

Appendix C

The Test-Curriculum Matching Analysis: Science

To ensure that comparisons of student achievement across countries would be as fair and equitable as possible, TIMSS developed extensive assessment frameworks and specifications that addressed the important aspects of science in countries' curricula and instructional programs, and went to great lengths to develop assessment items that faithfully represented those specifications. Similar to the procedures used for developing the original TIMSS instruments, developing the TIMSS 2003 tests involved a series of reviews by representatives of the participating countries, experts in science, and testing specialists.¹ The National Research Coordinators (NRCs) from each country formally approved the TIMSS 2003 tests, thus accepting them as being sufficiently fair to compare their students' science achievement with that of students from other countries.

Although the tests were developed to represent an agreed-upon framework and were intended to have as much in common across countries as possible, it was inevitable that the match between the TIMSS 2003 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

¹ See Appendix A for more information on test development.

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 appropriateness of the TIMSS 2003 science test for the eighth- and fourth-grade students in the participating countries. TCMA also shows how student performance for individual countries varies when based only on the test questions that are judged to be relevant to their own curricula.²

To gather data about the extent to which the TIMSS 2003 tests were relevant to the curricula of the participating countries, each NRC reported whether each item was in that country's intended curriculum at the grade tested (eighth or fourth grade in most countries). The NRC was asked to choose a person or persons who were very familiar with the curriculum at these grades to make this determination. Since an item might be in the curriculum for some but not all students in a country, an item was to be determined appropriate if it was in the intended curriculum for more than 50 percent of the students. The NRCs had considerable flexibility in selecting items and may have considered items inappropriate for other reasons. All participants returned the information for analysis except Syria at eighth grade and Yemen at fourth grade.

Exhibits C.1 and C.2 present the TCMA results for the TIMSS 2003 tests at eighth and fourth grades. Exhibit C.1 shows the average percent correct on the science items selected as appropriate by each country. Exhibit C.2 shows the standard errors corresponding to the percentages presented in Exhibit C.1.

In Exhibit C.1, the last row of the exhibit shows the number of items (score points) identified as appropriate in each country.³ At the eighth grade, the percentage of score points ranged from 100 percent (206 score points) in Israel and Saudi Arabia to 31 percent (63 score points) in Belgium (Flemish). Generally, the proportion of items judged

2 Because there may also be curriculum areas covered in some countries that are not covered by the TIMSS 2003 tests, the TCMA does not provide complete information about how well the tests cover the curricula of the countries.

3 Some items were assigned more score points than others. In particular, some items had two parts, and some extended-response items were scored on a two-point scale. The TCMA uses score points in order to give the same weight to items given them in test scoring.

appropriate was high, with 40 of the 50 participants indicating that items representing three-quarters or more of the score points (154 out of a possible 206) were appropriate. Only Belgium (Flemish) and Chile considered less than 50 percent of the score points appropriate. At the fourth grade, the percentage of score points ranged from 98 percent (161 score points) in Hungary, the United States, Latvia, Lithuania, Moldova, and Armenia to 27 percent (44 score points) in Japan. Eighteen of the 28 fourth-grade participants indicated that items representing three-quarters or more of the score points (124 out of a possible 165) were appropriate.

Since most countries indicated that 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 in Exhibit C.1 shows the average percent correct on all test items for each participant. 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 this exhibit, reading across a row provides the average percent correct for the students in that country on the items selected by each of the countries listed across the top of the exhibit. For example, at the eighth-grade, Singapore, where the average percent correct was 65 percent on its own set of items, had 63 percent correct for the items selected by Chinese Taipei, 63 percent for the items selected by Estonia, 62 percent for the items selected by Korea, and so forth. The column for a country listed across the top shows how each of the other participants performed on the subset of items selected as appropriate for that country's students. Using the set of items selected by Bulgaria as an example, on average, 63 percent of these items were answered correctly by students in Singapore, 62 percent by students in Chinese Taipei, 58 percent by students in Estonia, 58 percent by those in Korea, and so forth. The shaded

4 It should be noted that the science achievement presented in Exhibit C.1 is based on average percent correct, which is different from the average scale scores that are presented in Chapter 1.

Exhibit C.1: Average Percent Correct for Test-Curriculum Matching Analysis – Science

Based on Subset of Items Specially Identified by Each Country as Addressing its Curriculum
(See Exhibit C.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 different country based on its own decisions about the test items to include.

Countries	Average Percent Correct on All Items	Singapore	Chinese Taipei	Estonia	Korea, Rep. of	Hong Kong, SAR	Japan	Hungary	England	Sweden	Netherlands	United States	Australia	Slovak Republic	Lithuania	Russian Federation	Slovenia	New Zealand	Latvia	Belgium (Flemish)	Scotland	Malaysia	Italy	Israel	Norway	Jordan	
Singapore	62 (0.9)	65	63	63	62	63	63	62	62	63	64	63	62	62	63	64	62	65	63	67	62	64	63	62	61	61	
Chinese Taipei	61 (0.8)	63	61	61	61	61	63	61	60	61	63	62	61	60	62	62	61	63	62	65	60	62	63	62	61	61	59
Estonia	58 (0.6)	59	57	58	58	58	58	58	58	58	60	59	59	58	59	60	59	61	59	62	57	59	59	58	58	57	
Korea, Rep. of	57 (0.4)	58	57	57	58	58	59	58	58	57	60	58	57	57	58	57	57	60	58	61	57	59	59	57	58	56	
Hong Kong, SAR	57 (0.7)	58	57	57	58	58	58	57	57	57	60	58	57	56	58	58	57	59	58	64	57	58	58	57	57	56	
Japan	57 (0.5)	57	57	57	57	57	61	57	57	57	59	58	57	57	58	59	56	59	58	59	56	57	58	57	57	55	
Hungary	56 (0.6)	57	56	56	56	56	58	57	56	57	57	57	57	56	57	59	57	59	57	61	55	57	58	56	57	55	
England	55 (0.9)	56	54	55	55	55	55	55	55	55	57	56	55	55	56	55	55	58	56	61	54	56	56	55	55	53	
Sweden	53 (0.7)	53	52	52	52	53	54	52	52	53	54	54	53	53	54	54	53	56	54	58	52	54	54	53	53	51	
Netherlands	52 (0.8)	53	51	53	53	53	52	53	53	52	56	54	53	52	54	52	53	56	54	59	51	54	55	52	53	50	
United States	52 (0.7)	52	51	52	52	52	49	52	52	52	53	52	52	52	52	52	52	55	52	58	51	53	54	52	52	50	
Australia	51 (0.8)	51	50	51	51	52	49	51	52	51	54	52	52	51	52	52	52	55	52	57	50	52	53	51	52	50	
Slovak Republic	51 (0.7)	51	51	51	51	51	52	51	51	52	52	52	51	52	52	54	52	54	52	56	50	52	53	51	51	50	
Lithuania	51 (0.5)	52	50	52	51	51	52	51	51	52	53	52	51	51	52	52	51	54	52	56	50	52	52	51	51	50	
Russian Federation	50 (0.7)	51	50	51	50	50	51	51	50	51	52	51	51	51	51	55	51	53	51	55	49	52	52	50	51	49	
Slovenia	50 (0.4)	50	49	50	49	49	51	50	49	50	51	51	51	50	51	51	51	52	51	55	49	51	51	50	50	48	
New Zealand	50 (1.1)	50	49	50	50	50	48	50	50	49	51	51	50	49	51	49	50	53	51	54	49	51	51	50	50	48	
Latvia	49 (0.6)	49	48	49	49	49	49	49	49	49	51	50	49	49	50	50	50	52	50	53	47	50	50	49	49	47	
Belgium (Flemish)	49 (0.5)	49	48	48	49	49	48	49	49	49	52	50	49	49	50	48	49	51	50	56	47	49	51	49	49	47	
Scotland	49 (0.8)	49	48	48	49	49	48	48	49	49	51	50	49	48	50	48	49	51	50	54	47	49	50	49	49	47	
Malaysia	48 (0.8)	49	48	48	48	48	49	47	48	48	50	49	48	48	49	49	48	51	49	53	47	50	49	48	48	47	
Italy	46 (0.6)	46	45	46	46	46	45	46	46	46	47	46	46	46	46	47	46	48	46	54	45	47	47	46	46	44	
Israel	45 (0.6)	46	45	46	45	45	47	45	45	46	46	46	46	46	46	47	46	48	46	51	45	47	47	45	46	44	
Norway	45 (0.5)	44	44	45	45	46	44	45	45	45	46	46	46	45	46	46	46	48	46	49	44	46	47	45	45	43	
Jordan	42 (0.7)	42	42	43	42	42	44	42	42	42	43	43	42	42	43	44	42	44	43	46	42	43	43	42	42	43	
Bulgaria	42 (1.0)	42	41	42	41	41	44	42	42	42	43	42	42	42	42	44	42	44	42	47	41	42	43	42	41	41	
Romania	41 (0.9)	41	40	41	40	40	42	41	40	41	43	41	41	41	41	43	41	43	41	45	40	42	42	41	41	40	
Serbia	40 (0.5)	40	39	40	39	39	41	39	39	40	40	40	40	40	40	42	40	42	40	45	39	41	41	40	40	39	
Moldova, Rep. of	40 (0.7)	40	39	40	39	39	42	40	40	41	41	40	40	40	40	42	40	41	40	44	39	41	41	40	40	39	
Armenia	38 (0.7)	39	38	38	37	37	40	38	38	39	40	39	38	38	39	40	38	40	39	41	37	39	40	38	38	38	
Iran, Islamic Rep. of	38 (0.5)	38	37	38	38	37	38	38	38	37	39	39	38	38	39	40	38	40	39	41	37	39	39	38	38	37	
Macedonia, Rep. of	37 (0.7)	38	37	37	37	36	40	37	37	38	38	37	38	37	37	39	38	39	37	42	36	38	38	37	37	36	
Bahrain	36 (0.3)	37	36	36	36	36	40	36	36	37	38	37	36	36	37	37	35	38	37	41	36	37	38	36	37	36	
Palestinian Nat'l Auth.	35 (0.6)	36	35	36	35	35	38	35	35	35	36	36	36	35	36	37	35	37	36	38	35	37	36	35	35	36	
Cyprus	35 (0.3)	36	34	35	35	34	37	35	35	35	36	36	35	35	36	36	35	37	36	40	34	36	36	35	35	33	
Chile	34 (0.5)	34	33	34	33	33	32	33	33	34	34	34	33	34	34	35	34	35	34	39	33	35	35	34	34	32	
Egypt	33 (0.6)	34	33	33	33	33	35	33	33	33	34	34	33	33	34	34	33	34	34	36	33	34	34	33	33	33	
Indonesia	32 (0.5)	33	31	32	32	32	32	31	32	32	34	32	32	32	32	33	32	34	32	38	31	33	33	32	32	31	
Lebanon	29 (0.6)	31	29	29	29	28	33	29	29	30	29	30	30	29	30	31	30	30	30	33	29	31	30	29	29	29	
Tunisia	29 (0.3)	30	30	29	29	29	32	29	29	30	31	30	30	29	30	31	30	31	30	35	29	31	30	29	29	28	
Morocco	28 (0.3)	30	28	28	28	27	32	28	28	29	29	29	29	29	29	29	28	30	29	33	27	29	29	28	29	27	
Saudi Arabia	28 (0.6)	29	28	29	28	28	31	28	28	29	29	29	29	28	29	29	28	30	29	33	28	29	30	28	28	28	
Philippines	28 (0.8)	28	27	28	28	27	28	27	28	28	28	28	28	28	28	28	28	29	28	34	27	29	29	28	28	27	
Botswana	26 (0.4)	27	25	25	26	25	27	25	25	25	26	26	26	26	26	26	26	26	26	30	25	26	27	26	26	25	
South Africa	19 (0.7)	20	18	19	19	18	20	19	19	19	18	19	19	19	19	20	19	19	19	21	19	20	19	19	19	19	
Ghana	19 (0.4)	19	18	19	18	17	20	18	18	19	17	18	19	19	18	20	19	18	18	21	18	19	19	19	19	18	
International Avg.	43 (0.1)	44	43	43	43	43	44	43	43	43	45	44	43	43	44	44	43	45	44	48	42	44	44	43	43	42	
Benchmarking Participants																											
Basque Country, Spain	44 (0.7)	44	43	44	44	44	44	44	44	44	46	45	44	44	45	45	44	47	45	51	43	45	46	44	44	42	
Indiana State, US	52 (1.1)	52	51	52	52	52	49	52	52	52	54	53	52	52	53	52	52	55	53	59	51	53	54	52	52	50	
Ontario Province, Can.	53 (0.7)	53	52	53	54	54	50	53	53	53	56	54	53	53	54	53	54	57	54	60	52	54	55	53	54	52	
Quebec Province, Can.	53 (0.7)	53	52	53	53	54	52	53	53	53	55	54	53	53	54	54	54	57	54	59	52	55	55	53	54	51	
Number of Items (Score Points) Identified*	206	176	164	184	194	177	104	198	199	192	145	195	179	190	195	129	176	167	195	63	180	189	188	206	202	181	

SOURCE: IEA's Trends in International Mathematics and Science Study (TIMSS) 2003

* Of the 189 items in the Science test, some extended-response items were scored on a two-point scale, resulting in 211 total score points. Following item review, response categories were combined for a number of items, resulting in 206 total score points. () Standard errors appear in parentheses.

Exhibit C.1: Average Percent Correct for Test-Curriculum Matching Analysis – Science

Based on Subset of Items Specially Identified by Each Country as Addressing its Curriculum
(See Exhibit C.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 different country based on its own decisions about the test items to include.

Countries	Average Percent Correct on All Items	Singapore	Chinese Taipei	Japan	Hungary	England	United States	Latvia	Hong Kong, SAR	Russian Federation	Australia	Netherlands	New Zealand	Italy	Belgium (Flemish)	Lithuania	Scotland	Slovenia	Moldova, Rep. of	Cyprus	Norway	Iran, Islamic Rep. of	Armenia	Philippines	Tunisia	Morocco	Indiana State, US	Ontario Province, Can.	Quebec Province, Can.
Singapore	67 (1.1)	70	73	69	68	69	67	67	69	70	68	70	69	71	69	67	70	68	67	69	68	71	67	67	70	70	70	69	66
Chinese Taipei	66 (0.3)	66	72	64	66	65	65	65	66	70	67	67	66	69	66	65	69	67	65	66	67	68	65	66	68	67	70	66	65
Japan	65 (0.3)	67	71	69	66	65	65	65	67	68	66	68	66	68	67	65	67	66	65	67	66	68	65	64	67	67	71	67	64
Hungary	64 (0.6)	65	67	60	65	64	64	64	64	70	65	66	65	69	66	64	68	65	64	65	66	67	64	63	66	67	67	66	64
England	63 (0.7)	64	69	64	64	64	63	63	64	68	65	66	65	67	64	63	67	65	63	64	65	66	63	64	65	64	69	65	63
United States	63 (0.5)	64	67	61	64	64	63	63	64	68	64	65	64	67	64	63	66	64	63	64	65	66	63	62	65	65	69	65	63
Latvia	63 (0.5)	63	66	60	63	63	62	62	64	68	64	65	63	67	64	62	65	64	62	63	65	65	62	62	65	64	67	64	62
Hong Kong, SAR	62 (0.7)	63	68	61	62	62	62	62	63	65	63	64	63	66	63	62	65	63	62	62	63	65	62	62	64	63	65	63	61
Russian Federation	61 (0.9)	62	65	59	62	61	61	61	62	67	62	63	63	66	63	61	65	63	61	62	63	64	61	61	65	63	66	63	61
Australia	61 (0.8)	61	65	59	61	61	61	61	61	66	62	63	62	65	62	61	64	61	61	61	62	63	61	60	63	63	67	62	60
Netherlands	60 (0.5)	61	63	53	61	60	60	60	61	65	61	63	60	65	63	60	62	61	60	61	62	62	60	57	62	62	64	62	60
New Zealand	59 (0.5)	60	63	58	60	60	59	59	60	64	61	61	60	64	61	59	63	60	59	60	61	62	59	58	62	61	65	61	59
Italy	59 (0.7)	61	63	58	60	59	59	59	61	64	60	62	61	65	61	59	63	60	59	61	61	63	59	58	62	62	64	61	59
Belgium (Flemish)	59 (0.4)	60	63	56	60	60	59	59	60	64	60	62	59	64	62	59	62	60	59	60	62	62	59	57	61	61	65	61	59
Lithuania	59 (0.5)	60	63	56	60	59	59	59	60	64	60	61	60	63	61	59	62	61	59	60	61	62	59	59	61	61	64	61	59
Scotland	56 (0.6)	57	61	54	57	57	56	56	57	61	58	59	57	61	58	56	60	57	56	57	58	59	56	55	58	57	62	58	56
Slovenia	54 (0.5)	55	59	54	55	55	54	54	55	59	55	58	55	59	57	54	57	56	54	55	56	57	54	54	56	56	60	56	54
Moldova, Rep. of	53 (0.8)	53	57	49	53	53	53	53	53	58	53	55	54	57	54	53	56	54	53	53	54	56	53	53	56	55	58	55	53
Cyprus	53 (0.4)	54	59	49	53	53	52	52	54	56	53	55	53	57	54	52	56	54	52	54	54	57	52	53	56	54	56	55	53
Norway	52 (0.4)	52	55	47	52	51	51	51	53	56	52	53	53	56	53	51	55	53	51	50	54	54	51	50	53	53	55	53	51
Iran, Islamic Rep. of	41 (0.7)	42	47	39	42	42	41	41	42	45	42	43	43	45	42	41	45	43	41	42	42	45	41	42	44	43	45	43	41
Armenia	41 (0.7)	42	44	34	41	41	41	41	42	45	42	43	42	45	42	41	45	43	41	41	43	45	41	42	44	43	45	43	41
Philippines	31 (1.3)	31	35	29	31	32	31	31	31	32	31	32	32	34	30	31	34	32	31	31	32	34	31	34	32	33	34	32	30
Tunisia	27 (0.6)	28	33	27	28	28	27	27	28	30	28	28	29	30	28	27	30	29	27	28	28	30	27	26	30	29	30	29	27
Morocco	27 (0.6)	27	31	22	27	27	26	26	27	28	27	27	27	29	26	26	29	27	26	27	28	30	26	27	30	29	30	28	26
International Avg.	55 (0.1)	55	59	52	55	55	55	55	56	59	56	57	56	59	56	55	58	56	55	55	56	58	55	54	57	56	59	56	54
Benchmarking Participants																													
Indiana State, US	66 (0.8)	67	70	65	67	67	66	66	67	71	68	68	67	70	67	66	69	67	66	67	68	68	66	65	68	68	72	68	66
Ontario Province, Can.	65 (0.7)	66	69	63	66	66	65	65	66	69	66	67	66	69	66	65	68	65	65	66	66	67	65	63	67	66	71	67	64
Quebec Province, Can.	58 (0.5)	58	61	54	58	58	57	57	58	63	59	60	58	62	59	57	61	59	57	58	59	60	57	55	60	59	63	59	58
Number of Items (Score Points) Identified*	165	130	69	44	161	134	161	161	140	108	155	115	136	134	126	161	122	136	161	130	150	123	161	66	89	100	68	149	148

SOURCE: IEA's Trends in International Mathematics and Science Study (TIMSS) 2003

* Of the 152 items in the Science test, some extended-response items were scored on a two-point scale, resulting in 168 total score points. Following item review, some items were deleted and response categories were combined for a number of items, resulting in 150 items and 165 total score points.

() Standard errors appear in parentheses.

diagonal element in the exhibit shows how each country performed on the subset of items that it selected based on its own curriculum. Thus, Bulgarian students averaged 42 percent correct on the set of items identified by Bulgaria for the analysis.

The international averages on each country's selected items are presented in the lower part of the exhibit. They show that the selection of items for the participating countries varied somewhat in average difficulty, ranging from 42 percent for those chosen by Scotland and Jordan at the eighth grade to 48 percent for those chosen by Belgium (Flemish). Similarly at the fourth grade, the average percent correct ranged from 52 percent for those items chosen by Japan to 59 percent for those chosen by Chinese Taipei, Italy, Russian Federation, and Indiana State. Despite these differences in the difficulty of the selected items, the overall message of Exhibit C.1 is that different item selections do not make a major difference in how well countries perform relative to one another. The items selected by some countries were more difficult than those selected by others. The relative performance of countries on various item selections did vary somewhat, but generally not in a statistically significant manner.⁵

Comparing the diagonal element for a country with the overall average percent correct shows the difference between performance on the subset of items chosen as appropriate and performance on the test as a whole. In general, there were only small increases in each country's performance on its own subset of items. To illustrate, the average percent correct for Singapore across all eighth-grade science items was 62 percent. The diagonal element shows that Singaporean students had a slightly greater average percent correct (65 percent) across the set of items selected as appropriate for Singapore than they did overall. Almost all participants had a difference of one or two percentage points between the two performance measures, with the largest difference – seven percent – for Belgium (Flemish) (49 percent compared with 56 percent).

5 Small differences in performance shown in this exhibit are not statistically significant. The standard errors for the estimated average percent correct statistics are in Exhibit C.2. It can be said with 95 percent confidence that the value for the entire population falls between the sample estimate plus or minus two standard errors.

Exhibit C.2: Standard Errors for the Test-Curriculum Matching Analysis – Science



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 different country based on its own decisions about the test items to include.

Countries	Average Percent Correct on All Items	Singapore	Chinese Taipei	Japan	Hungary	England	United States	Latvia	Hong Kong, SAR	Russian Federation	Australia	Netherlands	New Zealand	Italy	Belgium (Flemish)	Lithuania	Scotland	Slovenia	Moldova, Rep. of	Cyprus	Norway	Iran, Islamic Rep. of	Armenia	Philippines	Tunisia	Morocco	Indiana State, US	Ontario Province, Can.	Quebec Province, Can.			
Singapore	67 (1.1)	1.1	1.1	1.2	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1		
Chinese Taipei	66 (0.3)	0.3	0.3	0.5	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	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	
Japan	65 (0.3)	0.3	0.3	0.4	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	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	
Hungary	64 (0.6)	0.6	0.5	0.6	0.6	0.6	0.6	0.6	0.5	0.6	0.6	0.5	0.5	0.6	0.6	0.6	0.6	0.5	0.6	0.6	0.6	0.5	0.6	0.6	0.5	0.5	0.5	0.5	0.6	0.6	0.6	
England	63 (0.7)	0.7	0.7	0.8	0.7	0.7	0.7	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	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	
United States	63 (0.5)	0.5	0.5	0.6	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	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	
Latvia	63 (0.5)	0.5	0.5	0.7	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	0.5	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	
Hong Kong, SAR	62 (0.7)	0.7	0.7	0.9	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	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	
Russian Federation	61 (0.9)	0.9	0.9	1.1	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8	0.9	0.9	0.9	0.9	0.9	
Australia	61 (0.8)	0.8	0.8	0.9	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.7	0.8	0.8	0.8	0.7	0.8	0.8	0.8	0.8	0.7	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Netherlands	60 (0.5)	0.5	0.6	0.7	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	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
New Zealand	59 (0.5)	0.5	0.5	0.6	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	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	
Italy	59 (0.7)	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	0.7	0.7	0.7	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	0.7	0.7
Belgium (Flemish)	59 (0.4)	0.4	0.4	0.5	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	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
Lithuania	59 (0.5)	0.5	0.5	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.4	0.5	0.5	0.5	0.5	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Scotland	56 (0.6)	0.6	0.6	0.8	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	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.6	0.6	0.6	0.6	0.6	0.6	0.6
Slovenia	54 (0.5)	0.5	0.4	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	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	0.5
Moldova, Rep. of	53 (0.8)	0.8	0.7	0.9	0.8	0.8	0.8	0.8	0.9	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	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Cyprus	53 (0.4)	0.5	0.4	0.6	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.5	0.4	0.5	0.5	0.4	0.4	0.4	0.4	0.4	0.5	0.4	0.4	0.4	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Norway	52 (0.4)	0.4	0.5	0.5	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	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
Iran, Islamic Rep. of	41 (0.7)	0.8	0.9	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	0.7	0.8	0.8	0.7	0.8	0.8	0.7	0.9	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Armenia	41 (0.7)	0.8	0.8	0.8	0.7	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.7	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.7	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Philippines	31 (1.3)	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.3	1.3	1.3	1.3	1.4	1.4	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.3	1.4	1.3	1.3	1.3	
Tunisia	27 (0.6)	0.6	0.7	0.7	0.6	0.6	0.6	0.6	0.7	0.6	0.6	0.7	0.6	0.7	0.7	0.6	0.7	0.6	0.6	0.6	0.7	0.6	0.7	0.6	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Morocco	27 (0.6)	0.7	0.7	0.6	0.6	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	0.6	0.6	0.6	0.7	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
International Avg.	55 (0.1)	0.1	0.1	0.2	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	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
Benchmarking Participants																																
Indiana State, US	66 (0.8)	0.8	0.8	0.9	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	0.8	0.8	0.7	0.8	0.8	0.7	0.8	0.8	0.7	0.8	0.8	0.8	0.8	0.8
Ontario Province, Can.	65 (0.7)	0.7	0.7	0.9	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	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Quebec Province, Can.	58 (0.5)	0.6	0.6	0.6	0.5	0.6	0.5	0.5	0.5	0.6	0.5	0.6	0.5	0.6	0.6	0.5	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	0.5	0.5	0.5
Number of Items (Score Points) Identified*	165	130	69	44	161	134	161	161	140	108	155	115	136	134	126	161	122	136	161	130	150	123	161	66	89	100	68	149	148			

SOURCE: IEA's Trends in International Mathematics and Science Study (TIMSS) 2003

* Of the 152 items in the Science test, some extended-response items were scored on a two-point scale, resulting in 168 total score points. Following item review, some items were deleted and response categories were combined for a number of items, resulting in 150 items and 165 total score points.

() 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 Table C.1.

It is clear that the selection of items does not have a major effect on the general relationship among countries. Countries 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 on the test as a whole and on all of the different item selections, with Chinese Taipei, Estonia, and Korea 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. As an example, consider the 195 score points selected by Armenia. The students in Armenia did a little better on these items than on the test as a whole, with 39 percent correct on these items, on average, compared with 38 percent correct on all items. However, most other countries also did better on these particular items, with an international average of 44 percent correct on the items selected by Armenia. All 29 participants that performed better than Armenia on the overall test also performed better on the items selected by Armenia.

The TCMA results provide evidence that the TIMSS 2003 science test provides a reasonable basis for comparing achievement of the participating countries and benchmarking entities. This result is not unexpected, since making the test 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 essentially 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 results is largely unaffected.