

Appendix F

The Test-Curriculum Matching Analysis—Mathematics

TIMSS went to great lengths to ensure that comparisons of student achievement across countries would be as fair and equitable as possible. The TIMSS 2011 Assessment Frameworks were designed to specify the important aspects of mathematics that participating countries agreed should be the focus of an international assessment of mathematics achievement, and the assessment items were developed through a collaborative process with national representatives to faithfully represent the specifications in the frameworks and field tested extensively in participating countries. Finalizing the TIMSS 2011 assessments involved a series of reviews by representatives of the participating countries, experts in mathematics, and testing specialists. At the end of this process, the National Research Coordinators (NRCs) from each country formally approved the TIMSS 2011 assessments, thus accepting them as being sufficiently fair to compare their students' mathematics achievement with that of students from other countries.

Although the assessments were developed to represent an agreed-upon framework and were intended to have as much in common across countries as possible, it was unavoidable that the match between the TIMSS 2011 assessment (or test) and the mathematics curriculum would not be the same in all countries. To restrict test items to just those topics included in the curricula of all participating countries and covered in the same sequence would severely limit test coverage and restrict the research questions that the study is designed to address. The tests, therefore, inevitably have some items measuring topics unfamiliar to some students in some countries.

The Test-Curriculum Matching Analysis (TCMA) was conducted to investigate the extent to which the TIMSS 2011 mathematics assessment was relevant to each country's curriculum. The TCMA also investigates the impact on a country's performance of including only achievement items that were judged to be relevant to its own curriculum.¹

To gather data about the extent to which the TIMSS 2011 tests were relevant to the curricula of the TIMSS countries and benchmarking participants, NRCs were asked to examine each achievement item and indicate whether the item was in their country's intended curriculum at the grade tested (fourth or eighth grade). The NRCs were asked to choose persons very familiar with the curriculum at these grades to make this determination. In some countries, the curriculum was prescribed for a range of grades and was not explicit about what was to be covered by the end of the fourth or eighth grades. For example, in Sweden the curriculum specifies the curricular goals to be achieved by the end of

¹ Because there also may be curriculum areas covered in some countries that are not covered by the TIMSS 2011 tests, the TCMA does not provide complete information about how well the tests cover the curricula of the countries.

the fifth and ninth grades, but does not provide a grade-by-grade specification. In such situations, coordinators were asked to make the best judgment possible.² Because an item might be in the curriculum for some but not all students in a country, NRCs were asked to consider an item included if it was in the intended curriculum for more than 50 percent of the students. All TIMSS 2011 participants took part in the TCMA analysis except Bahrain, Georgia, Saudi Arabia, Honduras (sixth grade participant) and the US benchmarking states at the fourth grade, and Bahrain, Georgia, Ghana, Indonesia, Saudi Arabia, Syrian Arab Republic, Honduras (ninth grade participant), and the US benchmarking states at the eighth grade.

Exhibits F.1 through F.4 present the TCMA results for the TIMSS 2011 mathematics test at fourth and eighth grades. Exhibits F.1 and F.2 show the average percent correct on the mathematics items judged appropriate by each country at the fourth and eighth grades, respectively. Exhibits F.3 and F.4 show the standard errors corresponding to the percentages presented in Exhibits F.1 and F.2.

In Exhibit F.1, the bottom row of the exhibit shows the number of items, in terms of score points, identified as appropriate in each country. At the fourth grade, the maximum number of score points in the assessment was 184 points.³ Generally, the proportion of items judged appropriate was fairly high. Reading along the bottom row, it can be seen that two of the 47 countries that took part in the TCMA analysis judged 100 percent of the items to be included in their curricula. Another 38 countries and all of the sixth grade and benchmarking participants judged 75 percent or more (138 score points) to be appropriate. Only the Russian Federation judged less than half of the mathematics items to be included in their curricula.

At the eighth grade, the percentage of items judged appropriate was somewhat higher; four of the 36 countries and one of the five benchmarking participants judged 100 percent of the items to be appropriate (all 230 score points), and an additional 30 countries, two ninth grade, and four benchmarking participants judged 75 percent or more (173 score points) to be appropriate. For all participants, the majority of the eighth grade mathematics items were judged to be appropriate to their curricula.

Because most countries indicated that at least some items were not included in their intended curriculum at the grade tested, the data were

2 Exhibit 5 of the *TIMSS 2011 Encyclopedia* provides information on the grade-to-grade structure of the curriculum for each TIMSS 2011 participant.

3 The TIMSS 2011 fourth grade mathematics assessment contained 175 items, yielding 185 score points. However, following item review, response categories for one of the items were combined, resulting in data for 184 score points. Similarly, following item review, the 217 items and 232 score points in the eighth grade assessment were reduced to 230 score points.

Appendix F.2: Average Percent Correct for the Test-Curriculum Matching Analysis (Continued)

Read across the row to compare that country's performance based on the test items included by each of the countries across the top. Read down the column under a country name to compare the performance of the country down the left on the items included by the country listed on the top. Read along the diagonal to compare performance for each different country based on its own decisions about the test items to include.

										Benchmarking Participants										Average Percent Correct on All Items	Country
										Quebec, Canada	Ontario, Canada	Alberta, Canada	Dubai, UAE	Abu Dhabi, UAE							
Tunisia	Palestinian Nat'l Auth.	Jordan	Chile	Oman	Morocco	Botswana (9)	South Africa (9)														
74	75	76	74	74	75	73	73	75	75	74	74	73	73	74	73	74	(0.5)	Korea, Rep. of			
75	75	76	74	74	75	73	73	74	74	74	73	73	73	74	73	73	(0.9)	Singapore			
73	73	74	72	72	73	72	72	72	72	72	72	72	72	72	72	72	(0.6)	Chinese Taipei			
69	70	71	69	69	69	68	68	69	69	69	68	68	68	69	69	68	(0.9)	Hong Kong SAR			
65	65	67	65	65	65	64	64	65	65	65	64	64	64	65	65	64	(0.6)	Japan			
57	58	60	55	57	58	56	56	57	57	56	56	56	56	57	57	56	(0.9)	Russian Federation			
52	53	55	52	52	52	50	51	51	52	52	51	51	51	52	52	51	(1.0)	Israel			
50	51	53	53	50	51	48	49	51	52	51	49	49	49	51	52	49	(0.7)	Finland			
50	50	52	51	49	50	48	48	49	50	50	49	49	49	49	50	49	(0.8)	Hungary			
49	50	52	50	49	49	48	48	49	51	50	48	49	48	49	51	48	(0.7)	United States			
49	50	51	52	49	49	47	48	49	51	50	48	48	48	49	51	48	(1.4)	England			
49	49	51	51	49	48	46	47	49	51	49	48	48	48	49	51	48	(1.3)	Australia			
48	49	51	50	48	48	46	47	48	50	48	47	47	47	48	50	47	(0.5)	Slovenia			
48	49	50	48	48	48	46	46	47	48	48	47	47	47	49	50	47	(0.6)	Lithuania			
47	48	50	48	47	47	45	46	46	48	47	46	46	46	48	48	46	(0.6)	Italy			
45	45	47	47	45	44	43	44	44	45	47	45	44	44	45	47	44	(1.4)	New Zealand			
43	44	46	41	43	44	42	42	43	43	43	43	43	43	43	43	43	(1.0)	Kazakhstan			
42	43	46	41	43	43	41	42	42	42	42	42	42	42	42	43	42	(0.9)	Ukraine			
42	44	45	46	42	43	41	41	41	42	44	43	41	41	41	44	41	(0.5)	Sweden			
40	41	42	44	40	40	38	39	39	40	42	41	39	39	40	42	39	(0.6)	Norway			
40	40	41	36	39	40	38	38	38	39	38	38	38	38	38	39	38	(0.6)	Armenia			
38	39	41	39	39	38	37	38	38	38	39	39	38	38	38	39	38	(0.8)	Turkey			
38	39	41	37	38	39	37	38	38	38	38	38	38	38	38	39	38	(0.8)	Romania			
38	38	41	37	38	37	36	37	37	37	38	37	37	37	37	38	37	(0.5)	United Arab Emirates			
35	36	38	33	35	36	34	34	34	35	34	34	34	34	34	35	34	(0.8)	Lebanon			
35	36	38	36	35	36	34	34	34	35	36	35	34	34	34	35	34	(1.0)	Malaysia			
33	34	36	32	33	33	32	32	32	33	33	32	33	33	33	33	32	(0.9)	Macedonia, Rep. of			
31	32	35	32	32	32	30	31	31	32	33	32	31	31	31	32	31	(0.9)	Thailand			
31	32	34	31	31	31	30	30	30	31	31	31	30	30	30	31	30	(0.5)	Qatar			
30	31	33	30	30	30	29	29	29	30	31	30	30	30	30	30	30	(0.8)	Iran, Islamic Rep. of			
30	30	33	30	30	30	29	29	29	30	31	30	29	29	29	30	29	(0.6)	Tunisia			
29	30	33	29	29	29	28	29	29	29	30	29	29	29	29	29	29	(0.6)	Palestinian Nat'l Auth.			
29	30	33	28	29	29	28	29	29	29	29	29	29	29	29	29	29	(0.5)	Jordan			
29	30	32	31	29	29	28	28	28	29	30	29	28	29	28	29	28	(0.4)	Chile			
24	25	27	24	25	24	24	24	24	24	25	24	24	24	24	24	24	(0.3)	Oman			
23	23	25	23	23	23	22	22	22	23	23	22	22	22	22	23	22	(0.2)	Morocco			
44	45	47	44	44	44	43	43	43	44	45	44	43	43	43	44	43	(0.1)	International Avg.			
26	26	29	27	26	26	25	25	25	26	27	26	25	25	25	26	25	(0.4)	Botswana (9)			
21	21	23	21	21	21	20	20	20	21	21	21	20	20	20	21	20	(0.4)	South Africa (9)			
Benchmarking Participants																					
56	56	57	57	55	56	53	54	54	56	57	56	54	54	54	54	54	(0.7)	Quebec, Canada			
50	51	52	53	50	50	48	49	49	50	53	51	49	49	49	49	49	(0.6)	Ontario, Canada			
48	48	50	51	48	48	46	47	47	48	50	49	47	47	47	47	47	(0.7)	Alberta, Canada			
43	44	46	42	43	43	41	42	42	42	43	43	42	42	42	42	42	(0.5)	Dubai, UAE			
36	37	39	36	36	36	35	35	35	36	37	36	35	35	35	35	35	(0.8)	Abu Dhabi, UAE			
191	203	182	172	214	197	216	228	230	215	200	206	230	224	230	215	230		Number of Items (Score Points) Identified*			

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2011

analyzed to determine whether the inclusion of these items had any effect on the international performance comparisons.⁴

The first column of data in Exhibits F.1 and F.2 show the average percent correct on all test items for each participant, together with its standard error. Subsequent columns show the performance of each participant on those items judged appropriate by the participant listed at the head of the column. Participants are presented in order of their performance based on average percent correct on all items, from highest to lowest. To interpret these exhibits, choosing a country and reading across its row provides the average percent correct for the students in that country on the items selected by each of the countries listed along the top of the exhibit. For example, at the fourth grade, Singapore, where the average percent correct was 75 percent on its own set of items, had 74 percent correct on the items selected by Korea and Hong Kong SAR, 75 percent on the items selected by Chinese Taipei, 74 percent on the items selected by Japan, and so forth. The column for a country listed at the top shows how each of the other participants performed on the set of items selected as appropriate for that country's students. Using the set of items selected by the England as an example, 75 percent of these items, on average, were answered correctly by students in Singapore, 74 percent by students in Korea and Hong Kong SAR, 71 percent by students in Chinese Taipei, 70 percent by students in Japan, 65 percent by those in Northern Ireland, and so forth. The shaded diagonal element in the exhibit shows how each country performed on the set of items that it selected based on its own curriculum. Thus, students from the England averaged 60 percent correct on the set of items identified by England for the analysis.

For each country's selected items, the international averages across participating countries are presented in the lower part of the exhibit. These show that the selections of items by the participating countries varied somewhat in average difficulty, ranging at the fourth grade from 49 percent correct, for those chosen by Hong Kong SAR, to 53 percent correct for those chosen by Austria. At the eighth grade, the average percent correct ranged from 43 percent, for many participants, to 47 percent for those chosen by Jordan.

Comparing the diagonal element for a country with the overall average percent correct shows the difference between performance on the set of items chosen as appropriate for that country and performance on the test as a whole. In general, countries performed better on their own item sets

4 It should be noted that the mathematics achievement presented in Exhibits F.1 and F.2 is based on average percent correct (the percentage of students in a country, averaged across all items), which is different from the average scale scores that are presented in Chapter 1.

than on the items overall, although not by much. To illustrate, the average percent correct for Singapore across all fourth grade mathematics items was 74 percent. The diagonal element shows that students from Singapore had a slightly greater average percent correct (75 percent) across the set of items selected as appropriate for Singapore than they did overall. Most participants had a difference of one or two percentage points between the two performance measures, with the largest differences in the Slovak Republic (5 percentage points). At the eighth grade, the differences were generally smaller; the largest being in Jordan and the province of Ontario (4 percentage points).

It is clear that the selection of items does not have a major effect on the relative performance among TIMSS participants. Participants that had relatively high or low performance across all the mathematics items also had relatively high or low performance on each of the various sets of items selected for the TCMA. For example, at the eighth grade, Korea had the highest average percent correct, not only on the test as a whole, but also on all of the different item selections (with some ties), with Singapore, Chinese Taipei, Hong Kong SAR and Japan next in order of performance on practically all selections of items. Although there are some changes in the ordering of countries based on the items selected for the TCMA, most of these differences are within the boundaries of sampling error.⁵

Even when countries performed better on the items judged by them to be included in their curriculum than they did overall, their performance relative to other participants was changed little. As an example, consider the 162 score points selected by the Slovenia at the fourth grade. The students in the Slovenia did better on these items (56% correct) than on the test as a whole (53% correct). However, most other countries also did better on these particular items, with an international average of 51 percent correct compared with 50 percent correct overall. The countries that performed better than the Slovenia on the overall test also performed as well or better on the items selected by the Slovenia.

The TCMA results provide evidence that the TIMSS 2011 mathematics assessment provides a reasonable basis for comparing achievement of the participating countries and benchmarking entities. This result is not unexpected; making the assessment as fair as possible was a major consideration in test development. The fact that the majority of countries indicated that most items were appropriate for their students means that the different average percent correct estimates were based on many of the same items. Insofar as countries

5 Small differences in performance between adjacent countries shown in this exhibit usually are not statistically significant. The standard errors for the average percent correct statistics based on the TIMSS 2011 sample are provided in Exhibits F.3 and F.4. For any sample average shown in Exhibits F.1 and F.2, it can be said with 95 percent confidence that the corresponding value in the population falls between the sample estimate plus or minus two standard errors.

Appendix F.4: Standard Errors for the Test-Curriculum Matching Analysis (Continued)

Read across the row to compare that country's performance based on the test items included by each of the countries across the top. Read down the column under a country name to compare the performance of the country down the left on the items included by the country listed on the top. Read along the diagonal to compare performance for each different country based on its own decisions about the test items to include.

		Tunisia	Palestinian Nat'l Auth.	Jordan	Chile	Oman	Morocco	Botswana (9)	South Africa (9)	Benchmarking Participants					Average Percent Correct on All Items	Country
		Quebec, Canada	Ontario, Canada	Alberta, Canada	Dubai, UAE	Abu Dhabi, UAE										
0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	74 (0.5)	Korea, Rep. of	
0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	73 (0.9)	Singapore	
0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	72 (0.6)	Chinese Taipei	
0.9	0.9	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	68 (0.9)	Hong Kong SAR	
0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	64 (0.6)	Japan	
0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	56 (0.9)	Russian Federation	
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	51 (1.0)	Israel	
0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	49 (0.7)	Finland	
0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	49 (0.8)	Hungary	
0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	48 (0.7)	United States	
1.4	1.4	1.4	1.5	1.5	1.5	1.5	1.5	1.5	1.4	1.4	1.4	1.4	1.4	48 (1.4)	England	
1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	48 (1.3)	Australia	
0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	47 (0.5)	Slovenia	
0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	47 (0.6)	Lithuania	
0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	46 (0.6)	Italy	
1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	44 (1.4)	New Zealand	
1.1	1.0	1.0	1.0	1.1	1.1	1.1	1.1	1.1	1.0	1.0	1.0	1.0	1.0	43 (1.0)	Kazakhstan	
0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	42 (0.9)	Ukraine	
0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	41 (0.5)	Sweden	
0.6	0.6	0.6	0.7	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	39 (0.6)	Norway	
0.6	0.6	0.6	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	38 (0.6)	Armenia	
0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	38 (0.8)	Turkey	
0.9	0.9	0.8	0.8	0.8	0.9	0.9	0.9	0.8	0.8	0.8	0.8	0.8	0.8	38 (0.8)	Romania	
0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	37 (0.5)	United Arab Emirates	
0.9	0.9	0.8	0.8	0.8	0.9	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	34 (0.8)	Lebanon	
1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.0	1.0	1.0	1.0	1.0	34 (1.0)	Malaysia	
0.9	1.0	1.0	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	32 (0.9)	Macedonia, Rep. of	
0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	31 (0.9)	Thailand	
0.5	0.6	0.5	0.5	0.5	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	30 (0.5)	Qatar	
0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	30 (0.8)	Iran, Islamic Rep. of	
0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	29 (0.6)	Tunisia	
0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	29 (0.6)	Palestinian Nat'l Auth.	
0.5	0.6	0.6	0.5	0.6	0.6	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.5	29 (0.5)	Jordan	
0.4	0.5	0.4	0.5	0.4	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	28 (0.4)	Chile	
0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	24 (0.3)	Oman	
0.2	0.3	0.3	0.2	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.2	22 (0.2)	Morocco	
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	43 (0.1)	International Avg.	
0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	25 (0.4)	Botswana (9)	
0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	20 (0.4)	South Africa (9)	
Benchmarking Participants																
0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	54 (0.7)	Quebec, Canada	
0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	49 (0.6)	Ontario, Canada	
0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	47 (0.7)	Alberta, Canada	
0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	42 (0.5)	Dubai, UAE	
0.9	0.9	0.8	0.9	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	35 (0.8)	Abu Dhabi, UAE	
191	203	182	172	214	197	216	228	215	200	206	230	224	230	Number of Items (Score Points) Identified*		

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2011

rejected items that would be difficult for their students, these items tended to be difficult for students in other countries as well. The analysis shows that omitting such items tends to improve the results for that country, but also tends to improve the results for all other countries, so that the overall pattern of relative performance is largely unaffected.

