

CHAPTER 10

Reviewing the TIMSS 2019 Achievement Item Statistics

Bethany Fishbein
Pierre Foy
Lillian Tyack

Overview

Prior to scaling and reporting the results for an assessment, the TIMSS & PIRLS International Study Center reviews key diagnostic statistics for each achievement item to evaluate its psychometric characteristics across the participating countries. This item-by-item, country-by-country review to detect unusual item properties or anomalous patterns plays a crucial role in the quality assurance of the achievement data. Finding a faulty item this late in the process is rare, but an uncharacteristically difficult item, or one with unusually low discriminating power, could indicate a potential problem with either translation or printing. Similarly, a human-scored constructed response item with low scoring reliability could indicate a problem in the translation of the scoring guide for a particular country. If such an item is found, the country's translation verification documents, printed booklets, and digital item archives can be examined for flaws or inaccuracies and, if necessary, the item can be removed from the international database for that country.

The TIMSS 2019 assessment cycle marked the beginning of the transition to eTIMSS—a digital version of the TIMSS assessment that was administered to students on computers and tablets. eTIMSS also included a novel section consisting of problem solving and inquiry tasks (PSIs), which were designed to capitalize on the digital environment to its fullest. About half the participating countries chose to transition to eTIMSS. eTIMSS trend countries also administered the paper version of their trend items to a sample of schools, providing a “bridge” that helped link the two test-taking modes. For TIMSS 2019, comparing the item statistics for eTIMSS and paperTIMSS was integral in identifying items that were psychometrically invariant (equivalent) under the IRT scaling.

The TIMSS 2019 Item Review

The TIMSS & PIRLS International Study Center computed item statistics for all achievement items in the 2019 fourth and eighth grade assessments, including both eTIMSS and paperTIMSS versions, including the eTIMSS PSIs and the paper “bridge” booklets administered in eTIMSS trend countries. TIMSS

fourth grade included 175 mathematics items and 175 science items in both paper and digital formats, 131 less difficult mathematics items in paper only, and 39 mathematics items from three PSIs and 19 science items from two PSIs.¹ The fourth grade paper bridge booklets consisted of only trend items—92 mathematics items and 111 science items. TIMSS eighth grade included 211 mathematics items and 220 science items in both paper and digital formats, plus 25 mathematics items from three PSIs and 29 science items from two PSIs. The bridge booklets consisted of 117 mathematics and 122 science trend items at the eighth grade.

In addition to evaluating the performance of each item, the TIMSS & PIRLS International Study Center looked for any changes in the measurement properties of trend items from the 2015 assessment, and examined differences between items common to eTIMSS and the paper bridge booklets for mode effects. Item position effects were evaluated to ensure student performance remained steady throughout the assessment. Finally, using the item statistics, extensive analyses of each country’s data were conducted to detect any anomalous patterns relative to previous cycles or the pool of participating countries on average.

Although reviewing item statistics took place over several months, staff at the TIMSS & PIRLS International Study Center met for four consecutive working days in March 2020 to conduct a formal adjudication of the achievement data in preparation for IRT scaling. During these four days, decisions were made about any modifications needed to the data or if further analyses were required. The review was conducted item-by-item simultaneously for eTIMSS and paperTIMSS. Both versions of an item and its scoring guide were displayed while staff reviewed the item statistics as well as accompanying graphical displays. Country reports about translation errors, printing issues, or other technical problems were also referenced. In addition, graphical displays of item statistics were reviewed to detect any anomalous and systematic patterns in a particular country’s data that may warrant further investigation.

Following item review, some National Research Coordinators from the participating countries and benchmarking entities were contacted to inquire about concerns or anomalies detected in the data. Decisions about item deletions or recodes were communicated to IEA Hamburg to make edits to the international data files.

Item Review Statistics

The item statistics for each of the TIMSS 2019 participating countries were computed and combined to produce “item almanacs” for eTIMSS and paperTIMSS respectively. Each item almanac page included unweighted statistics for all countries that administered the particular item. The paperTIMSS item almanacs included data from paperTIMSS countries, countries participating in the fourth grade less

¹ Two fourth grade mathematics items involving an on-screen ruler tool were only included in the eTIMSS assessment. Forty-eight fourth grade mathematics items were also included in the less difficult version of the assessment.

Exhibit 10.2: Example International Item Statistics for a TIMSS 2019 Constructed Response Item

Trends in International Mathematics and Science Study - eTIMSS 2019 Assessment Results - 8th Grade
International Item Review Statistics (Unweighted)

Science: Biology / Knowing (SE06_02 - SE62274) - 2 Points
Label: Raw materials for photosynthesis

Country	N	DIFF	DISC	P_0	Percentages					Point Biserials					RDIFF	Reliability			Flags	
					P_1	P_2	P_OM	P_NR	PB_0	PB_1	PB_2	PB_OM	PB_NR	N		Score	Code			
*Chile	588	38.8	0.45	17.2	59.1	9.2	14.5	0.2	0.2	-0.19	0.26	0.24	-0.36	-0.02	0.11	197	97.5	97.5	EAF	B
*Chinese Taipei	700	35.7	0.59	39.6	39.7	15.9	4.9	0.0	0.0	-0.45	0.24	0.42	-0.23	.	1.21	195	94.4	94.4	H	B
*England	492	37.1	0.40	25.8	51.0	11.6	11.6	0.0	0.0	-0.11	0.13	0.30	-0.36	.	0.56	191	99.5	99.5	H	
Finland	677	50.6	0.43	14.5	61.8	19.7	4.0	0.3	0.3	-0.20	-0.06	0.37	-0.25	-0.07	0.05	200	99.5	99.5	E	
France	557	28.1	0.35	25.4	48.9	3.6	22.1	0.2	0.2	-0.05	0.31	0.10	-0.37	-0.10	1.02	200	99.5	99.5	HAF	
Georgia	472	34.3	0.48	17.2	51.6	8.5	22.7	0.2	0.2	-0.12	0.32	0.25	-0.43	-0.01	0.03	183	98.9	98.9	EAF	
*Hong Kong SAR	469	39.5	0.57	30.3	44.4	17.3	7.9	0.2	0.2	-0.36	0.12	0.46	-0.25	0.00	0.12	195	100.0	100.0	E	
*Hungary	652	38.7	0.46	24.5	53.5	12.0	10.0	0.0	0.0	-0.21	0.18	0.32	-0.34	.	0.66	199	94.5	94.5	H	B
*Israel	533	38.6	0.54	21.2	54.8	11.3	12.8	0.0	0.0	-0.22	0.24	0.35	-0.42	.	0.50	196	97.4	97.4	H	
*Italy	529	38.1	0.24	20.6	64.8	5.7	8.9	0.0	0.0	-0.03	0.17	0.11	-0.32	.	0.60	200	100.0	100.0	HAF	
*Korea, Rep. of	553	39.2	0.50	27.1	55.7	11.4	5.8	0.0	0.0	-0.32	0.24	0.31	-0.31	.	0.92	200	97.0	97.0	H	
*Lithuania	546	46.4	0.53	14.7	61.4	15.8	8.2	0.0	0.0	-0.19	-0.02	0.46	-0.33	.	0.10	200	98.0	98.0	E	
*Malaysia	1018	38.1	0.50	30.9	58.5	8.8	1.7	0.0	0.0	-0.43	0.27	0.29	-0.14	.	0.31	200	99.0	99.0	F	
Norway (9)	627	43.9	0.48	13.4	62.6	12.6	11.4	0.3	0.3	-0.21	0.18	0.31	-0.37	-0.08	0.02	198	96.0	96.0	E	
Portugal	480	40.7	0.37	19.4	61.9	9.8	9.0	0.0	0.0	-0.09	0.15	0.24	-0.39	.	0.41	193	99.0	99.0	F	
Qatar	546	37.8	0.50	27.9	52.8	11.4	7.9	0.2	0.2	-0.28	0.20	0.35	-0.31	-0.02	0.14	200	97.5	97.5	E	
*Russian Federation	555	47.5	0.41	17.7	58.5	18.2	5.6	0.2	0.2	-0.20	0.06	0.30	-0.31	0.00	0.25	200	98.0	98.0	H	
*Singapore	689	54.5	0.54	14.7	59.0	25.0	1.3	0.1	0.1	-0.39	-0.04	0.42	-0.21	-0.03	0.41	200	100.0	100.0	H	
*Sweden	548	53.0	0.50	14.5	46.3	29.9	9.3	0.4	0.4	-0.21	-0.06	0.43	-0.33	0.06	-0.20	198	99.0	99.0	E	
*Turkey	582	41.8	0.58	25.8	46.2	18.7	9.3	0.0	0.0	-0.30	0.08	0.48	-0.33	.	0.21	197	98.5	98.5	E	
United Arab Emirates	3185	36.4	0.57	32.5	45.5	13.6	8.3	0.0	0.0	-0.39	0.26	0.38	-0.27	0.02	0.18	199	97.5	97.5	E	
*United States	1249	38.8	0.45	26.9	61.8	7.9	3.4	0.2	0.2	-0.31	0.22	0.28	-0.25	-0.03	0.69	198	99.5	99.5	H	F
*Reference Avg (15)	9703	41.7	0.48	23.4	54.3	14.6	7.7	0.1	0.1	-0.26	0.14	0.35	-0.30	-0.00	0.43	198	98.1	98.1	B	
International Avg (22)	116247	40.8	0.47	22.8	54.5	13.5	9.1	0.1	0.1	-0.24	0.16	0.33	-0.31	-0.02	0.38	197	98.2	98.2	B	
Ontario, Canada	535	37.1	0.35	22.6	64.9	4.7	7.9	0.0	0.0	-0.16	0.25	0.16	-0.33	.	0.79	117	100.0	100.0	HAF	
Quebec, Canada	455	42.2	0.42	22.7	61.5	11.5	4.4	0.2	0.2	-0.30	0.19	0.25	-0.24	-0.03	0.56	80	96.3	96.3	H	
Moscow City, Russian Fed	536	45.9	0.37	16.8	63.8	14.0	5.4	0.0	0.0	-0.22	0.17	0.20	-0.30	.	0.75	200	99.5	99.5	H	
Abu Dhabi, UAE	1166	27.6	0.65	43.7	35.2	10.0	11.1	0.0	0.0	-0.44	0.36	0.43	-0.27	.	0.26	70	100.0	100.0	E	
Dubai, UAE	817	46.9	0.48	21.4	53.9	20.0	4.7	0.1	0.1	-0.33	0.12	0.33	-0.28	0.02	0.25	67	98.5	98.5	E	

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= Ability not ordered/Attractive distractor; B= Boys outperform girls; C= Difficulty less than chance; D= Negative/low discrimination; E= Easier than average; F= Score obtained by less than 10%; G= Girls outperform boys; H= Harder than average; R= Scoring reliability less than 85%; V= Difficulty greater than 95%.

For all items, regardless of format (i.e., selected response or constructed response) or administration mode, 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).² 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—and a reference average—based on a pool of countries that have participated regularly in the TIMSS assessments—for each statistic. The reference countries are shown with an asterisk next to their names. The international and reference averages of the item difficulties and item discriminations served as guides to the overall statistical properties of the items. The item review outputs also listed the benchmarking participants.

Statistics displayed for selected response 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

² For computing point-biserial correlations, the total score is the percentage of points a student has scored on the items they were administered. In the context of TIMSS, a separate total score is computed for mathematics and for science. Not-reached responses are not included in the total score.

agreement between the scorers, and—because TIMSS has a 2-digit scoring scheme for diagnostic coding—the percentage of code agreement between scorers. Diagnostic codes between 20 and 29 are worth 2 score points, between 10 and 19 worth 1 score points, and between 70 and 79 worth 0 score points.

During item review, “not reached” responses (i.e., items toward the end of the booklet that the student did not attempt)³ 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 grade and subject, and are listed in order of their appearance in the item review outputs:

- **N:** 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 selected response 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, and the other 25 percent scored 0 points, then the average percent correct would be 50 percent. For this statistic, not-reached responses were not included.
- **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.
- **Percentages (P_A, P_B, P_C, P_D, etc.):** Available for selected-response items. Each column indicates the percentage of students choosing the particular response option for the item (e.g., A, B, C, D, etc.). Not-reached responses were excluded from the denominator.
- **Percentages (P_0, P_1, and P_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.
- **Percentages (P_OM):** Percentage of students who, having reached the item, did not provide a response. Not reached responses were excluded from the denominator.
- **Percentages (P_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.

³ An item was considered “not reached” if the item itself and the item immediately preceding it were not answered and no subsequent items were answered. The decision as to whether an item was not reached was made separately for part 1 and part 2 of each assessment booklet.

- **Point Biserials (PB_A, PB_B, PB_C, PB_D, etc.):** Available for selected-response items. These columns show the point-biserial correlations between choosing each of the response options (e.g., A, B, C, D, etc.) 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.
- **Point Biserials (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.
- **Point Biserials (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.
- **Point Biserials (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):** Available for human-scored constructed response items. To provide a measure of the reliability of the scoring of the constructed response items, approximately 200 responses per item 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):** Available for human-scored constructed response items. This column contains the percentage of agreement on the score point value (0, 1, or 2) of the two-digit diagnostic codes assigned by the two independent TIMSS scorers.
- **Reliability (Code):** Available for human-scored constructed response items. This column contains the percentage of full agreement on the two-digit diagnostic codes assigned by the two independent TIMSS 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 flags rarely indicate and

actual problem, but serve to draw attention to potential sources of concern. The following conditions were flagged:

- The item discrimination (DISC) was less than 0.10 (flag D)
- The item difficulty (DIFF) was less than 0.25 for selected response items (flag C)
- The item difficulty (DIFF) exceeded 0.95 (flag V)
- The Rasch difficulty estimate (RDIFF) for a given country showed 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 selected response 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 selected response 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).

Scoring Reliability for Human Scored Items

Constructed-response items comprised about half the assessment score points in TIMSS 2019, with most of them requiring human scoring, especially for paper TIMSS and the paper bridge. To ensure that the items requiring human scoring were scored reliably in all countries, the TIMSS & PIRLS International Study Center developed detailed scoring guides for each constructed response item (that provided descriptions and examples of acceptable responses for each score point value) and provided extensive training in the application of the scoring guides. See [Chapter 1](#) for more information on developing the scoring guides and see [Chapter 6](#) for information on the human-scoring process.

For eTIMSS countries, the new mode of administration allowed for a substantial portion of the digital items to be machine scored, particularly in mathematics. For eTIMSS items suitable for machine scoring, the scoring guides served as the basis for developing machine scoring specifications for student responses that could be accurately applied without human judgment. Developing the machine scoring specifications involved testing each item in the eTIMSS Player, reviewing the output, and writing rules in terms of the output to classify all possible responses to a code in the item's scoring guide. The scoring unit at IEA Hamburg reviewed all specifications and provided feedback on an item-by-item basis, resulting in several rounds of revision until the rules for all items were clarified. The scoring unit at IEA Hamburg then applied the scoring rules for all machine-scored items and the data analysis team at the TIMSS & PIRLS International Study Center independently replicated the results to validate the scoring.

The following sections describe the three ways human-scoring reliability was assessed and documented in 2019: within-country, over-time (trend), and across countries.

Within-Country Scoring Reliability

To gather and document information about the within-country agreement among scorers for TIMSS 2019, a random sample of approximately 200 student responses per item were scored independently by two scorers. The inter-scorer agreement for each item in each country was examined as part of the item review process, with agreement below 75 percent giving cause for deleting the data for a particular country. Appendix 10A shows the average and range of the within-country percentages of score point agreement and diagnostic code agreement across all items. Exact percent agreement across items was high on average across countries—98 percent in mathematics and 95 percent in science across both the fourth and eighth grade countries. In TIMSS 2019 there also was high agreement at the diagnostic score level, where international average percent agreement ranged from 94 percent in eighth grade science to 97 percent in mathematics at both grades.

Trend Item Scoring Reliability Study

The TIMSS & PIRLS International Study Center also took steps to show that the 2019 human-scored constructed response items used in TIMSS 2015 were scored in the same way in both assessments. In anticipation of this, countries that participated in TIMSS 2015 sent samples of scored student booklets from the 2015 data collections to IEA Hamburg, 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 2019 data were asked also to score these 2015 responses using the CodingExpert Software developed by IEA Hamburg. Each country scored 200 responses for each of 11 mathematics items (13 items for countries that administered less difficult mathematics) and 11 science items at the fourth grade, and for 14 mathematics items and 13 science items at the eighth grade. The average and range of scoring consistency over time can be found in Appendix 10B.

There was a very high degree of scoring consistency in TIMSS 2019. The exact agreement between the scores awarded in 2015 and those given by the 2019 scorers ranged from 91 percent in science at both grades to 97 percent in mathematics at the fourth grade, on average internationally. There was similarly high agreement in TIMSS at the diagnostic code level.

Cross-Country Scoring Reliability Study

It also was important to document the consistency of scoring across countries. Because of the many different languages in use in TIMSS 2019, establishing the reliability of constructed response scoring across all countries was not feasible. However, the TIMSS & PIRLS International Study Center did conduct

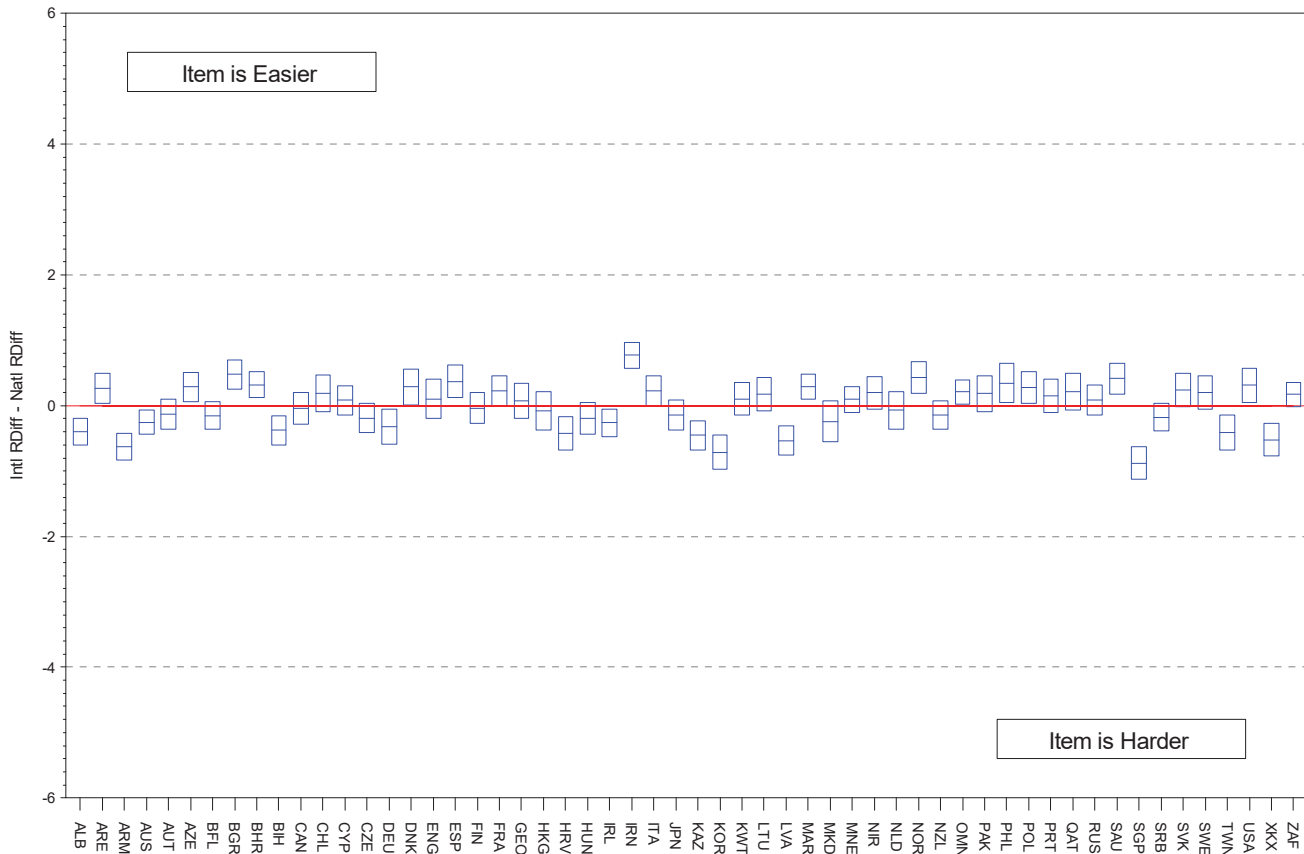
a cross-country study of scoring reliability among Northern Hemisphere countries that had scorers who were proficient in English. A sample of student responses was provided by the English-speaking Southern Hemisphere countries. Cross-country scoring included 200 student responses for each of 11 mathematics items (6 items for countries that administered less difficult mathematics) and 11 science items at the fourth grade, and for 14 mathematics items and 13 science items at the eighth grade. This same set of student responses in English was then scored independently in each country that had at least two scorers proficient in English, using IEA's CodingExpert Software. In all, scorers from 54 countries and 1 benchmarking participant at fourth grade and 35 countries and 2 benchmarking participants at the eighth grade participated in the study. Scoring for this study took place shortly after the other scoring reliability activities were completed. Making all possible comparisons among scorers gave 1,485 comparisons at fourth grade and 666 comparisons at eighth grade for each student response to each item. This resulted in more than 290,000 total comparisons at the fourth grade (200,000 for mathematics items not administered to less difficult countries) and more than 120,000 total comparisons at the eighth grade when aggregated across all 200 student responses to that item. Agreement across countries was defined in terms of the percentage of these comparisons that were in exact agreement.

Appendix 10C reports the results of the cross-country scoring reliability study. On average internationally, scorer reliability across countries in TIMSS 2019 was high. The exact agreement between the scores awarded across countries ranged from 89 percent in science to 96 percent in mathematics at the fourth grade and from 89 percent in science to 94 percent in mathematics at the eighth grade, on average internationally. There was similarly high agreement at the diagnostic code level.

Item-by-Country Interactions

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 10.3.

Exhibit 10.3: Example Plot of Item-by-Country Interaction for a TIMSS 2019 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:

$$\text{Upper Limit} = RDIFF_i - RDIFF_{ik} + SE(RDIFF_{ik}) \cdot Z_b \quad (10.1)$$

$$\text{Lower Limit} = RDIFF_i - RDIFF_{ik} - SE(RDIFF_{ik}) \cdot Z_b \quad (10.2)$$

where $RDIFF_{ik}$ 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 percent critical value from the Z distribution corrected for multiple comparisons using the Bonferroni procedure.

Review of Item Statistics for Measuring Trends

To measure trends, TIMSS assessments include achievement items from previous assessments as well as items developed for use for the first time in 2019. Accordingly, the TIMSS 2019 assessments included items developed in 2011, 2015, and 2019. Therefore, an important review step included checking that these “trend items” had statistical properties in 2019 similar to those they had in the previous assessments (e.g., a TIMSS item that was relatively easy in 2015 should still be relatively easy in 2019).

As shown in the example in Exhibit 10.4, the trend item review focused on statistics for paper trend items from the current and previous assessments (2019 and 2015) for countries that participated in both. This included statistics for the eTIMSS bridge samples. For each country, trend item statistics included the percentage of students in each score category (or response option for selected response 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 10.4: Example Item Statistics in 2019 and 2015 for a TIMSS 2019 Trend Item

Trends in International Mathematics and Science Study - TIMSS 2019 Assessment Results - 8th Grade
Trend Achievement Data Almanac for Mathematics Items (Weighted)

MP06_08 (MP62105): Algebra / Reasoning - 2 Points
Label: Area of rectangle with sides x and $2x + 1$

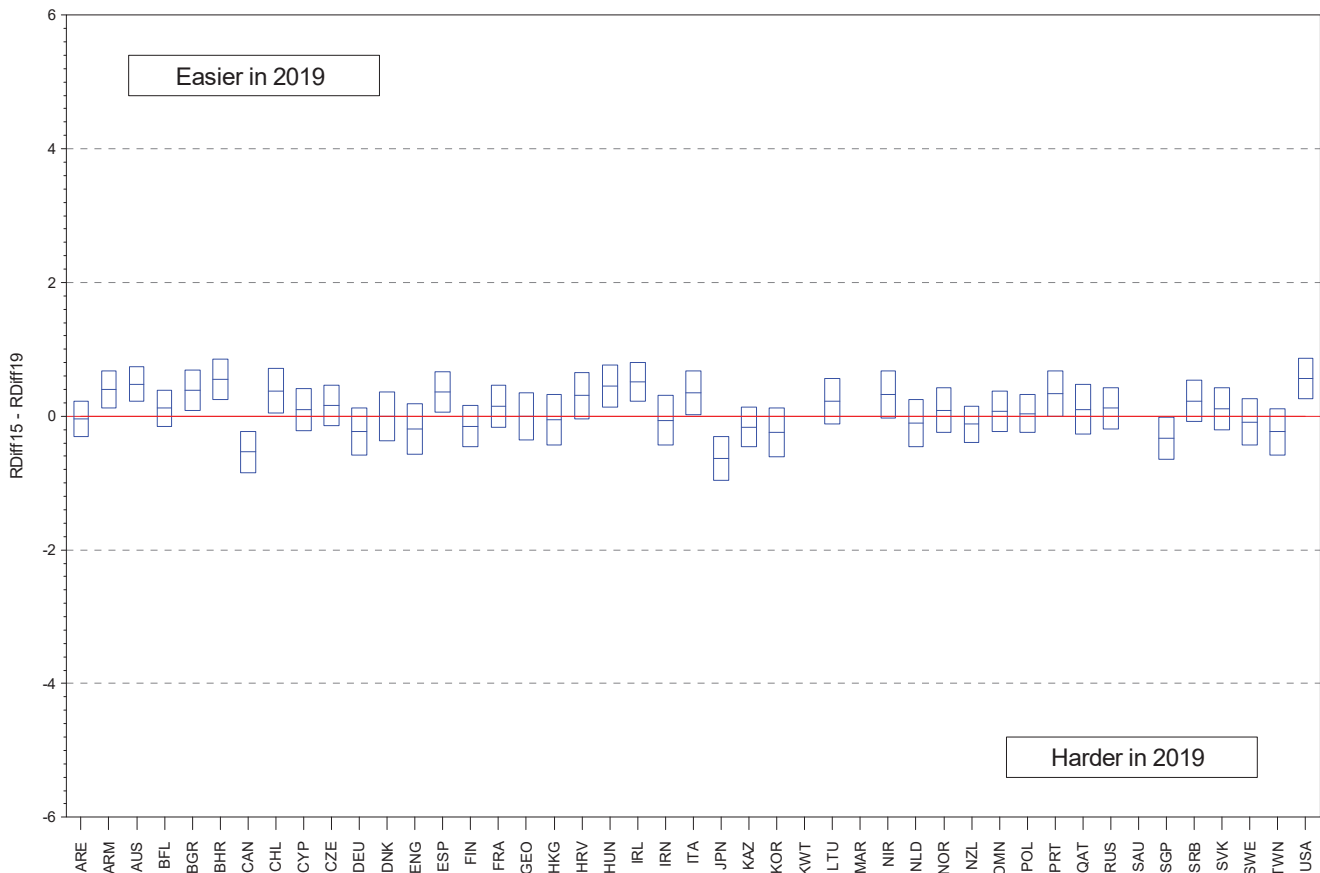
COUNTRY	Year	N	20 %	10 %	79 %	OMITTED %	NOT REACHED %	V1 %	V2 %	GIRL PCT RIGHT	BOY PCT RIGHT
Australia	2015	1476	23.5	1.4	60.1	14.6	0.5	24.8	23.5	23.9	23.0
	2019	1282	26.9	2.3	58.6	11.7	0.5	29.1	26.9	25.0	28.7
Bahrain	2015	710	8.7	5.8	72.2	13.0	0.3	14.5	8.7	8.4	8.9
	2019	814	26.0	3.0	44.2	26.8	0.0	29.0	26.0	22.9	28.8
Egypt	2015	1125	8.9	1.2	76.8	13.0	0.1	10.1	8.9	10.4	7.1
	2019	1036	12.9	2.9	76.3	7.7	0.2	15.8	12.9	12.7	13.1
Iran, Islamic Rep. of	2015	866	13.1	1.9	63.7	20.1	1.2	15.0	13.1	13.7	12.6
	2019	855	14.8	1.8	63.5	18.0	1.9	16.6	14.8	13.9	15.5
Ireland	2015	663	22.7	1.3	62.7	12.9	0.4	24.0	22.7	23.1	22.4
	2019	597	20.8	2.6	61.6	14.4	0.6	23.4	20.8	21.7	20.0
Japan	2015	672	43.4	1.8	46.6	8.1	0.1	45.2	43.4	42.5	44.3
	2019	639	41.4	0.8	49.7	8.1	0.0	42.2	41.4	35.2	47.5
Jordan	2015	1131	3.5	4.0	80.9	10.8	0.8	7.5	3.5	4.2	2.8
	2019	1010	7.7	5.2	78.9	7.5	0.7	12.9	7.7	8.2	7.3
Kuwait	2015	644	5.7	7.0	66.8	19.2	1.4	12.7	5.7	2.8	8.1
	2019	648	5.2	5.7	80.2	8.4	0.5	10.9	5.2	5.5	4.8
Lebanon	2015	547	4.6	4.2	67.6	22.4	1.1	8.8	4.6	5.4	3.5
	2019	671	2.7	3.3	69.9	22.2	1.8	6.0	2.7	2.7	2.7
Morocco	2015	1881	3.7	1.9	72.8	20.8	0.8	5.6	3.7	4.0	3.4
	2019	1213	5.1	0.3	72.0	21.9	0.7	5.4	5.1	3.0	7.4
New Zealand	2015	1122	19.3	1.1	64.0	15.0	0.7	20.3	19.3	19.7	18.8
	2019	862	17.6	2.7	65.4	13.2	1.1	20.3	17.6	20.1	15.4
Oman	2015	1267	6.3	6.5	79.6	6.4	1.3	12.7	6.3	6.2	6.4
	2019	952	7.1	4.3	80.2	7.0	1.3	11.4	7.1	9.5	4.8
Saudi Arabia	2015	538	0.7	4.9	81.7	11.8	0.9	5.6	0.7	0.4	1.1
	2019	814	5.7	4.9	83.0	5.8	0.6	10.6	5.7	7.1	4.3
South Africa (9)	2015	1788	3.3	5.9	85.3	4.7	0.8	9.3	3.3	4.4	2.2
	2019	2964	3.8	4.4	86.2	4.9	0.7	8.2	3.8	3.5	4.0
International Avg (n=14)	2015	14430	11.9	3.5	70.0	13.8	0.7	15.4	11.9	12.1	11.7
	2019	14357	14.1	3.2	69.3	12.7	0.8	17.3	14.1	13.6	14.6

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.

The TIMSS & PIRLS International Study Center used two different graphical displays to examine the differences in countries' Rasch item difficulties between 2019 and 2015. 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.

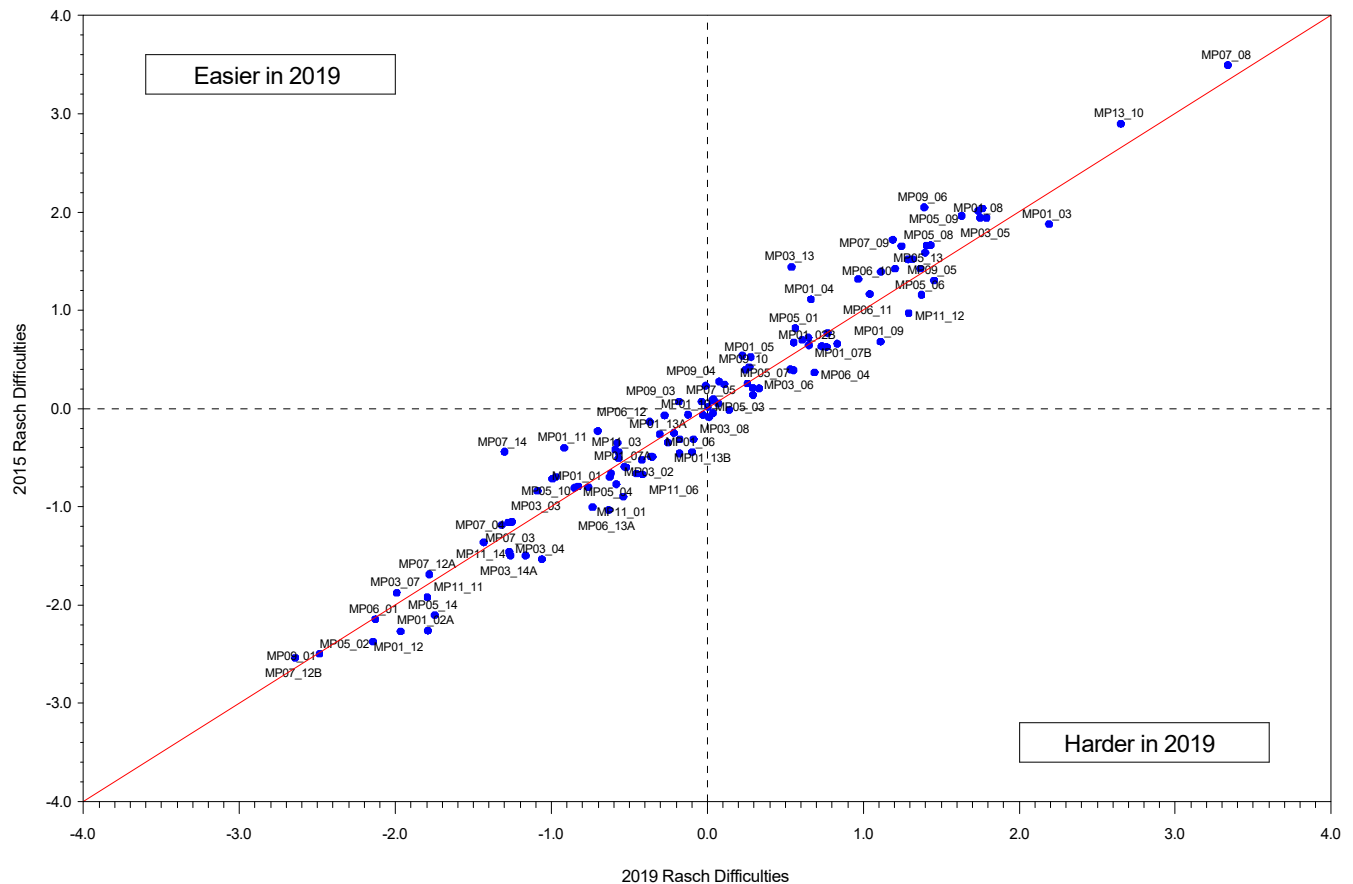
The first of these displays, shown for an example item in Exhibit 10.5, displays the difference in Rasch item difficulty of the item between 2019 and 2015 for each country. The difference in Rasch item difficulty for each country is displayed as a confidence interval, calculated using equations (10.1) and (10.2) but using each country's 2019 and 2015 Rasch difficulties and the standard error of their difference. A positive difference for a country indicates that the item was relatively easier in 2019, and a negative difference indicates that the item was relatively more difficult.

Exhibit 10.5: Example Plot of Differences in Rasch Item Difficulties Between 2019 and 2015 for a TIMSS 2019 Trend Item



The second graphical display, presented in Exhibit 10.6, shows the performance of a given country on all trend items simultaneously. For each country, the graph plots the 2019 Rasch difficulty of every trend item against its Rasch difficulty in 2015. When there were no differences between the difficulties in the two successive administrations, the data points aligned on or near the diagonal. Large deviations from the diagonal were noted for further investigation.

Exhibit 10.6: Example Plot of Rasch Trend Item Difficulties Across TIMSS 2019 and 2015 by Country



Item Position Effects

As described in the [TIMSS 2019 Assessment Design](#), assessment items for each grade and subject are arranged in 14 groups called “item blocks” (for paperTIMSS) or “item block combinations” (for eTIMSS), which were assembled into achievement booklets. Each item block appears in two booklets, with each item block appearing in the first half of one booklet and the second half of another. This counterbalancing helps to control for the impact of item position on the item statistics.

To examine the magnitude of item position effects on item percent correct and the percent of omitted and not-reached responses, block-level item statistics weighted by maximum score points were computed for each of the two positions that each block appears in the booklet design—either position 1 and position 4 or position 2 and position 3. The results are reported in Appendix 10D for each assessment averaged across countries, as well as for each country across item blocks. A summary of results with the average differences in item statistics between the booklet positions is provided in Exhibits 10.7 and 10.8 for the fourth and eighth grade, respectively.

Exhibit 10.7: Summary of International Average Item Block Statistics by Booklet Position (Weighted)—Grade 4

	Average Percent Correct Across Items			Average Percent Omitted Responses Across Items			Average Percent Not Reached Across Items		
	Positions 1&3	Positions 2&4	Difference	Positions 1&3	Positions 2&4	Difference	Positions 1&3	Positions 2&4	Difference
Mathematics									
eTIMSS	49.9	48.3	-1.6	3.3	4.5	1.1	0.1	2.5	2.5
paperTIMSS	49.4	48.1	-1.3	5.8	6.3	0.5	0.2	3.3	3.1
Less Difficult	43.4	42.2	-1.2	7.9	8.3	0.4	0.3	3.7	3.4
Science									
eTIMSS	53.3	52.0	-1.3	3.2	4.4	1.2	0.1	2.5	2.5
paperTIMSS	50.9	49.5	-1.3	5.7	6.7	1.0	0.1	2.5	2.4
Less Difficult	33.9	32.6	-1.3	10.9	12.2	1.3	0.3	4.9	4.6

Exhibit 10.8: Summary of International Average Item Block Statistics by Booklet Position (Weighted)—Grade 8

	Average Percent Correct Across Items			Average Percent Omitted Responses Across Items			Average Percent Not Reached Across Items		
	Positions 1&3	Positions 2&4	Difference	Positions 1&3	Positions 2&4	Difference	Positions 1&3	Positions 2&4	Difference
Mathematics									
eTIMSS	44.0	41.9	-2.1	5.5	7.0	1.5	0.1	1.8	1.7
paperTIMSS	34.9	33.2	-1.6	7.9	8.5	0.7	0.1	2.0	1.9
Science									
eTIMSS	47.8	46.3	-1.5	4.0	5.0	1.1	0.1	0.8	0.7
paperTIMSS	39.6	38.1	-1.5	6.7	7.5	0.8	0.1	1.6	1.6

The results indicate minimal impact of block position on the TIMSS 2019 item statistics. On average, item blocks appearing in the second half of a booklet part (positions 2 and 4) were slightly more difficult and had slightly more missing responses than item blocks appearing the first half of a booklet part (positions 1 and 3). However, the differences were small. Across countries, differences in average item percent correct between positions 2/4 and positions 1/3 ranged from -1.2 at the fourth grade for the less difficult mathematics items to -2.1 in eighth grade eTIMSS mathematics. Differences in average percent omitted ranged between 0.4 in fourth grade less difficult mathematics to 1.5 in eighth grade eTIMSS mathematics; and differences in average percent not reached ranged between 0.7 in eighth grade eTIMSS science and 4.2 in fourth grade science for less difficult countries.

Detecting Anomalies in the TIMSS 2019 Achievement Data

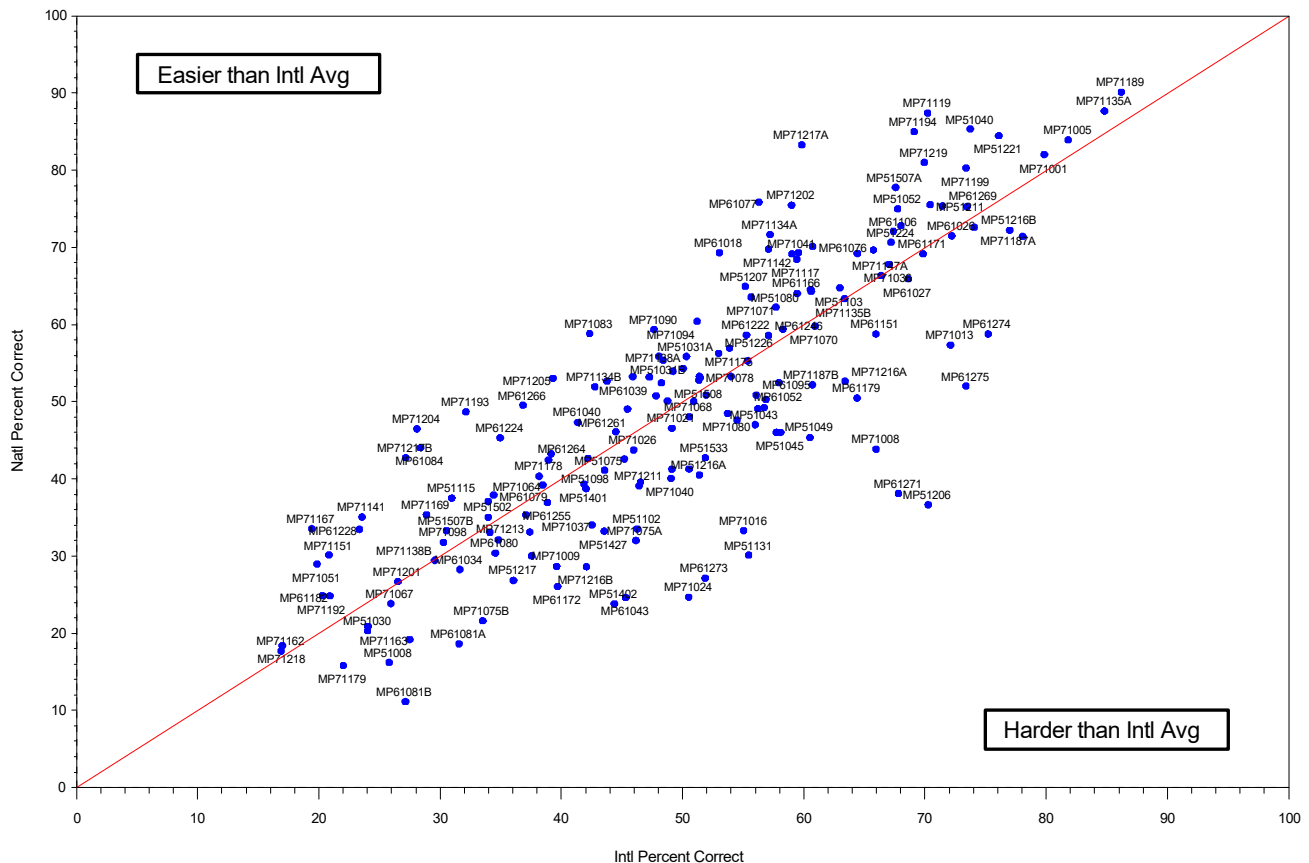
To ensure that each participating country and benchmarking entity had data adhering to TIMSS' quality standards, the TIMSS & PIRLS International Study Center conducted analyses of item statistics at the country level. Several graphical displays were produced for each TIMSS participant for item percent correct, item point-biserials, and percent omitted responses. The graphs were analyzed to detect any anomalous patterns in any particular country's data relative to the international average or to their previous TIMSS performance. Anomalous patterns may be indicative that systematic errors occur in a country's data, which may be due to errors in collecting and processing the data. For any anomalous patterns detected in the item statistics for a particular country, the National Research Coordinator was contacted to discuss how best to address any issues.

The first set of graphical displays compared each country's item performance to the international average for all items simultaneously, where item performance is defined in terms of item percent correct, item discrimination (point-biserial correlation), and item percent omitted. An example is shown in Exhibit 10.9 for item percent correct. For each country, the graph plots the 2019 item percent correct of all items against the 2019 international average. Typical patterns show data points along the range of the x - and y axis, with random deviations from the diagonal. There may be more points above the diagonal for higher performing countries and more points below for lower performing countries, but otherwise the points should align closely with the diagonal. The best-fit line should be approximately linear and parallel with the diagonal. Any patterns largely deviating from this were noted for further investigation. Plots comparing national and international item discrimination (point-biserial correlation) and percent omitted should have similar patterns, but points more tightly clustered together since there is a smaller range.

These plots of national versus international item statistics were also compared against the same plots produced in TIMSS 2015. If the patterns for both assessments were unusually different, it may indicate a problem in the 2019 data. The plots may also be examined separately for selected response and

constructed response formatted items to ensure similar patterns. The relationship between national and international statistics for both item types should also match that from TIMSS 2015.

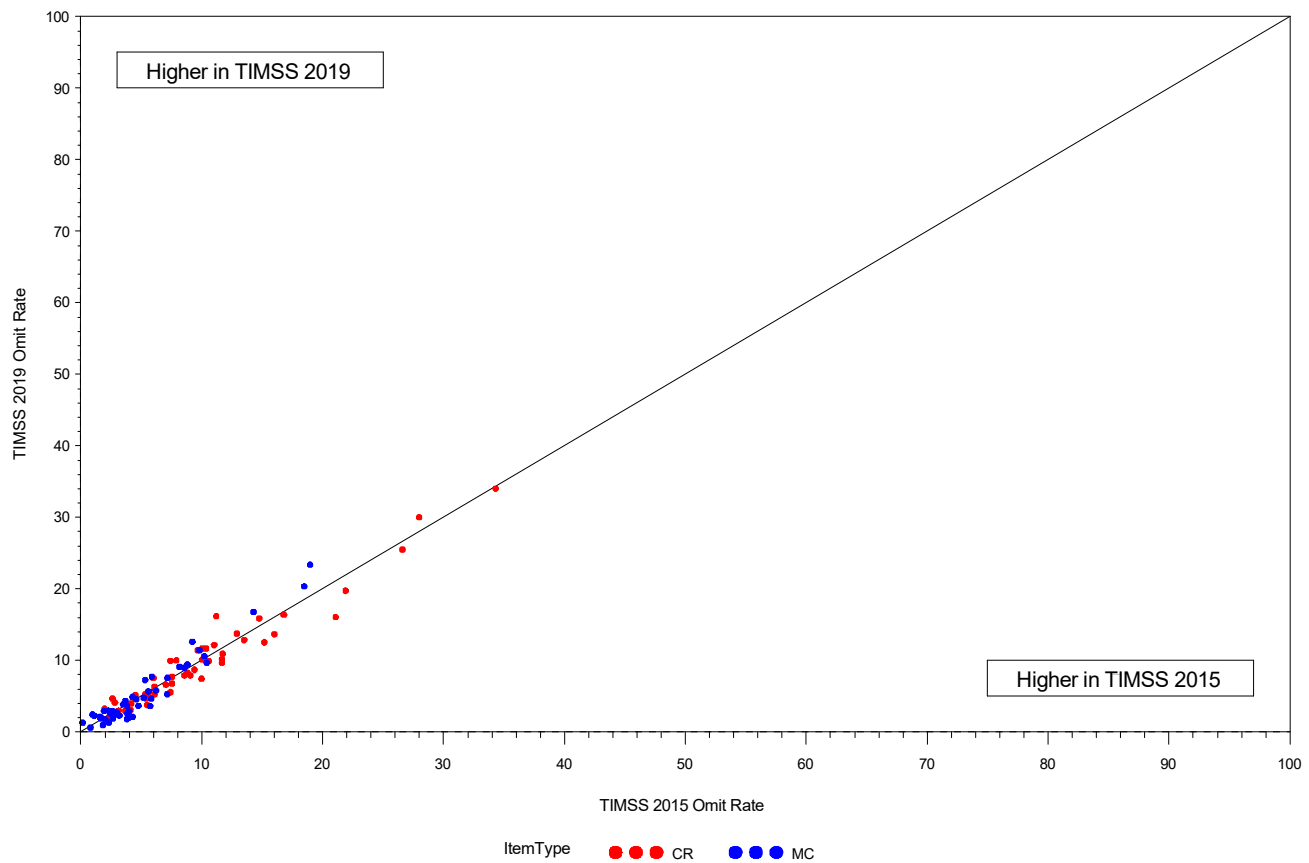
Exhibit 10.9: Example Plot of Item Percent Correct Across National and International by Country



The second set of graphical displays compares each country's TIMSS 2019 trend item performance with their TIMSS 2015 item performance for all items simultaneously, where item performance is defined in terms of percent correct, item discrimination (point-biserial correlation), and item percent omitted. An example is shown in Exhibit 10.10 for item percent omitted, displaying a typical pattern. For each country, the graphs plots the 2019 item percent omitted of every trend item against its item percent omitted in 2015, with points colored according to item type. When there were no differences between the difficulties in the two successive administrations, the data points aligned on or near the diagonal from the graph origin. While some changes were anticipated, as countries' overall achievement may have improved or declined, unusually large deviations from the diagonal were noted for further investigation. For all statistics plotted, comparisons should show similar patterns for both selected response and constructed response item types, and any differences should not relate to the difficulty of the item.

An additional set of plots were produced comparing each country's TIMSS 2019 item performance with their item performance from the field test conducted one year earlier. These plots were similar to the example in Exhibit 10.10 below comparing 2019 and 2015 performance, with the expectation of smaller differences. Large differences in item performance compared to the field test would be considered a implausible change in performance, warranting further review.

Exhibit 10.10: Example Plot of Item Percent Omitted Across TIMSS 2019 and 2015 by Country



Item Review Outcomes

Using all the information from the comprehensive collection of item analyses and reliability data that were computed and summarized for TIMSS 2019, the TIMSS & PIRLS International Study Center thoroughly reviewed all item statistics for every participating country and benchmarking participant to ensure that the items were performing comparably across countries and modes. In particular, items with the following problems were considered for possible deletion from the international database:

- An error was detected for a particular country during translation verification but was not corrected before test administration
- Data checking revealed a selected response item with more or fewer options than in the international version for a particular country
- 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
- For selected response items, the item review revealed a faulty distracter influencing the item statistics for all countries
- 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 75 percent
- For trend items, an item performed substantially differently in 2019 compared to the TIMSS 2015 administration, or an item was not included in the previous assessment for a particular country
- For eTIMSS trend items, a substantially larger than average difference in item difficulty or percent omitted between eTIMSS and bridge 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 2019 achievement data involved review of more than 1,800 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. A small number of items were identified as having severe differential item functioning after item review during IRT scaling. Diagnostic score codes for some constructed

response items may be recoded if the point-biserials did not behave. Decisions about deleting items for all countries were most often implemented for both eTIMSS and paperTIMSS versions, with a few exceptions.

Appendix 10E includes a list of deleted items, as well as a list of recodes made to constructed response items. There also were a number of items in each study that were combined, or derived, for scoring purposes. See Appendix 10F for details about how score points were awarded for each derived item.

Review of Item Statistics Between eTIMSS and paperTIMSS

To establish a link between eTIMSS and paperTIMSS, eTIMSS countries that also participated in TIMSS 2015 administered paper booklets of trend items to randomly equivalent “bridge” samples of students. To strengthen the link, an important review step for TIMSS 2019 included checking the extent that items had similar statistical properties between eTIMSS and paperTIMSS (e.g., an item that was relatively easy on paper should also be easy in digital format).

The review focused on eTIMSS item percent correct statistics for trend items administered in digital format to the regular sample of students compared to those in paper format for the bridge samples (see Exhibit 10.11 for an example of this type of item almanac). For each eTIMSS country, mode difference item statistics included the percentage of students in each score category (or response option for selected response 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 identify items that were likely to be found invariant under IRT, as well as to detect any unusual differences in item difficulties between modes for a particular country that might indicate a problem. Further item equivalence analyses performed during scaling are described in [Chapter 12](#) of this volume.

Exhibit 10.11: Example Item Statistics in eTIMSS and paperTIMSS (Bridge) for a TIMSS 2019 Trend Item

Trends in International Mathematics and Science Study - TIMSS 2019 Assessment Results - 8th Grade
Mode Differences Data Almanac for Science Items (Weighted)

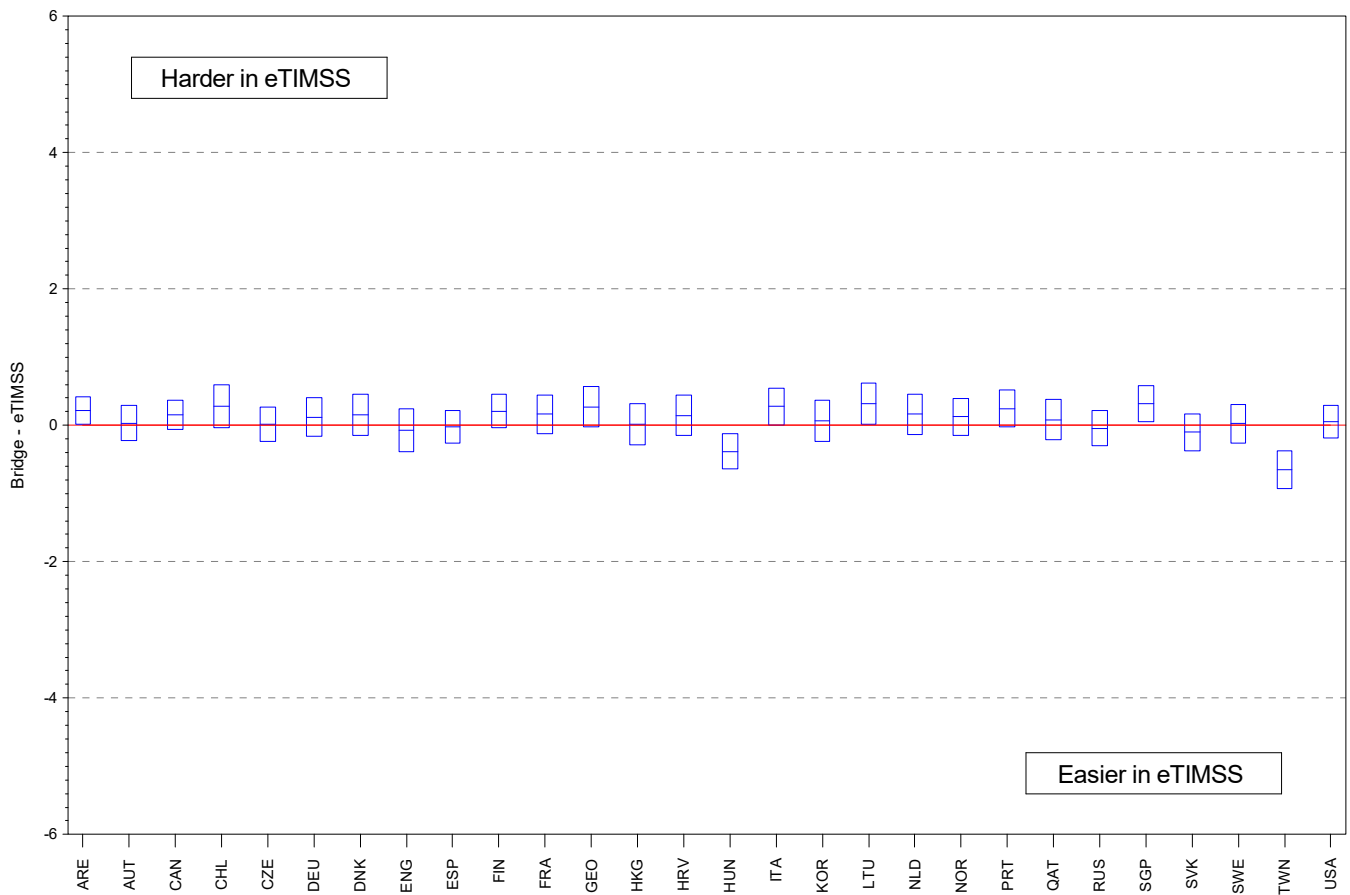
SE06_03 (SE62284): Biology / Applying - 1 Point - Key: B
Label: Hair color of young rabbits

COUNTRY	MODE	N	DIFF %	A %	B %	C %	D %	OMITTED %	NOT REACHED %	GIRL PCT RIGHT	BOY PCT RIGHT
Chile	eTIMSS	588	56.5	14.4	56.5	5.0	22.3	1.6	0.1	62.4	50.6
	pTIMSS	381	58.7	13.6	58.7	4.2	22.0	1.2	0.4	63.5	52.4
Chinese Taipei	eTIMSS	700	81.4	6.2	81.4	1.9	10.4	0.2	0.0	82.8	79.9
	pTIMSS	401	79.1	8.2	79.1	1.2	11.2	0.2	0.0	79.4	78.9
England	eTIMSS	492	60.7	11.6	60.7	5.4	21.8	0.3	0.2	65.9	54.8
	pTIMSS	408	58.1	12.2	58.1	3.0	25.1	1.3	0.2	67.6	50.6
Georgia	eTIMSS	472	46.9	21.0	46.9	7.9	22.5	1.4	0.3	52.5	42.0
	pTIMSS	324	49.9	14.4	49.9	9.3	25.7	0.8	0.0	60.3	40.2
Hong Kong SAR	eTIMSS	469	48.3	12.2	48.3	12.9	25.4	1.0	0.2	52.0	45.2
	pTIMSS	358	54.4	7.6	54.4	3.7	32.9	0.8	0.7	58.6	51.0
Hungary	eTIMSS	652	66.9	18.1	66.9	2.6	11.9	0.4	0.0	71.4	62.7
	pTIMSS	439	69.7	13.4	69.7	2.1	13.5	1.3	0.0	74.2	65.0
Israel	eTIMSS	533	56.9	11.9	56.9	4.3	25.7	1.2	0.1	63.9	49.0
	pTIMSS	452	55.4	9.1	55.4	3.7	29.8	0.9	1.0	59.6	51.1
Italy	eTIMSS	529	71.4	15.4	71.4	3.3	8.7	1.2	0.0	72.2	70.6
	pTIMSS	517	70.0	11.9	70.0	4.3	12.8	0.5	0.6	79.8	59.4
Korea, Rep. of	eTIMSS	553	55.0	6.6	55.0	3.4	35.0	0.0	0.0	61.3	49.7
	pTIMSS	424	51.4	6.5	51.4	2.6	39.2	0.3	0.0	59.4	46.2
Lithuania	eTIMSS	546	64.8	14.3	64.8	2.0	18.7	0.2	0.0	70.9	57.4
	pTIMSS	415	62.8	13.2	62.8	1.3	22.1	0.5	0.1	68.1	57.3
Malaysia	eTIMSS	1018	43.4	13.7	43.4	4.9	37.8	0.2	0.0	46.7	40.0
	pTIMSS	390	43.2	15.4	43.2	5.7	35.1	0.6	0.0	46.7	39.9
Norway (9)	eTIMSS	627	66.9	5.6	66.9	3.3	22.7	1.0	0.6	76.9	57.7
	pTIMSS	499	61.6	8.0	61.6	3.0	24.7	2.4	0.2	69.9	53.6
Qatar	eTIMSS	546	45.7	15.8	45.7	8.6	28.7	0.9	0.3	51.2	40.8
	pTIMSS	378	47.2	17.4	47.2	6.4	28.5	0.6	0.0	49.2	45.2
Russian Federation	eTIMSS	555	60.6	12.6	60.6	4.2	21.5	1.0	0.2	69.0	52.8
	pTIMSS	520	69.6	11.8	69.6	1.6	16.1	0.9	0.0	76.3	62.9
Singapore	eTIMSS	689	66.7	6.7	66.7	2.7	23.7	0.0	0.1	67.7	65.8
	pTIMSS	469	65.0	4.4	65.0	3.8	26.0	0.6	0.2	68.1	62.1
Sweden	eTIMSS	548	66.6	7.3	66.6	4.8	20.2	0.8	0.4	73.4	60.4
	pTIMSS	396	67.8	8.8	67.8	1.4	19.8	1.4	0.7	76.9	60.4
Turkey	eTIMSS	582	57.6	11.5	57.6	2.4	28.5	0.0	0.0	60.6	54.3
	pTIMSS	455	57.8	10.1	57.8	3.5	28.2	0.3	0.1	63.2	52.7
United Arab Emirates	eTIMSS	3186	49.9	15.4	49.9	11.3	22.3	1.0	0.0	53.4	46.6
	pTIMSS	526	50.4	12.9	50.4	7.3	28.0	1.4	0.1	47.9	53.0
United States	eTIMSS	1249	69.1	8.4	69.1	4.4	17.4	0.6	0.1	72.6	65.5
	pTIMSS	370	66.7	11.6	66.7	5.0	15.6	0.6	0.6	75.6	59.9
International Avg (n=19)	eTIMSS	14534	59.8	12.0	59.8	5.0	22.4	0.7	0.1	64.6	55.0
	pTIMSS	8122	59.9	11.1	59.9	3.8	24.0	0.9	0.3	65.5	54.8

DIFF = Percent correct
Because of missing gender information, some totals may appear inconsistent.

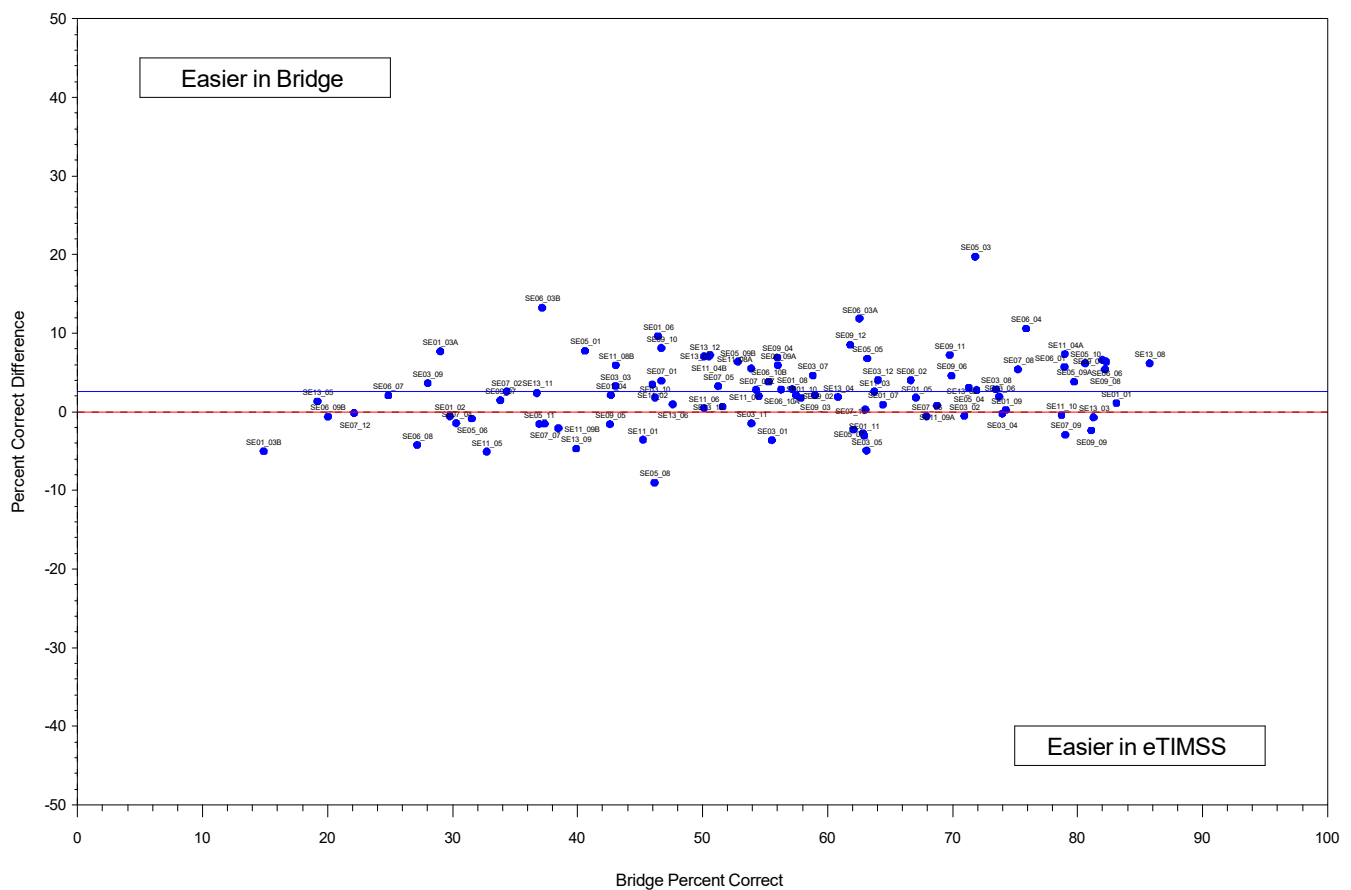
Two different graphical displays were produced for item review to assess the differences in item difficulty by mode of administration. Exhibit 10.12 shows the first of these for an example item. For each country, the difference in item percent correct between eTIMSS and bridge is displayed as a confidence interval. This was calculated using equations (10.1) and (10.2), but using each country’s 2019 bridge and eTIMSS percent correct values and the standard error of their difference. A positive difference for a country indicates that the item was relatively harder in eTIMSS, and a negative difference indicates that the item was relatively easier.

Exhibit 10.12: Example Plot of Differences in Item Percent Correct Between eTIMSS and paperTIMSS (Bridge) for a TIMSS 2019 Trend Item



The second graphical display, presented in Exhibit 10.13, shows the mode differences in percent correct for a given country on all items simultaneously. The blue horizontal line represents the country's average difference across all the items. Where there were no differences between the percent correct in the two modes, the data points aligned on or near the horizontal axis. A positive difference for an item indicates that it was relatively easier on paper, and a negative difference indicates that the item was relatively easier in eTIMSS. Any large or systematic deviations were flagged for further review.

Exhibit 10.13: Example Plot of Differences in Trend Item Percent Correct Between eTIMSS and paperTIMSS (Bridge) by Country



Appendix 10A: TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 4 Mathematics

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Albania	100	98	100	100	97	100
Armenia	100	100	100	100	99	100
Australia	99	95	100	99	94	100
Austria	98	95	100	98	95	100
Azerbaijan	100	98	100	100	98	100
Bahrain	92	78	100	91	75	99
Belgium (Flemish)	97	86	100	97	82	100
Bosnia and Herzegovina	99	91	100	98	91	100
Bulgaria	99	97	100	99	97	100
Canada	97	93	100	96	93	100
Chile	98	94	100	98	94	100
Chinese Taipei	97	89	100	97	89	100
Croatia	99	97	100	99	97	100
Cyprus	98	89	100	97	88	100
Czech Republic	98	94	100	97	92	100
Denmark	97	90	100	96	89	100
England	98	92	100	98	92	100
Finland	100	98	100	100	98	100
France	98	94	100	98	94	100
Georgia	96	89	100	96	88	100
Germany	98	93	100	97	93	100
Hong Kong SAR	100	100	100	100	100	100
Hungary	98	91	99	97	91	99
Iran, Islamic Rep. of	99	94	100	97	92	100
Ireland	100	97	100	99	97	100
Italy	99	94	100	98	94	100
Japan	99	87	100	98	87	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 4 Mathematics (continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Kazakhstan	92	81	99	89	75	97
Korea, Rep. of	99	96	100	99	96	100
Kosovo	96	70	100	94	70	100
Kuwait	100	98	100	99	96	100
Latvia	99	92	100	98	91	100
Lithuania	98	92	100	98	92	100
Malta	95	89	99	94	87	99
Montenegro	98	92	100	97	92	100
Morocco	95	72	100	94	72	100
Netherlands	96	89	100	95	89	100
New Zealand	99	92	100	98	90	100
North Macedonia	99	96	100	99	91	100
Northern Ireland	100	97	100	100	95	100
Norway (5)	98	92	100	97	92	100
Oman	98	92	100	96	88	100
Pakistan	100	100	100	100	100	100
Philippines	99	97	100	99	93	100
Poland	99	93	100	98	88	100
Portugal	98	95	100	97	94	99
Qatar	98	95	100	97	94	100
Russian Federation	98	91	100	97	91	100
Saudi Arabia	96	72	100	94	69	99
Serbia	98	95	100	98	93	100
Singapore	98	95	100	98	95	100
Slovak Republic	99	95	100	98	95	100
South Africa (5)	98	83	100	97	82	100
Spain	97	92	100	96	92	100
Sweden	98	91	100	98	91	100
Turkey (5)	99	96	100	99	96	100
United Arab Emirates	99	96	100	98	96	100
United States	98	96	100	98	96	100
International Average	98	92	100	97	91	100

**TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 4 Mathematics
(continued)**

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Ontario, Canada	97	88	100	96	86	100
Quebec, Canada	96	89	100	95	87	100
Moscow City, Russian Fed.	99	95	100	98	95	100
Madrid, Spain	97	92	100	97	92	100
Abu Dhabi, UAE	99	94	100	99	94	100
Dubai, UAE	98	92	100	98	92	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 4 Science

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Albania	100	99	100	99	98	100
Armenia	100	100	100	100	100	100
Australia	97	90	100	97	90	100
Austria	95	85	100	95	85	100
Azerbaijan	100	98	100	100	98	100
Bahrain	98	93	100	97	92	100
Belgium (Flemish)	93	74	99	92	74	99
Bosnia and Herzegovina	99	94	100	98	90	100
Bulgaria	98	94	100	97	92	100
Canada	92	80	100	92	80	100
Chile	95	90	100	94	90	100
Chinese Taipei	95	83	100	95	81	100
Croatia	94	87	100	94	87	100
Cyprus	93	85	100	92	85	100
Czech Republic	93	84	100	93	84	100
Denmark	93	84	100	93	84	100
England	94	90	100	93	90	100
Finland	97	91	100	96	91	100
France	95	90	99	95	90	99
Georgia	92	76	100	91	76	99
Germany	96	85	100	95	85	100
Hong Kong SAR	100	100	100	100	100	100
Hungary	92	76	99	91	71	99
Iran, Islamic Rep. of	96	84	100	94	75	100
Ireland	98	91	100	98	91	100
Italy	97	94	100	97	93	100
Japan	95	85	100	94	85	100
Kazakhstan	89	79	97	86	75	95
Korea, Rep. of	98	93	100	98	93	100
Kosovo	89	75	97	85	60	96
Kuwait	100	99	100	99	97	100

**TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 4 Science
(continued)**

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Latvia	94	81	100	93	79	100
Lithuania	95	87	99	94	87	99
Malta	92	85	99	91	85	99
Montenegro	98	90	100	98	90	100
Morocco	91	74	99	88	65	99
Netherlands	92	81	100	91	81	100
New Zealand	96	83	100	95	82	100
North Macedonia	98	94	100	98	90	100
Northern Ireland	96	90	100	95	90	100
Norway (5)	93	84	100	92	84	100
Oman	97	93	100	95	87	99
Pakistan	100	100	100	100	100	100
Philippines	99	96	100	98	93	100
Poland	94	72	100	93	71	100
Portugal	96	90	100	95	89	100
Qatar	96	91	100	95	91	100
Russian Federation	94	88	100	94	88	100
Saudi Arabia	94	73	100	92	55	100
Serbia	97	92	100	96	91	100
Singapore	96	86	100	96	86	100
Slovak Republic	97	93	100	97	93	100
South Africa (5)	98	91	100	97	90	100
Spain	93	86	100	92	85	100
Sweden	92	80	100	92	80	100
Turkey (5)	96	90	100	96	85	100
United Arab Emirates	94	90	99	94	90	99
United States	96	87	100	95	87	100
International Average	95	87	100	95	86	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 4 Science
(continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Ontario, Canada	92	77	100	91	77	100
Quebec, Canada	91	74	100	91	74	100
Moscow City, Russian Fed.	96	86	100	95	86	100
Madrid, Spain	92	82	100	92	81	100
Abu Dhabi, UAE	95	87	100	94	86	100
Dubai, UAE	92	81	100	92	81	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 8 Mathematics

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Australia	98	85	100	97	85	100
Bahrain	90	61	100	89	61	100
Chile	96	90	100	95	90	100
Chinese Taipei	97	87	100	97	87	100
Cyprus	98	89	100	97	88	100
Egypt	99	95	100	98	92	100
England	97	93	100	97	93	100
Finland	99	97	100	99	97	100
France	97	88	100	97	88	100
Georgia	96	89	100	95	87	100
Hong Kong SAR	100	100	100	100	100	100
Hungary	97	91	100	96	90	100
Iran, Islamic Rep. of	99	89	100	97	89	100
Ireland	99	91	100	98	86	100
Israel	98	90	100	97	90	100
Italy	98	91	100	97	91	100
Japan	99	90	100	98	90	100
Jordan	99	96	100	98	93	100
Kazakhstan	93	78	100	90	69	99
Korea, Rep. of	99	96	100	99	95	100
Kuwait	100	97	100	99	96	100
Lebanon	100	100	100	100	100	100
Lithuania	97	89	100	96	89	100
Malaysia	99	97	100	99	97	100
Morocco	96	68	100	90	30	100
New Zealand	98	87	100	97	85	100
Norway (9)	97	92	100	97	92	100
Oman	99	95	100	98	89	100
Portugal	97	89	100	97	89	100
Qatar	97	93	100	96	93	100
Romania	99	95	100	97	92	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 8 Mathematics
(continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Russian Federation	98	90	100	97	90	100
Saudi Arabia	98	76	100	95	29	100
Singapore	99	95	100	98	93	100
South Africa (9)	100	98	100	99	95	100
Sweden	97	85	100	97	85	100
Turkey	99	94	100	98	93	100
United Arab Emirates	97	93	100	97	93	100
United States	98	91	100	98	91	100
International Average	98	90	100	97	87	100
Ontario, Canada	96	83	100	95	83	100
Quebec, Canada	96	84	100	95	80	100
Moscow City, Russian Fed.	98	92	100	98	92	100
Gauteng, RSA (9)	100	95	100	99	91	100
Western Cape, RSA (9)	100	93	100	99	91	100
Abu Dhabi, UAE	97	91	100	97	91	100
Dubai, UAE	97	87	100	96	87	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 8 Science

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Australia	94	76	100	92	76	100
Bahrain	98	89	100	97	83	100
Chile	95	90	100	94	90	100
Chinese Taipei	95	81	100	94	81	100
Cyprus	94	86	100	93	86	100
Egypt	97	89	100	96	86	100
England	95	90	100	94	90	100
Finland	97	91	100	96	91	100
France	96	86	100	96	86	100
Georgia	93	73	100	92	73	100
Hong Kong SAR	100	100	100	100	100	100
Hungary	91	79	99	90	78	99
Iran, Islamic Rep. of	96	85	100	94	83	100
Ireland	96	83	100	95	82	100
Israel	93	81	100	92	81	100
Italy	97	92	100	97	92	100
Japan	95	86	100	95	86	100
Jordan	99	96	100	98	85	100
Kazakhstan	88	72	99	85	67	98
Korea, Rep. of	97	90	100	96	90	100
Kuwait	99	98	100	99	97	100
Lebanon	100	100	100	100	100	100
Lithuania	94	80	100	93	80	100
Malaysia	95	90	100	95	90	100
Morocco	91	70	100	87	38	99
New Zealand	95	78	99	94	77	99
Norway (9)	94	85	100	94	85	100
Oman	97	89	100	96	89	100
Portugal	95	87	100	95	87	100
Qatar	95	90	100	94	90	100
Romania	96	81	100	95	80	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 8 Science
(continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Russian Federation	94	87	100	94	87	100
Saudi Arabia	92	75	100	90	45	100
Singapore	96	87	100	95	87	100
South Africa (9)	99	96	100	99	96	100
Sweden	93	80	100	93	80	100
Turkey	96	88	100	96	88	100
United Arab Emirates	94	90	100	94	90	100
United States	95	90	100	95	89	100
International Average	95	86	100	94	84	100
Ontario, Canada	93	82	100	92	82	100
Quebec, Canada	92	78	100	90	78	100
Moscow City, Russian Fed.	97	87	100	96	87	100
Gauteng, RSA (9)	98	84	100	98	84	100
Western Cape, RSA (9)	99	94	100	99	94	100
Abu Dhabi, UAE	95	83	100	94	83	100
Dubai, UAE	93	84	100	93	83	100

Appendix 10B: TIMSS 2019 Trend Scoring Reliability for Human Scored Items

TIMSS 2019 Trend Scoring Reliability for Human Scored Items—Grade 4 Mathematics

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Australia	98	90	100	97	89	100
Bahrain	99	96	100	98	91	100
Belgium (Flemish)	96	87	99	96	86	99
Bulgaria	99	95	100	99	95	100
Canada	95	76	99	94	75	99
Chile	96	85	100	96	85	100
Chinese Taipei	98	93	100	97	92	100
Croatia	98	90	100	97	90	100
Cyprus	97	87	100	97	87	100
Czech Republic	97	89	100	97	89	100
Denmark	96	84	100	94	78	100
England	98	86	100	97	86	100
Finland	99	93	100	98	93	100
France	97	87	100	96	87	100
Georgia	95	67	100	95	67	100
Germany	98	88	100	97	88	100
Hong Kong SAR	96	81	100	96	81	100
Hungary	97	93	99	97	92	99
Iran, Islamic Rep. of	98	92	99	97	92	99
Ireland	98	85	100	97	85	100
Italy	97	87	100	97	86	100
Japan	97	89	100	97	89	100
Korea, Rep. of	99	94	100	99	94	100
Lithuania	98	89	100	97	84	100
Netherlands	98	84	100	97	84	100
New Zealand	97	86	100	96	86	100
Northern Ireland	98	87	100	98	87	100

TIMSS 2019 Trend Scoring Reliability for Human Scored Items—Grade 4 Mathematics (continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Norway (5)	98	91	100	97	90	100
Oman	95	87	100	93	72	100
Poland	97	89	100	97	89	100
Portugal	98	88	100	98	87	100
Qatar	97	89	100	95	77	100
Saudi Arabia	93	85	98	93	84	98
Serbia	96	79	100	95	79	100
Singapore	99	95	100	99	94	100
Slovak Republic	97	82	100	96	81	100
South Africa (5)	97	94	99	97	93	99
Spain	96	87	100	96	86	100
Sweden	96	78	100	96	77	100
United Arab Emirates	97	89	100	96	86	100
United States	97	85	100	97	85	100
International Average	97	87	100	96	86	100

TIMSS 2019 Trend Scoring Reliability for Human Scored Items—Grade 4 Science

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Australia	93	85	99	93	85	99
Bahrain	91	82	98	89	77	98
Belgium (Flemish)	92	82	99	91	82	99
Bulgaria	96	85	100	95	85	100
Canada	91	79	99	90	79	99
Chile	87	73	98	87	73	98
Chinese Taipei	93	81	100	93	76	100
Croatia	92	73	99	91	73	99
Cyprus	91	77	99	91	77	99
Czech Republic	91	69	98	90	69	98
Denmark	86	75	97	86	75	97
England	89	70	99	88	70	99
Finland	93	84	100	93	84	100
France	93	73	99	92	73	99
Georgia	88	69	97	87	69	95
Germany	93	82	99	92	82	99
Hong Kong SAR	89	82	96	88	80	96
Hungary	93	86	99	92	83	99
Iran, Islamic Rep. of	93	84	99	92	84	99
Ireland	91	74	99	91	74	99
Italy	93	81	100	93	81	100
Japan	91	83	99	90	83	99
Korea, Rep. of	95	88	100	95	88	100
Lithuania	93	80	99	93	80	99
Netherlands	94	78	99	93	78	99
New Zealand	90	85	97	90	81	97
Northern Ireland	92	82	99	91	82	99
Norway (5)	87	67	99	86	67	99
Oman	83	62	97	82	62	97
Poland	92	76	98	92	76	98
Portugal	95	87	98	95	87	98
Qatar	86	71	98	85	71	98

TIMSS 2019 Trend Scoring Reliability for Human Scored Items—Grade 4 Science (continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Saudi Arabia	81	67	95	80	67	95
Serbia	90	81	99	89	77	99
Singapore	95	90	100	95	90	100
Slovak Republic	91	73	98	91	73	98
Spain	86	66	100	86	66	100
Sweden	90	77	99	90	77	99
United Arab Emirates	91	80	99	90	80	99
United States	94	83	100	93	83	100
International Average	91	78	99	90	77	99

TIMSS 2019 Trend Scoring Reliability for Human Scored Items—Grade 8 Mathematics

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Australia	96	89	100	96	89	100
Bahrain	98	89	100	95	65	100
Canada *	88	74	93	85	74	93
Chile	95	86	100	94	78	100
Chinese Taipei	97	90	100	96	79	100
England	95	63	100	95	63	100
Georgia	94	79	100	91	52	100
Hong Kong SAR	92	69	100	92	63	100
Hungary	97	83	100	95	74	100
Iran, Islamic Rep. of	97	89	100	97	89	100
Ireland	97	84	100	97	84	100
Israel	97	89	100	96	89	100
Italy	97	89	100	97	89	100
Japan	96	68	100	94	68	100
Jordan	99	97	100	98	93	100
Korea, Rep. of	94	49	100	93	49	100
Lithuania	98	92	100	96	74	100
Malaysia	96	84	100	93	60	100
New Zealand	95	81	100	95	81	100
Norway (9)	94	71	100	93	71	100
Oman	96	83	99	94	68	99
Qatar	96	83	100	95	80	100
Russian Federation	97	84	100	96	81	100
Saudi Arabia	97	88	100	97	88	99
Singapore	97	78	100	97	78	100
South Africa (9)	97	83	100	92	27	100
Sweden	96	74	100	95	74	100
Turkey	96	75	100	94	75	100
United Arab Emirates	97	86	100	96	85	100
United States	97	84	100	97	84	100
International Average	96	81	100	95	74	100

* Canada participated in trend scoring reliability for the benchmarking participants Ontario and Quebec.

TIMSS 2019 Trend Scoring Reliability for Human Scored Items—Grade 8 Science

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Australia	92	78	97	91	78	97
Bahrain	89	74	98	87	65	98
Canada *	90	80	98	89	70	98
Chile	87	70	99	86	66	99
Chinese Taipei	92	81	99	90	74	99
England	91	81	98	90	81	98
Georgia	90	75	99	88	66	99
Hong Kong SAR	90	77	99	89	77	99
Hungary	92	83	97	91	78	97
Iran, Islamic Rep. of	93	81	100	92	81	100
Ireland	93	84	100	92	84	100
Israel	92	85	100	91	81	100
Italy	92	83	100	91	73	100
Japan	92	81	100	91	75	100
Jordan	98	94	100	98	94	100
Korea, Rep. of	95	87	99	95	83	99
Lithuania	95	78	100	94	78	100
Malaysia	90	79	98	89	79	98
New Zealand	91	82	98	90	82	98
Norway (9)	91	82	99	90	79	99
Oman	86	71	97	84	65	97
Qatar	89	80	99	87	73	99
Russian Federation	86	66	98	85	65	98
Saudi Arabia	88	78	99	87	72	99
Singapore	94	82	100	94	82	100
South Africa (9)	95	90	100	94	88	100
Sweden	91	78	100	91	78	100
Turkey	90	77	97	88	66	97
United Arab Emirates	92	88	98	91	84	98
United States	93	80	98	92	76	98
International Average	91	80	99	90	76	99

* Canada participated in trend scoring reliability for the benchmarking participants Ontario and Quebec.

Appendix 10C: TIMSS 2019 Cross-Country Scoring Reliability for Human Scored Items

TIMSS 2019 Cross-Country Scoring Reliability for Human Scored Items—Grade 4 Mathematics

Item	Total Valid Comparisons	Exact Percent Agreement	
		Score Point Agreement	Diagnostic Score Agreement
M03_03 – MP61034	295,299	99	99
M03_05 – MP61228	296,261	81	81
M03_06 – MP61166	297,000	97	97
M03_08 – MP61080	295,515	98	98
M03_10 – MP61076	296,892	100	100
M03_11 – MP61084	293,850	95	95
M05_01 – MP51206	215,204	97	97
M05_04 – MP51045	216,200	99	98
M05_06 – MP51030	215,248	97	97
M05_11 – MP51533	216,108	100	100
M05_12 – MP51080	211,916	95	91
Average Percent Agreement		96	96

TIMSS 2019 Cross-Country Scoring Reliability for Human Scored Items—Grade 4 Science

Item	Total Valid Comparisons	Exact Percent Agreement	
		Score Point Agreement	Diagnostic Score Agreement
S03_02 – SP61023	295,515	98	98
S03_03 – SP61054	293,491	87	87
S03_05 – SP61006	296,730	91	91
S03_09 – SP61088	295,245	91	91
S03_10 – SP61151	296,334	86	86
S03_11 – SP61150	294,925	82	82
S05_01 – SP51044	296,892	86	86
S05_04 – SP51168	296,152	88	86
S05_05 – SP51010	296,677	91	86
S05_07 – SP51059	295,138	75	75
S05_10 – SP51151	296,946	98	98
Average Percent Agreement		89	88

TIMSS 2019 Cross-Country Scoring Reliability for Human Scored Items—Grade 8 Mathematics

Item	Total Valid Comparisons	Exact Percent Agreement	
		Score Point Agreement	Diagnostic Score Agreement
M03_02 – MP62139	125,965	100	100
M03_04 – MP62142	125,895	98	98
M03_08 – MP62027	125,701	99	99
M03_10 – MP62244	125,518	97	97
M03_12 – MP62300	125,825	94	93
M03_13 – MP62254	125,350	71	71
M03_14 – MP62132A	126,000	100	100
M05_05 – MP52174A	126,000	98	98
M05_05 – MP52174B	125,965	99	98
M05_08 – MP52110	125,791	100	100
M05_09 – MP52105	124,740	88	88
M05_11 – MP52036	125,755	86	86
M05_12 – MP52502	125,721	96	96
M05_13 – MP52117	125,057	90	75
Average Percent Agreement		94	93

TIMSS 2019 Cross-Country Scoring Reliability for Human Scored Items—Grade 8 Science

Item	Total Valid Comparisons	Exact Percent Agreement	
		Score Point Agreement	Diagnostic Score Agreement
S03_03 – SP62275	125,965	93	93
S03_05 – SP62111	124,565	93	93
S03_06 – SP62116A	125,790	90	90
S03_06 – SP62116B	125,090	90	90
S03_06 – SP62116C	125,790	76	76
S03_10 – SP62162	125,965	85	85
S05_02 – SP52272	125,930	92	86
S05_03 – SP52085A	125,791	80	72
S05_03 – SP52085B	125,755	83	83
S05_04 – SP52094	126,000	96	96
S05_06 – SP52146	124,775	94	92
S05_10 – SP52214	126,000	98	98
S05_12 – SP52101	125,301	88	88
Average Percent Agreement		89	88

Appendix 10D: TIMSS 2019 Item Statistics by Booklet Position

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 4 Mathematics (eTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
ME01 - Positions 1 & 4	11,670	11,603	53.1	48.3	2.7	4.7	0.1	2.8
ME02 - Positions 2 & 3	11,541	11,670	51.4	51.8	4.3	4.5	0.0	1.9
ME03 - Positions 1 & 4	11,596	11,542	50.3	49.4	4.4	5.9	0.1	1.2
ME04 - Positions 2 & 3	11,598	11,596	49.2	48.3	4.6	5.4	0.2	4.2
ME05 - Positions 1 & 4	11,634	11,598	51.2	48.2	2.6	4.5	0.1	2.9
ME06 - Positions 2 & 3	11,584	11,634	48.8	47.7	2.5	3.8	0.1	2.6
ME07 - Positions 1 & 4	11,635	11,585	49.7	48.7	2.5	2.9	0.0	0.8
ME08 - Positions 2 & 3	11,594	11,636	46.0	45.7	2.9	3.3	0.1	3.1
ME09 - Positions 1 & 4	11,614	11,594	48.7	45.7	2.5	4.2	0.1	2.8
ME10 - Positions 2 & 3	11,578	11,614	49.2	48.5	3.7	4.9	0.1	3.4
ME11 - Positions 1 & 4	11,613	11,577	51.0	48.8	4.2	5.5	0.1	2.0
ME12 - Positions 2 & 3	11,634	11,613	48.6	46.3	4.0	5.2	0.1	3.5
ME13 - Positions 1 & 4	11,605	11,633	52.1	50.1	2.5	3.3	0.0	1.8
ME14 - Positions 2 & 3	11,603	11,605	49.3	48.6	3.2	4.3	0.1	2.8
Overall	162,499	162,500	49.9	48.3	3.3	4.5	0.1	2.5

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 4 Mathematics (paperTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
MP01 - Positions 1 & 4	5,984	5,977	52.7	48.8	3.6	5.1	0.1	4.1
MP02 - Positions 2 & 3	5,997	5,984	49.7	51.3	7.3	6.3	0.3	1.9
MP03 - Positions 1 & 4	5,990	5,997	49.0	46.6	6.8	8.4	0.1	3.4
MP04 - Positions 2 & 3	5,984	5,990	50.1	49.0	6.3	6.7	0.3	4.6
MP05 - Positions 1 & 4	5,960	5,983	51.7	48.8	4.7	5.9	0.1	3.8
MP06 - Positions 2 & 3	5,974	5,960	49.6	48.7	5.3	5.9	0.2	1.8
MP07 - Positions 1 & 4	5,967	5,974	50.9	49.6	3.8	5.1	0.1	1.4
MP08 - Positions 2 & 3	6,002	5,967	42.2	42.7	8.0	6.6	0.3	3.6
MP09 - Positions 1 & 4	5,994	6,002	50.0	46.4	6.2	7.2	0.1	4.1
MP10 - Positions 2 & 3	5,976	5,994	47.5	47.6	5.9	5.4	0.3	3.5
MP11 - Positions 1 & 4	5,973	5,976	49.4	47.6	6.1	8.0	0.1	3.6
MP12 - Positions 2 & 3	6,005	5,973	47.9	47.0	5.6	5.8	0.3	4.1
MP13 - Positions 1 & 4	5,983	6,005	51.7	49.3	5.3	6.1	0.1	3.9
MP14 - Positions 2 & 3	5,978	5,982	49.2	49.3	6.2	6.0	0.2	2.8
Overall	83,767	83,764	49.4	48.1	5.8	6.3	0.2	3.3

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 4 Less Difficult Mathematics (paperTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
MN01 - Positions 1 & 4	4,399	4,375	48.0	47.2	6.1	6.7	0.2	2.2
MN03 - Positions 1 & 4	4,434	4,389	48.6	44.6	5.0	6.6	0.2	4.5
MN04 - Positions 2 & 3	4,419	4,434	49.2	49.9	7.1	7.2	0.3	4.1
MN05 - Positions 1 & 4	4,435	4,419	53.4	50.2	5.3	7.1	0.2	2.8
MN07 - Positions 1 & 4	4,440	4,413	53.6	51.6	7.3	7.9	0.3	2.9
MN09 - Positions 1 & 4	4,397	4,405	53.3	50.0	4.3	7.3	0.2	4.3
MN11 - Positions 1 & 4	4,407	4,373	44.3	42.9	6.5	7.5	0.1	2.9
MN12 - Positions 2 & 3	4,391	4,407	49.3	48.8	6.5	6.5	0.2	2.9
MN13 - Positions 1 & 4	4,420	4,391	48.2	46.8	5.2	5.6	0.2	2.6
MN14 - Positions 2 & 3	4,375	4,420	50.8	50.6	7.5	7.2	0.2	2.7
MP02 - Positions 2 & 3	4,389	4,398	28.3	28.0	13.5	12.8	0.4	4.3
MP03 - Positions 2 & 3	4,412	4,435	27.8	27.9	13.8	13.6	0.3	4.0
MP08 - Positions 2 & 3	4,405	4,440	23.6	23.0	13.4	12.6	0.7	7.5
MP13 - Positions 2 & 3	4,373	4,397	28.5	29.3	9.5	8.2	0.5	3.4
Overall	61,696	61,696	43.4	42.2	7.9	8.3	0.3	3.7

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 4 Mathematics

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Albania	4,417	4,417	62.3	61.8	4.1	5.0	0.2	2.8
Armenia	5,380	5,380	43.9	42.2	13.4	13.8	0.7	7.4
Australia	5,879	5,879	48.9	47.4	3.0	3.8	0.2	2.6
Austria	4,463	4,463	51.4	50.3	4.7	6.1	0.0	1.2
Azerbaijan	5,220	5,220	50.5	50.0	10.0	9.0	0.3	2.8
Bahrain	5,758	5,758	40.2	38.4	7.9	8.6	0.1	1.9
Belgium (Flemish)	4,646	4,646	52.4	51.7	3.9	3.8	0.0	1.6
Bosnia and Herzegovina	5,612	5,612	50.3	49.3	12.0	11.8	0.1	3.4
Bulgaria	4,267	4,267	52.9	51.2	8.6	8.3	0.1	2.8
Canada	13,576	13,576	43.2	41.6	2.7	4.0	0.2	4.0
Chile	4,161	4,161	32.4	28.8	5.1	8.0	0.4	10.0
Chinese Taipei	3,763	3,763	67.9	67.5	0.8	1.0	0.0	0.1
Croatia	3,783	3,783	43.9	43.4	3.8	4.5	0.0	1.1
Cyprus	4,061	4,061	52.7	51.2	3.6	4.3	0.1	3.8
Czech Republic	4,689	4,689	50.5	48.6	4.4	5.5	0.0	1.5
Denmark	3,213	3,213	50.0	46.9	4.7	6.8	0.2	6.4
England	3,393	3,393	55.1	55.0	2.6	3.3	0.1	1.0
Finland	4,723	4,723	50.3	48.9	4.2	5.1	0.1	1.8
France	4,179	4,179	39.7	38.2	6.9	8.7	0.1	4.1
Georgia	3,765	3,765	37.3	35.5	7.8	9.7	0.1	4.7
Germany	3,434	3,434	47.0	45.4	5.0	6.4	0.1	2.6
Hong Kong SAR	2,964	2,964	68.4	67.9	0.9	1.1	0.0	0.2
Hungary	4,569	4,569	50.0	49.3	2.3	2.8	0.0	1.1
Iran, Islamic Rep. of	5,989	5,989	34.8	32.5	10.6	11.8	0.6	10.8
Ireland	4,566	4,566	56.7	55.7	2.4	2.5	0.0	0.9
Italy	3,738	3,738	45.8	43.2	4.6	6.5	0.1	4.6
Japan	4,193	4,193	68.5	68.2	1.7	2.0	0.0	1.1
Kazakhstan	4,786	4,786	46.3	46.0	4.6	5.3	0.1	2.9
Korea, Rep. of	3,893	3,893	68.6	67.7	1.0	1.3	0.0	0.2
Kosovo	4,490	4,491	50.6	49.9	5.2	5.4	0.2	2.6
Kuwait	4,417	4,417	39.5	38.0	5.6	5.9	0.4	3.4
Latvia	4,474	4,473	57.5	55.9	3.4	4.0	0.1	1.8
Lithuania	3,739	3,739	51.5	50.2	2.8	3.4	0.1	0.7

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 4 Mathematics (continued)

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Malta	3,626	3,626	43.6	42.1	2.2	2.9	0.0	1.1
Montenegro	5,060	5,060	51.4	50.2	13.8	14.0	0.2	4.2
Morocco	7,712	7,712	37.1	34.9	7.0	7.8	0.2	3.9
Netherlands	3,336	3,336	51.1	49.1	3.0	4.0	0.1	1.8
New Zealand	5,002	5,002	41.9	40.7	3.7	4.6	0.3	3.8
North Macedonia	3,264	3,264	56.9	55.6	5.9	6.8	0.3	5.1
Northern Ireland	3,491	3,491	61.5	60.9	2.6	3.0	0.1	0.9
Norway (5)	3,938	3,938	54.9	52.8	3.7	5.0	0.3	4.1
Oman	6,801	6,801	30.6	28.9	4.3	5.7	0.3	4.5
Pakistan	3,942	3,942	27.2	25.9	18.5	17.7	0.6	5.8
Philippines	5,495	5,495	26.2	25.5	5.8	6.8	0.5	3.3
Poland	4,881	4,881	50.0	48.2	7.7	8.3	0.1	2.2
Portugal	4,297	4,297	49.0	46.9	2.7	4.0	0.0	3.2
Qatar	4,929	4,929	32.9	30.6	3.9	5.2	0.1	5.0
Russian Federation	4,022	4,022	59.8	59.0	2.2	2.9	0.0	1.0
Saudi Arabia	5,445	5,445	43.2	42.1	5.0	5.6	0.2	2.8
Serbia	4,373	4,373	48.6	47.6	7.9	8.5	0.2	4.4
Singapore	5,983	5,983	72.5	72.2	0.5	0.7	0.0	0.2
Slovak Republic	4,243	4,243	45.0	43.2	4.4	5.1	0.0	1.6
South Africa (5)	11,842	11,842	36.3	35.0	3.7	4.4	0.1	3.0
Spain	9,543	9,543	44.8	43.1	3.5	4.7	0.1	2.1
Sweden	3,958	3,958	49.3	46.4	4.8	7.2	0.3	5.7
Turkey (5)	4,028	4,028	48.1	47.5	1.9	2.6	0.0	0.9
United Arab Emirates	25,785	25,785	39.3	37.4	2.7	3.9	0.1	3.3
United States	8,769	8,769	50.8	49.2	1.1	1.9	0.1	2.9
International Average	307,965	307,965	48.5	47.1	4.9	5.7	0.2	3.0
Ontario, Canada	3,810	3,810	45.4	43.6	2.4	3.7	0.1	4.2
Quebec, Canada	3,816	3,816	50.3	48.9	2.6	3.8	0.0	2.7
Moscow City, Russian Fed.	3,842	3,842	66.4	65.6	1.6	1.9	0.0	0.4
Madrid, Spain	3,385	3,385	46.7	44.8	3.3	4.0	0.0	1.5
Abu Dhabi, UAE	9,001	8,998	31.3	29.4	2.9	3.9	0.1	3.0
Dubai, UAE	7,262	7,262	52.6	51.2	1.6	2.5	0.0	2.4

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 4 Science (eTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
SE01 - Positions 2 & 3	11,666	11,617	51.0	51.1	2.5	3.5	0.0	4.2
SE02 - Positions 1 & 4	11,554	11,667	55.0	52.6	3.5	4.5	0.1	1.9
SE03 - Positions 2 & 3	11,585	11,554	53.0	52.1	2.9	4.1	0.0	2.5
SE04 - Positions 1 & 4	11,603	11,584	58.8	57.3	2.5	4.0	0.1	1.1
SE05 - Positions 2 & 3	11,630	11,603	53.7	53.0	3.6	4.8	0.1	3.4
SE06 - Positions 1 & 4	11,594	11,630	50.6	47.7	5.5	8.4	0.2	2.2
SE07 - Positions 2 & 3	11,624	11,594	46.7	46.0	3.2	4.2	0.1	3.6
SE08 - Positions 1 & 4	11,603	11,625	56.1	54.9	1.9	3.3	0.1	1.5
SE09 - Positions 2 & 3	11,603	11,603	53.2	52.6	4.4	5.0	0.1	2.5
SE10 - Positions 1 & 4	11,596	11,603	53.6	52.2	1.5	2.9	0.1	1.6
SE11 - Positions 2 & 3	11,592	11,596	51.0	50.5	6.1	6.6	0.1	3.5
SE12 - Positions 1 & 4	11,647	11,592	56.0	53.1	2.2	3.7	0.1	1.9
SE13 - Positions 2 & 3	11,596	11,647	49.9	50.1	2.9	3.4	0.1	4.1
SE14 - Positions 1 & 4	11,617	11,596	57.1	54.8	1.8	3.1	0.1	1.4
Overall	162,510	162,511	53.3	52.0	3.2	4.4	0.1	2.5

**TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 4 Science
(paperTIMSS)**

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
SP01 - Positions 2 & 3	5,977	5,987	48.3	48.5	6.5	7.0	0.2	3.6
SP02 - Positions 1 & 4	6,005	5,977	51.0	49.4	7.5	7.8	0.2	2.1
SP03 - Positions 2 & 3	5,984	6,005	48.9	48.1	6.2	6.0	0.0	2.6
SP04 - Positions 1 & 4	5,994	5,984	55.5	52.9	4.9	6.6	0.1	1.7
SP05 - Positions 2 & 3	5,948	5,994	51.2	51.2	5.5	5.5	0.1	2.5
SP06 - Positions 1 & 4	5,974	5,948	50.1	46.6	6.9	9.7	0.2	2.2
SP07 - Positions 2 & 3	5,952	5,974	43.6	43.5	6.2	6.1	0.1	3.3
SP08 - Positions 1 & 4	6,017	5,952	55.6	53.6	4.3	5.5	0.1	1.7
SP09 - Positions 2 & 3	5,983	6,017	48.6	49.2	6.7	6.4	0.2	2.4
SP10 - Positions 1 & 4	6,010	5,983	52.4	49.5	3.4	5.8	0.1	2.7
SP11 - Positions 2 & 3	5,959	6,010	50.0	49.0	7.4	7.9	0.2	5.0
SP12 - Positions 1 & 4	6,011	5,959	53.6	51.6	5.4	6.0	0.0	2.8
SP13 - Positions 2 & 3	5,979	6,011	47.5	46.8	5.8	8.7	0.0	0.4
SP14 - Positions 1 & 4	5,987	5,979	56.1	53.6	3.4	5.3	0.2	2.6
Overall	83,780	83,780	50.9	49.5	5.7	6.7	0.1	2.5

**TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 4 Science
(Less Difficult paperTIMSS)**

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
SP01 - Positions 2 & 3	4,381	4,396	32.2	31.3	12.0	12.0	0.4	6.3
SP02 - Positions 1 & 4	4,406	4,382	33.9	32.9	13.9	13.3	0.6	4.8
SP03 - Positions 2 & 3	4,408	4,406	35.9	33.8	9.7	11.4	0.1	5.6
SP04 - Positions 1 & 4	4,429	4,408	38.4	37.5	10.0	10.8	0.4	3.5
SP05 - Positions 2 & 3	4,413	4,428	33.0	32.4	9.3	11.3	0.0	4.6
SP06 - Positions 1 & 4	4,431	4,413	31.3	30.3	15.3	16.9	0.7	4.5
SP07 - Positions 2 & 3	4,420	4,431	30.3	30.4	12.6	12.1	0.3	6.7
SP08 - Positions 1 & 4	4,421	4,420	35.7	34.0	9.4	11.2	0.4	3.6
SP09 - Positions 2 & 3	4,379	4,421	34.1	32.9	9.3	11.2	0.3	4.9
SP10 - Positions 1 & 4	4,409	4,378	33.5	32.4	9.0	10.7	0.5	4.3
SP11 - Positions 2 & 3	4,391	4,408	30.4	29.2	13.7	15.6	0.2	6.9
SP12 - Positions 1 & 4	4,411	4,392	36.1	34.1	10.6	10.5	0.1	4.6
SP13 - Positions 2 & 3	4,398	4,410	31.9	29.8	10.3	14.5	0.0	3.2
SP14 - Positions 1 & 4	4,396	4,399	37.9	35.6	7.8	9.4	0.4	4.6
Overall	61,693	61,692	33.9	32.6	10.9	12.2	0.3	4.9

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 4 Science

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Albania	4,422	4,422	48.7	47.0	6.0	6.9	0.3	3.6
Armenia	5,391	5,391	43.0	40.2	13.0	15.9	0.3	5.0
Australia	5,878	5,878	56.9	55.7	2.9	3.7	0.1	2.0
Austria	4,464	4,464	52.1	51.2	4.6	5.6	0.0	1.7
Azerbaijan	5,185	5,185	38.0	37.4	13.6	13.2	0.4	3.1
Bahrain	5,760	5,760	49.5	46.9	4.5	6.5	0.1	2.0
Belgium (Flemish)	4,653	4,653	47.9	47.5	4.8	4.6	0.0	1.6
Bosnia and Herzegovina	5,611	5,611	41.1	40.0	12.6	13.8	0.2	4.3
Bulgaria	4,267	4,267	59.1	58.2	6.6	7.1	0.0	1.7
Canada	13,579	13,579	51.9	50.4	2.7	3.8	0.1	3.2
Chile	4,158	4,158	44.9	42.0	3.9	6.5	0.4	8.1
Chinese Taipei	3,764	3,764	60.8	60.0	2.0	2.6	0.0	0.5
Croatia	3,784	3,784	52.3	51.5	2.7	3.5	0.0	0.8
Cyprus	4,062	4,062	52.1	50.3	4.4	6.0	0.1	4.4
Czech Republic	4,688	4,688	54.6	53.6	4.0	5.0	0.0	1.1
Denmark	3,220	3,220	52.6	51.0	3.5	4.8	0.1	3.3
England	3,387	3,387	54.9	54.1	2.3	3.0	0.0	0.7
Finland	4,711	4,711	59.3	58.6	2.9	3.6	0.0	0.7
France	4,184	4,184	47.0	45.3	6.3	8.7	0.1	5.0
Georgia	3,764	3,764	39.1	36.7	8.2	10.8	0.2	6.3
Germany	3,432	3,432	52.8	50.7	4.3	6.3	0.1	4.2
Hong Kong SAR	2,968	2,968	54.9	53.6	2.2	3.4	0.0	0.8
Hungary	4,570	4,570	55.8	54.8	2.2	3.1	0.0	1.7
Iran, Islamic Rep. of	5,994	5,994	41.1	39.7	9.7	11.5	0.4	6.9
Ireland	4,576	4,576	55.1	54.0	2.7	3.2	0.0	0.9
Italy	3,740	3,740	50.4	48.7	4.2	6.4	0.1	4.7
Japan	4,192	4,192	62.6	61.1	2.2	2.5	0.0	1.1
Kazakhstan	4,791	4,791	47.4	46.9	5.0	6.0	0.0	2.3
Korea, Rep. of	3,891	3,891	66.7	67.0	1.1	1.3	0.0	0.1
Kosovo	4,486	4,484	34.8	33.6	9.3	10.0	0.2	3.2
Kuwait	4,412	4,412	34.0	32.8	9.2	10.1	0.3	3.4
Latvia	4,476	4,476	59.0	57.6	2.8	3.3	0.1	1.2
Lithuania	3,738	3,738	54.4	54.1	2.7	3.3	0.0	0.6

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 4 Science (continued)

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Malta	3,625	3,625	47.4	46.2	2.5	3.4	0.0	1.1
Montenegro	5,068	5,068	41.7	39.4	14.3	15.8	0.4	7.7
Morocco	7,714	7,714	29.9	28.1	10.5	12.7	0.2	4.7
Netherlands	3,337	3,337	50.9	49.7	2.8	3.4	0.1	1.1
New Zealand	5,003	5,003	49.8	48.8	3.3	4.8	0.2	2.2
North Macedonia	3,262	3,262	38.7	36.8	10.4	12.8	0.6	8.8
Northern Ireland	3,490	3,490	52.9	51.9	3.3	3.7	0.1	1.0
Norway (5)	3,940	3,940	56.8	55.8	2.9	3.4	0.1	1.7
Oman	6,811	6,811	39.9	38.3	6.0	7.7	0.1	3.2
Pakistan	3,919	3,919	21.9	21.9	23.7	23.7	0.6	4.9
Philippines	5,501	5,501	20.9	20.2	10.3	12.0	0.3	4.2
Poland	4,875	4,875	55.6	55.0	6.5	7.2	0.0	1.2
Portugal	4,297	4,297	49.4	46.9	3.0	5.2	0.0	5.1
Qatar	4,933	4,933	41.4	39.8	4.4	6.3	0.3	6.9
Russian Federation	4,021	4,021	62.9	62.2	2.3	3.2	0.1	1.2
Saudi Arabia	5,448	5,448	35.7	35.0	9.0	9.4	0.1	2.7
Serbia	4,376	4,376	55.4	52.8	6.3	7.7	0.2	4.2
Singapore	5,983	5,983	68.0	67.2	0.8	1.1	0.0	0.3
Slovak Republic	4,246	4,246	53.4	52.6	3.6	5.1	0.0	1.4
South Africa (5)	11,852	11,852	26.1	24.6	5.1	7.2	0.3	6.2
Spain	9,544	9,544	51.5	50.2	3.1	4.1	0.0	1.7
Sweden	3,951	3,951	56.7	55.2	3.4	4.7	0.1	2.1
Turkey (5)	4,028	4,028	54.9	53.4	2.2	3.0	0.0	1.7
United Arab Emirates	25,796	25,796	45.5	43.8	3.3	5.1	0.2	4.1
United States	8,770	8,770	56.5	55.1	1.4	2.4	0.2	3.3
International Average	307,988	307,986	48.9	47.6	5.4	6.6	0.1	3.0
Ontario, Canada	3,806	3,806	53.0	51.1	2.5	3.7	0.1	3.9
Quebec, Canada	3,828	3,828	52.3	50.8	2.7	3.7	0.1	2.1
Moscow City, Russian Fed.	3,841	3,841	68.8	67.5	1.7	2.3	0.0	0.8
Madrid, Spain	3,388	3,388	52.5	51.3	2.7	3.5	0.0	1.0
Abu Dhabi, UAE	9,004	9,004	36.7	35.1	3.8	5.7	0.2	4.6
Dubai, UAE	7,265	7,265	57.6	55.9	1.7	2.7	0.1	2.3

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 8 Mathematics (eTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
ME01 - Positions 1 & 4	8,099	8,152	46.0	44.9	3.9	5.4	0.1	1.2
ME02 - Positions 2 & 3	8,140	8,099	43.1	41.8	6.8	6.5	0.1	2.0
ME03 - Positions 1 & 4	8,134	8,140	46.0	43.3	4.1	6.4	0.1	1.9
ME04 - Positions 2 & 3	8,164	8,134	43.0	39.9	6.5	7.2	0.1	2.9
ME05 - Positions 1 & 4	8,153	8,164	47.0	45.1	3.3	5.6	0.0	1.1
ME06 - Positions 2 & 3	8,095	8,153	42.9	41.4	8.0	8.4	0.0	2.0
ME07 - Positions 1 & 4	8,090	8,095	49.6	47.5	3.7	5.3	0.0	0.6
ME08 - Positions 2 & 3	8,120	8,090	40.7	38.7	7.5	8.2	0.1	2.1
ME09 - Positions 1 & 4	8,138	8,120	37.0	35.5	5.3	8.3	0.1	1.0
ME10 - Positions 2 & 3	8,106	8,137	43.6	41.1	8.2	10.5	0.1	2.7
ME11 - Positions 1 & 4	8,101	8,106	46.4	44.4	4.0	5.8	0.1	1.2
ME12 - Positions 2 & 3	8,187	8,101	40.0	37.3	8.3	9.8	0.2	2.7
ME13 - Positions 1 & 4	8,188	8,187	47.2	42.9	3.4	5.6	0.0	1.4
ME14 - Positions 2 & 3	8,152	8,188	43.5	42.7	4.4	4.7	0.1	1.6
Overall	113,867	113,866	44.0	41.9	5.5	7.0	0.1	1.8

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 8 Mathematics (paperTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
MP01 - Positions 1 & 4	8,070	8,086	39.0	36.9	5.1	6.2	0.1	2.0
MP02 - Positions 2 & 3	8,026	8,070	32.0	31.9	9.2	8.7	0.1	2.3
MP03 - Positions 1 & 4	8,045	8,026	37.7	33.2	7.7	9.9	0.1	2.9
MP04 - Positions 2 & 3	8,063	8,045	35.0	34.0	9.9	9.3	0.2	3.2
MP05 - Positions 1 & 4	8,053	8,063	41.4	38.5	5.7	7.6	0.1	1.9
MP06 - Positions 2 & 3	8,083	8,053	31.4	31.9	8.8	8.4	0.1	1.7
MP07 - Positions 1 & 4	8,088	8,083	38.4	36.8	5.7	7.2	0.0	1.0
MP08 - Positions 2 & 3	8,052	8,088	31.3	30.5	10.0	9.7	0.1	1.4
MP09 - Positions 1 & 4	8,051	8,052	30.6	28.7	7.6	8.9	0.1	1.5
MP10 - Positions 2 & 3	8,107	8,051	34.7	33.8	11.0	11.7	0.2	1.9
MP11 - Positions 1 & 4	8,118	8,107	35.9	33.6	6.6	8.5	0.1	1.9
MP12 - Positions 2 & 3	8,113	8,118	30.3	29.1	9.9	10.4	0.1	2.9
MP13 - Positions 1 & 4	8,087	8,113	36.7	32.7	6.2	6.5	0.1	1.9
MP14 - Positions 2 & 3	8,086	8,088	33.6	33.5	6.7	6.7	0.2	2.0
Overall	113,042	113,043	34.9	33.2	7.9	8.5	0.1	2.0

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 8 Mathematics

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Australia	9,002	9,002	49.4	48.3	5.0	5.7	0.2	1.4
Bahrain	5,724	5,724	39.3	37.4	11.5	11.8	0.0	0.8
Chile	4,100	4,100	28.5	25.7	10.8	13.2	0.1	4.0
Chinese Taipei	4,914	4,914	66.3	65.7	1.5	2.0	0.0	0.2
Cyprus	3,515	3,515	42.5	40.6	6.6	7.6	0.0	1.1
Egypt	7,201	7,201	25.7	24.0	7.9	8.0	0.2	2.4
England	3,345	3,345	43.2	41.1	7.1	8.9	0.1	1.5
Finland	4,835	4,835	40.4	38.8	5.5	6.9	0.2	1.6
France	3,869	3,869	35.1	32.9	8.9	11.3	0.0	2.4
Georgia	3,309	3,309	31.5	29.1	11.5	13.5	0.1	2.3
Hong Kong SAR	3,255	3,255	58.9	58.3	2.8	3.2	0.2	0.8
Hungary	4,559	4,559	46.0	44.8	4.7	5.3	0.0	0.2
Iran, Islamic Rep. of	5,975	5,975	32.3	29.5	10.5	12.4	0.1	3.3
Ireland	4,109	4,109	48.2	47.2	5.4	6.1	0.3	1.5
Israel	3,725	3,725	45.2	42.0	5.8	7.4	0.0	2.0
Italy	3,618	3,618	38.8	36.2	7.3	9.3	0.1	2.5
Japan	4,444	4,444	66.0	65.0	2.2	2.8	0.1	0.6
Jordan	7,172	7,172	25.3	23.3	5.5	6.6	0.1	1.6
Kazakhstan	4,447	4,447	38.8	38.0	9.4	10.1	0.1	2.2
Korea, Rep. of	3,858	3,858	65.3	64.4	1.9	2.4	0.1	0.3
Kuwait	4,569	4,569	23.4	21.8	6.0	6.4	0.2	2.5
Lebanon	4,724	4,724	25.9	24.4	18.3	19.1	0.3	4.8
Lithuania	3,823	3,823	42.8	41.2	5.8	6.5	0.0	0.4
Malaysia	7,065	7,065	37.7	35.5	1.7	2.6	0.0	1.8
Morocco	8,431	8,431	19.0	18.2	15.3	14.7	0.1	2.4
New Zealand	6,025	6,025	42.4	41.3	6.0	6.8	0.2	1.7
Norway (9)	4,541	4,541	41.1	38.1	9.6	11.9	0.4	3.7
Oman	6,745	6,745	26.6	24.6	5.6	6.2	0.1	3.0
Portugal	3,369	3,369	40.2	37.2	5.8	7.6	0.0	1.4

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 8 Mathematics (continued)

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Qatar	3,882	3,882	29.7	27.3	4.8	6.5	0.1	2.5
Romania	4,485	4,485	41.9	38.9	9.9	10.8	0.1	2.2
Russian Federation	3,900	3,900	50.6	48.2	6.8	8.4	0.0	2.0
Saudi Arabia	5,680	5,680	23.7	21.5	4.8	5.6	0.0	1.0
Singapore	4,845	4,845	67.0	66.2	1.1	1.4	0.0	0.4
South Africa (9)	20,796	20,796	21.4	20.9	4.2	4.5	0.0	2.6
Sweden	3,970	3,970	41.6	37.7	7.5	10.7	0.2	4.2
Turkey	4,075	4,075	38.4	37.0	5.6	6.8	0.0	0.7
United Arab Emirates	22,327	22,326	34.1	31.2	3.3	4.5	0.0	1.6
United States	8,683	8,683	44.8	42.5	1.9	2.7	0.1	2.5
International Average	226,911	226,910	40.0	38.1	6.6	7.6	0.1	1.9
Ontario, Canada	3,764	3,764	46.6	43.5	4.0	5.7	0.2	3.5
Quebec, Canada	3,173	3,173	50.7	47.7	3.7	5.4	0.2	2.6
Moscow City, Russian Fed.	3,780	3,780	59.1	56.3	5.6	7.4	0.0	1.2
Gauteng, RSA (9)	5,621	5,621	23.0	22.4	3.4	3.6	0.0	2.1
Western Cape, RSA (9)	5,340	5,340	27.0	26.4	4.4	4.8	0.0	2.3
Abu Dhabi, UAE	8,201	8,201	29.0	26.2	2.9	3.9	0.1	1.5
Dubai, UAE	5,726	5,726	47.4	44.3	3.0	4.6	0.0	1.6

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 8 Science (eTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
SE01 - Positions 2 & 3	8,074	8,160	52.5	51.4	5.6	6.7	0.1	1.2
SE02 - Positions 1 & 4	8,155	8,074	48.9	46.8	2.9	4.2	0.1	0.7
SE03 - Positions 2 & 3	8,124	8,155	45.0	43.7	5.1	5.4	0.2	0.8
SE04 - Positions 1 & 4	8,176	8,124	47.8	45.0	5.3	7.4	0.1	0.7
SE05 - Positions 2 & 3	8,149	8,176	48.6	48.5	5.1	5.8	0.1	0.8
SE06 - Positions 1 & 4	8,100	8,149	43.1	40.7	8.6	10.4	0.0	0.7
SE07 - Positions 2 & 3	8,080	8,100	45.1	44.6	3.9	4.2	0.1	0.6
SE08 - Positions 1 & 4	8,123	8,080	47.6	45.9	2.6	4.5	0.0	0.8
SE09 - Positions 2 & 3	8,126	8,124	39.8	39.4	4.5	4.8	0.1	1.0
SE10 - Positions 1 & 4	8,124	8,126	48.1	46.9	2.2	3.3	0.1	0.5
SE11 - Positions 2 & 3	8,104	8,124	49.4	48.5	3.8	4.1	0.1	1.1
SE12 - Positions 1 & 4	8,195	8,104	51.3	48.4	1.5	2.9	0.0	0.8
SE13 - Positions 2 & 3	8,179	8,195	52.8	52.5	2.8	3.3	0.1	1.4
SE14 - Positions 1 & 4	8,159	8,179	48.8	45.8	2.0	3.7	0.1	0.6
Overall	113,868	113,870	47.8	46.3	4.0	5.0	0.1	0.8

**TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 8 Science
(paperTIMSS)**

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
SP01 - Positions 2 & 3	8,068	8,079	43.2	42.0	7.8	8.4	0.1	1.7
SP02 - Positions 1 & 4	8,039	8,068	40.7	38.4	6.0	6.8	0.1	1.3
SP03 - Positions 2 & 3	8,042	8,039	38.7	37.9	6.1	6.7	0.1	2.1
SP04 - Positions 1 & 4	8,063	8,042	39.1	36.4	8.9	10.4	0.1	1.9
SP05 - Positions 2 & 3	8,055	8,063	39.3	38.8	8.3	9.2	0.1	1.9
SP06 - Positions 1 & 4	8,080	8,055	36.9	34.9	10.9	12.5	0.0	1.7
SP07 - Positions 2 & 3	8,092	8,080	38.6	38.6	6.8	6.3	0.1	1.0
SP08 - Positions 1 & 4	8,064	8,092	39.9	37.3	5.5	7.0	0.1	1.1
SP09 - Positions 2 & 3	8,049	8,064	33.4	33.1	7.0	6.9	0.1	1.5
SP10 - Positions 1 & 4	8,105	8,049	40.9	39.7	5.1	5.7	0.0	1.3
SP11 - Positions 2 & 3	8,105	8,105	38.1	37.8	7.1	7.7	0.1	1.3
SP12 - Positions 1 & 4	8,115	8,104	42.5	39.5	4.0	4.8	0.0	1.3
SP13 - Positions 2 & 3	8,092	8,115	42.7	41.4	6.4	7.1	0.2	2.9
SP14 - Positions 1 & 4	8,080	8,092	40.3	37.0	4.0	5.8	0.0	2.1
Overall	113,049	113,047	39.6	38.1	6.7	7.5	0.1	1.6

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 8 Science

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Australia	9,002	9,002	53.7	52.2	3.8	4.0	0.2	1.1
Bahrain	5,719	5,719	43.3	41.4	4.9	5.9	0.0	0.7
Chile	4,097	4,097	38.8	36.5	6.5	8.6	0.1	1.6
Chinese Taipei	4,910	4,910	59.1	59.1	1.9	2.1	0.0	0.1
Cyprus	3,520	3,520	42.1	40.2	5.7	6.3	0.0	1.0
Egypt	7,200	7,199	28.9	27.4	9.3	9.6	0.1	2.0
England	3,358	3,358	48.0	46.0	4.0	5.4	0.2	1.0
Finland	4,843	4,843	53.1	52.1	3.2	3.8	0.2	0.7
France	3,870	3,870	42.5	40.3	5.4	7.2	0.1	1.2
Georgia	3,308	3,308	34.2	31.9	9.8	12.6	0.1	1.4
Hong Kong SAR	3,253	3,253	44.7	42.8	3.6	4.8	0.2	0.7
Hungary	4,558	4,558	52.0	51.2	3.8	4.2	0.0	0.0
Iran, Islamic Rep. of	5,976	5,976	37.3	34.9	8.1	9.5	0.1	2.1
Ireland	4,097	4,097	49.9	49.2	3.8	4.1	0.2	1.1
Israel	3,721	3,721	47.2	45.3	4.2	5.3	0.0	0.6
Italy	3,618	3,618	44.5	42.7	5.3	6.2	0.0	0.9
Japan	4,442	4,442	58.9	58.3	2.1	2.6	0.0	0.4
Jordan	7,174	7,174	36.6	34.8	4.8	5.9	0.1	1.1
Kazakhstan	4,453	4,453	40.9	39.7	8.3	9.5	0.1	1.9
Korea, Rep. of	3,858	3,858	55.5	54.7	2.0	2.5	0.0	0.1
Kuwait	4,569	4,569	35.5	34.4	5.3	6.0	0.1	1.5
Lebanon	4,714	4,714	26.3	24.6	17.1	18.6	0.1	3.6
Lithuania	3,823	3,823	48.9	47.9	3.5	4.2	0.0	0.0
Malaysia	7,064	7,064	42.5	41.3	1.1	1.6	0.0	0.8
Morocco	8,444	8,444	26.7	25.7	13.7	14.7	0.0	1.8
New Zealand	6,021	6,021	48.9	47.2	4.3	5.0	0.1	1.2
Norway (9)	4,538	4,538	44.2	42.6	6.3	8.1	0.3	1.7
Oman	6,745	6,745	38.9	37.4	4.4	4.9	0.1	1.9
Portugal	3,362	3,362	48.3	46.9	3.8	4.9	0.0	0.3
Qatar	3,881	3,881	41.2	39.0	3.7	5.0	0.1	1.2
Romania	4,489	4,489	42.0	39.6	8.6	9.9	0.1	1.4
Russian Federation	3,899	3,899	52.6	51.5	4.7	5.6	0.0	0.7

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 8 Science (continued)

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Saudi Arabia	5,678	5,678	35.2	32.9	5.0	5.8	0.0	0.7
Singapore	4,848	4,848	65.1	64.5	0.8	0.9	0.0	0.1
South Africa (9)	20,807	20,807	27.6	26.5	4.1	5.5	0.1	4.3
Sweden	3,974	3,974	50.2	48.1	5.1	6.9	0.2	1.9
Turkey	4,077	4,077	46.4	45.2	4.2	5.2	0.0	0.5
United Arab Emirates	22,322	22,324	40.5	38.7	3.1	4.2	0.0	0.9
United States	8,686	8,686	50.1	48.3	1.5	2.1	0.2	1.6
International Average	226,918	226,919	44.2	42.6	5.1	6.1	0.1	1.2
Ontario, Canada	3,767	3,767	47.1	45.5	2.8	3.8	0.1	1.6
Quebec, Canada	3,170	3,170	51.5	49.8	2.2	2.9	0.0	0.5
Moscow City, Russian Fed.	3,783	3,783	57.4	55.9	3.4	4.1	0.0	0.3
Gauteng, RSA (9)	5,629	5,629	30.6	29.4	2.8	3.6	0.1	2.5
Western Cape, RSA (9)	5,339	5,339	33.5	32.2	3.4	4.4	0.0	2.9
Abu Dhabi, UAE	8,197	8,198	34.3	32.1	3.4	4.7	0.0	0.9
Dubai, UAE	5,726	5,726	52.9	51.6	1.9	2.8	0.0	0.8

Appendix 10E: Modifications to the TIMSS 2019 Achievement Data

Grade 4 Mathematics	
Items Deleted for All Countries	
M02_10B – ME71217B, MP71217B	(severe differential item functioning)
M04_10A – ME71135A, MP71135A	(severe differential item functioning)
M08_09 – ME71199, M08_08 – MP71199	(severe differential item functioning)
M10_01 – ME71005, MP71005	(severe differential item functioning)
Items Recoded for All Countries	
M05_12 – ME51080, MP51080	(20 to 10, 10 to 71, 11 to 72)
M10_11 – ME71189, MP71189	(20 to 10, 10 to 79, 11 to 79)
M12_11 – ME71190, M12_10 – MP71190	(20 to 10, 10 to 70)
M13_02 – ME61254, MP61254	(20 to 10, 10 to 70)
M13_08 – ME61224, MP61224	(70 to 12)
M14_09 – ME71177, MP71177	(20 to 10, 10 to 70)
Items Deleted by Country	
Chile	
M05_12 – ME51080	(poor discrimination)
Croatia	
M12_03 – ME71062	(negative discrimination)
Hungary	
M12_03 – ME71062	(negative discrimination)
Japan	
M08_12 – MP71194	(translation error)
M12_04B – MP71216B	(translation error)
M12_10 – MP71202	(translation error)
Korea	
M11_01 – ME61178, MP61178	(severe item-by-country interaction)
Netherlands	
M10_08 – ME71179	(derived item, poor discrimination)

Items beginning with “ME” are eTIMSS items. Items beginning with “MP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also modified for eTIMSS bridge samples.

Grade 4 Mathematics – Less Difficult

Items Deleted for All Countries

MP02_10B – MP71217B (severe differential item functioning)

MP08_08 – MP71199 (severe differential item functioning)

Items Recoded for All Countries

MP13_02 – MP61254 (20 to 10, 10 to 70)

MP13_08 – MP61224 (70 to 12)

Items Deleted by Country

Bosnia and Herzegovina (Cyrillic language only)

MP13_03, MP61244 (translation error)

Morocco

MN11_09 – MN11158 (poor reliability)

Saudi Arabia

MP13_01 – MP61240 (derived item, translation error)

Items beginning with “MP” are items shared with the regular fourth grade mathematics assessment. Items beginning with “MN” are items unique to less difficult mathematics.

Grade 4 Science

Items Deleted for All Countries

S05_02 – SE51020, SP51020 (poor discrimination)
S06_05 – SE61166, SP61166 (poor discrimination)
S07_03C – SE51138C, SP51138C (poor discrimination)
S08_01 – SE71091, SP71091 (severe differential item functioning)
S10_11 – SP71921 (poor discrimination)
S12_09 – SE71910, SP71910 (severe differential item functioning)
S13_01 – SE61125, SP61125 (poor discrimination)
S14_03 – SE71021, SP71021 (severe differential item functioning)

Items Recoded for All Countries

S12_01 – SE71031, SP71031 (11 to 70)
S13_02 – SE61014, SP61014 (20 to 10, 10 to 70)

Items Deleted by Country

Azerbaijan (Azerbaijani language only)

S10_06 – SP71080 (translation error)

Bosnia and Herzegovina (Serbian language only)

S04_08 – SP71102 (translation error)

Chile

S10_11 – SE71921 (negative discrimination)
--

France

S07_02 – SE51051 (negative discrimination)
--

Georgia

S13_03 – SE61056 (poor discrimination)
--

Germany

S01_03A – SE51132A, SP51132A (translation error)
--

Kosovo

S03_03 – SP61054 (poor reliability)
S13_02 – SP61014 (poor reliability)

Morocco

S03_03 – SP61054 (poor reliability)

Netherlands

S01_06 – SP51063 (negative discrimination)
--

Portugal

S10_11 – SE71921 (negative discrimination)
--

Russian Federation

S07_07 – SE51200 (severe item-by-country interaction)

Saudi Arabia

S11_06 – SP61093 (poor reliability)
S13_11 – SP61124 (derived item, poor discrimination)

Items beginning with “SE” are eTIMSS items. Items beginning with “SP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also modified for eTIMSS bridge samples.

Grade 8 Mathematics

Items Deleted for All Countries

M06_07 – ME62342, MP62342 (poor discrimination)

M09_12B – ME62345B, MP62345B (derived item, poor discrimination)

M10_03 – ME72038, MP72038 (severe differential item functioning)

M12_14B – ME72211B, MP72211B (severe differential item functioning)

M13_12 – ME62048, MP62048 (derived item, poor discrimination)

Items Recoded for All Countries

M03_13 – ME62254, MP62254 (20 to 10)

M07_08 – ME52087, MP52087 (20 to 10, 10 to 70)

M08_09B – ME72128B, MP72128B (10 to 20, 70 to 10)

Items Deleted by Country

Georgia

M02_03 – ME72017 (poor discrimination)

M03_06 – MP62351 (negative discrimination)

Kazakhstan (Kazakh language only)

M10_15 – MP72206 (translation error)

M12_09A – MP72110A (translation error)

M12_09B – MP72110B (translation error)

M12_12 – MP72229 (translation error)

Lebanon

M09_07 – MP62350 (negative discrimination)

Saudi Arabia

M03_02 – MP62139 (translation error)

M05_12 – MP52502 (translation error)

Items beginning with “ME” are eTIMSS items. Items beginning with “MP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also modified for eTIMSS bridge samples.

Grade 8 Science

Items Deleted for All Countries

S01_06 – SE52134 (severe differential item functioning)
S03_12 – SE62272, SP62272 (poor discrimination)
S04_02 – SP72403 (severe differential item functioning)
S05_11 – SE52221, SP52221 (poor discrimination)
S08_09 – SE72133, SP72133 (severe differential item functioning)
S10_07 – SE72048, SP72048 (severe differential item functioning)
S11_12 – SE62036, SP62036 (attractive distracter)
S11_15C – SE62242C, SP62242C (poor discrimination)
S12_04 – SE72906, SP72906 (derived item, severe differential item functioning)
S12_15 – SE72329, SP72329 (severe differential item functioning)
S13_05 – SE62266, SP62266 (attractive distracter)

Items Recoded for All Countries

S12_09 – SE72523, SP72523 (10 to 20, 11 to 10)
S12_13A – SE72280A, SP72280A (20 to 10, 10 to 70)

Items Deleted by Country

England

S09_03 – SE62106 (translation error)

Egypt

S10_16 – SP72720 (negative discrimination)
--

Iran, Islamic Rep. of

S05_05 – SP52248 (negative discrimination)
--

Japan

S10_09 – SP72116 (translation error)
S14_16 – SP72303 (translation error)

Jordan

S01_06 – SP52134 (negative discrimination)
--

Morocco

S01_06 – SP52134 (negative discrimination)
S10_14 – SP72220 (negative discrimination)

Saudi Arabia

S02_06 – SP72103 (printing error)
S03_04 – SP62225 (item not administered)
S04_08B – SP72141B (low reliability)
S06_04A – SP62098A (low reliability)

South Africa, including Gauteng and Western Cape

S06_05 – SP62032 (poor discrimination)
--

Items beginning with “SE” are eTIMSS items. Items beginning with “SP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also modified for eTIMSS bridge samples.

Appendix 10F: Derived Items in TIMSS 2019

Grade 4 Mathematics

M01_01 – ME51043: Item parts A, B, C, D, E, F, G, and H are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M01_05 – ME51508: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct

M02_03 – ME71167: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M02_05 – ME71162, MP71162: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both parts are correct and 1 score point is awarded if 1 part is correct

M02_06 – ME71078: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M02_08 – ME71151, MP71151: 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

M02_11 – ME71142: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct

M02_12 – ME71204, MP71024: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M04_03 – ME71036, MP71036: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct

M04_09 – ME71178, MP71178: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M04_12 – ME71175, MP71175: 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 1 or 2 are correct

M06_01 – ME61018, MP61018: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M06_10 – ME61266: Item parts A, B, C, D, E, and F 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 5 parts are correct

M08_11 – ME71141, M08_10 – MP71141: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M08_12 – ME71194: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct

M08_13 – ME71193, M08_12 – MP71193: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both are correct and 1 score point is awarded if 1 part is correct

M10_05 – ME71213: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M10_08 – ME71179, MP71179: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M10_12A – ME71187A: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

Grade 4 Mathematics (continued)

M11_08 – ME61095: Item parts B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)

M12_04A – ME71216A, MP71216A: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct

M12_05 – ME71117: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M12_10 – ME71202: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M13_01 – ME61240, MP61240: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M13_02 – ME61254: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M14_11A – ME71138A, MP71128A: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M14_13 – ME71205, MP71205: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

Items beginning with “ME” are eTIMSS items. Items beginning with “MP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also derived for eTIMSS bridge samples.

Grade 4 Mathematics – Less Difficult

MN04_14 – MN21003: Item parts A, B, C, and D are combined to create a 2-point item, where 2 score points are awarded if all parts correct and 1 score point is awarded if 3 parts are correct

MN14_10 – MN21057: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

MP02_05 – MP71162: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both parts are correct and 1 score point is awarded if 1 part is correct

MP02_08 – MP71151: 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

MP02_12 – MP71024: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

MP08_10 – MP71141: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

MP08_12 – MP71193: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both are correct and 1 score point is awarded if 1 part is correct

MP13_01 – MP61240: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

Items beginning with “MP” are items shared with the regular fourth grade mathematics assessment. Items beginning with “MN” are items unique to less difficult mathematics.

Grade 4 Science

S02_03 – SE71017, SP71017: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S04_02 – SE71902, SP71902: Item parts B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)

S04_04 – SE71041, SP71041: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both parts are correct and 1 score point is awarded if 1 part is correct

S04_05 – SE71046, SP71046: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S05_10 – SE51151: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S06_06 – SE61083, SP61083: Item parts B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)

S06_09A – SE61142A, SP61142A: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S07_03 – SE51138Z, SP51138Z: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct (part C was deleted)

S09_08 – SE61160: Item parts B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)

S10_01 – SE71009, SP71009: Item parts A, B, C, D, E, and F 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 or 5 parts are correct

S10_09 – SE71106, SP71006: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S10_13 – SE71254: Item parts A, B, C, D, E, F, G, and H are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S13_11 – SE61124, SP61124: Item parts B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)

S13_12 – SE61116, SP61116: Item parts B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)

S14_01 – SE71063: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S14_08 – SE71114: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

Items beginning with “SE” are eTIMSS items. Items beginning with “SP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also derived for eTIMSS bridge samples.

Grade 8 Mathematics

M02_01 – ME72007, MP72007: 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

M02_11 – ME72180: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M02_12 – ME72198, MP72198: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct

M02_14 – ME72170: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M03_10 – ME62244: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct

M04_01 – ME72178: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M04_03 – ME72020: 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

M04_05 – ME72052, MP72052: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct

M04_11 – ME72164, MP72164: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M05_12 – ME52502: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M06_10 – ME62288: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both parts are correct and 1 score point is awarded if 1 part is correct

M07_08 – ME52087: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both parts are correct and 1 score point is awarded if 1 part is correct

M08_04 – ME72055: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M09_06 – ME62317: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M09_12A – ME62345A: 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 2 or 3 parts are correct

M10_09 – ME72095, MP72095: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct,

M10_14 – ME72232, MP72232: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M11_03 – ME62215: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both are correct and 1 score point is awarded if 1 part is correct

M12_08 – ME72225, MP72225: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct

M13_09 – ME62170: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both parts are correct and 1 score point is awarded if 1 part is correct

Grade 8 Mathematics (continued)

M14_09 – ME72081: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

M14_10 – ME72140, MP72140: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

Items beginning with “ME” are eTIMSS items. Items beginning with “MP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also derived for eTIMSS bridge samples.

Grade 8 Science

S01_05 – SE52095Z, SP52095Z: Item parts B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)

S02_08 – SE72130, SP72130: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S02_11 – SE72232, SP72232: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S04_02 – SE72403: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S04_13 – SE72345, SP72345: Item parts A, B, C, D, E, F, and G 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 5 or 6 parts are correct

S06_13A – SE62173A, SP62173A: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S07_05 – SE52015Z, SP52015Z: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S08_02 – SE72400: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S08_13 – SE72260: Item parts A, B, C, D, E, F, and G are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S08_14 – SE72265, SP72265: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S09_08 – SE62018, SP62018: 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

S10_01 – SE72033: 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

S10_05 – SE72086: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S10_13 – SE72261, SP72261: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S11_06 – SE62006: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S11_15 – SE62242, SP62242: Item parts A, B, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part C was deleted)

Grade 8 Science (continued)

S12_03 – SE72000: 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

S12_08 – SE72143: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S13_04 – SE62101: 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 2 or 3 parts are correct

S13_07 – SE62047, SP62047: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S13_08 – SE62042: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S13_14 – SE62022, SP62022: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S13_15 – SE62243: 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 2 or 3 parts are correct

S14_02 – SE72905: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

S14_04 – SE72016, SP72016: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both are correct and 1 score point is awarded if 1 part is correct

Items beginning with “SE” are eTIMSS items. Items beginning with “SP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also derived for eTIMSS bridge samples.