

Appendix F: Test–Curriculum Matching Analysis

TIMSS went to great lengths to ensure that comparisons of student achievement across countries would be as fair and equitable as possible. The [TIMSS 2015 Assessment Frameworks](#) were designed to specify the important aspects of science that participating countries agreed should be the focus of an international assessment of science 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 2015 assessments involved a series of reviews by representatives of the participating countries, experts in science, and testing specialists. At the end of this process, the National Research Coordinators (NRCs) from each country formally approved the TIMSS 2015 assessments, thus accepting them as being sufficiently fair to compare their students' science 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 2015 assessment (or test) and the science 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 2015 science assessment matched 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 2015 tests matched 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 Poland the curriculum specifies the curricular goals to be achieved by the end of the sixth and ninth grades, but does not provide a grade-by-grade specification. In such

¹ Because there also may be curriculum areas covered in some countries that are not covered by the TIMSS 2015 tests, the TCMA does not provide complete information about how well the tests cover the curricula of the countries.

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 2015 participants took part in the TCMA analysis except Norway (4) and Buenos Aires at the fourth grade and Egypt, Norway (8), and Buenos Aires at the eighth grade.

Exhibits F.1 through F.4 present the TCMA results for the TIMSS 2015 science test at the fourth and eighth grades. Exhibits F.1 and F.2 show the average percent correct on the science 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 180 points.³ Generally, the proportion of items judged appropriate was fairly high. Reading along the bottom row, it can be seen that 2 of the 47 countries that took part in the TCMA analysis judged 100 percent of the items to be included in their curricula. A further 21 countries and 2 of the 5 benchmarking participants judged 75 percent or more (135 score points) to be appropriate.

At the eighth grade, the percentage of items judged appropriate was similar; 2 of the 38 countries judged 100 percent of the items to be appropriate (all 233 score points), and an additional 26 countries and 3 of the 5 benchmarking participants judged 75 percent or more (175 score points) to be appropriate. All but two of the countries and two of the benchmarking participants concurred that more than half of the science items were included in 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 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 81 percent on its own set of items, also had 70 percent correct on the items selected by Korea, 74 percent on the items selected by Japan, 67 percent on the items selected by the Russian Federation, and so forth.

2 Exhibits 5 and 6 of the TIMSS 2015 Encyclopedia provide information on the grade-to-grade structure of the curriculum for each TIMSS 2015 participant.
 3 The TIMSS 2015 fourth grade science assessment contained 176 items, yielding 188 score points. However, following item review, eight items were deleted, resulting in 168 items and 180 score points. Similarly, following item review, the 220 items and 239 score points in the eighth grade assessment were reduced to 215 items and 233 score points by deleting five items and reducing the point value of one item.
 4 It should be noted that the science achievement presented in Exhibits F.1 and F.2 is based on average percent correct (the percentage of students in a country answering each item correctly, averaged across all items), which is different from the average scale scores that are presented in main tables of the report.

Exhibit F.2: Average Percent Correct for the Test-Curriculum Matching Analysis, Eighth Grade (Continued)

Based on a subset of items specifically identified by each country as addressing its curriculum

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.

Country	Average Percent Correct on All Items	Benchmarking Participants							
		Georgia	Jordan	Kuwait	Lebanon	Botswana (9)	Saudi Arabia	Morocco	South Africa (9)
Singapore	64 (0.7)	64	66	64	63	65	65	65	64
Chinese Taipei	59 (0.4)	59	62	59	58	59	59	58	59
Japan	59 (0.4)	58	60	59	59	58	59	58	59
Korea, Rep. of	56 (0.5)	55	58	56	54	56	56	54	56
Slovenia	55 (0.5)	55	58	56	54	55	56	55	55
Russian Federation	54 (0.9)	54	56	54	54	54	54	53	54
Hong Kong SAR	53 (0.8)	52	55	53	52	53	53	51	53
Kazakhstan	51 (1.0)	51	52	51	54	51	51	51	51
England	51 (0.8)	50	53	51	51	50	51	48	51
United States	50 (0.6)	50	52	50	48	49	50	48	50
Hungary	50 (0.7)	50	52	50	50	50	50	48	50
Ireland	50 (0.5)	50	52	50	48	49	50	47	50
Canada	49 (0.4)	49	51	49	47	48	49	47	49
Sweden	49 (0.7)	48	50	49	48	47	49	46	49
Lithuania	48 (0.6)	48	50	48	48	47	48	46	48
New Zealand	47 (0.6)	46	49	47	44	46	47	45	47
Australia	47 (0.5)	46	49	46	45	46	47	44	47
Israel	46 (0.7)	46	48	46	47	46	46	45	46
Norway (9)	46 (0.5)	45	47	45	45	44	46	43	46
Italy	44 (0.4)	44	46	44	44	43	44	42	44
Turkey	43 (0.8)	43	45	43	43	43	44	43	43
Malta	42 (0.3)	42	44	42	40	41	42	40	42
United Arab Emirates	41 (0.4)	41	43	41	41	41	41	41	41
Malaysia	40 (0.7)	40	42	40	39	41	40	38	40
Bahrain	39 (0.3)	39	41	39	39	39	39	39	39
Qatar	38 (0.5)	38	40	38	37	38	38	37	38
Iran, Islamic Rep. of	37 (0.7)	37	39	37	39	37	37	36	37
Oman	37 (0.4)	37	39	37	36	37	37	37	37
Thailand	37 (0.8)	36	39	36	35	36	37	35	37
Chile	36 (0.5)	35	38	36	35	35	36	34	36
Georgia	35 (0.5)	35	36	34	35	34	35	34	35
Jordan	33 (0.4)	32	35	33	33	32	33	32	33
Kuwait	31 (0.8)	30	32	30	31	30	31	30	31
Lebanon	29 (0.7)	29	30	29	31	29	29	28	29
Botswana (9)	28 (0.3)	28	30	29	27	29	28	27	28
Saudi Arabia	28 (0.6)	28	30	28	27	28	28	28	28
Morocco	27 (0.3)	26	28	27	27	26	27	27	27
South Africa (9)	24 (0.7)	24	26	24	23	24	24	23	24
International Avg.	44 (0.1)	43	45	43	43	43	44	42	44
Benchmarking Participants									
Dubai, UAE	50 (0.4)	50	52	50	50	50	50	50	50
Quebec, Canada	50 (0.9)	50	52	49	49	48	50	48	50
Ontario, Canada	49 (0.5)	48	50	48	46	48	49	46	49
Florida, US	46 (1.2)	46	48	46	45	46	46	45	46
Abu Dhabi, UAE	38 (0.9)	37	39	37	37	37	38	37	38
Number of Items (Score Points) Identified*	233	192	213	216	111	197	232	152	233
		67	176	108	224	180			

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

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 England at the fourth grade as an example, 69 percent of these items, on average, were answered correctly by students in Singapore, 67 percent by students in Korea, 63 percent by students in Japan, 62 percent by students in the Russian Federation, 59 percent by those in Hong Kong SAR, 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 England averaged 56 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 and benchmarking entities are presented in the lower part of the exhibit. These show that the selections of items by the participating countries and benchmarking entities varied somewhat in average difficulty, ranging at the fourth grade from 48 percent correct (the most difficult) for those chosen by Chinese Taipei, the United Arab Emirates, and Saudi Arabia to 53 percent correct (the least difficult) for those chosen by Denmark. At the eighth grade, the average percent correct ranged from 42 percent for Japan and Morocco to 46 percent for those chosen by New Zealand.

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 than on the items overall, although not by much. To illustrate, the average percent correct for Chinese Taipei across all fourth grade mathematics items was 59 percent. The diagonal element shows that students from Chinese Taipei had a slightly greater average percent correct (61 percent) across the set of items selected as appropriate for Chinese Taipei than they did overall. Most participants had a difference of one or two percentage points between the two performance measures, with the largest difference in Singapore (14 percentage points). At the eighth grade, the differences were generally smaller; the largest being in Singapore (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 science 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, Singapore had the highest average percent correct, not only on the test as a whole, but also on all of the different item selections, with Chinese Taipei, Japan, Korea, and Slovenia next in order of performance (with some ties) 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

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 2015 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.

little. As an example, consider the 149 score points selected by Denmark at the fourth grade. The students in Denmark did better on these items (58% 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 53 percent correct compared with 50 percent correct overall. The countries that performed better than Denmark on the overall test also performed about as well or better on the items selected by Denmark.

The TCMA results provide evidence that the TIMSS 2015 science 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 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.

Exhibit F.4: Standard Errors for the Test-Curriculum Matching Analysis, Eighth Grade (Continued)

Based on a subset of items specifically identified by each country as addressing its curriculum

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.

Country	Average Percent Correct on All Items	Benchmarking Participants												
		Georgia	Jordan	Kuwait	Lebanon	Botswana (9)	Saudi Arabia	Morocco	South Africa (9)	Dubai, UAE	Quebec, Canada	Ontario, Canada	Florida, US	Abu Dhabi, UAE
Singapore	64 (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	
Chinese Taipei	59 (0.4)	0.4	0.4	0.4	0.4	0.4	0.4	0.5	0.4	0.5	0.4	0.4	0.4	
Japan	59 (0.4)	0.4	0.4	0.4	0.4	0.4	0.4	0.5	0.4	0.5	0.4	0.4	0.4	
Korea, Rep. of	56 (0.5)	0.5	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.4	0.5	0.5	
Slovenia	55 (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	
Russian Federation	54 (0.9)	0.9	0.9	0.9	1.0	0.9	0.9	0.9	0.9	1.0	0.9	0.9	0.9	
Hong Kong SAR	53 (0.8)	0.8	0.8	0.8	0.8	0.9	0.8	0.9	0.8	0.8	0.9	0.8	0.9	
Kazakhstan	51 (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	
England	51 (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	
United States	50 (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	
Hungary	50 (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	
Ireland	50 (0.5)	0.5	0.5	0.6	0.6	0.6	0.5	0.6	0.5	0.6	0.5	0.6	0.5	
Canada	49 (0.4)	0.4	0.4	0.4	0.5	0.4	0.4	0.5	0.4	0.5	0.4	0.4	0.4	
Sweden	49 (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	
Lithuania	48 (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	
New Zealand	47 (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	
Australia	47 (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	
Israel	46 (0.7)	0.7	0.7	0.7	0.8	0.7	0.7	0.8	0.7	0.8	0.7	0.8	0.7	
Norway (9)	46 (0.5)	0.5	0.5	0.5	0.6	0.5	0.5	0.6	0.5	0.6	0.5	0.6	0.5	
Italy	44 (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	
Turkey	43 (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	
Malta	42 (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	
United Arab Emirates	41 (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	
Malaysia	40 (0.7)	0.7	0.7	0.7	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	
Bahrain	39 (0.3)	0.3	0.3	0.3	0.4	0.3	0.3	0.4	0.3	0.4	0.3	0.3	0.3	
Qatar	38 (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	
Iran, Islamic Rep. of	37 (0.7)	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	
Oman	37 (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	
Thailand	37 (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	
Chile	36 (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	
Georgia	35 (0.5)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.5	0.5	
Jordan	33 (0.4)	0.4	0.5	0.4	0.5	0.4	0.4	0.4	0.4	0.4	0.5	0.4	0.4	
Kuwait	31 (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	
Lebanon	29 (0.7)	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	
Botswana (9)	28 (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	
Saudi Arabia	28 (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	
Morocco	27 (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	
South Africa (9)	24 (0.7)	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	
International Avg.	44 (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	
Benchmarking Participants														
Dubai, UAE	50 (0.4)	0.4	0.4	0.4	0.5	0.4	0.4	0.4	0.4	0.5	0.4	0.4	0.4	0.4
Quebec, Canada	50 (0.9)	0.9	0.9	1.0	1.0	0.9	0.9	1.0	0.9	1.0	1.0	0.9	0.9	0.9
Ontario, Canada	49 (0.5)	0.5	0.5	0.5	0.6	0.5	0.5	0.5	0.5	0.6	0.5	0.5	0.5	0.5
Florida, US	46 (1.2)	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.1	1.2	1.2	1.2	1.2
Abu Dhabi, UAE	38 (0.9)	0.9	1.0	0.9	1.0	1.0	0.9	0.9	0.9	1.0	1.0	1.0	0.9	0.9
Number of Items (Score Points) Identified*	233	192	213	216	111	197	232	152	233	67	176	108	224	180

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015