

CHAPTER 17

Scaling and Reporting the TIMSS 2019 Problem Solving and Inquiry (PSI) Data

Bethany Fishbein
Pierre Foy

Introduction

TIMSS 2019 marked the beginning of the transition to eTIMSS—a digital version of the TIMSS assessment that was administered to fourth and eighth grade students on computers and tablets. An important step in TIMSS’ evolution to fully digital assessment included developing innovative achievement measures to include better coverage of problem solving and inquiry processes designed to capitalize on the digital environment to its fullest. eTIMSS 2019 included eight Problem Solving and Inquiry tasks (PSIs)—two in mathematics and two in science at each of the fourth and eighth grades. At each grade, two additional “eBooklets” containing the four PSIs were administered together with eTIMSS according to a rotated design. Over half the TIMSS 2019 participating countries administered eTIMSS—30 countries and 6 benchmarking entities at the fourth grade, and 22 countries and 5 benchmarking entities at the eighth grade. The rest administered paperTIMSS, consistent with previous TIMSS cycles.

The mathematics and science achievement results published for eTIMSS countries in the *TIMSS 2019 International Results and Mathematics and Science* report were based on only the “regular” eTIMSS items which had paper-equivalent versions in the paperTIMSS assessment. [Chapter 12](#) of this volume documents the psychometric modeling and analysis steps that enabled linking the eTIMSS and paperTIMSS results to the TIMSS trend reporting scales. The current chapter describes the additional analysis efforts by the TIMSS & PIRLS International Study Center to calibrate the PSI items on the same scale and provide achievement results for the students who were administered the TIMSS 2019 PSIs. The “eTIMSS with PSI” achievement results are reported in *Findings from the TIMSS 2019 Problem Solving and Inquiry Tasks*.

Following the same approach described in [Chapter 11](#) and [Chapter 12](#), scaling the PSI data and providing accurate measures of PSI student proficiency relied on Item Response Theory (IRT) scaling and latent regression population models with multiple imputation. This process was preceded by an in-depth analysis of the item response data collected through the PSI items.

Characteristics of National eTIMSS Samples with Problem Solving and Inquiry

Including the PSI items on the eTIMSS scale and reporting the mathematics and science achievement results included in *Findings from the TIMSS 2019 Problem Solving and Inquiry Tasks* required a different set of sampling weights to those included in the [TIMSS 2019 data files](#) and used to produce the results published in *TIMSS 2019 International Results and Mathematics and Science*. The “TIMSS+PSI weights” incorporate students who took the PSI eBooklets into the results, while the “TIMSS weights” include only students who took regular items. [Chapter 3](#) describes how both “TIMSS weights” and “TIMSS+PSI weights” were calculated based on the same student samples.

In this chapter, Appendix 17A presents the national sampling outcomes based on the “TIMSS+PSI weights” using the full samples of students assessed in each participating eTIMSS country—both students administered regular eTIMSS items and students administered PSIs. [Chapter 9](#) of this technical report gives a summary of the major characteristics of the national samples for TIMSS 2019, with detailed descriptions of the sample design for each country including information about population coverage and exclusions, stratification variables, and schools’ sampling allocations. For countries that participated in eTIMSS, the summary statistics in Chapter 9 are based only on the students who were administered eBooklets with regular eTIMSS items—using the “TIMSS weight”—consistent with the results in the *TIMSS 2019 International Results and Mathematics and Science* report.

As might be anticipated, the addition of the PSI students did not have any substantial impact on the sampling outcomes. Readers should refer back to [Chapter 9](#) for additional, in-depth information about the national samples including interpretation of the outcomes.

Reviewing the Problem Solving and Inquiry Item Statistics

Before putting the TIMSS 2019 Problem Solving and Inquiry (PSI) data on the TIMSS 2019 scale, the TIMSS & PIRLS International Study Center conducted an extensive review of descriptive item statistics for all countries. As described in [Chapter 10](#), this item-by-item, country-by-country review plays a crucial role in the quality assurance of the TIMSS achievement measures and helps identify any unexpected or problematic item properties that may be cause for excluding an item from achievement scaling. As part of this process for all TIMSS 2019 items, the TIMSS & PIRLS International Study Center computed item statistics for 29 mathematics PSI items and 19 science PSI items at the fourth grade, and for 25 mathematics PSI items and 29 science PSI items at the eighth grade.

For each country, the inter-scorer agreement for each human-scored PSI item 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 17B presents the average and range of the within-country percentages

of score point agreement and diagnostic code agreement across the PSI items. Exact percent agreement across PSI items was high on average across countries—97 percent in mathematics at the fourth grade, 98 percent in mathematics at the eighth grade, and 93 percent in science at both grades. There also was high agreement at the diagnostic score level across the grades and subjects, where international average percent agreement ranged from 93 percent in fourth grade science to 98 percent in eighth grade mathematics. Outcomes for the regular eTIMSS items as well as paperTIMSS items can be found in [Chapter 10](#).

The item review identified a small number of PSI items with poor psychometric properties, warranting exclusion from the achievement scaling. While the items did not contribute to student proficiency, their response data remain in the [TIMSS 2019 International Database](#), providing additional opportunities for analysis, such as exploring what types of interactive items were effective for assessing fourth and eighth grade students. Appendix 17C includes the list of PSI items that were excluded for creating the “eTIMSS with PSI” mathematics and science scale scores at each grade, as well as a list of recodes made to human-scored constructed response items. Appendix 17D provides details about how score points were awarded for “derived items” that were combined for scoring purposes. The same information for regular eTIMSS items can be found in [Chapter 10](#) in Appendices 10E and 10F, respectively.

Examining Non-Response and Item Position Effects for the Problem Solving and Inquiry Tasks

As described in the [TIMSS 2019 Assessment Design](#) (Martin, Mullis, & Foy, 2017), the eTIMSS assessment items (including PSIs) for each grade and subject are arranged in 16 item blocks which were assembled into digital achievement “eBooklets.” eBooklets consist of two parts (halves)—one part with two blocks of mathematics items and the other with two blocks of science items—and students get a break in between. Each item block appears in two of the 16 eBooklets, 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. eTIMSS 2019 included 14 eBooklets with blocks of regular eTIMSS items (i.e., with paperTIMSS equivalents) and two eBooklets with PSI tasks formed into item blocks.

Compared to regular, non-PSI items in eBooklets 1 to 14, the item statistics for the PSIs showed that students assigned PSI tasks (eBooklets 15 and 16) were not always completing their booklets at the same high rates. Following TIMSS’ standard data cleaning procedures, if a student omitted two items in a row and all subsequent items in the booklet part were also blank, the second omitted item and all subsequent items were coded in the data as “not reached.”

Because eBooklets 15 and 16 contain the same PSI blocks but in different orders, it was possible to investigate whether the lower completion rates were related to the block position in the booklet, or more specifically if a block presented earlier in a session had higher completion than a block later in a session. 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 for the PSI tasks are reported in Appendix 17E, averaged across countries for each item block, and averaged across item blocks for each country (Appendix 10D in [Chapter 10](#) includes results for regular eTIMSS items). A summary of results with the average differences in item statistics between the booklet positions is provided in Exhibits 17.1 and 17.2 for the fourth grade and eighth grade, respectively, comparing the regular eTIMSS averages with the PSI averages.

Exhibit 17.1: Summary of International Average Item Block Statistics by Booklet Position—Grade 4

	Average Percent Correct Across Items (Weighted)			Average Percent Omitted Across Items (Weighted)			Average Percent Not Reached Across Items (Weighted)		
	Positions 1 & 3	Positions 2 & 4	Difference	Positions 1 & 3	Positions 2 & 4	Difference	Positions 1 & 3	Positions 2 & 4	Difference
Mathematics									
Regular eTIMSS	49.9	48.3	-1.6	3.3	4.5	1.1	0.1	2.5	2.5
PSI	40.6	36.0	-4.6	10.4	12.0	1.6	0.4	11.9	11.5
Science									
Regular eTIMSS	53.3	52.0	-1.3	3.2	4.4	1.2	0.1	2.5	2.5
PSI	45.8	42.5	-3.3	6.6	8.1	1.5	0.4	12.4	11.9

Exhibit 17.2: Summary of International Average Item Block Statistics by Booklet Position—Grade 8

	Average Percent Correct Across Items (Weighted)			Average Percent Omitted Across Items (Weighted)			Average Percent Not Reached Across Items (Weighted)		
	Positions 1 & 3	Positions 2 & 4	Difference	Positions 1 & 3	Positions 2 & 4	Difference	Positions 1 & 3	Positions 2 & 4	Difference
Mathematics									
Regular eTIMSS	44.0	41.9	-2.1	5.5	7.0	1.5	0.1	1.8	1.7
PSI	30.6	28.1	-2.5	11.8	12.9	1.1	0.3	5.5	5.2
Science									
Regular eTIMSS	47.8	46.3	-1.5	4.0	5.0	1.1	0.1	0.8	0.7
PSI	45.0	42.5	-2.5	7.0	8.7	1.7	0.2	2.4	2.2

Compared to the results for regular eTIMSS items which indicate minimal impact of block position on the item statistics, the results for PSI items indicate a definite presence of position effects occurring in the PSI booklets, particularly in mathematics. On average, PSI item blocks appearing in the second half of a booklet part (positions 2 and 4) were more difficult and had more not-reached responses than item blocks appearing the first half of a booklet part (positions 1 and 3). Across fourth grade countries, differences in average PSI item percent correct between positions 2/4 and positions 1/3 were -3.3 for science and -4.6 in mathematics (compared to -1.3 and -1.6 for regular items, respectively). At the eighth grade, both mathematics and science PSIs had an average percent correct difference of -2.5 .

Differences in the average percent of not-reached responses between block positions were of greatest concern for the PSI items. At the fourth grade, PSI items had 11.5 percent more not-reached responses in positions 2/4 than in positions 1/3 for mathematics, and 11.9 percent more for science, on average. In comparison, regular items had differences of 2.5 percent in both subjects. The differences were smaller at the eighth grade, with average differences for PSI items of 5.2 percent in mathematics and 2.2 percent in science, and average differences for regular items of 1.7 percent and 0.7 percent, respectively. The [Findings from the TIMSS 2019 Problem Solving and Inquiry Tasks](#) (Appendices A and B) report describes an in-depth investigation into the lower completion rates by students with PSI eBooks, which incorporated process data to understand students' test-taking behaviors.

Treatment of Omitted and Not Reached Items in Scaling and Reporting

TIMSS typically treats not-reached item responses differently in estimating item parameters and in generating student plausible values. In estimating the item parameters, items in the assessment booklets that are considered not to have been reached by students are treated as if they were not administered. Subsequently, when estimating expected a posteriori (EAP) ability estimates and generating plausible values, not-reached items are recoded and treated as incorrect. This treatment is known to introduce an underestimation of achievement if a large percentage of responses is affected (Glas & Pimentel, 2006; Moustaki & Knott, 2000; Rose, von Davier, & Xu, 2010).

However, considering the relatively large percentages of students not reaching all of the items in the PSI blocks, it was clear that not all students had an opportunity to answer all of the items and that the usual TIMSS practice of treating not reached item responses as incorrect could introduce bias into the reported results. To avoid this situation, all not-reached responses on PSI items were considered to be “not administered” rather than as incorrect. Accordingly, the base of EAP ability estimates and plausible values generated for students who took PSIs include responses only for items which each student had an opportunity to answer, rather than the full set of items each student was administered. The procedure of treating not-reached responses as not administered also was adopted in reporting item percent correct statistics in the [Findings from the TIMSS 2019 Problem Solving and Inquiry Tasks](#) report.

Scaling the Problem Solving and Inquiry Data

The main objective of the effort by the TIMSS & PIRLS International Study Center to scale the TIMSS 2019 Problem Solving and Inquiry data was to derive TIMSS 2019 student plausible values (e.g., von Davier, Gonzalez, & Mislevy, 2009) from the eTIMSS assessment data for students who took PSI eBooklets, suitable for reporting and publication. Scaling the data collected from the TIMSS 2019 Problem Solving and Inquiry tasks involved four major steps: calibrating the achievement items (estimating model parameters for each PSI item), creating new principal components from the student questionnaire data for use in conditioning, generating plausible values for mathematics and science, and placing these plausible values on the metrics used to report trend results from previous assessments. Item parameters for the PSI items were estimated based on the data from both regular eTIMSS and PSI eBooklets, collected from the same classrooms of students in each country. After the item calibration and estimation of student proficiency, the scale transformations that placed paper TIMSS and eTIMSS on the TIMSS trend scales was applied to the PSI data.

The scaling procedures produced plausible values for the mathematics and science scales at both the fourth and eighth grades for all PSI students (i.e., students assigned eBooklets 15 and 16). The resulting posterior ability score distributions were compared and found to be virtually indistinguishable from those generated in the original eTIMSS scaling described in [Chapter 12](#) of this volume. The IRT models and population models used are described in [Chapter 11](#).

Linking the PSI data to the eTIMSS scale relied on having a majority of students with regular eTIMSS data—sampled from the same classrooms as students with PSI data—and a majority of regular eTIMSS items as the basis for the link. The regular eTIMSS assessment consisted of 14 mathematics item blocks and 14 science item blocks at each grade which formed the basis for linking the PSI data to the TIMSS achievement scales. Exhibits 17.3 through 17.6 show the numbers of items present in the eTIMSS with PSI 2019 calibration by item type for both mathematics and science and both fourth and eighth grades, respectively.

Exhibit 17.3: Mathematics Items for the eTIMSS with PSI 2019 Calibration—Grade 4

Item Type	Points	Regular eTIMSS Items		PSI Items		Total	
		Items	Points	Items	Points	Items	Points
Multiple Choice	1	68	68	2	2	70	70
Constructed Response	1	91	91	22	22	113	113
	2	12	24	5	10	17	34
Total		171	183	29	34	200	217

Exhibit 17.4: Science Items for the eTIMSS with PSI 2019 Calibration—Grade 4

Item Type	Points	Regular eTIMSS Items		PSI Items		Total	
		Items	Points	Items	Points	Items	Points
Multiple Choice	1	88	88	1	1	89	89
Constructed Response	1	76	76	11	11	87	87
	2	5	10	6	12	11	22
Total		169	174	18	24	187	198

Exhibit 17.5: Mathematics Items for the eTIMSS with PSI 2019 Calibration—Grade 8

Item Type	Points	Regular eTIMSS Items		PSI Items		Total	
		Items	Points	Items	Points	Items	Points
Multiple Choice	1	90	90	1	1	91	91
Constructed Response	1	105	105	21	21	126	126
	2	11	22	3	6	14	28
Total		206	217	25	28	231	245

Exhibit 17.6: Science Items for the eTIMSS with PSI 2019 Calibration—Grade 8

Item Type	Points	Regular eTIMSS Items		PSI Items		Total	
		Items	Points	Items	Points	Items	Points
Multiple Choice	1	107	107	7	7	114	114
Constructed Response	1	82	82	15	15	97	97
	2	22	44	5	10	27	54
Total		211	233	27	32	238	265

Exhibits 17.7 and 17.8 show the sample sizes for scaling the eTIMSS with PSI 2019 data, both for item calibration and for proficiency estimation at the fourth grade and eighth grade, respectively. All student samples were weighted so that each country contributed equally to the item calibration. Twenty-eight countries contributed data to the PSI item calibration at the fourth grade and 18 contributed at the eighth grade—the same set of countries that contributed to regular eTIMSS 2019 item calibrations. About 12 percent of each country’s student sample was administered the PSI tasks. The 28 countries at the fourth grade provided 115,038 “regular students” and 16,449 “PSI students.” At the eighth grade, the 18 countries provided 98,699 regular students and 14,095 PSI students.

Exhibit 17.7: Sample Sizes for Scaling the eTIMSS with PSI 2019 Data—Grade 4

Country	Item Calibration			Proficiency Estimation		
	Regular eTIMSS Students	PSI Students	Total	Regular eTIMSS Students	PSI Students	Total
Austria	4,464	633	5,097	4,464	633	5,097
Canada	13,653	1,929	15,582	13,653	1,929	15,582
Chile	4,174	601	4,775	4,174	601	4,775
Chinese Taipei	3,765	530	4,295	3,765	530	4,295
Croatia	3,785	550	4,335	3,785	550	4,335
Czech Republic	4,692	666	5,358	4,692	666	5,358
Denmark	3,227	466	3,693	3,227	466	3,693
England	3,396	476	3,872	3,396	476	3,872
Finland	4,730	667	5,397	4,730	667	5,397
France	4,186	606	4,792	4,186	606	4,792
Georgia	3,787	529	4,316	3,787	529	4,316
Germany	3,437	496	3,933	3,437	496	3,933
Hong Kong SAR	—	—	—	2,968	418	3,386
Hungary	4,571	656	5,227	4,571	656	5,227
Italy	3,741	528	4,269	3,741	528	4,269
Korea, Rep. of	3,893	555	4,448	3,893	555	4,448
Lithuania	3,741	524	4,265	3,741	524	4,265
Malta	—	—	—	3,630	524	4,154
Netherlands	3,355	476	3,831	3,355	476	3,831
Norway (5)	3,951	576	4,527	3,951	576	4,527
Portugal	4,300	615	4,915	4,300	615	4,915
Qatar	4,933	713	5,646	4,933	713	5,646
Russian Federation	4,022	574	4,596	4,022	574	4,596
Singapore	5,986	853	6,839	5,986	853	6,839
Slovak Republic	4,247	615	4,862	4,247	615	4,862
Spain	9,555	1,391	10,946	9,555	1,391	10,946
Sweden	3,965	570	4,535	3,965	570	4,535
Turkey (5)	4,028	571	4,599	4,028	571	4,599
United Arab Emirates	25,834	3,681	29,515	25,834	3,681	29,515
United States	8,776	1,253	10,029	8,776	1,253	10,029
Benchmarking Participants						
Ontario, Canada	—	—	—	3,830	530	4,360
Quebec, Canada	—	—	—	3,837	547	4,384
Moscow City, Russian Fed.	—	—	—	3,843	549	4,392
Madrid, Spain	—	—	—	3,390	489	3,879
Abu Dhabi, UAE	—	—	—	9,037	1,291	10,328
Dubai, UAE	—	—	—	7,265	1,034	8,299
TOTAL	115,038	16,449	131,487	152,838	21,831	174,669

Exhibit 17.8: Sample Sizes for Scaling the eTIMSS with PSI 2019 Data—Grade 8

Country	Item Calibration			Proficiency Estimation		
	Regular eTIMSS Students	PSI Students	Total	Regular eTIMSS Students	PSI Students	Total
Chile	4,115	582	4,697	4,115	582	4,697
Chinese Taipei	4,915	695	5,610	4,915	695	5,610
England	3,365	493	3,858	3,365	493	3,858
Finland	—	—	—	4,874	696	5,570
France	—	—	—	3,874	552	4,426
Georgia	3,315	474	3,789	3,315	474	3,789
Hong Kong SAR	—	—	—	3,265	465	3,730
Hungary	4,569	650	5,219	4,569	650	5,219
Israel	3,731	537	4,268	3,731	537	4,268
Italy	3,619	519	4,138	3,619	519	4,138
Korea, Rep. of	3,861	548	4,409	3,861	548	4,409
Lithuania	3,826	540	4,366	3,826	540	4,366
Malaysia	7,065	1,012	8,077	7,065	1,012	8,077
Norway (9)	4,575	640	5,215	4,575	640	5,215
Portugal	—	—	—	3,377	490	3,867
Qatar	3,884	552	4,436	3,884	552	4,436
Russian Federation	3,901	555	4,456	3,901	555	4,456
Singapore	4,853	693	5,546	4,853	693	5,546
Sweden	3,996	569	4,565	3,996	569	4,565
Turkey	4,077	585	4,662	4,077	585	4,662
United Arab Emirates	22,334	3,205	25,539	22,334	3,205	25,539
United States	8,698	1,246	9,944	8,698	1,246	9,944
Benchmarking Participants						
Ontario, Canada	—	—	—	3,776	554	4,330
Quebec, Canada	—	—	—	3,178	460	3,638
Moscow City, Russian Fed.	—	—	—	3,783	541	4,324
Abu Dhabi, UAE	—	—	—	8,204	1,176	9,380
Dubai, UAE	—	—	—	5,728	816	6,544
TOTAL	98,699	14,095	112,794	138,758	19,845	158,603

Calibrating the PSI Data

The first step of the “eTIMSS with PSI 2019” scaling involved estimating item parameters for the PSI items on the TIMSS scales by combining the PSI data with the regular eTIMSS data, and fixing item parameters for the regular items at their values from the eTIMSS 2019 final item calibration. Item calibration was conducted by the TIMSS & PIRLS International Study Center using Parscale software (Muraki & Bock, 1997) and included regular eTIMSS data collected through eBooklets 1 to 14 and PSI data collected through eBooklets 15 and 16 for all eTIMSS 2019 countries that contributed to the original

eTIMSS 2019 item calibration—28 countries at the fourth grade and 18 countries at the eighth grade. The item calibration used all available item response data from each country’s eTIMSS student samples.

The item parameters estimated from the eTIMSS with PSI calibration at the fourth and eighth grades and for mathematics and science are presented in Appendix 17F. The regular eTIMSS item parameters that were fixed for the PSI calibration are reported in Appendices 12K through 12N in [Chapter 12](#).

Evaluating Fit of IRT Models to the eTIMSS with PSI Assessment Data

After the PSI item calibration was completed, extensive checks were performed to evaluate the fit of the estimated item characteristic curves to the empirical response data. This included a graphical model fit assessment and computation of root mean square difference (RMSD) item fit statistics (see more details about checks in [Chapter 12](#)). In addition, item discrimination (slope) parameters were compared between regular items and PSI items to examine the extent that PSI items measure the same mathematics and science constructs as the regular items.

The item parameters shown in Appendix 17F include RMSD statistics, and plots of the RMSD for mathematics and science at the fourth and eighth grades are shown in Exhibits 17.9, 17.10, 17.11, and 17.12, respectively. In each exhibit, the items are sorted from smallest to largest RMSD values. All items have RMSD values less than 0.10, with the vast majority less than 0.05, indicating good model fit. In addition, PSI items had RMSD values in about the same range as the regular items.

Exhibit 17.9: RMSD Statistics for Items in the eTIMSS with PSI 2019 Calibration—Grade 4 Mathematics

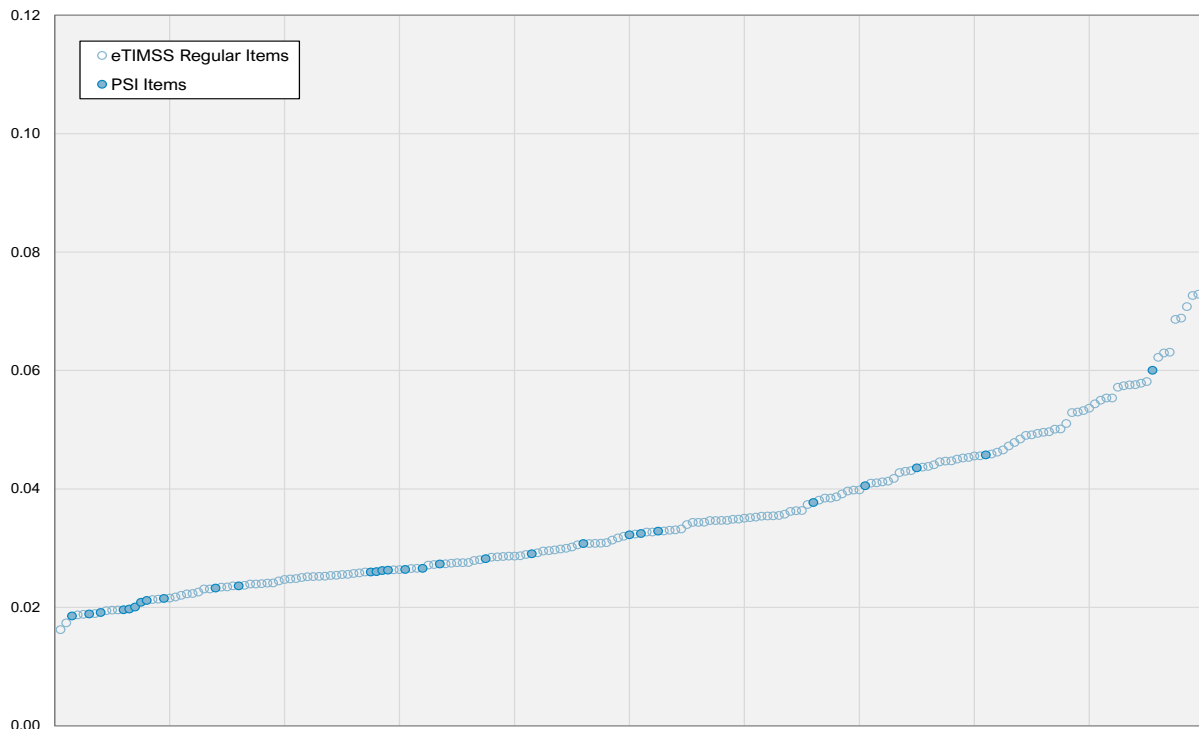


Exhibit 17.10: RMSD Statistics for Items in the eTIMSS with PSI 2019 Calibration—Grade 4 Science

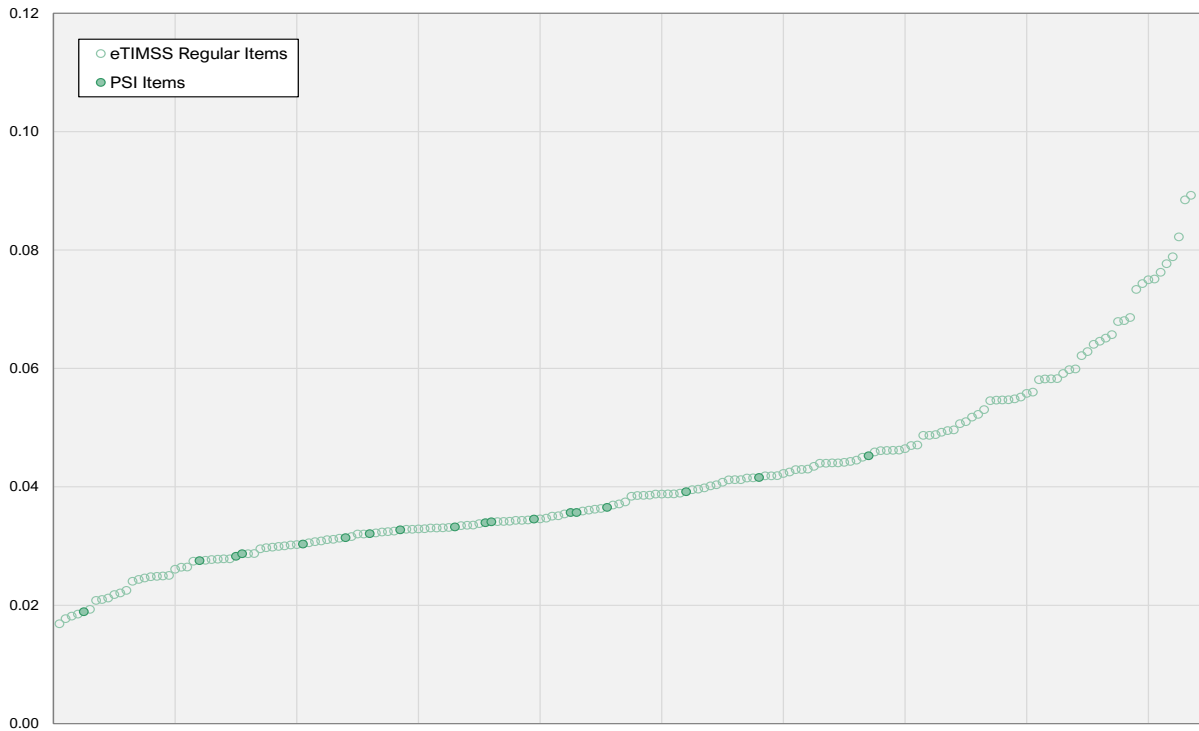


Exhibit 17.11: RMSD Statistics for Items in the eTIMSS with PSI 2019 Calibration—Grade 8 Mathematics

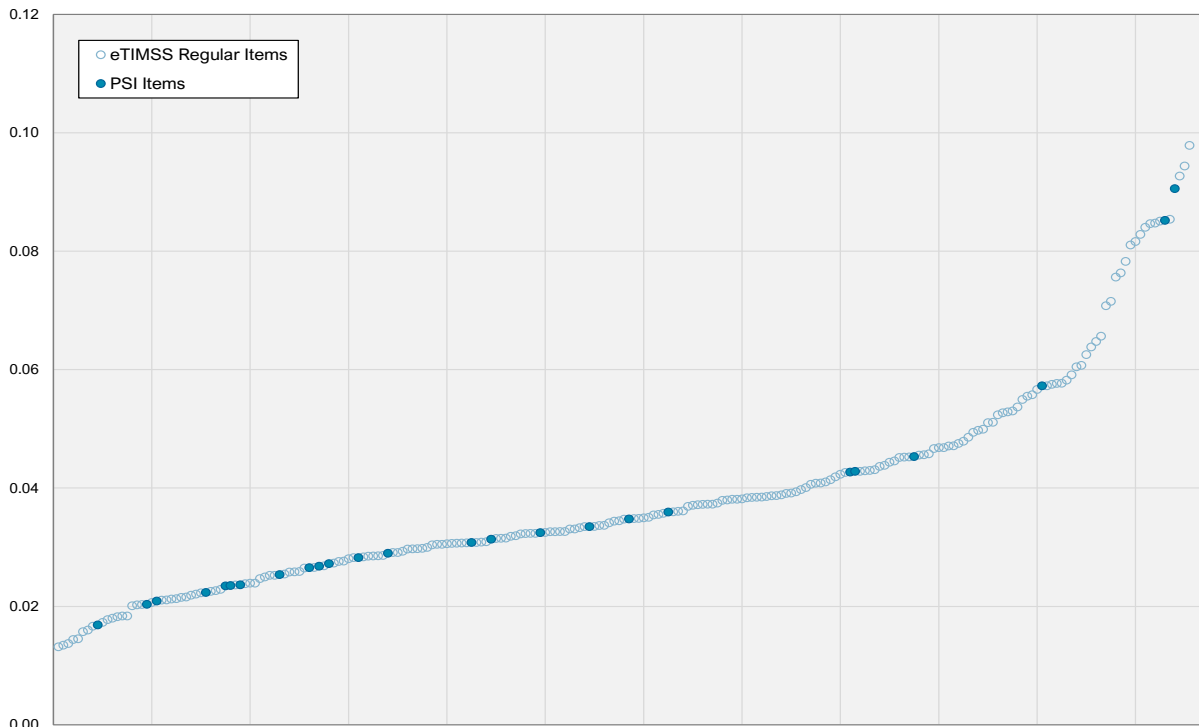
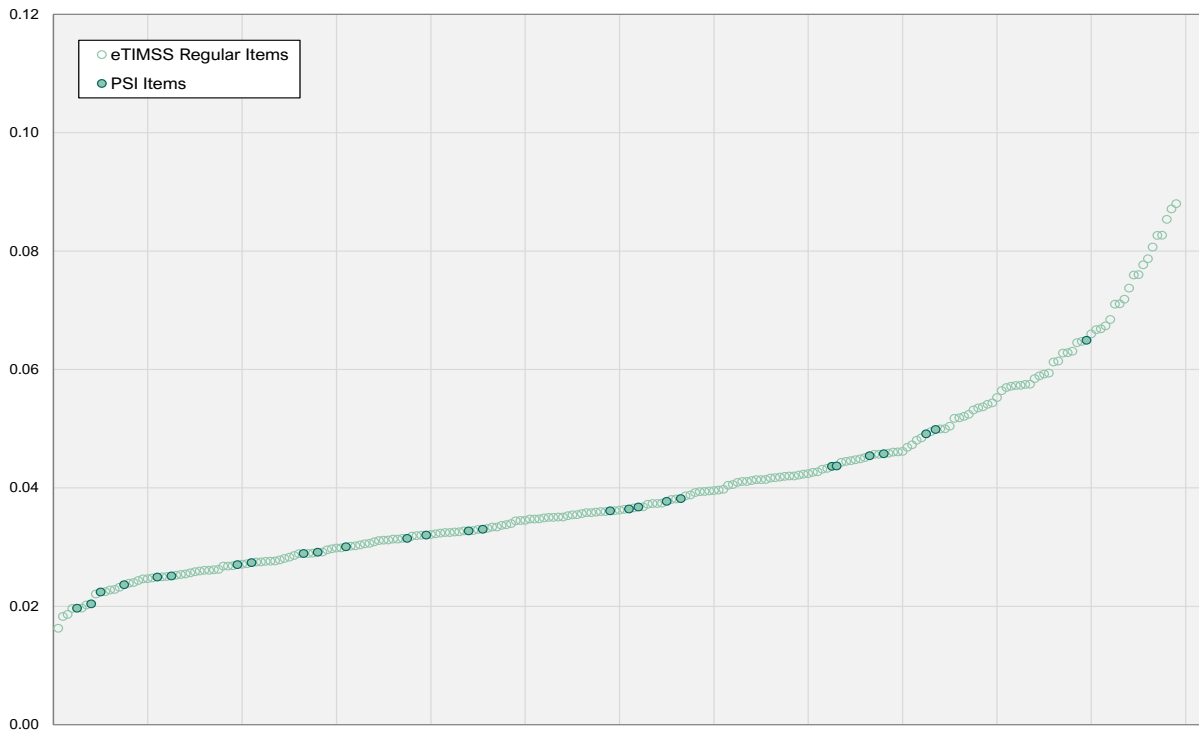


Exhibit 17.12: RMSD Statistics for Items in the eTIMSS with PSI 2019 Calibration—Grade 8 Science



Exhibits 17.13 through 17.16 present summary statistics and plots of the estimated PSI item discrimination (slope) parameters and the fixed regular item discrimination parameters used in the joint eTIMSS with PSI scaling. The discrimination parameters estimated for the PSI items have about the same range as the regular (fixed) item discriminations, indicating that the two groups of items measure the same construct. If the PSI items were measuring a different construct than the regular items, the estimated PSI item discrimination parameters would be considerably smaller in value when aligned with the fixed regular eTIMSS scale.

Exhibit 17.13: Item Discrimination (Slope) Parameters for Items in the eTIMSS with PSI 2019 Calibration—Grade 4 Mathematics

	Count	Mean	Minimum	Maximum
Regular eTIMSS	200	0.943	0.300	1.785
PSI	29	0.916	0.300	1.400

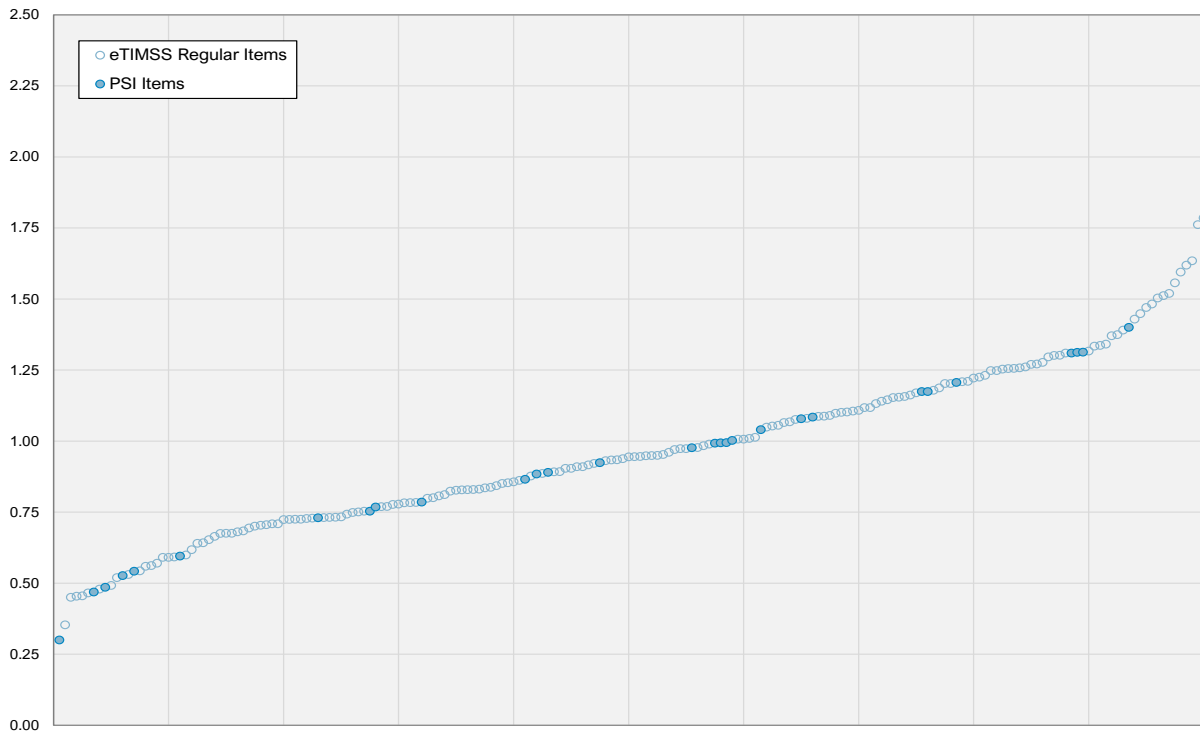


Exhibit 17.14: Item Discrimination (Slope) Parameters for Items in the eTIMSS with PSI 2019 Calibration—Grade 4 Science

	Count	Mean	Minimum	Maximum
Regular eTIMSS	169	0.848	0.270	1.701
PSI	18	0.679	0.424	1.028

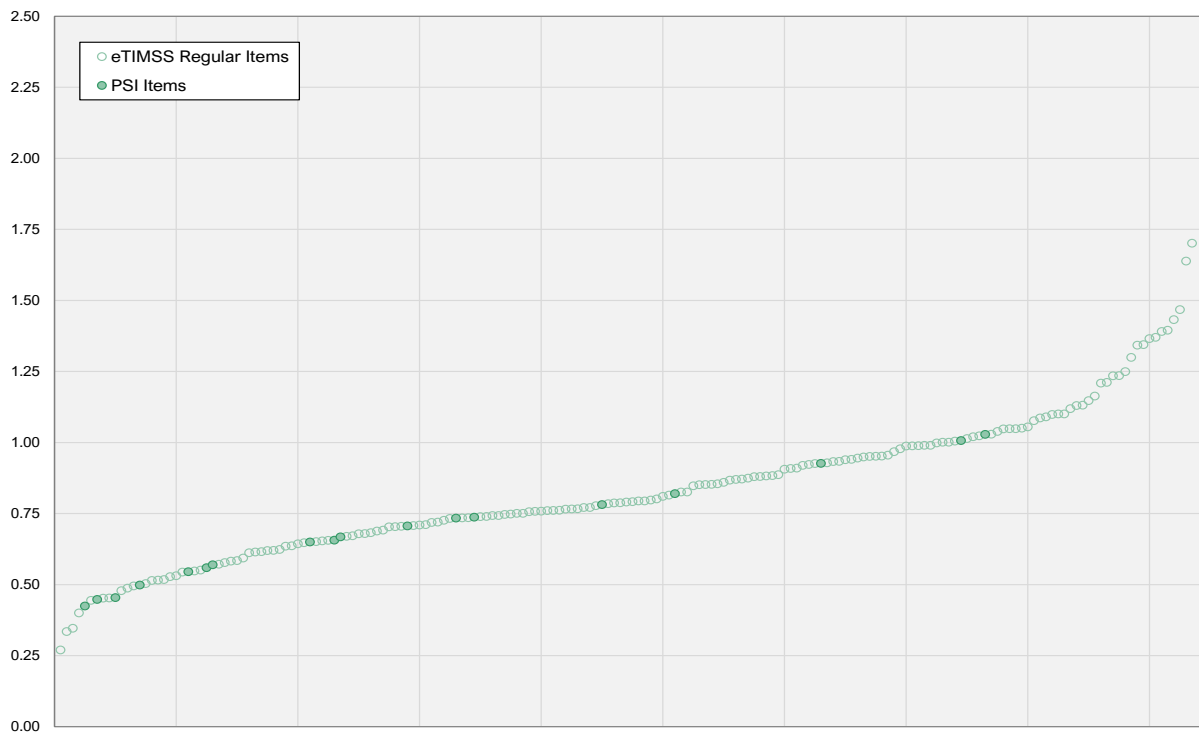


Exhibit 17.15: Item Discrimination (Slope) Parameters for Items in the eTIMSS with PSI 2019 Calibration—Grade 8 Mathematics

	Count	Mean	Minimum	Maximum
Regular eTIMSS	206	1.206	0.447	2.286
PSI	25	1.242	0.594	2.461

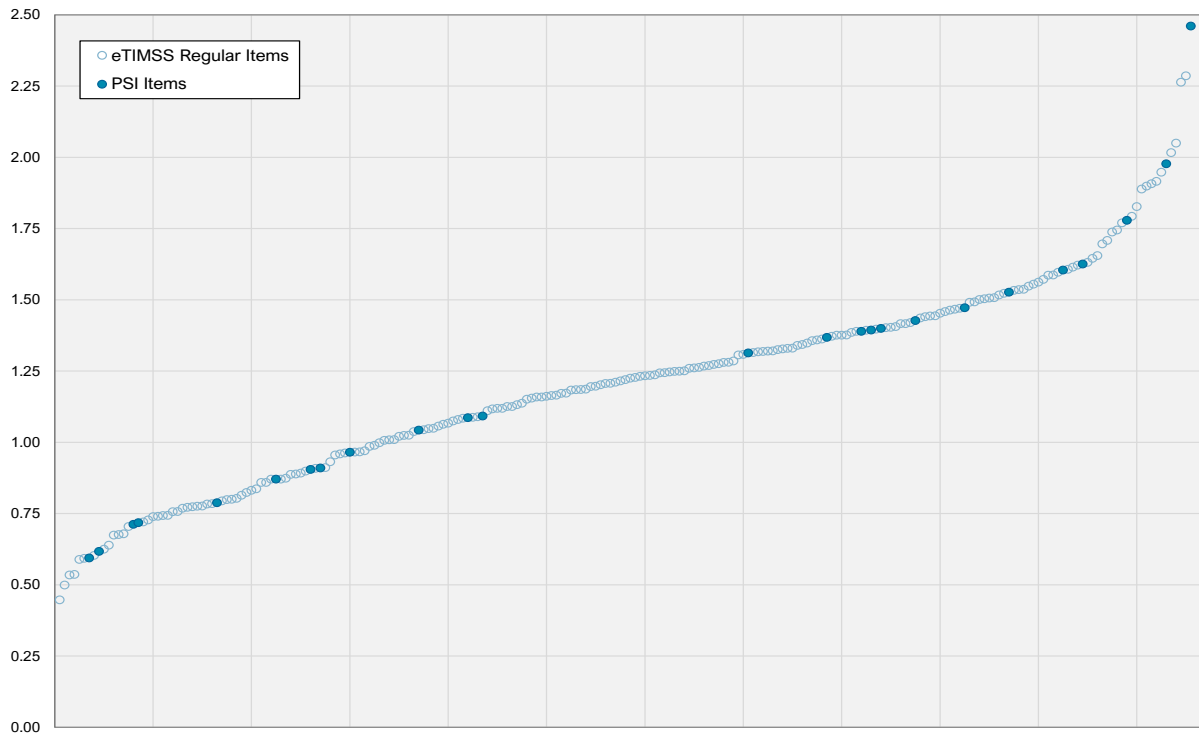
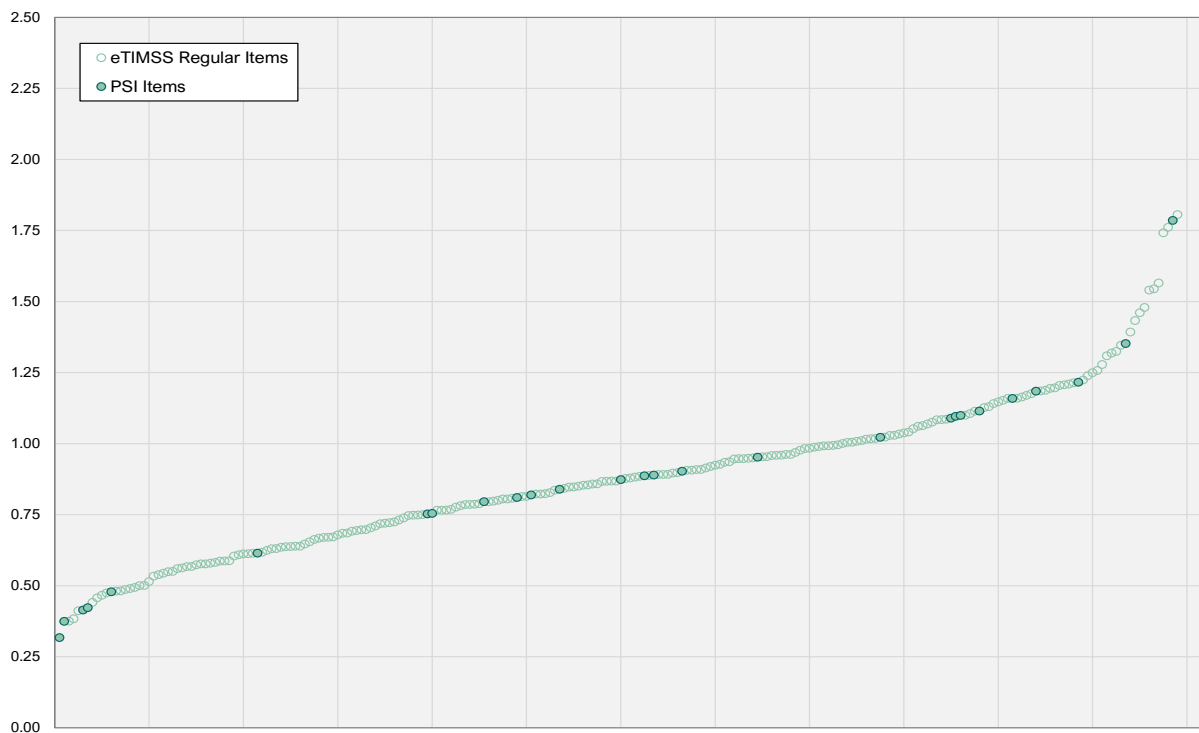


Exhibit 17.16: Item Discrimination (Slope) Parameters for Items in the eTIMSS with PSI 2019 Calibration—Grade 8 Science

	Count	Mean	Minimum	Maximum
Regular eTIMSS	211	0.882	0.375	1.806
PSI	27	0.890	0.317	1.785



Generating Plausible Values for the eTIMSS with PSI Data

Drawing student plausible values for the eTIMSS with PSI 2019 data followed the same general approach as for the eTIMSS 2019 data. A latent regression (conditioning) model utilized the available student and parent context variables together with the achievement data to improve the psychometric properties of student plausible values, as described for paperTIMSS 2019 in [Chapter 12](#). Exhibits 17.17 and 17.18 provide details on the conditioning models used for PSI proficiency estimation at the fourth grade and eighth grade, respectively.

Exhibit 17.17: Variable Selection for Conditioning Models Combining the eTIMSS and PSI 2019 Data—Grade 4

Country	eTIMSS with PSI 2019			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Austria	2	618	254	83
Canada	5	605	315	90
Chile	2	598	238	80
Chinese Taipei	2	622	214	78
Croatia	2	619	216	80
Czech Republic	2	618	267	86
Denmark	2	615	184	74
England	2	389	193	87
Finland	3	622	269	86
France	2	623	239	81
Georgia	2	620	215	77
Germany	2	623	196	77
Hong Kong SAR	3	623	169	73
Hungary	2	599	261	84
Italy	2	617	213	76
Korea, Rep. of	2	613	222	81
Lithuania	4	608	213	78
Malta	2	603	207	75
Netherlands	2	383	191	87
Norway (5)	4	543	226	84
Portugal	2	623	245	81
Qatar	3	622	282	85
Russian Federation	2	597	229	80
Singapore	2	599	301	90
Slovak Republic	3	623	243	81
Spain	6	616	319	90
Sweden	2	597	226	81
Turkey (5)	2	599	229	80
United Arab Emirates	5	623	327	90
United States	10	387	221	90
Benchmarking Participants				
Ontario, Canada	3	605	218	80
Quebec, Canada	3	605	219	80
Moscow City, Russian Fed.	2	591	219	79
Madrid, Spain	2	616	193	73
Abu Dhabi, UAE	3	623	320	90
Dubai, UAE	3	623	308	90

Exhibit 17.18: Variable Selection for Conditioning Models Combining the eTIMSS and PSI 2019 Data—Grade 8

Country	eTIMSS with PSI 2019			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Chile	2	638	234	83
Chinese Taipei	2	638	274	90
England	2	637	192	77
Finland	3	979	278	80
France	2	952	221	77
Georgia	2	978	189	66
Hong Kong SAR	3	639	186	81
Hungary	2	980	260	75
Israel	3	573	213	83
Italy	2	639	206	78
Korea, Rep. of	2	626	220	85
Lithuania	4	973	218	70
Malaysia	3	633	306	90
Norway (9)	4	595	260	88
Portugal	2	957	193	74
Qatar	3	639	221	81
Russian Federation	2	980	222	71
Singapore	2	615	277	90
Sweden	2	865	228	77
Turkey	2	639	233	81
United Arab Emirates	5	639	314	90
United States	10	634	301	90
Benchmarking Participants				
Ontario, Canada	3	639	216	81
Quebec, Canada	3	639	181	75
Moscow City, Russian Fed.	2	972	216	70
Abu Dhabi, UAE	3	639	312	90
Dubai, UAE	3	639	297	90

Mathematics proficiency and science proficiency for the eTIMSS with PSI data at both grades were estimated using the same latent regression (conditioning) models as for the eTIMSS 2019 data, incorporating the eTIMSS and PSI response data, item parameters, and context variables. A two-dimensional latent regression model was used to estimate simultaneously overall mathematics proficiency and overall science proficiency using Educational Testing Service's MGROUP program (Rogers, Tang, Lin, & Kandathil, 2006; Sheehan, 1985).

Because the PSI item calibrations were anchored to the eTIMSS 2019 calibrations using fixed parameter linking, the scale transformations calculated and applied to the paperTIMSS and eTIMSS data were appropriate for placing the estimated PSI student plausible values in mathematics and science on the TIMSS trend scales (see Exhibits 12.12 and 12.13 in [Chapter 12](#)).

Relative to regular eTIMSS, the PSI data comprised a small number of students and small number of items. These additional plausible values produced for PSI students could be combined with the original eTIMSS 2019 plausible values for producing average eTIMSS with PSI achievement scores. To ensure that the eTIMSS with PSI scaling accurately recreated regular eTIMSS achievement and did not impact the mathematics and science constructs measured, the underlying posterior ability distributions from which the plausible values were generated were compared between the eTIMSS scaling and the eTIMSS with PSI scaling. Normal Q-Q plots were produced for each country and benchmarking participant, plotting the quantiles of a normal distribution against the quantiles for the regular eTIMSS ability distribution, and the same for the eTIMSS with PSI ability distribution overlaid.

Exhibits 17.19 through 17.22 show the plots produced using posterior ability scores from all countries for mathematics and science at the fourth grade and eighth grade, respectively. The overlaid distributions are identical for most of the range of ability, with only slight mismatch occurring at the extreme ends where there are very few students.

Based on these results and the results obtained country-by-country, it was concluded that the addition of PSI data to the eTIMSS 2019 scale through the procedures described was successful in measuring achievement of all students who participated in eTIMSS 2019. The small performance differences observed between regular eTIMSS students and PSI students could be reasonably explained by larger sampling variance due to the small size of the PSI sample and smaller number of booklets, as well as the somewhat different nature of the PSI assessment booklets.

Exhibit 17.19: Normal Q-Q Plot of Posterior Ability from eTIMSS 2019 Scaling and eTIMSS with PSI 2019 Scaling—Grade 4 Mathematics

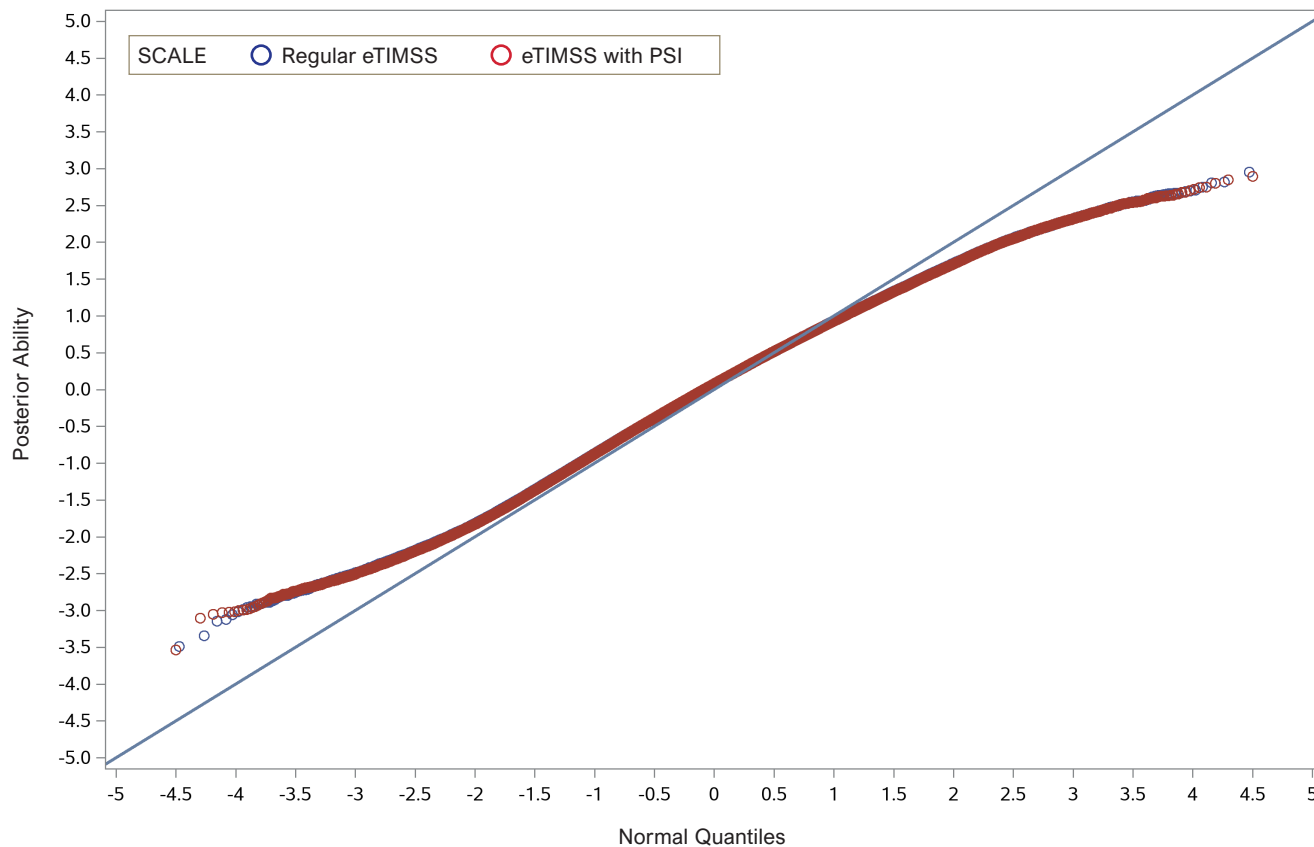


Exhibit 17.20: Normal Q-Q Plot of Posterior Ability from eTIMSS 2019 Scaling and eTIMSS with PSI 2019 Scaling—Grade 4 Science

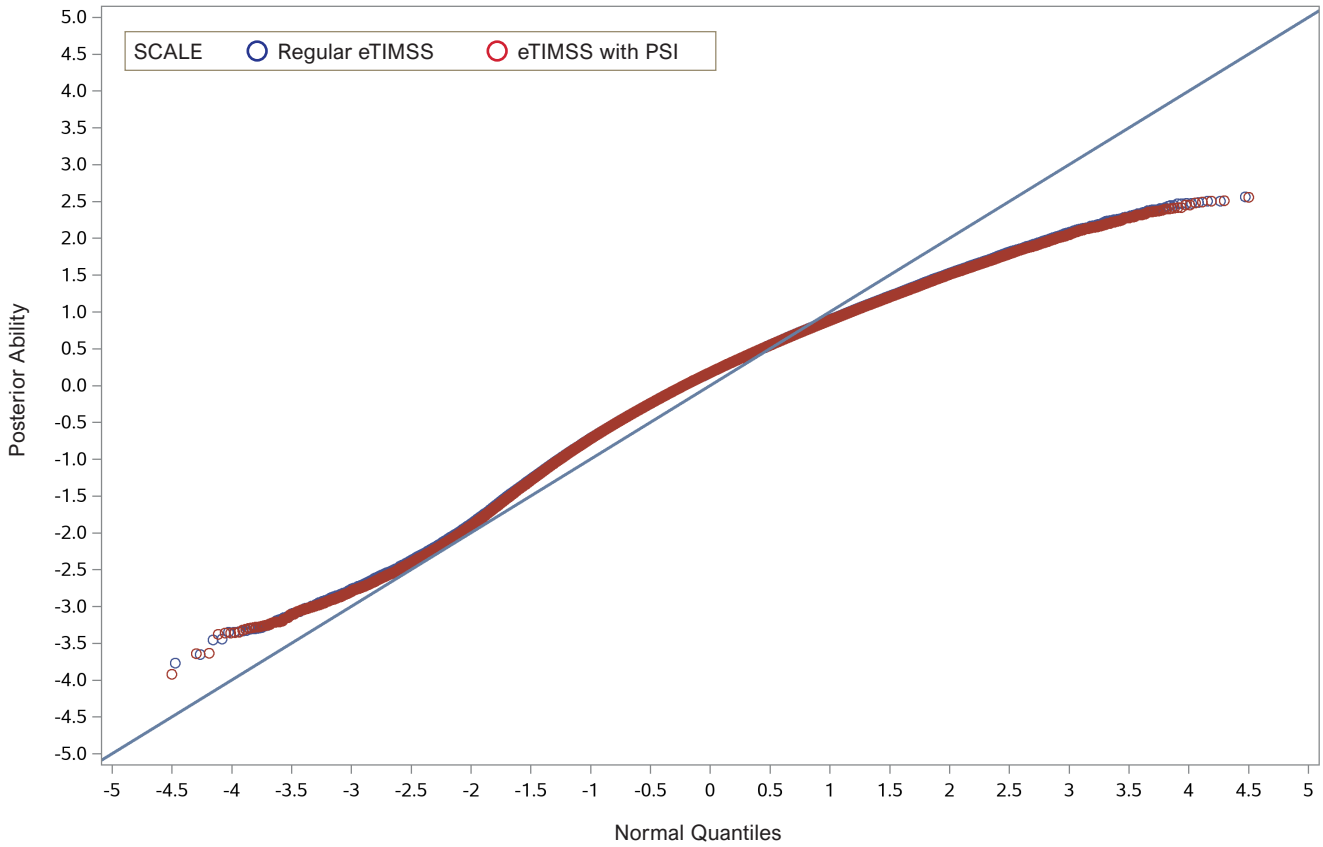


Exhibit 17.21: Normal Q-Q Plot of Posterior Ability from eTIMSS 2019 Scaling and eTIMSS with PSI 2019 Scaling—Grade 8 Mathematics

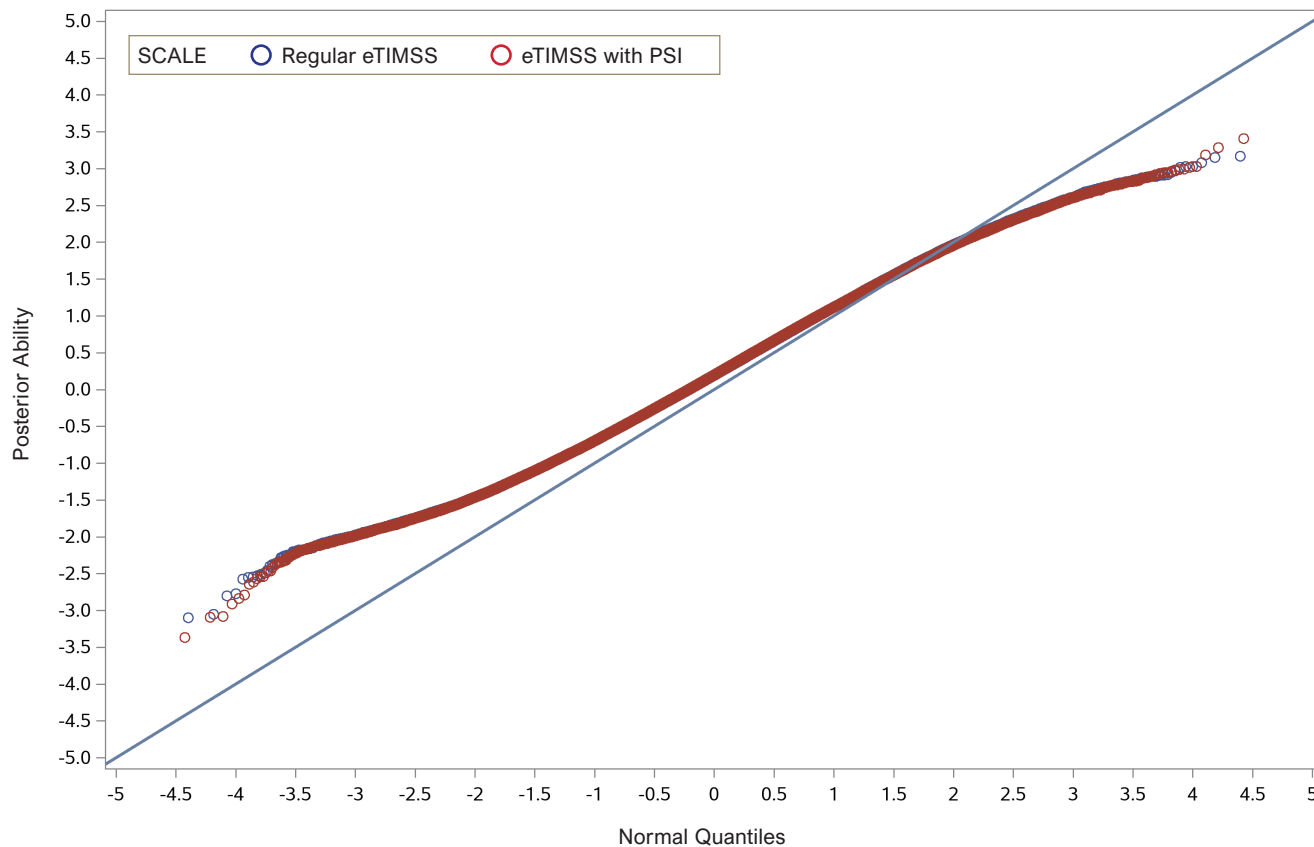
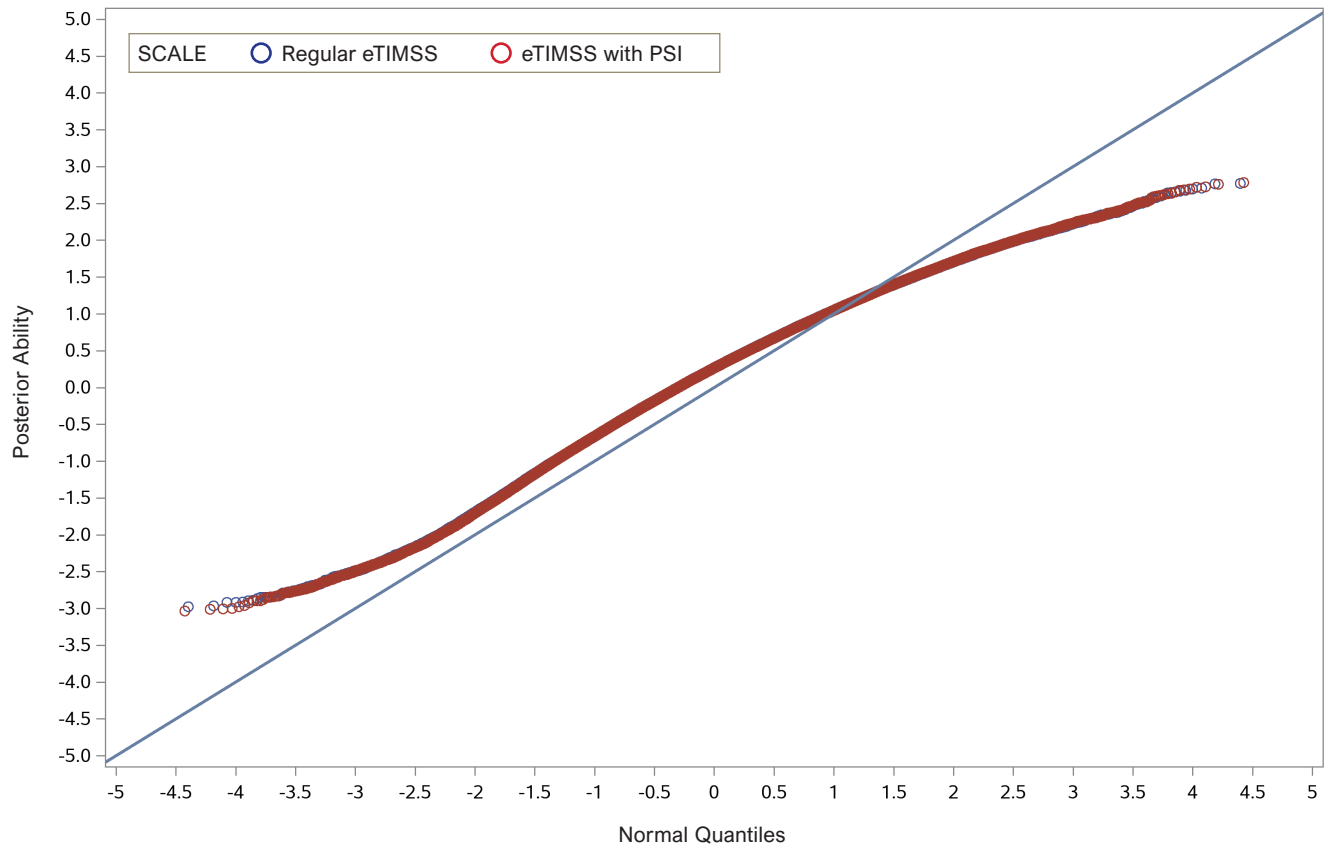


Exhibit 17.22: Normal Q-Q Plot of Posterior Ability from eTIMSS 2019 Scaling and eTIMSS with PSI 2019 Scaling—Grade 8 Science



References

- Glas, C. A. W., & Pimentel, J. (2006). *Modeling nonignorable missing data processes in item calibration*. (LSAC Research Report Series; No. 04-07). Newton, PA: Law School Admission Council.
- Martin, M. O., Mullis, I. V. S., & Foy, P. (2017). TIMSS 2019 assessment design. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 assessment frameworks* (pp. 79–91). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/timss2019/frameworks/>
- Moustaki, I., & Knott, M. (2000). Weighted for item non-response in attitude scales by using latent variable models with covariates. *Journal of the Royal Statistical Society*, 163(3), 445–459.
- Muraki, E., & Bock, R. D. (1997). PARSCALE: IRT item analysis and test scoring for rating scale data [computer software]. Chicago: Scientific Software.
- Rogers, A., Tang, C., Lin, M.-J., & Kandathil, M. (2006). DGROUPE [computer software]. Princeton, NJ: Educational Testing Service.
- Rose, N., von Davier, M., & Xu, X. (2010). *Modeling nonignorable missing data with Item Response Theory (IRT)*. (ETS Research Report No. RR-10-11). Princeton, NJ: Educational Testing Service.
- Sheehan, K. M. (1985). M-Group: Estimation of group effects in multivariate models [computer software, Version 3.2] Princeton, NJ: Educational Testing Service.
- von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? In M. von Davier & D. Hastedt (Eds.), *IERI Monograph Series: Issues and Methodologies in Large Scale Assessments* (Vol. 2, pp. 9–36). Retrieved from https://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02_Chapter_01.pdf

Appendix 17A: Characteristics of the eTIMSS 2019 with Problem Solving and Inquiry Samples

Target Population Coverage – eTIMSS 2019 with Problem Solving and Inquiry—Grade 4

Country	International Target Population		Exclusions from National Target Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Austria	100%		0.9%	4.5%	5.4%
^{1,2} Canada	79%	Students from the provinces of Alberta, Manitoba, Newfoundland, Ontario, and Quebec	3.1%	3.9%	7.0%
Chile	100%		1.2%	2.6%	3.8%
Chinese Taipei	100%		0.3%	1.6%	2.0%
Croatia	100%		1.1%	3.1%	4.2%
Czech Republic	100%		2.5%	2.2%	4.7%
Denmark	100%		1.6%	1.5%	3.1%
² England	100%		2.2%	3.6%	5.8%
Finland	100%		1.8%	1.5%	3.3%
France	100%		2.5%	1.9%	4.4%
¹ Georgia	92%	Students taught in Georgian	2.8%	1.8%	4.7%
Germany	100%		1.7%	2.2%	3.9%
Hong Kong SAR	100%		1.1%	2.4%	3.5%
Hungary	100%		2.1%	2.0%	4.1%
Italy	100%		0.9%	4.1%	4.9%
Korea, Rep. of	100%		0.9%	1.5%	2.3%
² Lithuania	100%		2.6%	4.1%	6.7%
Malta	100%		1.4%	3.1%	4.5%
Netherlands	100%		2.6%	0.9%	3.5%
Norway (5)	100%		1.4%	3.3%	4.7%
² Portugal	100%		0.9%	6.9%	7.8%
Qatar	100%		1.2%	1.0%	2.2%
² Russian Federation	100%		2.4%	3.9%	6.3%
³ Singapore	100%		12.5%	0.4%	12.8%
² Slovak Republic	100%		3.6%	1.9%	5.5%
Spain	100%		1.6%	3.8%	5.4%
Sweden	100%		1.6%	3.8%	5.4%
² Turkey (5)	100%		1.0%	5.9%	7.0%
United Arab Emirates	100%		1.1%	2.0%	3.2%
² United States	100%		0.0%	7.2%	7.2%

Benchmarking Participants

² Ontario, Canada	100%		2.3%	4.7%	7.0%
Quebec, Canada	100%		3.3%	1.2%	4.4%
Moscow City, Russian Fed.	100%		0.7%	1.4%	2.1%
Madrid, Spain	100%		0.5%	3.1%	3.6%
Abu Dhabi, UAE	100%		1.1%	2.5%	3.6%
² Dubai, UAE	100%		2.6%	3.0%	5.6%

1 National Target Population does not include all of the International Target Population.

2 National Defined Population covers 90% to 95% of National Target Population.

3 National Defined Population covers less than 90% of National Target Population (but at least 77%).

Target Population Coverage – eTIMSS 2019 with Problem Solving and Inquiry—Grade 8

Country	International Target Population		Exclusions from National Target Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Chile	100%		0.3%	1.9%	2.2%
Chinese Taipei	100%		0.1%	1.3%	1.5%
England	100%		2.9%	2.0%	4.8%
Finland	100%		1.5%	1.5%	3.1%
France	100%		2.8%	1.0%	3.8%
¹ Georgia	91%	Students taught in Georgian	2.2%	2.1%	4.3%
Hong Kong SAR	100%		1.2%	2.1%	3.3%
Hungary	100%		2.5%	1.9%	4.4%
³ Israel	100%		19.5%	3.8%	23.2%
Italy	100%		0.8%	3.6%	4.3%
Korea, Rep. of	100%		0.7%	0.9%	1.6%
Lithuania	100%		3.2%	2.0%	5.3%
Malaysia	100%		1.9%	1.3%	3.2%
Norway (9)	100%		1.4%	2.5%	4.0%
Portugal	100%		1.0%	4.5%	5.5%
Qatar	100%		1.3%	0.9%	2.2%
² Russian Federation	100%		2.8%	2.9%	5.7%
² Singapore	100%		10.1%	0.2%	10.3%
² Sweden	100%		1.7%	4.6%	6.3%
Turkey	100%		1.1%	2.4%	3.4%
United Arab Emirates	100%		1.1%	1.3%	2.4%
United States	100%		0.0%	3.9%	3.9%
Benchmarking Participants					
Ontario, Canada	100%		2.1%	3.4%	5.5%
Quebec, Canada	100%		3.3%	0.9%	4.2%
Moscow City, Russian Fed.	100%		0.7%	0.8%	1.5%
Abu Dhabi, UAE	100%		0.9%	0.8%	1.7%
² Dubai, UAE	100%		3.0%	2.5%	5.5%

1 National Target Population does not include all of the International Target Population.

2 National Defined Population covers 90% to 95% of National Target Population.

3 National Defined Population covers less than 90% of National Target Population (but at least 77%).

Target Population and Sample Sizes – eTIMSS 2019 with Problem Solving and Inquiry—Grade 4

Country	Population		Sample		
	Schools	Students	Schools	Students	Student Population Size Estimated from Sample
Austria	3,095	81,406	193	5,097	82,158
Canada	9,796	304,798	704	15,582	306,230
Chile	6,081	252,190	169	4,775	250,230
Chinese Taipei	2,476	190,975	162	4,295	188,886
Croatia	1,571	39,244	153	4,335	39,860
Czech Republic	3,578	114,774	152	5,358	113,904
Denmark	1,644	66,225	166	3,693	66,950
England	15,349	644,127	139	3,872	667,451
Finland	1,840	59,755	158	5,397	59,198
France	31,716	822,438	155	4,792	827,474
Georgia	1,678	42,980	154	4,316	40,185
Germany	17,584	716,091	203	3,933	725,273
Hong Kong SAR	564	60,786	139	3,386	60,761
Hungary	2,888	94,673	149	5,227	89,198
Italy	6,809	556,298	162	4,269	549,275
Korea, Rep. of	5,478	472,130	151	4,448	453,918
Lithuania	827	28,035	207	4,265	28,383
Malta	098	4,429	98	4,154	4,461
Netherlands	6,291	178,200	112	3,831	181,849
Norway (5)	1,945	62,012	150	4,527	63,745
Portugal	1,245	99,927	181	4,915	96,042
Qatar	247	25,506	242	5,646	24,518
Russian Federation	40,575	1,414,240	200	4,596	1,602,928
Singapore	187	39,934	187	6,839	40,099
Slovak Republic	2,000	52,222	157	4,862	51,506
Spain	12,861	489,765	501	10,946	493,083
Sweden	3,276	114,494	145	4,535	114,323
Turkey (5)	16,205	1,239,900	180	4,599	1,195,922
United Arab Emirates	754	85,609	688	29,515	85,132
United States	72,902	4,153,454	287	10,029	4,056,773

Benchmarking Participants

Ontario, Canada	3,683	147,295	163	4,360	144,324
Quebec, Canada	1,764	85,132	148	4,384	87,936
Moscow City, Russian Fed.	695	92,630	150	4,392	102,549
Madrid, Spain	1,343	70,232	167	3,879	72,520
Abu Dhabi, UAE	285	29,938	247	10,328	29,215
Dubai, UAE	184	22,567	199	8,299	23,893

Target Population and Sample Sizes – eTIMSS 2019 with Problem Solving and Inquiry—Grade 8

Country	Population		Sample		
	Schools	Students	Schools	Students	Student Population Size Estimated from Sample
Chile	5,767	246,120	164	4,697	238,684
Chinese Taipei	931	214,516	203	5,610	205,439
England	3,706	584,697	136	3,858	591,308
Finland	693	57,591	154	5,570	56,237
France	6,977	814,850	150	4,426	813,845
Georgia	1,837	45,339	145	3,789	44,727
Hong Kong SAR	478	54,160	136	3,730	55,130
Hungary	2,724	87,805	154	5,219	89,223
Israel	979	106,971	157	4,269	108,184
Italy	5,775	566,636	158	4,138	553,839
Korea, Rep. of	3,006	465,626	168	4,409	444,287
Lithuania	706	25,394	194	4,366	25,427
Malaysia	2,565	423,150	177	8,077	412,165
Norway (9)	1,012	60,847	157	5,216	62,287
Portugal	1,039	108,807	156	3,867	106,814
Qatar	156	19,513	152	4,437	18,715
Russian Federation	37,308	1,326,933	204	4,456	1,392,266
Singapore	153	38,517	153	5,546	38,595
Sweden	1,600	108,164	150	4,565	110,810
Turkey	16,179	1,204,063	181	4,662	1,158,547
United Arab Emirates	685	68,113	623	25,539	68,388
United States	48,557	4,059,757	273	9,944	3,800,419
Benchmarking Participants					
Ontario, Canada	2,896	143,484	158	4,330	138,206
Quebec, Canada	539	80,005	124	3,638	74,939
Moscow City, Russian Fed.	704	85,856	150	4,324	92,180
Abu Dhabi, UAE	266	24,654	230	9,380	23,805
Dubai, UAE	153	17,560	163	6,544	18,752

Participation Rates (Weighted) – eTIMSS 2019 with Problem Solving and Inquiry—Grade 4

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Austria	99%	99%	100%	100%	97%	96%
Canada	86%	90%	100%	100%	94%	81%
Chile	89%	99%	100%	100%	95%	85%
Chinese Taipei	95%	99%	100%	100%	98%	94%
Croatia	95%	97%	99%	99%	90%	84%
Czech Republic	99%	100%	100%	100%	95%	94%
† Denmark	70%	95%	99%	99%	86%	60%
England	86%	93%	100%	100%	95%	82%
Finland	99%	100%	100%	100%	96%	96%
France	100%	100%	100%	100%	97%	97%
Georgia	97%	99%	100%	100%	97%	93%
Germany	97%	100%	100%	100%	96%	94%
† Hong Kong SAR	67%	88%	100%	100%	89%	60%
Hungary	93%	99%	100%	100%	96%	89%
Italy	96%	100%	100%	100%	96%	92%
Korea, Rep. of	99%	99%	100%	100%	98%	97%
Lithuania	100%	100%	100%	100%	93%	93%
Malta	100%	100%	100%	100%	96%	96%
≡ Netherlands	46%	75%	100%	100%	97%	45%
† Norway (5)	70%	90%	100%	100%	93%	65%
Portugal	87%	100%	99%	99%	94%	81%
Qatar	100%	100%	100%	100%	97%	97%
Russian Federation	99%	99%	100%	100%	98%	97%
Singapore	100%	100%	100%	100%	97%	97%
Slovak Republic	97%	99%	100%	100%	96%	93%
Spain	97%	99%	100%	100%	95%	92%
Sweden	100%	100%	100%	100%	94%	94%
Turkey (5)	99%	100%	100%	100%	99%	98%
United Arab Emirates	100%	100%	100%	100%	96%	95%
† United States	76%	88%	100%	100%	95%	73%

Benchmarking Participants

Ontario, Canada	93%	95%	100%	100%	94%	87%
Quebec, Canada	82%	86%	100%	100%	96%	79%
Moscow City, Russian Fed.	99%	100%	100%	100%	97%	96%
Madrid, Spain	100%	100%	100%	100%	96%	96%
Abu Dhabi, UAE	100%	100%	100%	100%	95%	95%
Dubai, UAE	100%	100%	100%	100%	96%	96%

TIMSS guidelines for sampling participation: The minimum acceptable participation rates were 85 percent of both schools and students, or a combined rate (the product of school and student participation) of 75 percent.

Participants not meeting these guidelines were annotated as follows:

- † Met guidelines for sample participation rates only after replacement schools were included
- ‡ Nearly satisfied guidelines for sample participation rates after replacement schools were included
- ≡ Did not satisfy guidelines for sample participation rates

Participation Rates (Weighted) – eTIMSS 2019 with Problem Solving and Inquiry—Grade 8

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Chile	90%	99%	100%	100%	95%	86%
Chinese Taipei	98%	99%	100%	100%	98%	95%
England	83%	90%	100%	100%	94%	78%
Finland	100%	100%	100%	100%	95%	95%
France	100%	100%	100%	100%	96%	96%
Georgia	90%	92%	100%	100%	97%	87%
† Hong Kong SAR	70%	86%	100%	100%	93%	65%
Hungary	95%	99%	100%	100%	96%	91%
Israel	95%	98%	100%	100%	92%	87%
Italy	97%	100%	100%	100%	97%	93%
Korea, Rep. of	100%	100%	100%	100%	97%	97%
Lithuania	99%	99%	100%	100%	92%	92%
Malaysia	99%	100%	100%	100%	98%	97%
† Norway (9)	79%	95%	99%	99%	88%	68%
Portugal	95%	99%	99%	99%	95%	89%
Qatar	100%	100%	100%	100%	97%	97%
Russian Federation	99%	100%	100%	100%	97%	97%
Singapore	100%	100%	100%	100%	96%	96%
Sweden	98%	99%	100%	100%	91%	89%
Turkey	100%	100%	100%	100%	98%	98%
United Arab Emirates	100%	100%	100%	100%	95%	95%
† United States	72%	85%	100%	100%	93%	66%

Benchmarking Participants

Ontario, Canada	93%	93%	100%	100%	94%	87%
‡ Quebec, Canada	74%	77%	99%	99%	94%	69%
Moscow City, Russian Fed.	99%	100%	100%	100%	97%	95%
Abu Dhabi, UAE	100%	100%	100%	100%	95%	95%
Dubai, UAE	100%	100%	100%	100%	96%	96%

TIMSS guidelines for sampling participation: The minimum acceptable participation rates were 85 percent of both schools and students, or a combined rate (the product of school and student participation) of 75 percent.

Participants not meeting these guidelines were annotated as follows:

- † Met guidelines for sample participation rates only after replacement schools were included
- ‡ Nearly satisfied guidelines for sample participation rates after replacement schools were included
- ≡ Did not satisfy guidelines for sample participation rates

Participation Rates (Unweighted) – eTIMSS 2019 with Problem Solving and Inquiry—Grade 4

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Austria	99%	99%	100%	100%	97%	96%
Canada	86%	91%	100%	100%	94%	81%
Chile	88%	98%	100%	100%	95%	84%
Chinese Taipei	95%	99%	100%	100%	98%	93%
Croatia	95%	97%	98%	98%	89%	84%
Czech Republic	99%	100%	100%	100%	95%	94%
Denmark	71%	95%	99%	99%	86%	60%
England	86%	93%	99%	99%	95%	81%
Finland	99%	100%	100%	100%	96%	96%
France	100%	100%	100%	100%	97%	97%
Georgia	96%	98%	100%	100%	97%	93%
Germany	98%	100%	100%	100%	96%	94%
Hong Kong SAR	69%	87%	100%	100%	89%	61%
Hungary	93%	99%	100%	100%	96%	89%
Italy	94%	100%	100%	100%	96%	91%
Korea, Rep. of	99%	99%	100%	100%	98%	97%
Lithuania	100%	100%	100%	100%	93%	93%
Malta	100%	100%	100%	100%	96%	96%
Netherlands	48%	75%	100%	100%	97%	46%
Norway (5)	71%	90%	100%	100%	93%	66%
Portugal	87%	100%	99%	99%	93%	81%
Qatar	100%	100%	100%	100%	97%	97%
Russian Federation	99%	99%	100%	100%	97%	96%
Singapore	100%	100%	100%	100%	97%	97%
Slovak Republic	97%	99%	100%	100%	96%	93%
Spain	98%	100%	100%	100%	95%	94%
Sweden	99%	100%	100%	100%	94%	94%
Turkey (5)	99%	100%	100%	100%	98%	98%
United Arab Emirates	100%	100%	100%	100%	95%	95%
United States	77%	88%	100%	100%	95%	73%

Benchmarking Participants

Ontario, Canada	94%	96%	100%	100%	94%	88%
Quebec, Canada	81%	86%	100%	100%	96%	78%
Moscow City, Russian Fed.	98%	99%	100%	100%	97%	95%
Madrid, Spain	100%	100%	100%	100%	96%	96%
Abu Dhabi, UAE	100%	100%	100%	100%	95%	95%
Dubai, UAE	100%	100%	100%	100%	96%	96%

Participation Rates (Unweighted) – eTIMSS 2019 with Problem Solving and Inquiry—Grade 8

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Chile	88%	98%	100%	100%	95%	84%
Chinese Taipei	98%	99%	100%	100%	98%	95%
England	83%	90%	100%	100%	94%	78%
Finland	100%	100%	100%	100%	95%	95%
France	100%	100%	100%	100%	96%	96%
Georgia	90%	92%	100%	100%	97%	87%
Hong Kong SAR	71%	86%	100%	100%	93%	66%
Hungary	94%	99%	100%	100%	96%	91%
Israel	94%	98%	100%	100%	92%	87%
Italy	97%	100%	100%	100%	97%	94%
Korea, Rep. of	100%	100%	100%	100%	97%	97%
Lithuania	99%	99%	100%	100%	92%	92%
Malaysia	99%	100%	100%	100%	98%	97%
Norway (9)	80%	95%	98%	98%	89%	70%
Portugal	94%	99%	99%	99%	95%	88%
Qatar	100%	100%	100%	100%	97%	97%
Russian Federation	100%	100%	100%	100%	97%	97%
Singapore	100%	100%	100%	100%	96%	96%
Sweden	99%	99%	100%	100%	91%	89%
Turkey	99%	100%	100%	100%	98%	98%
United Arab Emirates	100%	100%	100%	100%	95%	95%
United States	72%	85%	100%	100%	93%	67%
Benchmarking Participants						
Ontario, Canada	92%	93%	100%	100%	93%	86%
Quebec, Canada	74%	77%	99%	99%	94%	69%
Moscow City, Russian Fed.	97%	99%	100%	100%	97%	94%
Abu Dhabi, UAE	100%	100%	100%	100%	95%	95%
Dubai, UAE	100%	100%	100%	100%	96%	96%

School Sample Sizes – eTIMSS 2019 with Problem Solving and Inquiry—Grade 4

Country	Number of Schools in Original Sample	Number of Eligible Schools	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
Austria	197	194	193	0	193
Canada	788	777	669	35	704
Chile	174	172	151	18	169
Chinese Taipei	163	163	155	7	162
Croatia	159	158	150	3	153
Czech Republic	156	152	151	1	152
Denmark	175	174	123	43	166
England	150	150	129	10	139
Finland	159	158	157	1	158
France	156	155	155	0	155
Georgia	158	157	151	3	154
Germany	206	203	198	5	203
Hong Kong SAR	159	159	109	30	139
Hungary	151	150	139	10	149
Italy	162	162	153	9	162
Korea, Rep. of	152	152	151	0	151
Lithuania	208	207	207	0	207
Malta	99	98	98	0	98
Netherlands	151	149	71	41	112
Norway (5)	167	167	119	31	150
Portugal	182	181	158	23	181
Qatar	242	242	242	0	242
Russian Federation	202	202	200	0	200
Singapore	187	187	187	0	187
Slovak Republic	159	158	153	4	157
Spain	502	502	494	7	501
Sweden	150	145	144	1	145
Turkey (5)	181	180	179	1	180
United Arab Emirates	697	688	688	0	688
United States	329	325	249	38	287

Benchmarking Participants

Ontario, Canada	171	170	160	3	163
Quebec, Canada	172	172	140	8	148
Moscow City, Russian Fed.	152	151	148	2	150
Madrid, Spain	167	167	167	0	167
Abu Dhabi, UAE	249	247	247	0	247
Dubai, UAE	205	199	199	0	199

School Sample Sizes – eTIMSS 2019 with Problem Solving and Inquiry—Grade 8

Country	Number of Schools in Original Sample	Number of Eligible Schools	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
Chile	169	167	147	17	164
Chinese Taipei	206	205	200	3	203
England	151	151	125	11	136
Finland	158	154	154	0	154
France	150	150	150	0	150
Georgia	158	157	142	3	145
Hong Kong SAR	158	158	112	24	136
Hungary	155	155	146	8	154
Israel	161	161	152	5	157
Italy	158	158	153	5	158
Korea, Rep. of	168	168	168	0	168
Lithuania	195	195	194	0	194
Malaysia	178	177	175	2	177
Norway (9)	166	165	132	25	157
Portugal	158	158	149	7	156
Qatar	152	152	152	0	152
Russian Federation	204	204	203	1	204
Singapore	153	153	153	0	153
Sweden	153	151	149	1	150
Turkey	181	181	180	1	181
United Arab Emirates	631	623	623	0	623
United States	325	321	231	42	273
Benchmarking Participants					
Ontario, Canada	172	170	157	1	158
Quebec, Canada	166	161	119	5	124
Moscow City, Russian Fed.	152	151	147	3	150
Abu Dhabi, UAE	230	230	230	0	230
Dubai, UAE	171	163	163	0	163

Student Sample Sizes – eTIMSS 2019 with Problem Solving and Inquiry—Grade 4

Country	Within-School Student Participation (Weighted Percentage)	Number of Students Sampled in Participating Schools	Number of Students Withdrawn from Class/School	Number of Students Excluded	Number of Eligible Students	Number of Students Absent	Number of Students Assessed
Austria	97%	5,551	33	256	5,262	165	5,097
Canada	94%	17,231	199	429	16,603	1021	15,582
Chile	95%	5,209	77	112	5,020	245	4,775
Chinese Taipei	98%	4,500	65	65	4,370	75	4,295
Croatia	90%	5,003	8	148	4,847	512	4,335
Czech Republic	95%	5,767	48	53	5,666	308	5,358
Denmark	86%	4,414	67	48	4,299	606	3,693
England	95%	4,268	78	127	4,063	191	3,872
Finland	96%	5,685	37	45	5,603	206	5,397
France	97%	5,079	35	104	4,940	148	4,792
Georgia	97%	4,575	28	83	4,464	148	4,316
Germany	96%	4,220	51	89	4,080	147	3,933
Hong Kong SAR	89%	3,932	18	101	3,813	427	3,386
Hungary	96%	5,551	34	89	5,428	201	5,227
Italy	96%	4,664	22	199	4,443	174	4,269
Korea, Rep. of	98%	4,674	50	63	4,561	113	4,448
Lithuania	93%	4,763	12	186	4,565	300	4,265
Malta	96%	4,453	17	115	4,321	167	4,154
Netherlands	97%	4,056	69	27	3,960	129	3,831
Norway (5)	93%	5,029	27	149	4,853	326	4,527
Portugal	94%	5,673	35	366	5,272	357	4,915
Qatar	97%	5,980	127	60	5,793	147	5,646
Russian Federation	98%	4,871	8	144	4,719	123	4,596
Singapore	97%	7,090	22	0	7,068	229	6,839
Slovak Republic	96%	5,111	26	24	5,061	199	4,862
Spain	95%	11,946	48	421	11,477	531	10,946
Sweden	94%	5,007	31	160	4,816	281	4,535
Turkey (5)	99%	5,134	142	319	4,673	74	4,599
United Arab Emirates	96%	31,887	414	564	30,909	1394	29,515
United States	95%	11,271	152	601	10,518	489	10,029
Benchmarking Participants							
Ontario, Canada	94%	4,825	83	95	4,647	287	4,360
Quebec, Canada	96%	4,618	9	37	4,572	188	4,384
Moscow City, Russian Fed.	97%	4,558	11	35	4,512	120	4,392
Madrid, Spain	96%	4,177	17	123	4,037	158	3,879
Abu Dhabi, UAE	95%	11,189	38	239	10,912	584	10,328
Dubai, UAE	96%	9,205	362	213	8,630	331	8,299

Student Sample Sizes – eTIMSS 2019 with Problem Solving and Inquiry—Grade 8

Country	Within-School Student Participation (Weighted Percentage)	Number of Students Sampled in Participating Schools	Number of Students Withdrawn from Class/School	Number of Students Excluded	Number of Eligible Students	Number of Students Absent	Number of Students Assessed
Chile	95%	5,085	68	76	4,941	244	4,697
Chinese Taipei	98%	5,901	106	42	5,753	143	5,610
England	94%	4,310	140	70	4,100	242	3,858
Finland	95%	5,979	62	62	5,855	285	5,570
France	96%	4,700	53	49	4,598	172	4,426
Georgia	97%	4,031	37	73	3,921	132	3,789
Hong Kong SAR	93%	4,115	12	73	4,030	300	3,730
Hungary	96%	5,540	23	86	5,431	212	5,219
Israel	92%	4,737	36	51	4,650	381	4,269
Italy	97%	4,456	22	153	4,281	143	4,138
Korea, Rep. of	97%	4,593	18	37	4,538	129	4,409
Lithuania	92%	4,843	19	98	4,726	360	4,366
Malaysia	98%	8,354	120	0	8,234	157	8,077
Norway (9)	88%	6,069	41	141	5,887	671	5,216
Portugal	95%	4,268	32	152	4,084	217	3,867
Qatar	97%	4,767	138	32	4,597	160	4,437
Russian Federation	97%	4,692	28	76	4,588	132	4,456
Singapore	96%	5,797	19	0	5,778	232	5,546
Sweden	91%	5,313	64	213	5,036	471	4,565
Turkey	98%	4,971	111	123	4,737	75	4,662
United Arab Emirates	95%	27,322	251	315	26,756	1217	25,539
United States	93%	11,270	307	242	10,721	777	9,944
Benchmarking Participants							
Ontario, Canada	94%	4,779	63	75	4,641	311	4,330
Quebec, Canada	94%	3,895	28	7	3,860	222	3,638
Moscow City, Russian Fed.	97%	4,517	21	19	4,477	153	4,324
Abu Dhabi, UAE	95%	9,997	41	86	9,870	490	9,380
Dubai, UAE	96%	7,164	199	141	6,824	280	6,544

Appendix 17B: Within-Country Scoring Reliability for Human Scored Items in the TIMSS 2019 Problem Solving and Inquiry Tasks

Within-Country Scoring Reliability for Human Scored Items in the TIMSS 2019 Problem Solving and Inquiry Tasks—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
Austria	98	96	99	96	96	97
Canada	93	88	98	93	88	97
Chile	97	97	98	96	96	97
Chinese Taipei	95	93	97	94	93	95
Croatia	98	98	99	98	98	98
Czech Republic	96	93	99	95	93	97
Denmark	97	95	98	96	95	98
England	97	97	97	96	95	97
Finland	100	100	100	100	100	100
France	99	99	99	98	97	99
Georgia	96	94	97	96	94	97
Germany	96	95	98	94	93	95
Hong Kong SAR	100	100	100	100	100	100
Hungary	96	95	96	96	95	96
Italy	98	98	99	98	98	98
Korea, Rep. of	98	97	98	98	97	98
Lithuania	97	95	100	97	95	98
Malta	93	91	95	90	88	91
Netherlands	92	88	96	91	88	93
Norway (5)	97	97	98	97	97	97
Portugal	96	94	97	94	94	94
Qatar	97	96	97	97	96	97
Russian Federation	97	95	98	97	95	98
Singapore	99	99	100	99	99	99
Slovak Republic	97	96	99	96	96	97
Spain	97	94	99	97	94	99
Sweden	99	98	99	99	98	99
Turkey (5)	99	99	99	99	98	99
United Arab Emirates	99	98	99	98	98	98
United States	98	96	99	97	96	98
International Average	97	96	98	96	95	97
Benchmarking Participants						
Ontario, Canada	94	88	100	94	88	100
Quebec, Canada	95	93	98	94	93	96
Moscow City, Russian Fed.	99	98	99	99	98	99
Madrid, Spain	97	95	99	96	95	97
Abu Dhabi, UAE	100	100	100	98	97	100
Dubai, UAE	98	98	99	98	97	98

Within-Country Scoring Reliability for Human Scored Items in the TIMSS 2019 Problem Solving and Inquiry Tasks—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
Austria	95	93	98	95	93	97
Canada	89	81	97	88	81	96
Chile	94	91	97	93	91	96
Chinese Taipei	94	88	100	93	88	99
Croatia	92	85	97	92	85	97
Czech Republic	91	86	97	90	84	96
Denmark	93	85	98	93	85	98
England	93	88	98	93	88	97
Finland	97	93	100	97	92	99
France	93	89	97	93	89	97
Georgia	91	79	98	91	76	98
Germany	94	88	97	94	82	97
Hong Kong SAR	100	100	100	100	100	100
Hungary	89	79	98	89	79	97
Italy	97	96	99	97	95	99
Korea, Rep. of	97	94	100	97	94	100
Lithuania	94	89	98	93	88	97
Malta	91	85	99	91	85	98
Netherlands	89	83	96	89	82	94
Norway (5)	91	80	96	90	80	95
Portugal	94	90	98	94	90	98
Qatar	95	91	100	94	91	98
Russian Federation	94	90	99	94	90	98
Singapore	95	89	100	94	86	100
Slovak Republic	96	93	99	96	93	99
Spain	91	86	98	90	86	97
Sweden	90	86	96	89	84	96
Turkey (5)	95	90	99	95	90	98
United Arab Emirates	93	91	97	93	91	96
United States	96	93	99	96	93	99
International Average	93	88	98	93	88	98
Benchmarking Participants						
Ontario, Canada	89	81	100	88	81	97
Quebec, Canada	90	79	100	89	79	100
Moscow City, Russian Fed.	97	92	99	96	92	99
Madrid, Spain	91	84	99	90	84	97
Abu Dhabi, UAE	93	88	99	92	88	96
Dubai, UAE	91	84	97	90	84	96

Within-Country Scoring Reliability for Human Scored Items in the TIMSS 2019 Problem Solving and Inquiry Tasks—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
Chile	97	94	99	97	94	99
Chinese Taipei	98	94	100	98	94	100
England	98	96	99	98	95	99
Finland	99	99	100	99	99	100
France	97	93	99	97	93	99
Georgia	95	89	99	95	89	99
Hong Kong SAR	100	100	100	100	100	100
Hungary	96	93	99	96	93	99
Israel	98	96	99	98	96	99
Italy	97	93	99	97	93	99
Korea, Rep. of	98	96	99	98	96	99
Lithuania	97	92	100	97	92	100
Malaysia	99	97	100	99	97	100
Norway (9)	96	92	99	96	91	99
Portugal	98	96	100	98	96	100
Qatar	97	96	99	97	96	99
Russian Federation	99	96	100	99	96	100
Singapore	98	96	100	98	96	100
Sweden	98	95	99	98	95	99
Turkey	99	97	100	99	97	100
United Arab Emirates	98	93	100	98	93	100
United States	98	94	100	98	94	100
International Average	98	95	100	98	95	100
Benchmarking Participants						
Ontario, Canada	95	93	98	95	93	98
Quebec, Canada	96	91	100	96	90	100
Moscow City, Russian Fed.	97	94	99	97	93	99
Abu Dhabi, UAE	99	98	100	99	98	100
Dubai, UAE	96	82	100	96	82	100

Within-Country Scoring Reliability for Human Scored Items in the TIMSS 2019 Problem Solving and Inquiry Tasks—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
Chile	94	91	99	94	90	99
Chinese Taipei	93	85	98	92	85	98
England	93	90	98	93	90	98
Finland	95	90	99	95	89	99
France	95	86	99	94	86	99
Georgia	94	88	99	93	88	99
Hong Kong SAR	100	100	100	100	100	100
Hungary	87	78	94	86	78	94
Israel	90	84	96	89	77	96
Italy	95	92	99	95	91	99
Korea, Rep. of	95	88	99	94	88	99
Lithuania	90	83	99	90	83	99
Malaysia	93	88	97	93	88	97
Norway (9)	92	87	99	91	84	99
Portugal	92	88	97	92	88	97
Qatar	94	92	98	94	91	98
Russian Federation	94	88	100	93	88	96
Singapore	95	88	100	94	88	100
Sweden	89	82	96	89	82	96
Turkey	95	88	100	95	86	100
United Arab Emirates	93	90	96	92	90	96
United States	93	87	100	93	87	99
International Average	93	88	98	93	87	98
Benchmarking Participants						
Ontario, Canada	91	80	98	90	80	98
Quebec, Canada	88	80	98	87	80	98
Moscow City, Russian Fed.	98	93	100	97	93	100
Abu Dhabi, UAE	94	89	99	94	89	99
Dubai, UAE	92	85	98	91	85	98

Appendix 17C: Modifications to the TIMSS 2019 Problem Solving and Inquiry Data

Grade 4

Items Excluded from Scaling for All Countries*

MI02_04B – MQ11P04B (severe differential item functioning)

SI02_06 – SQ11S06 (severe differential item functioning)

Items Recoded for All Countries

SI01_02 – SQ11F02 (20 to 10)

SI01_05 – SQ11F05 (20 to 10, 10 to 70, 11 to 71)

Items Deleted by Country

Hong Kong SAR

SI01_07A – SQ11F07A (severe item-by-country interaction)

Grade 8

Items Excluded from Scaling for All Countries*

SI02_07 – SQ12S07 (item dependency)

SI02_08 – SQ12S08 (item dependency)

Appendix 17D: Derived Items in the TIMSS 2019 Problem Solving and Inquiry Tasks

Grade 4

MI01_04A – MQ11A04A: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct

MI01_05B – MQ11A05B: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct

MI01_09 – MQ11R03: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct

MI02_01B – MQ11P01B: 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 all parts are correct except for either part C or part E

MI02_04 – MQ11P04: Item parts AA and AB 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 (part B is excluded)

SI01_03 – SQ11F03: 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

SI01_04 – SQ11F04: 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

SI01_06 – SQ11F06: 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

SI01_08 – SQ11F08: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

SI02_03 – SQ11S03: 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 only part A is correct

SI02_04 – SQ11S04: 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

MI02_02A – MQ12D02A: 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

MI02_04 – MQ12D04: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct or if only part A is correct

SI01_09 – SQ12S09: 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

Appendix 17E: Item Statistics by Booklet Position for the TIMSS 2019 Problem Solving and Inquiry Tasks

International Average Item Block Statistics by Booklet Position for the TIMSS 2019 Problem Solving and Inquiry Tasks—Grade 4

PSI Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted 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	
Mathematics									
MI01 - Positions 1 & 4	11,605	11,597	47.0	41.5	11.8	13.4	0.7	8.4	
MI02 - Positions 2 & 3	11,597	11,605	34.2	30.5	9.0	10.6	0.2	15.4	
Overall	23,202	23,202	40.6	36.0	10.4	12.0	0.4	11.9	
Science									
SI01 - Positions 2 & 3	11,596	11,614	48.9	43.9	6.5	7.4	0.1	16.8	
SI02 - Positions 1 & 4	11,613	11,596	42.6	41.0	6.8	8.8	0.8	7.9	
Overall	23,209	23,210	45.8	42.5	6.6	8.1	0.4	12.4	

International Average Item Block Statistics by Booklet Position for the TIMSS 2019 Problem Solving and Inquiry Tasks—Grade 8

PSI Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted 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	
Mathematics									
MI01 - Positions 1 & 4	8,120	8,156	28.1	25.8	11.5	10.8	0.2	5.7	
MI02 - Positions 2 & 3	8,156	8,120	33.2	30.5	12.2	15.0	0.3	5.4	
Overall	16,276	16,276	30.6	28.1	11.8	12.9	0.3	5.5	
Science									
SI01 - Positions 2 & 3	8,106	8,157	46.7	46.5	8.3	6.8	0.2	3.1	
SI02 - Positions 1 & 4	8,157	8,106	43.2	38.4	5.7	10.6	0.1	1.7	
Overall	16,263	16,263	45.0	42.5	7.0	8.7	0.2	2.4	

Country Average Item Statistics by Booklet Position for the TIMSS 2019 Problem Solving and Inquiry Tasks—Grade 4 Mathematics

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted 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
Austria	633	633	43.4	38.4	13.4	15.5	0.2	10.4
Canada	1,919	1,919	35.5	29.9	9.6	12.4	0.8	15.5
Chile	599	599	24.5	19.6	13.9	17.1	1.7	24.4
Chinese Taipei	530	530	54.7	51.7	8.2	7.5	0.0	2.1
Croatia	550	550	36.8	33.3	11.4	11.5	0.1	6.1
Czech Republic	665	665	43.7	38.8	11.6	13.3	0.2	9.0
Denmark	465	465	45.3	38.9	9.8	13.3	1.3	23.4
England	475	475	49.9	47.0	6.9	7.7	0.2	5.0
Finland	664	664	42.6	37.3	11.0	12.5	0.2	8.9
France	606	606	32.6	26.9	15.1	18.9	1.0	18.7
Georgia	525	525	29.5	23.9	20.1	21.1	0.6	15.4
Germany	495	495	39.0	34.7	11.1	13.8	0.5	16.1
Hong Kong SAR	418	418	57.7	55.0	8.5	7.4	0.1	1.9
Hungary	656	656	43.8	39.0	8.9	9.6	0.2	7.5
Italy	527	527	36.1	30.8	13.9	15.8	1.0	21.7
Korea, Rep. of	555	555	57.1	53.2	5.9	6.7	0.0	3.1
Lithuania	524	524	44.0	39.8	9.2	11.0	0.0	6.2
Malta	522	522	37.7	32.4	7.3	8.9	0.2	7.3
Netherlands	474	474	46.2	41.6	7.9	9.2	0.2	7.4
Norway (5)	575	575	47.4	42.8	10.7	13.5	0.4	14.7
Portugal	614	614	38.9	31.7	10.2	13.1	0.7	15.5
Qatar	713	713	24.1	20.3	14.6	16.8	1.1	17.8
Russian Federation	574	574	51.7	47.1	9.7	10.1	0.1	7.6
Singapore	853	853	59.6	55.7	6.5	7.6	0.0	3.1
Slovak Republic	614	614	41.2	35.3	10.7	12.2	0.1	7.4
Spain	1,390	1,390	38.4	32.2	9.2	11.1	0.4	10.9
Sweden	567	567	42.7	36.7	11.9	14.7	1.2	22.7
Turkey (5)	571	571	36.9	32.4	11.6	11.3	0.1	6.9
United Arab Emirates	3,677	3,677	29.6	25.7	13.7	15.3	0.6	10.9
United States	1,252	1,252	40.3	36.3	6.6	9.1	0.4	11.7
International Average	23,202	23,202	41.7	36.9	10.6	12.3	0.5	11.3
Benchmarking Participants								
Ontario, Canada	528	528	37.5	31.4	9.0	12.1	0.7	15.0
Quebec, Canada	546	546	40.9	32.4	9.1	12.9	0.5	14.6
Moscow City, Russian Fed.	549	549	56.3	52.6	9.0	9.0	0.0	4.5
Madrid, Spain	488	488	40.3	34.0	8.5	10.3	0.3	9.0
Abu Dhabi, UAE	1,287	1,287	23.1	19.4	14.7	16.1	0.4	10.0
Dubai, UAE	1,034	1,034	42.6	37.3	8.8	12.1	0.2	7.4

Country Average Item Statistics by Booklet Position for the TIMSS 2019 Problem Solving and Inquiry Tasks—Grade 4 Science

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted 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
Austria	633	633	45.3	40.6	8.8	12.1	0.1	12.6
Canada	1,924	1,924	44.1	40.4	5.1	7.2	0.9	17.1
Chile	599	599	36.6	32.9	9.1	10.7	1.3	28.1
Chinese Taipei	530	530	50.3	49.4	7.6	10.0	0.1	6.7
Croatia	550	550	45.9	41.7	6.3	5.8	0.0	5.5
Czech Republic	666	666	48.8	44.8	8.1	8.3	0.2	8.9
Denmark	465	465	50.5	48.8	4.8	5.1	0.2	12.2
England	475	475	52.7	52.5	3.8	4.6	0.3	4.3
Finland	665	665	51.6	48.4	5.1	6.0	0.1	5.7
France	605	605	40.7	35.3	9.8	14.1	1.3	23.5
Georgia	526	526	31.6	25.7	12.5	15.5	0.4	18.0
Germany	496	496	43.9	38.4	7.2	10.1	1.4	24.4
Hong Kong SAR	418	418	43.7	40.6	8.1	9.3	0.4	7.8
Hungary	656	656	47.0	43.5	6.2	6.1	0.3	10.9
Italy	528	528	42.6	34.3	6.5	10.9	0.8	24.2
Korea, Rep. of	555	555	61.0	60.4	5.6	4.4	0.1	1.7
Lithuania	524	524	47.0	44.7	6.5	7.7	0.1	4.0
Malta	523	523	41.2	38.6	5.1	5.9	0.3	9.1
Netherlands	474	474	44.7	44.5	4.2	5.1	0.1	3.7
Norway (5)	575	575	48.8	47.1	5.7	6.8	0.4	7.7
Portugal	615	615	43.2	38.0	6.9	9.5	0.6	25.1
Qatar	712	712	27.9	24.6	11.8	13.5	1.4	24.5
Russian Federation	574	574	55.9	52.0	5.2	5.6	0.2	5.7
Singapore	851	851	59.3	56.4	3.7	4.6	0.0	4.4
Slovak Republic	615	615	48.0	45.4	6.0	7.4	0.0	6.6
Spain	1,390	1,390	45.0	42.9	5.2	7.2	0.1	9.5
Sweden	569	569	50.1	46.9	5.5	8.2	0.5	15.0
Turkey (5)	571	571	47.1	44.5	6.2	7.1	0.1	10.3
United Arab Emirates	3,673	3,674	31.0	28.0	8.9	10.2	0.6	16.8
United States	1,252	1,252	47.1	42.8	3.5	4.3	1.2	16.2
International Average	23,209	23,210	45.8	42.5	6.6	8.1	0.4	12.4
Benchmarking Participants								
Ontario, Canada	529	529	45.2	40.8	5.1	7.2	0.8	17.8
Quebec, Canada	547	547	45.0	39.8	5.0	7.7	0.9	16.4
Moscow City, Russian Fed.	549	549	60.0	54.7	5.1	6.4	0.0	5.2
Madrid, Spain	489	489	45.5	44.5	5.1	5.2	0.0	8.7
Abu Dhabi, UAE	1,287	1,287	22.9	19.4	9.1	12.1	0.7	18.5
Dubai, UAE	1,034	1,034	43.0	40.1	5.5	5.7	0.4	12.7

Country Average Item Statistics by Booklet Position for the TIMSS 2019 Problem Solving and Inquiry Tasks—Grade 8 Mathematics

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted 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
Chile	579	579	17.9	14.6	17.8	18.8	0.3	10.8
Chinese Taipei	695	695	47.7	44.2	4.6	6.2	0.1	1.1
England	492	492	30.0	29.0	15.1	16.0	0.1	5.8
Finland	689	689	27.5	25.7	12.4	13.4	0.5	5.9
France	552	552	23.1	20.1	16.4	18.5	0.1	8.5
Georgia	473	473	16.3	15.2	22.8	22.1	1.2	10.2
Hong Kong SAR	465	465	43.7	42.8	6.1	6.3	0.1	1.4
Hungary	650	650	33.5	33.2	12.7	11.6	0.0	1.7
Israel	537	537	29.4	24.5	11.4	12.5	0.4	7.0
Italy	519	519	24.6	21.1	17.2	18.2	0.3	5.7
Korea, Rep. of	548	548	42.4	42.0	7.8	8.5	0.2	2.2
Lithuania	540	540	31.1	29.2	12.8	13.7	0.1	3.5
Malaysia	1,012	1,012	26.0	24.5	3.7	4.7	0.0	3.1
Norway (9)	638	638	29.8	27.9	21.2	21.3	1.5	12.9
Portugal	489	489	28.0	24.2	13.2	14.6	0.2	5.5
Qatar	552	552	18.4	15.4	10.1	11.3	0.1	5.1
Russian Federation	555	555	36.1	33.3	14.5	16.0	0.1	5.6
Singapore	693	693	55.2	52.1	3.7	4.1	0.0	1.9
Sweden	567	567	33.6	27.4	14.5	18.5	0.7	11.5
Turkey	585	585	22.9	21.5	10.9	11.5	0.1	2.4
United Arab Emirates	3,203	3,203	22.0	18.7	7.3	7.7	0.0	3.8
United States	1,243	1,243	31.3	28.3	3.8	4.8	0.3	5.8
International Average	16,276	16,276	30.5	28.0	11.8	12.7	0.3	5.5
Benchmarking Participants								
Ontario, Canada	553	553	35.5	31.2	8.8	10.3	0.1	7.7
Quebec, Canada	458	458	38.1	32.4	8.9	10.1	0.1	5.8
Moscow City, Russian Fed.	541	541	44.9	41.4	13.1	13.3	0.0	5.2
Abu Dhabi, UAE	1,175	1,175	17.6	15.0	5.9	6.6	0.0	2.9
Dubai, UAE	816	816	34.1	30.7	7.6	8.5	0.1	4.8

Country Average Item Statistics by Booklet Position for the TIMSS 2019 Problem Solving and Inquiry Tasks—Grade 8 Science

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted 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
Chile	579	579	38.1	34.9	8.5	10.1	0.0	4.2
Chinese Taipei	695	695	53.5	53.2	4.0	4.5	0.0	0.7
England	491	491	47.2	46.6	6.6	7.6	0.6	1.9
Finland	691	691	51.5	50.2	4.2	5.4	0.3	1.3
France	552	552	45.7	42.8	7.1	10.0	0.4	6.0
Georgia	473	473	32.3	28.7	14.4	17.5	0.4	4.1
Hong Kong SAR	465	465	46.8	43.2	6.5	7.6	0.0	1.2
Hungary	648	648	47.8	47.0	6.2	7.4	0.0	0.2
Israel	537	537	45.6	42.1	9.0	12.1	0.1	3.3
Italy	519	519	38.0	34.7	9.5	10.8	0.0	2.4
Korea, Rep. of	548	548	53.6	51.1	4.4	5.4	0.1	0.8
Lithuania	540	540	45.4	44.5	6.0	6.3	0.0	0.2
Malaysia	1,012	1,012	41.1	38.3	4.1	5.3	0.0	3.0
Norway (9)	630	630	43.3	40.5	9.3	10.9	1.3	4.9
Portugal	490	490	40.6	36.9	7.7	9.5	0.1	2.0
Qatar	552	552	34.2	30.7	9.1	12.0	0.2	3.0
Russian Federation	555	555	46.1	45.9	8.1	7.8	0.0	2.1
Singapore	692	692	64.2	63.4	2.6	2.8	0.1	0.5
Sweden	560	560	48.5	44.6	7.6	10.2	0.3	4.6
Turkey	585	585	42.4	40.0	9.0	11.0	0.1	2.2
United Arab Emirates	3,205	3,205	33.4	30.0	8.7	10.6	0.0	2.7
United States	1,244	1,244	52.0	50.0	2.7	3.9	0.2	3.2
International Average	16,263	16,263	45.1	42.7	7.1	8.6	0.2	2.5
Benchmarking Participants								
Ontario, Canada	553	553	49.5	46.0	4.9	6.0	0.3	3.5
Quebec, Canada	459	459	52.9	50.3	4.4	5.5	0.0	1.5
Moscow City, Russian Fed.	540	540	55.2	53.6	6.2	5.9	0.0	1.8
Abu Dhabi, UAE	1,176	1,176	28.3	24.9	8.5	11.4	0.0	2.0
Dubai, UAE	816	816	47.0	43.1	6.0	7.3	0.0	2.5

Appendix 17F: Item Parameters from the eTIMSS with PSI 2019 Calibrations

Problem Solving and Inquiry Item Parameters from the eTIMSS with PSI 2019 Calibration—Grade 4 Mathematics

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MI01_01	MQ11A01	0.020	1.085 (0.132)	1.226 (0.063)	0.121 (0.019)	
MI01_02A	MQ11A02A	0.033	0.890 (0.050)	-0.499 (0.045)		
MI01_02B	MQ11A02B	0.026	0.786 (0.046)	0.238 (0.041)		
MI01_02C	MQ11A02C	0.027	0.994 (0.053)	-0.195 (0.036)		
MI01_03A	MQ11A03A	0.038	0.485 (0.038)	0.480 (0.066)		
MI01_03B	MQ11A03B	0.032	0.993 (0.053)	-0.194 (0.036)		
MI01_04A	MQ11A04A	0.027	0.595 (0.027)	0.817 (0.039)	-0.419 (0.066)	0.419 (0.078)
MI01_04B	MQ11A04B	0.032	0.995 (0.054)	0.328 (0.034)		
MI01_05A	MQ11A05A	0.022	1.313 (0.180)	1.257 (0.062)	0.203 (0.019)	
MI01_05B	MQ11A05B	0.026	0.924 (0.050)	-0.096 (0.037)		
MI01_06A	MQ11A06A	0.026	1.310 (0.065)	0.087 (0.028)		
MI01_06B	MQ11A06B	0.031	1.400 (0.081)	1.030 (0.037)		
MI01_07A	MQ11R01A	0.041	0.527 (0.041)	-1.035 (0.095)		
MI01_07B	MQ11R01B	0.029	0.731 (0.045)	-0.373 (0.051)		
MI01_08A	MQ11R02A	0.028	1.174 (0.067)	0.837 (0.037)		
MI01_08B	MQ11R02B	0.026	1.079 (0.059)	0.134 (0.034)		
MI01_09	MQ11R03	0.019	1.313 (0.074)	0.714 (0.033)		
MI01_10	MQ11R04	0.019	0.542 (0.023)	0.943 (0.044)	-1.311 (0.101)	1.311 (0.111)
MI02_01A	MQ11P01A	0.020	1.002 (0.055)	0.421 (0.035)		
MI02_01B	MQ11P01B	0.044	0.469 (0.025)	1.413 (0.070)	-0.660 (0.088)	0.660 (0.115)
MI02_02	MQ11P02	0.060	0.753 (0.046)	0.263 (0.043)		
MI02_03	MQ11P03	0.046	0.300 (0.014)	-0.730 (0.069)	-2.194 (0.166)	2.194 (0.152)
MI02_04	MQ11P04	0.021	0.768 (0.031)	1.058 (0.035)	-1.018 (0.078)	1.018 (0.088)
MI02_05A	MQ11P05A	0.026	0.977 (0.054)	-0.373 (0.041)		
MI02_05B	MQ11P05B	0.023	0.884 (0.069)	1.651 (0.089)		
MI02_06A	MQ11P06A	0.024	0.866 (0.069)	1.659 (0.092)		
MI02_06B	MQ11P06B	0.019	1.041 (0.067)	1.121 (0.052)		
MI02_07A	MQ11P07A	0.020	1.175 (0.083)	1.368 (0.059)		
MI02_07B	MQ11P07B	0.021	1.207 (0.096)	1.617 (0.074)		

The concurrent calibration model included regular eTIMSS items with item parameters fixed to equal the values from the final eTIMSS 2019 adjusted model calibration (see Chapter 12, Appendix 12K).

Problem Solving and Inquiry Item Parameters from the eTIMSS with PSI 2019 Calibration—Grade 4 Science

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SI01_01 SQ11F01	0.028	0.454 (0.025)	0.762 (0.046)		-0.514 (0.082)	0.514 (0.093)
SI01_02 SQ11F02	0.033	1.007 (0.058)	0.360 (0.033)			
SI01_03 SQ11F03	0.034	0.650 (0.045)	-0.194 (0.054)			
SI01_04 SQ11F04	0.037	0.782 (0.050)	0.055 (0.043)			
SI01_05 SQ11F05	0.032	0.656 (0.047)	0.099 (0.050)			
SI01_06 SQ11F06	0.042	0.424 (0.040)	-0.683 (0.106)			
SI01_07A SQ11F07A	0.036	0.570 (0.077)	0.061 (0.185)	0.178 (0.059)		
SI01_07B SQ11F07B	0.019	0.820 (0.079)	1.793 (0.122)			
SI01_08 SQ11F08	0.036	0.546 (0.044)	-0.663 (0.086)			
SI01_09 SQ11F09	0.034	0.737 (0.042)	0.678 (0.033)		0.218 (0.051)	-0.218 (0.060)
SI02_01 SQ11S01	0.035	1.028 (0.057)	-0.103 (0.035)			
SI02_02 SQ11S02	0.031	0.927 (0.055)	0.339 (0.035)			
SI02_03 SQ11S03	0.039	0.448 (0.019)	0.805 (0.043)		-1.580 (0.105)	1.580 (0.113)
SI02_04 SQ11S04	0.030	0.668 (0.036)	0.547 (0.032)		0.157 (0.054)	-0.157 (0.060)
SI02_05 SQ11S05	0.033	0.706 (0.039)	0.731 (0.034)		0.405 (0.049)	-0.405 (0.060)
SI02_07 SQ11S07	0.045	0.498 (0.030)	0.520 (0.045)		0.768 (0.070)	-0.768 (0.079)
SI02_08 SQ11S08	0.028	0.559 (0.047)	0.630 (0.062)			
SI02_09 SQ11S09	0.029	0.734 (0.051)	0.235 (0.045)			

The concurrent calibration model included regular eTIMSS items with item parameters fixed to equal the values from the final eTIMSS 2019 adjusted model calibration (see Chapter 12, Appendix 12L).

Problem Solving and Inquiry Item Parameters from the eTIMSS with PSI 2019 Calibration—Grade 8 Mathematics

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MI01_01 MQ12B01	0.032	1.043 (0.066)	0.467 (0.042)			
MI01_02 MQ12B02	0.024	1.779 (0.255)	1.528 (0.057)	0.123 (0.014)		
MI01_03 MQ12B03	0.027	0.617 (0.031)	0.932 (0.047)		-0.880 (0.100)	0.880 (0.111)
MI01_04A MQ12B04A	0.035	1.605 (0.102)	0.846 (0.034)			
MI01_04B MQ12B04B	0.036	1.394 (0.092)	1.023 (0.040)			
MI01_04C MQ12B04C	0.043	0.871 (0.062)	0.921 (0.056)			
MI01_05A MQ12B05A	0.031	1.390 (0.087)	0.771 (0.036)			
MI01_05B MQ12B05B	0.025	1.472 (0.093)	0.816 (0.036)			
MI01_05C MQ12B05C	0.017	0.965 (0.069)	2.084 (0.077)		-0.761 (0.114)	0.761 (0.152)
MI01_06 MQ12R01	0.020	1.400 (0.099)	1.227 (0.046)			
MI01_07A MQ12R02A	0.027	1.092 (0.074)	0.877 (0.046)			
MI01_07B MQ12R02B	0.031	1.086 (0.075)	0.866 (0.047)			
MI01_07C MQ12R02C	0.021	1.626 (0.121)	1.249 (0.043)			
MI02_01 MQ12D01	0.033	1.368 (0.084)	0.629 (0.035)			
MI02_02A MQ12D02A	0.091	0.718 (0.055)	-0.683 (0.077)			
MI02_02B MQ12D02B	0.027	1.428 (0.097)	1.111 (0.042)			
MI02_03 MQ12D03	0.022	1.527 (0.100)	0.977 (0.037)			
MI02_04 MQ12D04	0.043	0.712 (0.053)	0.558 (0.058)			
MI02_05 MQ12D05	0.028	0.905 (0.065)	1.007 (0.057)			
MI02_06A MQ12D06A	0.045	1.978 (0.162)	1.507 (0.044)			
MI02_06B MQ12D06B	0.029	2.461 (0.212)	1.491 (0.038)			
MI02_07A MQ12D07A	0.023	1.314 (0.101)	1.437 (0.056)			
MI02_07B MQ12D07B	0.024	0.910 (0.050)	1.536 (0.049)		-0.616 (0.083)	0.616 (0.103)
MI02_08A MQ12D08A	0.085	0.788 (0.059)	-0.681 (0.072)			
MI02_08B MQ12D08B	0.057	0.594 (0.052)	0.967 (0.082)			

The concurrent calibration model included regular eTIMSS items with item parameters fixed to equal the values from the final eTIMSS 2019 adjusted model calibration (see Chapter 12, Appendix 12M).

Problem Solving and Inquiry Item Parameters from the eTIMSS with PSI 2019 Calibration—Grade 8 Science

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SI01_01	SQ12S01	0.044	0.615 (0.055)	-1.052 (0.114)		
SI01_02	SQ12S02	0.044	0.887 (0.069)	-1.004 (0.083)		
SI01_03	SQ12S03	0.027	1.096 (0.075)	0.775 (0.043)		
SI01_04	SQ12S04	0.038	1.090 (0.071)	0.193 (0.039)		
SI01_05	SQ12S05	0.030	0.810 (0.112)	0.345 (0.137)	0.211 (0.054)	
SI01_06	SQ12S06	0.045	1.100 (0.075)	0.781 (0.043)		
SI01_07A	SQ12S07A	0.036	1.115 (0.120)	-0.512 (0.125)	0.234 (0.062)	
SI01_07B	SQ12S07B	0.033	1.785 (0.253)	1.303 (0.052)	0.140 (0.017)	
SI01_07C	SQ12S07C	0.029	0.840 (0.115)	0.536 (0.114)	0.181 (0.046)	
SI01_08	SQ12S08	0.046	0.375 (0.045)	-0.335 (0.118)		
SI01_09	SQ12S09	0.029	0.423 (0.054)	1.812 (0.198)		
SI01_10	SQ12S10	0.050	0.889 (0.063)	-0.035 (0.049)		
SI01_11	SQ12S11	0.020	0.753 (0.072)	1.590 (0.108)		
SI01_12	SQ12S12	0.033	1.022 (0.070)	0.603 (0.043)		
SI01_13A	SQ12S13A	0.031	0.796 (0.110)	0.597 (0.114)	0.160 (0.045)	
SI01_13B	SQ12S13B	0.025	0.873 (0.061)	1.494 (0.059)	0.131 (0.058)	-0.131 (0.093)
SI02_01	SQ12P01	0.027	0.903 (0.111)	-0.483 (0.175)	0.277 (0.073)	
SI02_02	SQ12P02	0.038	0.317 (0.024)	1.609 (0.121)	-1.020 (0.150)	1.020 (0.188)
SI02_03	SQ12P03	0.024	0.820 (0.067)	1.157 (0.070)		
SI02_04	SQ12P04	0.037	1.184 (0.069)	0.937 (0.030)	0.298 (0.040)	-0.298 (0.053)
SI02_05	SQ12P05	0.032	0.414 (0.026)	-0.433 (0.068)	-0.557 (0.130)	0.557 (0.112)
SI02_06	SQ12P06	0.065	0.952 (0.068)	0.780 (0.049)		
SI02_09	SQ12P10	0.036	0.754 (0.061)	0.873 (0.062)		
SI02_10	SQ12P13	0.022	1.216 (0.088)	1.125 (0.049)		
SI02_11	SQ12P14	0.020	1.159 (0.089)	1.271 (0.058)		
SI02_12	SQ12P15	0.049	0.479 (0.025)	0.437 (0.047)	-0.891 (0.109)	0.891 (0.111)
SI02_13	SQ12P16	0.025	1.353 (0.206)	0.735 (0.084)	0.352 (0.035)	

The concurrent calibration model included regular eTIMSS items with item parameters fixed to equal the values from the final eTIMSS 2019 adjusted model calibration (see Chapter 12, Appendix 12N).