

Appendix PE: Test–Curriculum Matching Analysis

TIMSS Advanced 2015 went to great lengths to ensure that comparisons of student achievement across countries would be as fair and equitable as possible. The [TIMSS Advanced 2015 Assessment Frameworks](#) were designed to specify the important aspects of 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 were field tested extensively in participating countries. Finalizing the TIMSS Advanced 2015 physics assessment involved a series of reviews by representatives of the participating countries, experts in physics, and testing specialists. At the end of this process, the National Research Coordinators (NRCs) from each country formally approved the TIMSS Advanced 2015 physics assessment, thus accepting it as being sufficiently fair to compare their students' physics achievement with that of students from other countries.

Although the assessment was developed to represent agreed upon frameworks and was intended to have as much in common across countries as possible, it was unavoidable that the match between the physics assessment (or test) and the physics 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 test, therefore, inevitably has 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 2015 physics assessment matched 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.¹

To gather data about the extent to which the TIMSS Advanced 2015 physics test matched the curricula of the participating 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 physics program(s) or track(s) assessed by TIMSS Advanced. 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 2015 participants took part in the TCMA analysis.

Exhibits PE.1 and PE.2 present the TCMA results for the TIMSS Advanced 2015 physics test. Exhibit PE.1 shows the average percent correct on the physics items judged appropriate by each

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

country. Exhibit PE.2 shows the standard errors corresponding to the percentages presented in Exhibit PE.1.

In Exhibit PE.1, the bottom row of the exhibit shows the number of items, in terms of score points, identified as appropriate in each country. For physics, the maximum number of score points in the assessment was 115 points.² Generally, the proportion of items judged appropriate was fairly high. From the bottom row, it can be seen that Slovenia and Italy judged all of the items (115 score points) to be appropriate and the United States judged almost all of the items (111 score points) to be appropriate. Norway (107), Sweden (107), and the Russian Federation (104) judged over 90 percent of the items to be included in the curriculum, and Portugal (94), Lebanon (93), and France (92) judged at least 80 percent to be included.

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.³

The first data column in Exhibit PE.1 shows the average percent correct on all physics test items for each country, together with its standard error. Subsequent columns show the performance of every 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 physics 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, Slovenia, where the average percent correct was 52 percent on the set of physics items that it judged appropriate, had, on average, 54 percent correct on the items judged appropriate by the Russian Federation, 51 percent on the items selected by Norway, 52 percent on the items selected by Portugal, and so forth.

The column for a country listed at the top shows how each of the other countries performed on the set of items selected as appropriate for that country's students. Using the set of physics items selected by Portugal as an example, 52 percent of these items, on average, were answered correctly by students in Slovenia, 49 percent by students in the Russian Federation, 50 percent by students in Norway, 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 Portugal averaged 44 percent correct on the set of items identified by Portugal for the analysis.

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. These show that the selections of items by the participating countries varied only slightly in average difficulty, which is not surprising given that countries included most items in the physics assessment.

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

2 The TIMSS Advanced 2015 physics assessment contained 103 items yielding 117 score points. However, following item review, the 103 items and 117 score points in the physics assessment were reduced to 101 items and 115 score points.

3 It should be noted that the physics achievement presented in Exhibits PE.1 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 PE.1: Average Percent Correct for the Test-Curriculum Matching Analyses in Physics

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	Slovenia	Russian Federation	Norway	Portugal	Sweden	United States	Lebanon	Italy	France
	Slovenia	52 (0.5)	52	54	51	52	53	53	52	52
Russian Federation	50 (1.1)	50	51	49	49	50	50	50	50	49
Norway	49 (0.7)	49	49	51	50	49	50	50	49	49
Portugal	42 (0.6)	42	44	43	44	43	43	42	42	44
Sweden	42 (0.8)	42	43	42	42	42	42	42	42	42
United States	39 (1.1)	39	40	40	40	39	39	39	39	40
Lebanon	35 (0.4)	35	36	35	34	35	35	36	35	35
Italy	32 (0.6)	32	33	32	31	32	33	31	32	31
France	31 (0.4)	31	31	32	31	31	31	31	31	32
International Avg.	41 (0.2)	41	42	42	41	42	42	42	41	42
Number of Items (Score Points) Identified*	115	115	104	107	94	107	111	93	115	92

* Of the 103 items in the Physics assessment, some extended-response items were scored on a two-point scale, resulting in 117 total score points. Following item review, two items were deleted, resulting in 101 items and 115 score points.

() Standard errors appear in parentheses. Because of rounding some results may appear inconsistent.

performance on the test as a whole. Countries generally performed similarly or a little better on their own items when compared to their performance on all items. To illustrate, the average percent correct for the Russian Federation across all the physics items was 50 percent. The diagonal element shows that students from the Russian Federation had a slightly greater average percent correct (51 percent) across the set of items selected as appropriate for Russian students than they did overall. Norway and Portugal had the biggest differences between the two measures with Norwegian students achieving 51 percent correct on the items judged to be in their curriculum while achieving 49 percent correct on all items, and Portuguese students achieving 44 percent correct on the items judged to be in their curriculum while achieving 42 percent correct on all items.

It is clear that the selection of items did not have a major effect on the relative performance in physics among TIMSS Advanced 2015 countries. 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, Slovenia had the highest average percent correct not only on the assessment as a whole, but also on all of the different item

Exhibit PE.2: Standard Errors for the Test-Curriculum Matching Analyses in Physics

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	Slovenia	Russian Federation	Norway	Portugal	Sweden	United States	Lebanon	Italy	France
		Slovenia	52 (0.5)	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Russian Federation	50 (1.1)	1.1	1.1	1.0	1.1	1.1	1.1	1.0	1.1	
Norway	49 (0.7)	0.7	0.6	0.7	0.7	0.7	0.7	0.7	0.7	
Portugal	42 (0.6)	0.6	0.6	0.6	0.7	0.6	0.6	0.6	0.7	
Sweden	42 (0.8)	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	
United States	39 (1.1)	1.1	1.2	1.1	1.1	1.2	1.1	1.2	1.1	
Lebanon	35 (0.4)	0.4	0.5	0.4	0.5	0.5	0.4	0.4	0.5	
Italy	32 (0.6)	0.6	0.7	0.6	0.7	0.7	0.7	0.6	0.6	
France	31 (0.4)	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.5	
International Avg.	41 (0.2)	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.2	
Number of Items (Score Points) Identified*	115	115	104	107	94	107	111	93	115	92

* Of the 103 items in the Physics assessment, some extended-response items were scored on a two-point scale, resulting in 117 total score points. Following item review, two items were deleted, resulting in 101 items and 115 score points.

() Standard errors appear in parentheses. Because of rounding some results may appear inconsistent.

selections, with the Russian Federation, Norway, Portugal, and Sweden next in order (with some ties) on practically all selections of items.⁴

The TCMA results provide evidence that the TIMSS Advanced 2015 physics assessment constitutes a reasonable basis for comparing the achievement 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 do not greatly affect the overall pattern of relative performance.

4 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 2015 sample are provided in Exhibit PE.2. For any sample average shown in Exhibit PE.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.