

# Appendix B

## The Test-Curriculum Matching Analysis

TIMSS Advanced 2008 went to great lengths to ensure that comparisons of student achievement in advanced mathematics and physics across countries would be as fair and equitable as possible. The *TIMSS Advanced 2008 Assessment Frameworks* was designed to specify the important aspects of advanced mathematics and physics that participating countries agreed should be the focus of an international assessment of student achievement. The assessment items were developed through a collaborative process with national representatives to faithfully represent the specifications in the frameworks, and the items were field tested extensively in participating countries. Finalizing the TIMSS Advanced 2008 assessments involved a series of reviews by representatives of the participating countries, experts in mathematics and physics, and testing specialists. At the end of this process, the National Research Coordinators from each country formally approved the TIMSS Advanced 2008 assessments, accepting them as being sufficiently fair to compare their students' advanced mathematics and physics achievement with that of students from other countries.

Although the assessments were developed to represent agreed-upon frameworks and were intended to have as much in common across countries as possible, it was unavoidable that the match between the TIMSS Advanced 2008 assessments (or tests) and the advanced mathematics and physics curricula 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, include 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 Advanced 2008 mathematics and physics assessments were relevant to each country's curriculum. The TCMA also investigated the impact on a country's performance of including only achievement items that were judged to be relevant to its own curriculum.<sup>1</sup>

To gather data about the extent to which the TIMSS Advanced 2008 tests were relevant to the curricula of the TIMSS countries, National Research Coordinators were asked to examine each achievement item and indicate whether the item was in their country's intended curriculum for the advanced mathematics and physics programs or tracks assessed by TIMSS Advanced. The National Research Coordinator was asked to assemble a team familiar with these curricula in order to make this determination. Since an item might be in the curriculum for some but not all students in a country, coordinators were asked to consider an item included if it was in the intended curriculum for more than 50 percent of the students. All TIMSS Advanced 2008 participants took part in the TCMA analysis.

<sup>1</sup> Because there may also be curriculum areas covered in some countries that are not covered by the TIMSS Advanced 2008 tests, the TCMA does not provide complete information about how well the tests cover the curricula of the countries.

Exhibits B.1 and B.2 present the TCMA results for the TIMSS Advanced 2008 advanced mathematics and physics tests. Exhibit B.1 shows the average percent correct on the advanced mathematics and physics items judged appropriate by each country. Exhibit B.2 shows the standard errors corresponding to the percentages presented in Exhibit B.1.

In Exhibit B.1, the bottom row of the exhibit shows the number of items, in terms of score points, on the entire assessment and the number identified as appropriate in each country. For advanced mathematics, the maximum number of score points in the assessment was 79 points.<sup>2</sup> Generally, the match between the advanced mathematics assessment and the curricula of the countries was very good, with a high proportion of items judged appropriate in each country. Reading along the bottom row, it can be seen that the Russian Federation and the Philippines judged all of the items (all 79 score points) to be appropriate, and the Netherlands (72), Lebanon (76), Iran (76), Slovenia (73), Italy (74), and Norway (73), almost all. Armenia (66) and Sweden (64) had the fewest items judged to be appropriate, but still had more than 80 percent of the total.

In physics, the match between the assessment and the countries' curricula was very good as well, with almost all of the 77 item points<sup>3</sup> judged appropriate in the Netherlands (71), Norway (76), Slovenia (74), the Russian Federation (73), Armenia (77), Sweden (75), and Iran (74). Fewer items were judged appropriate in Italy (57) and especially in Lebanon (47).

Since 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.<sup>4</sup>

2 The TIMSS Advanced 2008 advanced mathematics assessment contained 72 items yielding 82 score points. However, following item review, one item was deleted and response categories were combined for a number of items, resulting in data for reporting on 71 items and 79 score points.

3 The TIMSS Advanced 2008 physics assessment contained 70 items yielding 82 score points. However, following item review, two items were deleted and response categories were combined for a number of items, resulting in data for reporting on 68 items and 77 score points.

4 The advanced mathematics and physics achievement presented in Exhibit B.1 is based on average percent correct, which is different from the average scale scores that are presented in Chapters 2 and 8.

### Exhibit B.1 Average Percent Correct for Test-Curriculum Matching Analysis in Advanced Mathematics and Physics

TIMSS Advanced 2008

Based on Subset of Items Identified by Each Country as Addressing its Curriculum (See Exhibit B.2 for corresponding standard errors)

Instructions: 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 country based on its decisions about the test items to include.

#### Advanced Mathematics

Country	Average Percent Correct on All Items	Russian Federation	Netherlands	Lebanon	Iran, Islamic Rep. of	Slovenia	Italy	Norway	Armenia	Sweden	Philippines
Russian Federation	57 (1.6)	57	58	57	57	57	57	58	58	58	57
Netherlands	54 (0.5)	54	56	53	55	54	54	55	54	56	54
Lebanon	53 (0.5)	53	53	54	52	53	54	54	53	53	53
Iran, Islamic Rep. of	43 (1.4)	43	44	43	44	43	43	44	47	40	43
Slovenia	36 (0.7)	36	37	36	36	37	36	37	37	35	36
Italy	35 (1.1)	35	36	35	35	35	36	36	36	35	35
Norway	33 (0.7)	33	34	33	34	34	32	34	34	33	33
Armenia	32 (0.7)	32	33	32	33	33	32	33	36	32	32
Sweden	31 (0.7)	31	31	30	31	31	31	31	31	33	31
Philippines	24 (0.6)	24	25	24	24	25	24	25	25	24	24
International Avg.	40 (0.3)	40	41	40	40	40	40	41	41	40	40
Number of Items (Score Points) Identified	79	79	72	76	76	73	74	73	66	64	79

SOURCE: IEA TIMSS Advanced 2008 ©

Of the 72 items in the Advanced Mathematics assessment, some extended-response items were scored on a 2-point scale, resulting in 82 total score points. Following item review, one item was deleted and response categories were combined for a number of items, resulting in 71 items and 79 score points.

#### Physics

Country	Average Percent Correct on All Items	Netherlands	Norway	Slovenia	Russian Federation	Armenia	Sweden	Iran, Islamic Rep. of	Lebanon	Italy
Netherlands	57 (0.7)	57	57	57	57	57	57	57	60	55
Norway	47 (0.7)	47	47	47	46	47	47	47	50	48
Slovenia	47 (0.5)	47	47	47	47	47	47	48	50	49
Russian Federation	46 (1.6)	45	46	46	46	46	45	46	49	47
Armenia	42 (0.7)	42	42	42	42	42	42	42	43	42
Sweden	42 (0.8)	41	42	41	41	42	41	42	46	41
Iran, Islamic Rep. of	37 (1.1)	37	38	37	37	37	37	38	38	40
Lebanon	33 (0.4)	33	33	33	33	33	32	33	40	32
Italy	32 (0.9)	31	32	32	31	32	31	32	35	33
International Avg.	42 (0.3)	42	42	42	42	42	42	43	46	43
Number of Items (Score Points) Identified	77	71	76	74	73	77	75	74	47	57

Of the 70 items in the Physics assessment, some extended-response items were scored on a 2-point scale, resulting in 82 total score points. Following item review, two items were deleted and response categories were combined for a number of items, resulting in 68 items and 77 score points.

(1) Standard errors for the average percent of correct responses on all items appear in parentheses.



**Exhibit B.2 Standard Errors for the Test-Curriculum Matching Analysis in Advanced Mathematics and Physics**

Instructions: 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 country based on its decisions about the test items to include.

Advanced Mathematics											
Country	Average Percent Correct on All Items	Russian Federation	Netherlands	Lebanon	Iran, Islamic Rep. of	Slovenia	Italy	Norway	Armenia	Sweden	Philippines
Russian Federation	57 (1.6)	1.6	1.5	1.6	1.5	1.6	1.6	1.5	1.5	1.6	1.6
Netherlands	54 (0.5)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Lebanon	53 (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	43 (1.4)	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4
Slovenia	36 (0.7)	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Italy	35 (1.1)	1.1	1.2	1.1	1.2	1.1	1.2	1.2	1.2	1.2	1.1
Norway	33 (0.7)	0.7	0.7	0.6	0.7	0.6	0.6	0.7	0.7	0.7	0.7
Armenia	32 (0.7)	0.7	0.7	0.7	0.7	0.7	0.6	0.7	0.7	0.7	0.7
Sweden	31 (0.7)	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7
Philippines	24 (0.6)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.6	0.6
International Avg.	40 (0.3)	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Number of Items (Score Points) Identified	79	79	72	76	76	73	74	73	66	64	79

SOURCE: IEA TIMSS Advanced 2008 ©

Of the 72 items in the Advanced Mathematics assessment, some extended-response items were scored on a 2-point scale, resulting in 82 total score points. Following item review, one item was deleted and response categories were combined for a number of items, resulting in 71 items and 79 score points.

Physics										
Country	Average Percent Correct on All Items	Netherlands	Norway	Slovenia	Russian Federation	Armenia	Sweden	Iran, Islamic Rep. of	Lebanon	Italy
Netherlands	57 (0.7)	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8
Norway	47 (0.7)	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7
Slovenia	47 (0.5)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Russian Federation	46 (1.6)	1.6	1.6	1.6	1.7	1.6	1.6	1.6	1.6	1.7
Armenia	42 (0.7)	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Sweden	42 (0.8)	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Iran, Islamic Rep. of	37 (1.1)	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.2
Lebanon	33 (0.4)	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.5	0.4
Italy	32 (0.9)	0.9	0.9	0.9	0.9	0.9	0.9	0.9	1	1
International Avg.	42 (0.3)	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Number of Items (Score Points) Identified	77	71	76	74	73	77	75	74	47	57

Of the 70 items in the Physics assessment, some extended-response items were scored on a 2-point scale, resulting in 82 total score points. Following item review, two items were deleted and response categories were combined for a number of items, resulting in 68 items and 77 score points.

(1) Standard errors for the average percent of correct responses on all items appear in parentheses. The matrix contains standard errors corresponding to the average percent correct responses based on TCMA subset of items, as displayed in Exhibit B.1.

The first data column in the advanced mathematics section of Exhibit B.1 shows the average percent correct on all advanced mathematics test items for each country, together with its standard error. Subsequent columns show the performance of every other country on those items judged appropriate by the country listed at the head of the column. Countries are presented in order of their performance based on average percent correct on all of the advanced mathematics items, from highest to lowest. To interpret this exhibit, 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, the Netherlands, where the average percent correct was 56 percent on the set of advanced mathematics items that it judged appropriate, had, on average, 54 percent of the items judged appropriate by the Russian Federation answered correctly by its students,<sup>5</sup> 53 percent of the items selected by Lebanon, 55 percent of the items selected by Iran, 54 percent of the items selected by Slovenia, and so forth.

The column for a country listed at the top of the matrix for advanced mathematics shows how each of the other countries performed on the set of items selected as appropriate for the students of the country listed at the top. Again using the set of advanced mathematics items selected by the Netherlands as an example, 58 percent of these items, on average, were answered correctly by students in the Russian Federation, 53 percent by students in Lebanon, 44 percent by students in Iran, 37 percent by students in Slovenia, 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 Netherlands averaged 56 percent correct on the set of items identified by the Netherlands for the analysis.

<sup>5</sup> The Russian Federation judged all of the advanced mathematics items to be appropriate to their curriculum, so results based on the Russian selection are identical to the results based on the entire item pool.

For each country's selected items, the international averages across the participating countries are presented in a row in the lower part of the exhibit for each subject. The advanced mathematics averages show that the selections of advanced mathematics items by the participating countries varied only slightly in average difficulty, which is not surprising given that countries included most items in the advanced mathematics assessment. The international averages for physics also did not vary much, although Lebanon's item selection resulted in an international average of 46 percent, some 4 percentage points above the average based on all of the physics items. Clearly, the physics items judged not appropriate for their curriculum were among the more difficult for all of the countries, and omitting them from the analysis resulted in higher achievement for all countries.

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 advanced mathematics, countries generally performed better on their own item sets than on the items overall, although not by much. To illustrate, the average percent correct for the Netherlands across all the advanced mathematics items was 54 percent. The diagonal element shows that students from the Netherlands had a slightly greater average percent correct (56 percent) across the set of items selected as appropriate for Dutch students than they did overall. Almost all participants had a difference of 1 or 2 percentage points between the two performance measures, with the largest difference in Armenia (4 percentage points). Armenia also was one of the two countries with relatively fewer advanced mathematics items judged appropriate to their curriculum.

In physics, the Netherlands, Norway, Slovenia, the Russian Federation, and Armenia, all of which rejected very few items, had

the same average achievement on their selected items as on the test as a whole. Sweden performed slightly less well on its own item selection, but Iran, Italy, and especially Lebanon performed better on their selected items than on the assessment as a whole.

It is clear that the selection of items did not have a major effect on the relative performance in advanced mathematics or physics among TIMSS Advanced 2008 countries. In both subjects, countries that had relatively high or low performance across all of the items in the assessment also had relatively high or low performance on each of the various sets of items selected for the TCMA. For example, in advanced mathematics, the Russian Federation had the highest average percent correct not only on the assessment as a whole, but also on all of the different item selections, with the Netherlands, Lebanon, and Iran next in order of performance on practically all selections of items.<sup>6</sup> The situation was similar in physics, with the order of average country performance preserved across all item selections.

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 little changed. As an example, consider the set of advanced mathematics items selected by the Netherlands (72 score points). The students in the Netherlands did better on these items (56% correct) than on the test as a whole (54% correct). However, most other countries also did better on those items, with an international average of 41 percent correct compared with 40 percent correct overall. A more extreme example may be found in physics, where Lebanon, which rejected more physics items than any other country, had an average percent correct of 40 percent on the physics items it selected, compared to 33 percent on the complete set of physics items. However, every other country also performed better

<sup>6</sup> 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 Advanced 2008 sample are provided in Exhibit B.2. For any sample average shown in Exhibit B.1, it can be said with 95 percent confidence that the corresponding value in the population falls between the sample estimate plus or minus 2 standard errors.

on the Lebanese item selection than on the complete item set, so that relative performance differences among countries were unchanged.

The TCMA results provide evidence that the TIMSS Advanced 2008 advanced assessments constitute a reasonable basis for comparing the advanced mathematics and physics achievements of the participating countries. This result is not unexpected, since making the assessment as fair as possible was a major consideration in test development. The fact that all 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.

