	1.87697	1.11
MSBM22F	0.41761	-1.13239	-0.45103	1.58342	0.95
MSBM22G	-0.24169	-1.34822	-0.66654	2.01476	1.01
MSBM22H	-0.10713	-1.24202	-0.52085	1.76287	1.10
MSBM22I	0.19995	-1.32373	-0.51026	1.83399	1.33

Scale Transformation Constants for the TIMSS Advanced 2015 Students' Sense of School Belonging Scale, Advanced Mathematics

Scale Transformation Constants	
A = 7.993523	Transformed Scale Score = 7.993523 + 1.272937 • Logit Scale Score
B = 1.272937	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Students' Sense of School Belonging Scale, Advanced Mathematics

Raw Score	Transformed Scale Score	Cutpoint
0	2.80092	
1	4.17041	
2	4.81527	
3	5.25552	
4	5.59943	
5	5.88979	
6	6.14726	
7	6.38344	
8	6.60565	
9	6.81902	
10	7.02747	
11	7.23422	
12	7.44214	
13	7.65463	7.7
14	7.87272	
15	8.10081	
16	8.34184	
17	8.59861	
18	8.87410	
19	9.17103	
20	9.49135	
21	9.83727	
22	10.21291	
23	10.62626	10.6
24	11.09778	
25	11.66982	
26	12.45741	
27	13.98330	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Students' Sense of School Belonging Scale, Advanced Mathematics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			MSBM22A	MSBM22B	MSBM22C	MSBM22D	MSBM22E	MSBM22F	MSBM22G	MSBM22H	MSBM22I
France	0.81	40	0.72	0.67	0.67	0.58	0.56	0.73	0.63	0.56	0.56
Italy	0.83	43	0.71	0.71	0.76	0.53	0.62	0.79	0.68	0.52	0.53
Lebanon	0.85	45	0.63	0.79	0.76	0.57	0.63	0.73	0.67	0.64	0.59
Norway	0.84	45	0.73	0.70	0.78	0.67	0.58	0.72	0.66	0.60	0.56
Portugal	0.82	42	0.73	0.69	0.73	0.58	0.52	0.77	0.67	0.54	0.56
Russian Federation	0.89	55	0.79	0.76	0.81	0.69	0.66	0.82	0.71	0.69	0.70
Russian Federation 6hr+	0.89	54	0.78	0.74	0.81	0.70	0.65	0.79	0.72	0.68	0.70
Slovenia	0.85	45	0.71	0.68	0.78	0.57	0.61	0.78	0.74	0.58	0.53
Sweden	0.85	48	0.79	0.70	0.80	0.65	0.58	0.75	0.69	0.61	0.60
United States	0.86	48	0.71	0.66	0.78	0.70	0.65	0.79	0.72	0.66	0.56
International Avg.	0.85	46	0.73	0.71	0.76	0.61	0.60	0.76	0.69	0.60	0.58

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Relationship Between the TIMSS Advanced 2015 Students' Sense of School Belonging Scale, and TIMSS Advanced 2015 Advanced Mathematics Achievement

Country	Pearson's Correlation with Advanced Mathematics Achievement		Variance in Advanced Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.20	0.04	0.04
Italy	0.05	0.00	0.01
Lebanon	0.08	0.01	0.01
Norway	0.16	0.03	0.02
Portugal	0.07	0.00	0.01
Russian Federation	0.12	0.02	0.01
Russian Federation 6hr+	0.09	0.01	0.01
Slovenia	0.22	0.05	0.04
Sweden	0.20	0.04	0.04
United States	0.07	0.01	0.01
International Median	0.12	0.02	0.01

Students Value Advanced Mathematics Scale

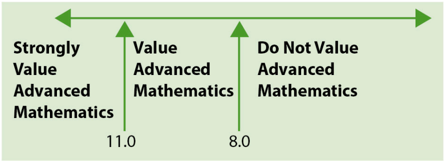
The Students Value Advanced Mathematics (SVM) scale was created based on students' degree of agreement with the nine statements described below.

Items in the TIMSS Advanced 2015 Students Value Advanced Mathematics Scale

How much do you agree with these statements about the mathematics you are studying?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
MSBM21A	1) Learning mathematics will help me get ahead in the world	○	○	○	○
MSBM21B	2) It is important to do well in my mathematics class	○	○	○	○
MSBM21C*	3) The mathematics I am studying is not useful for my future*	○	○	○	○
MSBM21D	4) My parents are pleased that I am taking advanced mathematics	○	○	○	○
MSBM21E	5) Doing well in mathematics will help me get into the university of my choice	○	○	○	○
MSBM21F*	6) Learning advanced mathematics does not seem to be a worthwhile exercise*	○	○	○	○
MSBM21G	7) My parents think that it is important that I do well in my mathematics class	○	○	○	○
MSBM21H	8) I like telling people I am studying advanced mathematics	○	○	○	○
MSBM21 I	9) Learning advanced mathematics will give me more job opportunities	○	○	○	○

*Reverse coded



Item Parameters for the TIMSS Advanced 2015 Students Value Advanced Mathematics Scale

Item	delta	tau_1	tau_2	tau_3	Infit
MSBM21A	0.07051	-1.13081	-0.21324	1.34405	0.94
MSBM21B	-0.33686	-1.13710	-0.21075	1.34785	0.84
MSBM21C*	0.58705	-1.02374	-0.04630	1.07004	1.35
MSBM21D	-0.34160	-0.88413	-0.47165	1.35578	0.93
MSBM21E	-0.16646	-0.40870	-0.14215	0.55085	0.90
MSBM21F*	0.03347	-0.91420	-0.19645	1.11065	1.57
MSBM21G	-0.65475	-0.83420	-0.49183	1.32603	0.93
MSBM21H	0.90365	-1.13939	0.07749	1.06190	1.25
MSBM21I	-0.09501	-0.78067	-0.26783	1.04850	0.81

*Reverse coded

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 Students Value Advanced Mathematics Scale

Scale Transformation Constants	
A = 8.202488	Transformed Scale Score = 8.202488 + 1.658016 • Logit Scale Score
B = 1.658016	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Students Value Advanced Mathematics Scale

Raw Score	Transformed Scale Score	Cutpoint
0	1.74486	
1	3.54182	
2	4.39260	
3	4.97121	
4	5.42447	
5	5.80430	
6	6.13612	
7	6.44041	
8	6.72231	
9	6.98816	
10	7.24296	
11	7.49073	
12	7.73488	
13	7.97837	8.0
14	8.22472	
15	8.47440	
16	8.73231	
17	9.00052	
18	9.28212	
19	9.57930	
20	9.89879	
21	10.24631	
22	10.63156	
23	11.07036	11.0
24	11.58834	
25	12.24488	
26	13.18781	
27	15.09728	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Students Value Advanced Mathematics Scale, Advanced Mathematics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			MSBM21A	MSBM21B	MSBM21C*	MSBM21D	MSBM21E	MSBM21F*	MSBM21G	MSBM21H	MSBM21I
France	0.81	41	0.65	0.62	0.63	0.59	0.68	0.65	0.58	0.58	0.75
Italy	0.85	46	0.79	0.74	0.63	0.61	0.69	0.69	0.53	0.58	0.79
Lebanon	0.67	32	0.41	0.71	0.42	0.67	0.64	0.31	0.63	0.45	0.65
Norway	0.73	35	0.69	0.68	0.39	0.63	0.67	0.65	0.52	0.35	0.62
Portugal	0.83	44	0.76	0.68	0.64	0.64	0.67	0.70	0.55	0.50	0.76
Russian Federation	0.83	45	0.72	0.76	0.36	0.75	0.75	0.51	0.73	0.58	0.78
Russian Federation 6hr+	0.82	45	0.74	0.77	0.30	0.75	0.76	0.50	0.72	0.52	0.80
Slovenia	0.73	34	0.72	0.44	0.46	0.70	0.57	0.08	0.63	0.60	0.77
Sweden	0.76	36	0.73	0.54	0.64	0.60	0.61	0.68	0.44	0.39	0.67
United States	0.81	42	0.74	0.73	0.60	0.63	0.66	0.62	0.58	0.48	0.76
International Avg.	0.78	39	0.69	0.66	0.53	0.65	0.66	0.54	0.58	0.50	0.73

*Reverse coded

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Relationship Between the TIMSS Advanced 2015 Students Value Advanced Mathematics Scale, and TIMSS Advanced 2015 Advanced Mathematics Achievement

Country	Pearson's Correlation with Advanced Mathematics Achievement		Variance in Advanced Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.36	0.13	0.11
Italy	0.24	0.06	0.05
Lebanon	0.25	0.06	0.04
Norway	0.23	0.05	0.04
Portugal	0.36	0.13	0.11
Russian Federation	0.29	0.08	0.07
Russian Federation 6hr+	0.28	0.08	0.08
Slovenia	0.40	0.16	0.14
Sweden	0.24	0.06	0.04
United States	0.24	0.06	0.05
International Median	0.25	0.06	0.05

Students’ Views on Engaging Teaching in Advanced Mathematics Lessons Scale

The Students’ Views on Engaging Teaching in Advanced Mathematics Lessons (EML) scale was created based on students’ degree of agreement with the fourteen statements described below.

Items in the TIMSS Advanced 2015 Students’ Views on Engaging Teaching in Advanced Mathematics Lessons Scale

How much do you agree with these statements about your advanced mathematics lessons?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
MSBP18A	1) The teacher clearly communicates the purpose of each mathematics lesson	○	○	○	○
MSBP18B	2) I know what my teacher expects me to do	○	○	○	○
MSBP18C	3) My teacher is easy to understand	○	○	○	○
MSBP18D	4) I am interested in what my teacher says	○	○	○	○
MSBP18E	5) My teacher gives me interesting things to do	○	○	○	○
MSBP18F	6) My teacher asks me thought provoking questions	○	○	○	○
MSBP18G	7) My teacher has clear answers to my questions	○	○	○	○
MSBP18H	8) My teacher links new content to what I already know	○	○	○	○
MSBP18 I	9) My teacher is good at explaining advanced mathematics	○	○	○	○
MSBP18J	10) My teacher provides the opportunity for me to show what I have learned	○	○	○	○
MSBP18K	11) My teacher encourages me to keep working on advanced mathematics problems until I solve them	○	○	○	○
MSBP18L	12) My teacher provides helpful feedback on my schoolwork (including homework)	○	○	○	○
MSBP18M	13) My teacher uses a variety of teaching methods, tasks, and activities to help us learn	○	○	○	○
MSBP18N	14) My teacher believes that I can learn difficult advanced mathematics material	○	○	○	○

Very Engaging Teaching | 10.4 | Engaging Teaching | 7.9 | Less than Engaging Teaching

Item Parameters for the TIMSS Advanced 2015 Students' Views on Engaging Teaching in Advanced Mathematics Lessons Scale, Advanced Mathematics

Item	delta	tau_1	tau_2	tau_3	Infit
MSBM18A	-0.27752	-1.49392	-0.31185	1.80577	0.88
MSBM18B	-0.52469	-1.55270	-0.46138	2.01408	1.13
MSBM18C	0.02085	-1.52655	-0.20957	1.73612	0.83
MSBM18D	0.15229	-1.49964	-0.38439	1.88403	1.21
MSBM18E	0.83288	-1.83818	-0.20095	2.03913	0.96
MSBM18F	0.32078	-1.44837	-0.23359	1.68196	1.18
MSBM18G	-0.08348	-1.31458	-0.29135	1.60593	0.96
MSBM18H	-0.47204	-1.41178	-0.47480	1.88658	1.03
MSBM18I	-0.40869	-1.16049	-0.28104	1.44153	0.79
MSBM18J	-0.25862	-1.91696	-0.35479	2.27175	1.06
MSBM18K	-0.07877	-1.57597	-0.14124	1.71721	1.07
MSBM18L	0.30429	-1.59816	-0.19349	1.79165	1.12
MSBM18M	0.53568	-1.70713	-0.02321	1.73034	1.14
MSBM18N	-0.06296	-1.38459	-0.35309	1.73768	1.20

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 Students' Views on Engaging Teaching in Advanced Mathematics Lessons Scale, Advanced Mathematics

Scale Transformation Constants	
A = 8.028837	Transformed Scale Score = 8.028837 + 1.168415 • Logit Scale Score
B = 1.168415	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Students' Views on Engaging Teaching in Advanced Mathematics Lessons Scale, Advanced Mathematics

Raw Score	Transformed Scale Score	Cutpoint
0	2.25920	
1	3.58079	
2	4.21671	
3	4.65055	
4	4.98578	
5	5.26436	
6	5.50518	
7	5.71976	
8	5.91528	
9	6.09775	
10	6.26742	
11	6.42988	
12	6.58608	
13	6.73753	
14	6.88552	
15	7.03109	
16	7.17519	
17	7.31924	
18	7.46246	
19	7.60644	
20	7.75182	
21	7.89911	7.9
22	8.04905	
23	8.20155	
24	8.35759	
25	8.51741	
26	8.68135	
27	8.84974	
28	9.02315	
29	9.20196	
30	9.38677	
31	9.57837	
32	9.77786	
33	9.98680	
34	10.20742	
35	10.44289	10.4
36	10.69789	
37	10.97836	
38	11.29738	
39	11.67448	
40	12.15065	
41	12.82864	
42	14.19398	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Students' Views on Engaging Teaching in Advanced Mathematics Lessons Scale, Advanced Mathematics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item													
			MSBM18A	MSBM18B	MSBM18C	MSBM18D	MSBM18E	MSBM18F	MSBM18G	MSBM18H	MSBM18I	MSBM18J	MSBM18K	MSBM18L	MSBM18M	MSBM18N
France	0.91	46	0.69	0.65	0.77	0.64	0.70	0.63	0.75	0.61	0.80	0.63	0.66	0.63	0.64	0.66
Italy	0.92	52	0.81	0.58	0.83	0.51	0.68	0.70	0.79	0.69	0.82	0.76	0.74	0.77	0.74	0.54
Lebanon	0.91	46	0.68	0.57	0.75	0.64	0.70	0.59	0.71	0.68	0.69	0.69	0.70	0.68	0.71	0.65
Norway	0.91	45	0.69	0.68	0.79	0.59	0.70	0.58	0.73	0.70	0.78	0.68	0.59	0.70	0.59	0.58
Portugal	0.93	52	0.79	0.66	0.80	0.57	0.73	0.62	0.78	0.69	0.81	0.70	0.75	0.74	0.73	0.66
Russian Federation	0.92	50	0.77	0.64	0.76	0.76	0.75	0.56	0.75	0.63	0.79	0.64	0.65	0.76	0.75	0.65
Russian Federation 6hr+	0.90	46	0.73	0.60	0.72	0.73	0.73	0.57	0.70	0.59	0.74	0.60	0.66	0.73	0.70	0.62
Slovenia	0.92	51	0.79	0.71	0.80	0.59	0.72	0.69	0.77	0.73	0.83	0.66	0.56	0.70	0.73	0.63
Sweden	0.93	53	0.75	0.65	0.82	0.68	0.77	0.74	0.75	0.75	0.81	0.71	0.71	0.74	0.64	0.60
United States	0.94	57	0.81	0.72	0.82	0.70	0.74	0.71	0.79	0.77	0.85	0.75	0.75	0.74	0.69	0.70
International Avg.	0.92	50	0.75	0.65	0.79	0.63	0.72	0.65	0.76	0.69	0.80	0.69	0.68	0.72	0.69	0.63

Relationship Between the TIMSS Advanced 2015 Students' Views on Engaging Teaching in Advanced Mathematics Lessons Scale, and TIMSS Advanced 2015 Advanced Mathematics Achievement

Country	Pearson's Correlation with Advanced Mathematics Achievement		Variance in Advanced Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.26	0.07	0.05
Italy	0.09	0.01	0.01
Lebanon	0.11	0.01	0.02
Norway	0.26	0.07	0.06
Portugal	0.22	0.05	0.06
Russian Federation	0.29	0.08	0.07
Russian Federation 6hr+	0.19	0.04	0.03
Slovenia	0.31	0.09	0.09
Sweden	0.30	0.09	0.08
United States	0.20	0.04	0.04
International Median	0.26	0.07	0.06

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Teacher Job Satisfaction Scale, Advanced Mathematics

The Teacher Job Satisfaction (TJS) scale was created based on how often teachers responded positively to the seven statements described below.

Items in the TIMSS Advanced 2015 Teacher Job Satisfaction Scale, Advanced Mathematics[1]

How often do you feel the following way about being a teacher?

		Very often	Often	Sometimes	Never or almost never
MTBG10A	1) I am content with my profession as a teacher	◯	◯	◯	◯
MTBG10B	2) I am satisfied with being a teacher at this school	◯	◯	◯	◯
MTBG10C	3) I find my work full of meaning and purpose	◯	◯	◯	◯
MTBG10D	4) I am enthusiastic about my job	◯	◯	◯	◯
MTBG10E	5) My work inspires me	◯	◯	◯	◯
MTBG10F	6) I am proud of the work I do	◯	◯	◯	◯
MTBG10G	7) I am going to continue teaching for as long as I can	◯	◯	◯	◯

Very Satisfied | 10.5 | **Satisfied** | 7.2 | **Less than Satisfied**

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Sometimes" and "Never or almost never" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

Item Parameters for the TIMSS Advanced 2015 Teacher Job Satisfaction Scale, Advanced Mathematics

Item	delta	tau_1	tau_2	Infit
MTBG10A	0.02623	-1.95127	1.95127	0.88
MTBG10B	-0.43330	-1.84194	1.84194	1.13
MTBG10C	-0.61209	-1.77591	1.77591	1.01
MTBG10D	0.27416	-1.46188	1.46188	0.86
MTBG10E	0.59966	-1.80856	1.80856	0.82
MTBG10F	-0.16519	-1.70772	1.70772	1.04
MTBG10G	0.31053	-1.31610	1.31610	1.30

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 Teacher Job Satisfaction Scale, Advanced Mathematics

Scale Transformation Constants	
A = 8.820804	Transformed Scale Score = 8.820804 + 0.846688 • Logit Scale Score
B = 0.846688	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Teacher Job Satisfaction Scale, Advanced Mathematics

Raw Score	Transformed Scale Score	Cutpoint
0	4.98060	
1	6.05555	
2	6.65040	
3	7.11867	7.2
4	7.54548	
5	7.96186	
6	8.39462	
7	8.84453	
8	9.28947	
9	9.70621	
10	10.10676	
11	10.51404	10.5
12	10.96709	
13	11.54630	
14	12.61485	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Teacher Job Satisfaction Scale, Advanced Mathematics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			MTBG10A	MTBG10B	MTBG10C	MTBG10D	MTBG10E	MTBG10F	MTBG10G
France	0.92	69	0.85	0.82	0.85	0.88	0.82	0.78	0.81
Italy	0.90	63	0.84	0.75	0.75	0.87	0.82	0.83	0.67
Lebanon	0.82	51	0.77	0.69	0.78	0.76	0.66	0.65	0.66
Norway	0.93	70	0.86	0.76	0.86	0.87	0.91	0.83	0.76
Portugal	0.87	56	0.79	0.67	0.68	0.84	0.79	0.76	0.70
Russian Federation	0.91	66	0.83	0.80	0.77	0.83	0.84	0.82	0.79
Russian Federation 6hr+	0.90	62	0.81	0.76	0.77	0.82	0.81	0.78	0.76
Slovenia	0.92	68	0.82	0.75	0.82	0.91	0.88	0.82	0.74
Sweden	0.89	61	0.78	0.72	0.70	0.83	0.87	0.78	0.78
United States	0.92	68	0.85	0.81	0.85	0.88	0.88	0.75	0.73
International Avg.	0.90	63	0.82	0.75	0.78	0.85	0.83	0.78	0.74

Relationship Between the TIMSS Advanced 2015 Teacher Job Satisfaction Scale, and TIMSS Advanced 2015 Advanced Mathematics Achievement

Country	Pearson's Correlation with Advanced Mathematics Achievement		Variance in Advanced Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.05	0.00	0.00
Italy	0.04	0.00	0.01
Lebanon	-0.01	0.00	0.00
Norway	0.18	0.03	0.04
Portugal	0.06	0.00	0.00
Russian Federation	0.22	0.05	0.03
Russian Federation 6hr+	0.08	0.01	0.00
Slovenia	0.13	0.02	0.02
Sweden	0.05	0.00	0.00
United States	-0.03	0.00	0.00
International Median	0.05	0.00	0.00

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Appendix 15B: TIMSS Advanced 2015 Context Questionnaire Scales, Physics

Home Educational Resources Scale, Physics

The Home Educational Resources (HER) scale was created based on students' responses concerning the availability of four resources described below.

Items in the TIMSS Advanced 2015 Home Educational Resources Scale, Physics

PSBG04

Number of books in the home:

1) 0-10
2) 11-25
3) 26-100
4) 101-200
5) More than 200

PSDG06S[1]

Number of home study supports:

1) None
2) Study desk/table or own room
3) Both

PSDGEDUP[1]

Highest level of education of either parent:

1) Finished some primary or lower secondary or did not go to school
2) Finished lower secondary
3) Finished upper secondary
4) Finished post-secondary education
5) Finished university or higher

PSDGOCCP[1]

Highest level of occupation of either parent:

1) Has never worked outside home for pay, general laborer, or semi-professional (skilled agricultural or fishery worker, craft or trade worker, plant or machine operator)
2) Clerical (clerk or service or sales worker)
3) Small business owner
4) Professional (corporate manager or senior official, professional, or technician or associate professional)

Many Resources | 11.4 | **Some Resources** | 5.8 | **Few Resources**

1 Derived variable. For more details, see Supplement 3 of the TIMSS Advanced 2015 User Guide for the International Database.

Item Parameters for the TIMSS Advanced 2015 Home Educational Resources Scale, Physics

Item	delta	tau_1	tau_2	tau_3	tau_4	Infit
PSBG04	0.64929	-0.89914	-0.53587	0.86039	0.57462	1.09
PSDG06S	-0.90816	-0.25972	0.25972			1.11
PSDGEDUP	-0.11749	-0.59426	-0.48752	0.88533	0.19645	1.01
PSDGOCCP	0.37636	-0.64586	1.24705	-0.60119		0.99

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 Home Educational Resources Scale, Physics

Scale Transformation Constants	
A = 7.525439	Transformed Scale Score = 7.525439 + 2.234962 • Logit Scale Score
B = 2.234962	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Home Educational Resources Scale, Physics

Raw Score	Transformed Scale Score	Cutpoint
0	1.31964	
1	3.68097	
2	4.88859	
3	5.77587	5.8
4	6.50614	
5	7.14369	
6	7.71485	
7	8.25370	
8	8.77829	
9	9.31522	
10	9.89707	
11	10.57858	
12	11.48758	11.4
13	13.18946	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Home Educational Resources Scale, Physics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item			
			PSBG04	PSDG06S	PSDGEDUP	PSDGOCCP
France	0.60	47	0.65	0.31	0.84	0.81
Italy	0.59	46	0.66	0.22	0.84	0.81
Lebanon	0.51	40	0.60	0.40	0.73	0.75
Norway	0.51	43	0.67	0.28	0.79	0.75
Portugal	0.67	51	0.75	0.08	0.87	0.84
Russian Federation	0.41	38	0.60	0.17	0.78	0.74
Slovenia	0.55	44	0.67	0.14	0.83	0.77
Sweden	0.58	46	0.70	0.42	0.76	0.77
United States	0.52	43	0.73	0.50	0.81	0.51
International Avg.	0.55	44	0.67	0.28	0.80	0.75

Relationship Between the TIMSS Advanced 2015 Home Educational Resources Scale, and TIMSS Advanced 2015 Physics Achievement

Country	Pearson's Correlation with Physics Achievement		Variance in Physics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.33	0.11	0.07
Italy	0.14	0.02	0.01
Lebanon	0.19	0.04	0.02
Norway	0.26	0.07	0.05
Portugal	0.31	0.09	0.07
Russian Federation	0.15	0.02	0.01
Slovenia	0.21	0.04	0.04
Sweden	0.36	0.13	0.10
United States	0.32	0.10	0.08
International Median	0.26	0.07	0.05

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Safe and Orderly School–Teachers' Reports Scale, Physics

The Safe and Orderly School–Teachers' Reports (SOS) scale was created based on teachers' degree of agreement with the eight statements described below.

Items in the TIMSS Advanced 2015 Safe and Orderly School–Teachers' Reports Scale, Physics[1]

Thinking about your current school, indicate the extent to which you agree or disagree with each of the following statements.

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
PTBG07A	1) This school is located in a safe neighborhood	◯	◯	◯	◯
PTBG07B	2) I feel safe at this school	◯	◯	◯	◯
PTBG07C	3) This school's security policies and practices are sufficient	◯	◯	◯	◯
PTBG07D	4) The students behave in an orderly manner	◯	◯	◯	◯
PTBG07E	5) The students are respectful of the teachers	◯	◯	◯	◯
PTBG07F	6) The students respect school property	◯	◯	◯	◯
PTBG07G	7) This school has clear rules about student conduct	◯	◯	◯	◯
PTBG07H	8) This school's rules are enforced in a fair and consistent manner	◯	◯	◯	◯

Very Safe and Orderly | 9.8 | **Safe and Orderly** | 6.4 | **Less than Safe and Orderly**

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Disagree a little" and "Disagree a lot" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

Item Parameters for the TIMSS Advanced 2015 Safe and Orderly School - Teachers' Reports Scale, Physics

Item	delta	tau_1	tau_2	Infit
PTBG07A	-0.79213	-1.28024	1.28024	1.62
PTBG07B	-1.54669	-1.28512	1.28512	0.91
PTBG07C	-0.40475	-1.47335	1.47335	0.97
PTBG07D	0.38110	-1.85404	1.85404	0.82
PTBG07E	0.25303	-2.01362	2.01362	0.81
PTBG07F	1.12283	-2.06225	2.06225	0.88
PTBG07G	0.11548	-1.62992	1.62992	1.02
PTBG07H	0.87113	-1.60812	1.60812	0.95

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 Safe and Orderly School - Teachers' Reports Scale, Physics

Scale Transformation Constants	
A = 8.10253	Transformed Scale Score = 8.10253 + 0.992158 • Logit Scale Score
B = 0.992158	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Safe and Orderly School - Teachers' Reports Scale, Physics

Raw Score	Transformed Scale Score	Cutpoint
0	3.45082	
1	4.70451	
2	5.37921	
3	5.89100	
4	6.33230	6.4
5	6.74144	
6	7.14044	
7	7.54539	
8	7.96637	
9	8.41273	
10	8.88463	
11	9.37552	
12	9.88022	9.8
13	10.40751	
14	10.99267	
15	11.73382	
16	13.03341	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Safe and Orderly School - Teachers' Reports Scale, Physics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			PTBG07A	PTBG07B	PTBG07C	PTBG07D	PTBG07E	PTBG07F	PTBG07G	PTBG07H
France	0.85	50	0.58	0.68	0.64	0.81	0.80	0.68	0.71	0.70
Italy	0.87	52	0.62	0.67	0.69	0.79	0.77	0.74	0.65	0.80
Lebanon	0.82	46	0.33	0.48	0.65	0.82	0.75	0.77	0.70	0.74
Norway	0.84	49	0.60	0.66	0.68	0.82	0.79	0.70	0.62	0.69
Portugal	0.84	49	0.34	0.63	0.67	0.76	0.84	0.80	0.71	0.73
Russian Federation	0.77	40	0.48	0.59	0.57	0.76	0.79	0.72	0.67	0.29
Slovenia	0.91	63	0.83	0.79	0.82	0.85	0.83	0.77	0.73	0.73
Sweden	0.83	48	0.42	0.59	0.75	0.79	0.80	0.76	0.58	0.76
United States	0.90	59	0.42	0.69	0.77	0.89	0.85	0.83	0.79	0.81
International Avg.	0.85	51	0.51	0.64	0.69	0.81	0.81	0.75	0.69	0.70

Relationship Between the TIMSS Advanced 2015 Safe and Orderly School - Teachers' Reports Scale, and TIMSS Advanced 2015 Physics Achievement

Country	Pearson's Correlation with Physics Achievement		Variance in Physics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.08	0.01	0.01
Italy	0.03	0.00	0.00
Lebanon	0.07	0.00	0.01
Norway	0.11	0.01	0.02
Portugal	0.04	0.00	0.01
Russian Federation	0.03	0.00	0.00
Slovenia	0.19	0.04	0.04
Sweden	0.12	0.01	0.02
United States	0.25	0.06	0.12
International Median	0.08	0.01	0.01

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

School Discipline Problems–Principals' Reports Scale, Physics

The School Discipline Problems–Principals' Reports (DAS) scale was created based on principals' responses concerning the eleven potential school problems described below.

Items in the TIMSS Advanced 2015 School Discipline Problems–Principals' Reports Scale, Physics

To what degree is each of the following a problem among twelfth grade students in your school?

		Not a problem	Minor problem	Moderate problem	Serious problem
PCBG11A	1) Arriving late at school	○	○	○	○
PCBG11B	2) Absenteeism (i.e., unjustified absences)	○	○	○	○
PCBG11C	3) Classroom disturbance	○	○	○	○
PCBG11D	4) Cheating	○	○	○	○
PCBG11E	5) Profanity	○	○	○	○
PCBG11F	6) Vandalism	○	○	○	○
PCBG11G	7) Theft	○	○	○	○
PCBG11H	8) Intimidation or verbal abuse among students (including texting, emailing, etc.)	○	○	○	○
PCBG11 I	9) Physical injury to other students	○	○	○	○
PCBG11J	10) Intimidation or verbal abuse of teachers or staff (including texting, emailing, etc.)	○	○	○	○
PCBG11K	11) Physical injury to teachers or staff	○	○	○	○

Hardly Any Problems | 10.0 | **Minor Problems** | 7.2 | **Moderate to Severe Problems**

Item Parameters for the TIMSS Advanced 2015 School Discipline Problems - Principals' Reports Scale, Physics

Item	delta	tau_1	tau_2	tau_3	Infit
PCBG11A	1.17877	-3.21012	-0.16690	3.37702	1.21
PCBG11B	1.46046	-2.19399	-0.42850	2.62249	1.13
PCBG11C	0.18891	-1.99828	-0.62084	2.61912	0.93
PCBG11D	0.80313	-2.53161	-0.66242	3.19403	1.11
PCBG11E	-0.22664	-1.90441	-0.40045	2.30486	0.84
PCBG11F	-0.50647	0.20063	-1.22273	1.02210	0.76
PCBG11G	-0.43015	-0.05140	-1.18166	1.23306	0.69
PCBG11H	-0.06170	-1.19025	-1.00895	2.19920	0.75
PCBG11I	-0.69241	0.65883	-1.47382	0.81499	0.64
PCBG11J	-0.67545	0.54715	-1.66947	1.12232	0.67

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 School Discipline Problems - Principals' Reports Scale, Physics

Scale Transformation Constants	
A = 7.709439	Transformed Scale Score = 7.709439 + 0.955134 • Logit Scale Score
B = 0.955134	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 School Discipline Problems - Principals' Reports Scale, Physics

Raw Score	Transformed Scale Score	Cutpoint
0	3.54143	
1	4.62379	
2	5.12325	
3	5.44481	
4	5.68073	
5	5.86766	
6	6.02294	
7	6.15980	
8	6.28303	
10	6.50367	
11	6.60698	
12	6.70877	
14	6.91459	
15	7.02423	
16	7.13973	7.2
17	7.26373	
18	7.39914	
19	7.54913	
20	7.71713	
21	7.90633	
22	8.12018	
23	8.35907	
24	8.62458	
25	8.91866	
26	9.24592	
27	9.61214	
28	10.02711	10.0
29	10.50216	
30	11.04762	
31	11.67530	
32	12.44148	
33	13.73438	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 School Discipline Problems - Principals' Reports Scale, Physics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item										
			PCBG11A	PCBG11B	PCBG11C	PCBG11D	PCBG11E	PCBG11F	PCBG11G	PCBG11H	PCBG11I	PCBG11J	PCBG11K
France	0.94	65	0.61	0.65	0.79	0.70	0.88	0.89	0.80	0.88	0.91	0.87	0.85
Italy	0.96	72	0.51	0.76	0.85	0.72	0.80	0.94	0.91	0.91	0.95	0.93	0.93
Lebanon	0.98	84	0.84	0.88	0.90	0.87	0.94	0.95	0.96	0.91	0.94	0.93	0.94
Norway	0.82	41	0.56	0.55	0.54	0.60	0.68	0.80	0.74	0.50	0.75	0.74	0.53
Portugal	0.95	66	0.55	0.77	0.84	0.72	0.77	0.90	0.89	0.87	0.89	0.87	0.81
Russian Federation	0.77	37	0.55	0.61	0.61	0.54	0.70	0.67	0.54	0.70	0.56	-	-
Slovenia	0.76	36	0.15	0.40	0.61	0.67	0.64	0.54	0.66	0.59	0.74	0.58	0.74
Sweden	0.84	42	0.50	0.54	0.61	0.59	0.67	0.77	0.70	0.68	0.69	0.78	0.50
United States	0.87	46	0.69	0.66	0.74	0.67	0.77	0.55	0.71	0.76	0.59	0.78	0.41
International Avg.	0.88	54	0.55	0.65	0.72	0.68	0.76	0.78	0.77	0.76	0.78	0.81	0.71

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS Advanced 2015 School Discipline Problems - Principals' Reports Scale, and TIMSS Advanced 2015 Physics Achievement

Country	Pearson's Correlation with Physics Achievement		Variance in Physics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.12	0.01	0.00
Italy	0.23	0.05	0.03
Lebanon	0.07	0.00	0.00
Norway	0.07	0.00	0.01
Portugal	0.05	0.00	0.00
Russian Federation	0.05	0.00	0.01
Slovenia	0.18	0.03	0.03
Sweden	0.18	0.03	0.02
United States	0.16	0.03	0.01
International Median	0.12	0.01	0.01

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

School Supports Advanced Mathematics and Physics Education–Principal Version Scale, Physics

The School Supports Advanced Mathematics and Physics Education–Principal Version (SMP) scale was created based on principals' responses characterizing the seven aspects described below.

Items in the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education–Principal Version Scale, Physics[1]

How much do you agree with these statements about advanced mathematics and physics education within your school?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
PCBG10A	1) The school encourages students to study advanced mathematics and physics	○	○	○	○
PCBG10B	2) The school promotes professional development for teachers of advanced mathematics and physics	○	○	○	○
PCBG10C	3) The school provides students with information about career options in advanced mathematics and physics	○	○	○	○
PCBG10D	4) The school has initiatives to promote student interest in advanced mathematics and physics (e.g., student clubs, competitions)	○	○	○	○
PCBG10E	5) The school has partnership initiatives with industry/businesses in advanced mathematics and physics	○	○	○	○
PCBG10F	6) Advanced mathematics and physics teachers are admired by other teachers in the school	○	○	○	○
PCBG10G	7) Students at this school respect students who excel in advanced mathematics and physics	○	○	○	○

Strongly Supportive | 10.8 | **Supportive** | 6.5 | **Less than Supportive**

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Disagree a little" and "Disagree a lot" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

Item Parameters for the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Principal Version Scale, Physics

Item	delta	tau_1	tau_2	Infit
PCBG10A	-0.96174	-1.03132	1.03132	0.84
PCBG10B	-0.48986	-1.17668	1.17668	0.91
PCBG10C	-0.82638	-1.18249	1.18249	0.91
PCBG10D	-0.45940	-0.92144	0.92144	0.92
PCBG10E	1.76069	-0.59088	0.59088	1.05
PCBG10F	1.04013	-0.61623	0.61623	1.03
PCBG10G	-0.06344	-0.95292	0.95292	1.08

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Principal Version Scale, Physics

Scale Transformation Constants	
A = 8.69507	Transformed Scale Score = 8.69507 + 1.384086 • Logit Scale Score
B = 1.384086	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Principal Version Scale, Physics

Raw Score	Transformed Scale Score	Cutpoint
0	3.03255	
1	4.78821	
2	5.74762	
3	6.48485	6.5
4	7.12106	
5	7.70041	
6	8.24387	
7	8.76353	
8	9.27161	
9	9.78476	
10	10.31709	
11	10.89611	10.8
12	11.57031	
13	12.46547	
14	14.14715	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Principal Version Scale, Physics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			PCBG10A	PCBG10B	PCBG10C	PCBG10D	PCBG10E	PCBG10F	PCBG10G
France	0.66	34	0.70	0.60	0.62	0.56	0.50	0.60	0.46
Italy	0.77	42	0.81	0.71	0.75	0.61	0.66	0.47	0.46
Lebanon	0.75	41	0.61	0.65	0.74	0.76	0.56	0.60	0.57
Norway	0.70	37	0.65	0.47	0.47	0.70	0.56	0.69	0.67
Portugal	0.76	41	0.53	0.68	0.62	0.65	0.58	0.73	0.67
Russian Federation	0.67	40	0.57	0.73	0.70	0.63	0.36	0.62	0.73
Slovenia	0.73	39	0.60	0.70	0.66	0.71	0.72	0.47	0.45
Sweden	0.73	40	0.76	0.79	0.78	0.75	0.44	0.21	0.40
United States	0.80	47	0.64	0.69	0.74	0.78	0.50	0.69	0.73
International Avg.	0.73	40	0.65	0.67	0.68	0.68	0.54	0.56	0.57

Relationship Between the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Principal Version Scale, and TIMSS Advanced 2015 Physics Achievement

Country	Pearson's Correlation with Physics Achievement		Variance in Physics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.03	0.00	0.00
Italy	0.05	0.00	0.01
Lebanon	0.04	0.00	0.01
Norway	0.07	0.01	0.00
Portugal	-0.02	0.00	0.00
Russian Federation	0.19	0.04	0.06
Slovenia	0.25	0.06	0.04
Sweden	0.14	0.02	0.02
United States	0.01	0.00	0.00
International Median	0.05	0.00	0.01

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

School Supports Advanced Mathematics and Physics Education–Teacher Version Scale, Physics

The School Supports Advanced Mathematics and Physics Education–Teacher Version (SMP) scale was created based on teachers' responses characterizing the seven aspects described below.

Items in the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education–Teacher Version Scale, Physics[1]

How much do you agree with these statements about advanced mathematics and physics education within your school?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
PTBG06A	1) The school encourages students to study advanced mathematics and physics	◯	◯	◯	◯
PTBG06B	2) The school promotes professional development for teachers of advanced mathematics and physics	◯	◯	◯	◯
PTBG06C	3) The school provides students with information about career options in advanced mathematics and physics	◯	◯	◯	◯
PTBG06D	4) Advanced mathematics and physics teachers are admired by other teachers in the school	◯	◯	◯	◯
PTBG06E	5) Teachers have high expectations for student achievement in advanced mathematics and physics	◯	◯	◯	◯
PTBG06F	6) Students at this school respect students who excel in advanced mathematics and physics	◯	◯	◯	◯
PTBG06G	7) Parents expect their children to study advanced mathematics and physics	◯	◯	◯	◯

Strongly Supportive | 11.6 | Supportive | 7.5 | Less than Supportive

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Disagree a little" and "Disagree a lot" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

Item Parameters for the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Teacher Version Scale, Physics

Item	delta	tau_1	tau_2	Infit
PTBG06A	-0.69327	-1.09016	1.09016	0.87
PTBG06B	0.42027	-0.81515	0.81515	1.05
PTBG06C	-0.23350	-1.24334	1.24334	0.96
PTBG06D	0.88624	-0.64947	0.64947	0.99
PTBG06E	-0.40883	-1.40288	1.40288	0.93
PTBG06F	-0.39632	-1.00752	1.00752	1.16
PTBG06G	0.42541	-1.28652	1.28652	1.01

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Teacher Version Scale, Physics

Scale Transformation Constants	
A = 9.52146	Transformed Scale Score = 9.52146 + 1.382519 • Logit Scale Score
B = 1.382519	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Teacher Version Scale, Physics

Raw Score	Transformed Scale Score	Cutpoint
0	4.00781	
1	5.74570	
2	6.68506	
3	7.40122	7.5
4	8.01543	
5	8.57240	
6	9.09310	
7	9.58929	
8	10.07291	
9	10.56143	
10	11.06907	
11	11.62470	11.6
12	12.28126	
13	13.16777	
14	14.85545	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Teacher Version Scale, Physics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			PTBG06A	PTBG06B	PTBG06C	PTBG06D	PTBG06E	PTBG06F	PTBG06G
France	0.63	32	0.56	0.47	0.42	0.51	0.65	0.66	0.62
Italy	0.78	44	0.72	0.60	0.70	0.64	0.73	0.51	0.70
Lebanon	0.77	43	0.64	0.73	0.65	0.51	0.68	0.65	0.71
Norway	0.60	31	0.69	0.38	0.62	0.66	0.52	0.60	0.26
Portugal	0.72	38	0.71	0.71	0.66	0.48	0.51	0.61	0.57
Russian Federation	0.77	44	0.76	0.79	0.65	0.61	0.54	0.73	0.48
Slovenia	0.70	36	0.71	0.70	0.57	0.69	0.65	0.38	0.40
Sweden	0.70	35	0.73	0.70	0.56	0.47	0.57	0.57	0.51
United States	0.84	51	0.66	0.80	0.78	0.69	0.68	0.58	0.80
International Avg.	0.72	39	0.69	0.65	0.62	0.58	0.62	0.59	0.56

Relationship Between the TIMSS Advanced 2015 School Supports Advanced Mathematics and Physics Education - Teacher Version Scale, and TIMSS Advanced 2015 Physics Achievement

Country	Pearson's Correlation with Physics Achievement		Variance in Physics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.05	0.00	0.00
Italy	-0.01	0.00	0.00
Lebanon	-0.03	0.00	0.00
Norway	0.02	0.00	0.00
Portugal	0.01	0.00	0.01
Russian Federation	0.08	0.01	0.02
Slovenia	0.23	0.05	0.03
Sweden	0.10	0.01	0.00
United States	0.23	0.05	0.01
International Median	0.05	0.00	0.00

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Students Like Learning Physics Scale

The Students Like Learning Physics (SLP) scale was created based on students' degree of agreement with the twelve statements described below.

Items in the TIMSS Advanced 2015 Students Like Learning Physics Scale

How much do you agree with these statements about the physics you are studying?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
PSBP20A	1 I enjoy conducting experiments or investigations in physics				
PSBP20B	2) I get a sense of satisfaction when I solve physics problems				
PSBP20C*	3) I feel bored when I do my physics schoolwork*				
PSBP20D	4) I like studying for my physics class outside of school				
PSBP20E	5) It is interesting to learn physics laws and principles				
PSBP20F*	6) I dread my physics class*				
PSBP20G	7) I am studying physics because I like to learn new things				
PSBP20H	8) I enjoy figuring out challenging physics				
PSBP20 I	9) Physics is one of my favorite subjects				
PSBP20J	10) Jobs that require physics skills seem interesting to me				
PSBP20K*	11) I wish I did not have to study physics*				
PSBP20L	12) I enjoy thinking about the world in terms of laws of physics				

* Reverse coded

Very Much Like Learning Physics — 11.4 — **Like Learning Physics** — 8.8 — **Do Not Like Learning Physics**

Item Parameters for the TIMSS Advanced 2015 Students Like Learning Physics Scale, Physics

Item	delta	tau_1	tau_2	tau_3	Infit
PSBP20A	-0.51423	-1.07025	-0.41990	1.49015	1.15
PSBP20B	-0.90801	-0.86799	-0.43946	1.30745	1.10
PSBP20C*	0.51315	-1.69256	-0.14245	1.83501	1.23
PSBP20D	1.03682	-1.72677	-0.08320	1.80997	1.11
PSBP20E	-0.21380	-1.38699	-0.26425	1.65124	0.90
PSBP20F*	-0.39443	-1.02035	0.04349	0.97686	1.82
PSBP20G	-0.18466	-1.35598	-0.31080	1.66678	0.81
PSBP20H	0.04092	-1.36190	-0.07935	1.44125	0.85
PSBP20I	0.54534	-0.86829	-0.06458	0.93287	0.75
PSBP20J	0.17389	-0.97760	-0.33432	1.31192	0.94
PSBP20K*	-0.07291	-0.60736	-0.22268	0.83004	1.02
PSBP20L	-0.02208	-1.18389	-0.23501	1.41890	1.02

* Reverse coded

Scale Transformation Constants for the TIMSS Advanced 2015 Students Like Learning Physics Scale, Physics

Scale Transformation Constants	
A = 8.81822	Transformed Scale Score = 8.81822 + 1.464843 • Logit Scale Score
B = 1.464843	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Students Like Learning Physics Scale, Physics

Raw Score	Transformed Scale Score	Cutpoint
0	2.29834	
1	3.92594	
2	4.70005	
3	5.22580	
4	5.63169	
5	5.97016	
6	6.26389	
7	6.52678	
8	6.76750	
9	6.99121	
10	7.20367	
11	7.40630	
12	7.60136	
13	7.79069	
14	7.97591	
15	8.15835	
16	8.33926	
17	8.51979	
18	8.70102	8.8
19	8.88404	
20	9.06991	
21	9.25980	
22	9.45491	
23	9.65655	
24	9.86615	
25	10.08468	
26	10.31566	
27	10.56116	
28	10.82464	
29	11.11077	
30	11.42601	11.4
31	11.77922	
32	12.18651	
33	12.67364	
34	13.29295	
35	14.17455	
36	15.92718	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Students Like Learning Physics Scale, Physics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item											
			PSBP20A	PSBP20B	PSBP20C*	PSBP20D	PSBP20E	PSBP20F*	PSBP20G	PSBP20H	PSBP20I	PSBP20J	PSBP20K*	PSBP20L
France	0.89	47	0.59	0.50	0.63	0.70	0.78	0.30	0.78	0.75	0.83	0.69	0.80	0.66
Italy	0.90	49	0.62	0.54	0.60	0.78	0.78	0.33	0.81	0.79	0.82	0.77	0.75	0.64
Lebanon	0.83	39	0.67	0.59	0.34	0.48	0.70	0.16	0.75	0.72	0.81	0.73	0.51	0.67
Norway	0.89	47	0.44	0.35	0.69	0.67	0.78	0.44	0.78	0.79	0.85	0.74	0.77	0.68
Portugal	0.89	47	0.58	0.67	0.61	0.70	0.76	0.40	0.78	0.78	0.82	0.68	0.73	0.66
Russian Federation	0.92	53	0.53	0.78	0.63	0.76	0.82	0.38	0.79	0.81	0.85	0.82	0.67	0.70
Slovenia	0.87	41	0.50	0.59	0.48	0.57	0.70	0.49	0.72	0.77	0.80	0.73	0.59	0.64
Sweden	0.91	51	0.52	0.61	0.69	0.72	0.79	0.51	0.79	0.81	0.84	0.74	0.78	0.69
United States	0.92	52	0.63	0.61	0.55	0.66	0.82	0.63	0.78	0.80	0.85	0.75	0.75	0.78
International Avg.	0.89	47	0.56	0.58	0.58	0.67	0.77	0.41	0.78	0.78	0.83	0.74	0.71	0.68

* Reverse coded

Relationship Between the TIMSS Advanced 2015 Students Like Learning Physics Scale, and TIMSS Advanced 2015 Physics Achievement

Country	Pearson's Correlation with Physics Achievement		Variance in Physics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.46	0.22	0.18
Italy	0.39	0.15	0.13
Lebanon	0.22	0.05	0.04
Norway	0.50	0.25	0.22
Portugal	0.43	0.18	0.17
Russian Federation	0.38	0.15	0.13
Slovenia	0.44	0.20	0.15
Sweden	0.44	0.20	0.16
United States	0.41	0.17	0.16
International Median	0.43	0.18	0.16

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Students' Sense of School Belonging Scale, Physics

The Students' Sense of School Belonging (SSB) scale was created based on students' degree of agreement with the nine statements described below.

Items in the TIMSS Advanced 2015 Students' Sense of School Belonging Scale, Physics

What do you think about your school? Tell how much you agree with these statements.

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
PSBP22A	1) I enjoy school	○	○	○	○
PSBP22B	2) I feel safe when I am at school	○	○	○	○
PSBP22C	3) I feel like I belong at this school	○	○	○	○
PSBP22D	4) I like to see my classmates at school	○	○	○	○
PSBP22E	5) Teachers at my school are fair to me	○	○	○	○
PSBP22F	6) I am proud to go to this school	○	○	○	○
PSBP22G	7) I learn a lot in school	○	○	○	○
PSBP22H	8) My classmates respect students who excel in school subjects	○	○	○	○
PSBP22 I	9) My classmates respect students who struggle learning school subjects	○	○	○	○

High Sense of School Belonging | 10.5 | **Sense of School Belonging** | 7.6 | **Little Sense of School Belonging**

Item Parameters for the TIMSS Advanced 2015 Students' Sense of School Belonging Scale, Physics

Item	delta	tau_1	tau_2	tau_3	Infit
PSBP22A	0.42482	-1.13862	-0.66108	1.79970	1.06
PSBP22B	-0.19651	-0.90823	-0.65897	1.56720	0.94
PSBP22C	0.27155	-1.09791	-0.37350	1.47141	0.87
PSBP22D	-0.77499	-0.66078	-0.69906	1.35984	1.04
PSBP22E	0.02649	-1.36739	-0.55367	1.92106	1.10
PSBP22F	0.36287	-1.16971	-0.42176	1.59147	0.93
PSBP22G	-0.30178	-1.41507	-0.65833	2.07340	0.98
PSBP22H	-0.08185	-1.20393	-0.55128	1.75521	1.09
PSBP22I	0.26940	-1.37985	-0.44086	1.82071	1.29

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 Students' Sense of School Belonging Scale, Physics

Scale Transformation Constants	
A = 7.921577	Transformed Scale Score = 7.921577 + 1.242992 • Logit Scale Score
B = 1.242992	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Students' Sense of School Belonging Scale, Physics

Raw Score	Transformed Scale Score	Cutpoint
0	2.83834	
1	4.17439	
2	4.80361	
3	5.23192	
4	5.56897	
5	5.85267	
6	6.10445	
7	6.33561	
8	6.55328	
9	6.76244	
10	6.96690	
11	7.16980	
12	7.37399	
13	7.58285	7.6
14	7.79734	
15	8.02192	
16	8.25960	
17	8.51313	
18	8.78559	
19	9.07945	
20	9.39640	
21	9.73823	
22	10.10871	
23	10.51544	10.5
24	10.97845	
25	11.53918	
26	12.31006	
27	13.80166	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Students' Sense of School Belonging Scale, Physics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			PSBP22A	PSBP22B	PSBP22C	PSBP22D	PSBP22E	PSBP22F	PSBP22G	PSBP22H	PSBP22I
France	0.82	42	0.72	0.68	0.69	0.59	0.54	0.73	0.65	0.59	0.58
Italy	0.83	43	0.73	0.73	0.76	0.51	0.64	0.79	0.68	0.51	0.48
Lebanon	0.87	49	0.68	0.78	0.80	0.64	0.65	0.76	0.70	0.66	0.61
Norway	0.83	44	0.70	0.69	0.79	0.65	0.56	0.71	0.64	0.62	0.56
Portugal	0.83	43	0.74	0.68	0.74	0.55	0.56	0.78	0.67	0.60	0.57
Russian Federation	0.89	54	0.79	0.74	0.79	0.70	0.66	0.81	0.72	0.67	0.70
Slovenia	0.86	47	0.69	0.67	0.80	0.60	0.67	0.81	0.77	0.62	0.50
Sweden	0.85	47	0.80	0.69	0.80	0.67	0.58	0.76	0.69	0.56	0.58
United States	0.87	49	0.67	0.69	0.79	0.68	0.64	0.79	0.71	0.71	0.62
International Avg.	0.85	47	0.73	0.70	0.77	0.62	0.61	0.77	0.69	0.62	0.58

Relationship Between the TIMSS Advanced 2015 Students' Sense of School Belonging Scale, and TIMSS Advanced 2015 Physics Achievement

Country	Pearson's Correlation with Physics Achievement		Variance in Physics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.25	0.06	0.05
Italy	0.06	0.00	0.01
Lebanon	0.02	0.00	0.00
Norway	0.16	0.03	0.03
Portugal	0.11	0.01	0.02
Russian Federation	0.07	0.01	0.00
Slovenia	0.22	0.05	0.05
Sweden	0.19	0.04	0.05
United States	0.09	0.01	0.01
International Median	0.11	0.01	0.02

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Students Value Physics Scale

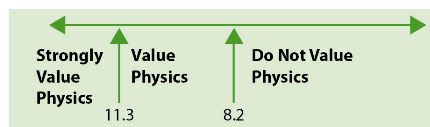
The Students Value Physics (SVP) scale was created based on students' degree of agreement with the nine statements described below.

Items in the TIMSS Advanced 2015 Students Value Physics Scale

How much do you agree with these statements about the physics you are studying?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
PSBP21A	1) Learning physics will help me get ahead in the world	○	○	○	○
PSBP21B	2) It is important to do well in my physics class	○	○	○	○
PSBP21C*	3) The physics I am studying is not useful for my future*	○	○	○	○
PSBP21D	4) My parents are pleased that I am taking physics	○	○	○	○
PSBP21E	5) Doing well in physics will help me get into the university of my choice	○	○	○	○
PSBP21F*	6) Learning physics does not seem to be a worthwhile exercise*	○	○	○	○
PSBP21G	7) My parents think that it is important that I do well in my physics class	○	○	○	○
PSBP21H	8) I like telling people I am studying physics	○	○	○	○
PSBP21 I	9) Learning physics will give me more job opportunities	○	○	○	○

*Reverse coded



Item Parameters for the TIMSS Advanced 2015 Students Value Physics Scale, Physics

Item	delta	tau_1	tau_2	tau_3	Infit
PSBP21A	0.12198	-1.22089	-0.28221	1.50310	0.91
PSBP21B	-0.11136	-1.22137	-0.16984	1.39121	1.07
PSBP21C*	0.21349	-0.87936	-0.23886	1.11822	1.30
PSBP21D	-0.34054	-1.18074	-0.59829	1.77903	1.02
PSBP21E	0.03096	-0.41265	-0.17841	0.59106	0.97
PSBP21F*	-0.16915	-0.85167	-0.33674	1.18841	1.43
PSBP21G	-0.43055	-0.94318	-0.50789	1.45107	1.05
PSBP21H	0.64924	-1.21577	-0.07682	1.29259	1.16
PSBP21I	0.03593	-0.88494	-0.30990	1.19484	0.88

*Reverse coded

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 Students Value Physics Scale, Physics

Scale Transformation Constants	
A = 8.418724	Transformed Scale Score = 8.418724 + 1.650896 • Logit Scale Score
B = 1.650896	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Students Value Physics Scale, Physics

Raw Score	Transformed Scale Score	Cutpoint
0	1.91155	
1	3.71438	
2	4.57065	
3	5.15343	
4	5.60982	
5	5.99212	
6	6.32606	
7	6.63211	
8	6.91547	
9	7.18258	
10	7.43842	
11	7.68700	
12	7.93169	
13	8.17550	8.2
14	8.42143	
15	8.67212	
16	8.93095	
17	9.20112	
18	9.48621	
19	9.78912	
20	10.11722	
21	10.47668	
22	10.87750	
23	11.33535	11.3
24	11.87577	
25	12.55759	
26	13.52803	
27	15.46562	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Students Value Physics Scale, Physics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			PSBP21A	PSBP21B	PSBP21C*	PSBP21D	PSBP21E	PSBP21F*	PSBP21G	PSBP21H	PSBP21I
France	0.83	43	0.70	0.57	0.69	0.58	0.75	0.67	0.52	0.61	0.77
Italy	0.85	46	0.79	0.73	0.58	0.61	0.70	0.68	0.57	0.63	0.77
Lebanon	0.73	36	0.58	0.70	0.31	0.66	0.69	0.20	0.67	0.59	0.74
Norway	0.77	37	0.75	0.70	0.49	0.63	0.65	0.69	0.46	0.46	0.56
Portugal	0.83	43	0.76	0.67	0.63	0.63	0.64	0.65	0.60	0.57	0.73
Russian Federation	0.86	51	0.76	0.79	0.42	0.80	0.82	0.43	0.79	0.61	0.83
Slovenia	0.63	29	0.68	0.44	0.20	0.67	0.45	-0.02	0.67	0.57	0.73
Sweden	0.79	38	0.75	0.57	0.66	0.59	0.62	0.67	0.49	0.48	0.66
United States	0.85	45	0.78	0.69	0.64	0.69	0.67	0.62	0.52	0.62	0.79
International Avg.	0.79	41	0.73	0.65	0.51	0.65	0.67	0.51	0.59	0.57	0.73

*Reverse coded

Relationship Between the TIMSS Advanced 2015 Students Value Physics Scale, and TIMSS Advanced 2015 Physics Achievement

Country	Pearson's Correlation with Physics Achievement		Variance in Physics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.38	0.14	0.10
Italy	0.28	0.08	0.07
Lebanon	0.19	0.04	0.03
Norway	0.28	0.08	0.08
Portugal	0.30	0.09	0.08
Russian Federation	0.31	0.10	0.10
Slovenia	0.22	0.05	0.03
Sweden	0.26	0.07	0.06
United States	0.33	0.11	0.10
International Median	0.28	0.08	0.08

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

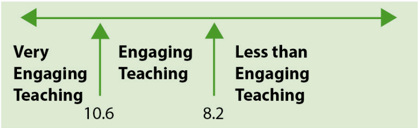
Students' Views on Engaging Teaching in Physics Lessons Scale

The Students' Views on Engaging Teaching in Physics Lessons (EPL) scale was created based on students' degree of agreement with the fourteen statements described below.

Items in the TIMSS Advanced 2015 Students' Views on Engaging Teaching in Physics Lessons Scale

How much do you agree with these statements about your physics lessons?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
PSBP18A	1) The teacher clearly communicates the purpose of each physics lesson	○	○	○	○
PSBP18B	2) I know what my teacher expects me to do	○	○	○	○
PSBP18C	3) My teacher is easy to understand	○	○	○	○
PSBP18D	4) I am interested in what my teacher says	○	○	○	○
PSBP18E	5) My teacher gives me interesting things to do	○	○	○	○
PSBP18F	6) My teacher asks me thought provoking questions	○	○	○	○
PSBP18G	7) My teacher has clear answers to my questions	○	○	○	○
PSBP18H	8) My teacher links new content to what I already know	○	○	○	○
PSBP18 I	9) My teacher is good at explaining physics	○	○	○	○
PSBP18J	10) My teacher provides the opportunity for me to show what I have learned	○	○	○	○
PSBP18K	11) My teacher encourages me to keep working on physics problems until I solve them	○	○	○	○
PSBP18L	12) My teacher provides helpful feedback on my schoolwork (including homework)	○	○	○	○
PSBP18M	13) My teacher uses a variety of teaching methods, tasks, and activities to help us learn	○	○	○	○
PSBP18N	14) My teacher believes that I can learn difficult physics material	○	○	○	○



Item Parameters for the TIMSS Advanced 2015 Students' Views on Engaging Teaching in Physics Lessons Scale, Physics

Item	delta	tau_1	tau_2	tau_3	Infit
PSBP18A	-0.14831	-1.38275	-0.35811	1.74086	0.87
PSBP18B	-0.43649	-1.60005	-0.45229	2.05234	1.13
PSBP18C	0.13061	-1.48920	-0.24496	1.73416	0.83
PSBP18D	-0.05456	-1.45103	-0.42519	1.87622	1.16
PSBP18E	0.63636	-1.76934	-0.24144	2.01078	1.02
PSBP18F	0.13739	-1.39531	-0.27005	1.66536	1.13
PSBP18G	0.14795	-1.29217	-0.29296	1.58513	0.95
PSBP18H	-0.32221	-1.50071	-0.42276	1.92347	1.00
PSBP18I	-0.19384	-1.15279	-0.30087	1.45366	0.79
PSBP18J	-0.27918	-1.98442	-0.36623	2.35065	1.06
PSBP18K	-0.01635	-1.57874	-0.18354	1.76228	1.16
PSBP18L	0.36382	-1.67707	-0.16971	1.84678	1.10
PSBP18M	0.21519	-1.60597	-0.15229	1.75826	1.13
PSBP18N	-0.18038	-1.41620	-0.42459	1.84079	1.34

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 Students' Views on Engaging Teaching in Physics Lessons Scale, Physics

Scale Transformation Constants	
A = 8.247393	Transformed Scale Score = 8.247393 + 1.153149 • Logit Scale Score
B = 1.153149	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Students' Views on Engaging Teaching in Physics Lessons Scale, Physics

Raw Score	Transformed Scale Score	Cutpoint
0	2.56831	
1	3.87485	
2	4.50403	
3	4.93321	
4	5.26459	
5	5.53959	
6	5.77697	
7	5.98812	
8	6.18044	
9	6.35812	
10	6.52518	
11	6.68393	
12	6.83632	
13	6.98389	
14	7.12790	
15	7.26945	
16	7.40948	
17	7.54955	
18	7.68865	
19	7.82857	
20	7.96996	
21	8.11342	8.2
22	8.25966	
23	8.40883	
24	8.56179	
25	8.71883	
26	8.88033	
27	9.04665	
28	9.21823	
29	9.39540	
30	9.57861	
31	9.76853	
32	9.96608	
33	10.17267	
34	10.39036	
35	10.62218	10.6
36	10.87268	
37	11.14749	
38	11.45962	
39	11.82813	
40	12.29340	
41	12.95657	
42	14.29669	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Students' Views on Engaging Teaching in Physics Lessons Scale, Physics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item													
			PSBP18A	PSBP18B	PSBP18C	PSBP18D	PSBP18E	PSBP18F	PSBP18G	PSBP18H	PSBP18I	PSBP18J	PSBP18K	PSBP18L	PSBP18M	PSBP18N
France	0.92	49	0.76	0.68	0.79	0.62	0.68	0.65	0.77	0.69	0.82	0.63	0.66	0.67	0.69	0.66
Italy	0.93	52	0.82	0.57	0.82	0.54	0.65	0.74	0.79	0.72	0.84	0.72	0.73	0.77	0.74	0.53
Lebanon	0.93	54	0.80	0.60	0.79	0.74	0.74	0.71	0.79	0.76	0.81	0.74	0.73	0.66	0.76	0.64
Norway	0.91	46	0.72	0.71	0.79	0.60	0.71	0.53	0.72	0.72	0.77	0.68	0.53	0.70	0.61	0.58
Portugal	0.94	55	0.82	0.67	0.83	0.64	0.76	0.65	0.81	0.76	0.84	0.72	0.75	0.71	0.73	0.60
Russian Federation	0.92	51	0.78	0.64	0.77	0.78	0.77	0.58	0.77	0.63	0.81	0.63	0.67	0.79	0.73	0.58
Slovenia	0.92	51	0.78	0.75	0.78	0.61	0.72	0.72	0.72	0.75	0.81	0.64	0.64	0.69	0.68	0.62
Sweden	0.92	50	0.75	0.65	0.82	0.66	0.75	0.70	0.75	0.74	0.83	0.66	0.66	0.69	0.65	0.58
United States	0.94	55	0.81	0.73	0.81	0.71	0.75	0.74	0.79	0.77	0.83	0.71	0.65	0.76	0.73	0.60
International Avg.	0.93	51	0.78	0.67	0.80	0.65	0.73	0.67	0.77	0.73	0.82	0.68	0.67	0.72	0.70	0.60

Relationship Between the TIMSS Advanced 2015 Students' Views on Engaging Teaching in Physics Lessons Scale, and TIMSS Advanced 2015 Physics Achievement

Country	Pearson's Correlation with Physics Achievement		Variance in Physics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.24	0.06	0.05
Italy	0.12	0.01	0.01
Lebanon	0.08	0.01	0.02
Norway	0.25	0.06	0.06
Portugal	0.04	0.00	0.00
Russian Federation	0.20	0.04	0.04
Slovenia	0.31	0.10	0.10
Sweden	0.24	0.06	0.06
United States	0.16	0.03	0.04
International Median	0.20	0.04	0.04

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Teacher Job Satisfaction Scale, Physics

The Teacher Job Satisfaction (TJS) scale was created based on how often teachers responded positively to the seven statements described below.

Items in the TIMSS Advanced 2015 Teacher Job Satisfaction Scale, Physics[1]

	How often do you feel the following way about being a teacher?	Very often	Often	Sometimes	Never or almost never
PTBG10A	1) I am content with my profession as a teacher	○	○	○	○
PTBG10B	2) I am satisfied with being a teacher at this school	○	○	○	○
PTBG10C	3) I find my work full of meaning and purpose	○	○	○	○
PTBG10D	4) I am enthusiastic about my job	○	○	○	○
PTBG10E	5) My work inspires me	○	○	○	○
PTBG10F	6) I am proud of the work I do	○	○	○	○
PTBG10G	7) I am going to continue teaching for as long as I can	○	○	○	○

Satisfied | 10.6 | Somewhat Satisfied | 7.4 | Less Than Satisfied

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Sometimes" and "Never or almost never" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

Item Parameters for the TIMSS Advanced 2015 Teacher Job Satisfaction Scale, Physics

Item	delta	tau_1	tau_2	Infit
PTBG10A	0.15383	-1.80976	1.80976	1.02
PTBG10B	-0.38606	-1.52386	1.52386	1.29
PTBG10C	-0.60185	-1.68715	1.68715	0.93
PTBG10D	0.30247	-1.45516	1.45516	0.87
PTBG10E	0.45983	-1.54534	1.54534	0.86
PTBG10F	-0.15541	-1.45758	1.45758	1.03
PTBG10G	0.22719	-1.16800	1.16800	1.47

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Scale Transformation Constants for the TIMSS Advanced 2015 Teacher Job Satisfaction Scale, Physics

Scale Transformation Constants	
A = 9.030432	Transformed Scale Score = 9.030432 + 0.897591 • Logit Scale Score
B = 0.897591	

Equivalence Table of the Raw Score and the Transformed Scale Score for the TIMSS Advanced 2015 Teacher Job Satisfaction Scale, Physics

Raw Score	Transformed Scale Score	Cutpoint
0	5.13977	
1	6.27326	
2	6.89329	
3	7.37441	7.4
4	7.80316	
5	8.21352	
6	8.62591	
7	9.04442	
8	9.46026	
9	9.86512	
10	10.26699	
11	10.68423	10.6
12	11.15621	
13	11.76054	
14	12.88244	

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS Advanced 2015 Teacher Job Satisfaction Scale, Physics

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			PTBG10A	PTBG10B	PTBG10C	PTBG10D	PTBG10E	PTBG10F	PTBG10G
France	0.91	66	0.83	0.70	0.85	0.84	0.86	0.84	0.74
Italy	0.90	63	0.84	0.70	0.72	0.89	0.86	0.84	0.67
Lebanon	0.91	65	0.77	0.74	0.81	0.85	0.81	0.83	0.82
Norway	0.90	63	0.83	0.76	0.80	0.84	0.90	0.75	0.66
Portugal	0.84	52	0.80	0.70	0.77	0.78	0.71	0.57	0.68
Russian Federation	0.90	64	0.84	0.81	0.77	0.83	0.85	0.81	0.68
Slovenia	0.89	61	0.79	0.62	0.82	0.84	0.91	0.81	0.63
Sweden	0.88	59	0.80	0.68	0.72	0.85	0.84	0.79	0.67
United States	0.92	69	0.84	0.77	0.83	0.88	0.87	0.84	0.77
International Avg.	0.89	62	0.81	0.72	0.79	0.85	0.85	0.79	0.70

Relationship Between the TIMSS Advanced 2015 Teacher Job Satisfaction Scale, and TIMSS Advanced 2015 Physics Achievement

Country	Pearson's Correlation with Physics Achievement		Variance in Physics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
France	0.02	0.00	0.00
Italy	-0.09	0.01	0.02
Lebanon	-0.02	0.00	0.01
Norway	0.01	0.00	0.00
Portugal	0.02	0.00	0.01
Russian Federation	0.07	0.00	0.00
Slovenia	0.20	0.04	0.06
Sweden	0.03	0.00	0.00
United States	0.11	0.01	0.03
International Median	0.02	0.00	0.01

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

IEA

International Association
for the Evaluation of
Educational Achievement

BOSTON
COLLEGE

timss.bc.edu

ISBN: 978-1-889938-34-9