The Test-Curriculum Matching Analysis: Mathematics

## C



C

When comparing student achievement across countries, it is important that the comparisons be as fair as possible. Lss has worked toward this goal in a number of ways, including providing detailed procedures for standardizing the population definitions, sampling, test translations, test administration, scoring, and database formation. Similar to the procedures used for developing the original TIMSS instruments, developing the TIMSS 1999 tests involved a series of reviews by representatives of the participating countries, experts in mathematics, and testing specialists.<sup>1</sup> The National Research Coordinators (NRCS) from each country formally approved the TIMSS 1999 tests, thus accepting them as being sufficiently fair to compare their students' mathematics achievement with that of students from other countries.

Although the tests were developed to represent a set of agreed-upon mathematics content, differences among the curricula of participating countries result in various topics being taught at different grades. To restrict test items to 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 appropriateness of the TIMSS 1999 mathematics test for the eighth-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.<sup>2</sup>

To gather data about the extent to which the TIMSS 1999 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. The NRC was asked to choose a person or persons who were very familiar with the curriculum at the grade tested to make this determination. Since an item might be in the curriculum for some but not all students in a country, an item was 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 participating countries returned the information for analysis.

Exhibits C.1 and C.2 present the TCMA results for the TIMSS 1999 tests. Exhibit C.1 shows the average percent correct for each country on items selected as appropriate and on the test as a whole. Exhibit C.2 shows the standard errors corresponding to the percentages presented in Exhibit C.1.



<sup>2</sup> Because there may also be curriculum areas covered in some countries that are not covered by the TIMSS 1999 tests, the TCMA does not provide complete information about how well the tests cover the curricula of the countries.

<sup>&</sup>lt;sup>1</sup> See Appendix A for more information on test development.

In Exhibit C.1, the last row of the exhibit indicates that the countries varied substantially in the number of items (score points) identified as appropriate.<sup>3</sup> The percentages ranged from 100 percent (169 score points) in Chinese Taipei, the Slovak Republic, Latvia (LSS), the United States, Lithuania, Moldova, and Indonesia to 58 percent (98 score points) in Chile. Thirty-four of the 38 countries indicated that the items representing three-quarters or more of the score points (127 out of a possible 169) 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.<sup>4</sup>

The first column in Exhibit C.1 shows the average percent correct on all test items for each country. The countries are presented in order of their overall performance based on overall percent correct, 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, Singapore, where the average percent correct was 77 percent on its own set of items, also had 78 percent correct for the items selected by Korea, 77 percent for the items selected by Hong Kong, and so forth. The column for a country listed across the top shows how each of the other countries performed on the subset of items selected as appropriate for its own students. Using the set of items selected by Finland as an example, on average 77 percent of these items were answered correctly by students in Singapore, 73 percent by students in Korea, 72 percent by those in Hong Kong, and so forth. The shaded diagonal element in the exhibit shows how each country performed on the subset of items that it selected based on its own curriculum. Thus, Finnish students averaged 56 percent correct on the set of items identified by Finland for the analysis.

The international averages of each country's selected items are presented across the second to the last row of the exhibit. They show that the selection of items for the participating countries varied somewhat in average difficulty, ranging from 48 to 54 percent. Despite these differences, the overall picture presented by Exhibit C.1 reveals 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.<sup>5</sup>

348

<sup>&</sup>lt;sup>3</sup> Of the 162 items in the test, 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 total number of score points available for analysis was 169. The TCMA uses score points in order to give the same weight to items given them in test scoring.

<sup>4</sup> It should be noted that the performance levels presented in Exhibit C.1 are based on average percents, which are different from the average scale scores that are presented in Chapter 1.

<sup>&</sup>lt;sup>5</sup> 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.

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 was 77 percent. The diagonal element shows that Singaporean students had the same percent correct (77 percent) based on the smaller set of items selected as they did overall. All countries had a difference of less than five percentage points between the two performance measures, with the largest difference four percent for the Netherlands (65 percent compared with 61 percent).

It is clear that the selection of items does not have a major effect on the general relationship among countries. Countries that had substantially higher or lower relative performance on all items also had higher or lower relative performance on the different sets of items selected for the TCMA. For example, Singapore had the highest average percent correct on the test as a whole and on most of the different item selections, with Korea, Hong Kong, and Chinese Taipei among the four highestperforming countries in all cases. 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 149 score points selected by Jordan. The Jordanian students did better on these items than on the test as a whole, with 41 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 54 percent correct on the items selected by Jordan. All 30 countries that performed better than Jordan on the overall test also performed better on the items selected by Jordan.

The TCMA results provide evidence that the TIMSS 1999 mathematics test provides a reasonable basis for comparing achievement of the participating countries. 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. the countries across the top. on the items included by the

left

the country down the

performance of

Read across the row to compare that country's performance based on the

down the column under a country name to compare the

Read o

Instructions:

test items included by each of the

C.1

## Exhibit C.1 Average Percent Correct for Test-Curriculum Matching Analysis – Mathematics Based on Subsets of Items Specially Identified by Each Country as Addressing its Curriculum

(See Exhibit C.2 for corresponding standard errors)



Schools only.

Because population coverage falls below 65% Latvia is annotated LSS for Latvian Speaking 141 Μοτοςςο 22 46 45 45 45 46 40 40 33 33 33 33 33 33 33 33 22 56 22 50 22 32 22 51 126 150 səuiddilind 47 46 40 41 39 337 337 335 337 332 332 24 24 24 22 52 52 98 47 46 45 169 eisəuopuj 2 2 22 22 45 39 39 5 country listed on the top. Read along the **diagona**l to compare performance for each different country based on its own decisions about 150 2 2 1 Iran, Islamic Rep. 72 666 662 662 660 660 660 660 660 660 649 449 449 449 449 449 449 45 446 440 337 337 337 335 335 335 335 227 227 224 224 224 51 165 5 Iurkey , 149 lordan 78 76 48 48 48 43 42 41 40 33 33 33 23 25 23 23 23 23 23 23 23 23 23 23 23 4 145 
 Solution
 134 47 45 46 39 37 35 35 35 31 27 21 21 Macedonia, Kep. of 74 41 41 5 128 147 45 46 39 37 37 37 37 27 27 21 21 <u>20 m</u> 41 51 puelledT 
 Number
 Numer
 Numer
 Numer</th 51 169 51 evobioM 2 2 2 47 46 45 ) Standard errors for the average percent of correct responses on all items appear in parenthe- 
 Somania

 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 33
 <td 153 ses. Because results are rounded to the nearest whole number, some totals may appear inconsistent. cyprus 20/ 47 45 42 41 38 39 35 35 33 33 32 32 28 22 22 22 22 22 22 48 66 5 72 72 66 66 61 61 61 61 61 61 65 55 55 55 55 55 55 55 55 55 74 47 47 46 45 40 38 37 35 35 31 27 24 21 161 Yistiy ~ m 2 5 169 445 445 339 337 337 337 337 337 337 335 335 335 337 224 224 224 224 51 156 46 46 41 41 40 40 33 33 33 33 33 33 33 32 27 22 22 22 22 22 22 52 
 Figure
 Signal
 Signal< 136 169 72 72 66 61 61 61 61 61 61 55 55 55 55 55 55 55 55 55 73 55 74 949 449 47 46 5 vatets betinU 73 45 45 40 39 38 37 35 35 31 26 26 21 21 169 (SSJ) EIVTEJ 2 2 2 46 45 45 39 39 38 37 35 35 35 31 26 21 21 21 5 135 51 140 
 Finland

 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 5
 3
 3
 3
 3
 3
 3
 3
 3
 3</ 162 141 39 37 37 35 35 31 27 21 21 47 45 46 40 41 5 46 46 40 40 38 37 35 35 35 27 21 21 21 Australia 73 73 72 667 667 662 662 662 662 655 558 558 553 553 553 553 553 553 51 50 47 5 165 126 159 5 
 77
 73
 73

 77
 72
 72

 77
 72
 72

 75
 55
 66

 66
 66
 66

 67
 55
 55

 55
 55
 55

 74
 49
 99

 45

 45

 45

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 33

 34

 35

 35

 36

 37

 37

 37

 37

 37

 37

 37

 37

 37

 37

 37

 37

 37

 37

 37

 37

 37

 37

 37

 37

 37
51 162 45 40 39 38 36 335 335 335 335 23 23 23 23 23 23 51 51 120 47 47 47 42 42 39 38 37 37 36 33 33 28 28 28 28 28 21 53 169 46 45 45 40 39 38 37 35 35 35 31 26 26 21 голак керирис 5 the test items to include 142 45 45 41 38 37 37 37 37 27 27 21 21 5 41 resulting in 169 total score 166 
 State
 <th 160 some ueder 51 47 46 45 45 40 39 38 37 35 35 35 31 26 21 21 21 51 169 some items had two parts and AA2 , gnoX gnoH 22 12 7 47 46 44 45 39 39 50 152 2 38 37 37 37 37 27 27 21 21 127 Korea, Kep. of 5 159 46 44 45 40 Singapore  $\begin{array}{c} (1,3)\\ (0,4)\\ (0,1)\\ (1,1)\\ ($ (0.2) sməti IIA no 169 Percent Correct 26 24 21 5 Average Of the 162 items in the mathematics test, Number of Items (Score Points) ٥f **Russian Federation** International Avg. Belgium (Flemish) Hong Kong, SAR Macedonia, Rep. Iran, Islamic Rep. Slovak Republic Czech Republic Korea, Rep. of Chinese Taipei United States New Zealand South Africa Netherlands Latvia (LSS) Philippines Singapore Indonesia Lithuania Countries Malaysia Hungary Australia Romania Moldova Thailand England Morocco ldentified\* Slovenia Bulgaria Finland Canada Jordan Turkey Cyprus Tunisia apan Israe

Italy

 $\mathbf{C}$ 

extended-response items were scored on a two-point scale,

points.

Chile

	The Test-Curriculum	Matching Analysis	Mathematics
--	---------------------	-------------------	-------------

								u										c				
Country	Average Percent Correct smsti IIA no	Singapore Korea, Rep. of	Aong Kong, SAR	Chinese Taipei nagan	Belgium (Flemish) Belgium (Flemish)	Slovak Republic	Netherlands Slovenia	Canada Russian Federation	ailsıtıala Biləysia	Czech Republic	Finland Bulgaria	Latvia (LSS) Vinted States	bnelgn3 bnelesZ w9N	Lithuania Ltaly	Cyprus Romania	evobloM	Israel Thailand	Macedonia, Rep. o Tunisia	Turkey Jordan	Iran, Islamic Rep. Indonesia	Philippines	Morocco South Africa
Singapore	77 (1.3)	1.3 1.3	1.3		1.3 1.3	Ľ	<b></b>	1.3 1.3	1.3 1.3		1.3	1.3	1.3 1.3	1.3 1.3	1.4	1.3	L		1.3	1.3		1.3 1.3
Korea, Rep. of	73 (0.4)	0.5 0.5	0.5	0.4 0.5		0.4		0.4 0.5	0.4 0.5	0.5 0	0.5	0.4	0.4 0.4	0.4 0.4	4 0.5 0.5		0.5 0.5 0	0.5 0.5	0.4 0.4	0.5 0.4 (	0.5 0.5	0.5 0.5
Hong Kong, SAR	72 (1.1)					1.1			1.1		1.1	1.1		1.1	1.2		1.1		1.1	1.1		
	(6.0) 21					0.7			0.0		0.7	0.9		0.7	2.7		0.9		0.9	0.7	÷	
Japan Balaium (Elemish)	/2 (0.4) 66 (0.8)	0.4 0.4	0.4	0.4 0.4	0.4 0.4	4.0 8	0.4 0.4	0.4 0.4	0.4 0.4		0.4 0.4 0	0.4 0.4 0	0.4 0.4	0.4 0.4	4 0.5 0.4		0.4 0.4 0	0.4 0.4	0.4 0.4	0.4 0.4 (	0.4 0.4	0.4 0.4
Hundary	61 (0.9)		6.0		_	0.9			0.8		6.0	0.0	1	6.0	6.0		6.0		6.0	6.0	1	
Slovak Republic	61 (1.1)					1.1				1	1.1	1.1	1	1.1	1.1	1	1.1		1.0	1.1	1	
Netherlands	61 (1.8)		1.8			1.8			1.8	1	1.8	1.8	5	1.8	1.9	5	1.8		1.9	1.8	5	
Slovenia	60 (0.7)		0.7			0.7			0.7		0.7	0.7		0.7	0.8		0.7		0.7	0.7		
Canada	59 (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	- 1	-
Russian Federation	58 (1.3) 58 (1.3)	1.3 1.3	<u>.</u> 5	1.3 1.3	1.3 1.3	(	1.3 1.3	1.3 1.3	1.3 1.3		1.3 1.4 1	1.3 1.3 1	1.3 1.3	1.3 1.3	1.4	2	1.3 1.3 1	1.3 1.3	1.3 1.3	1.3 1.3 .	1.4 1.3	1.3 1.3
Malavsia	57 (1.1)		1 1			11			1.1		12	11		1 1	1.2		11		11	1		
Czech Republic	57 (1.1)		1:1						1:1		1:	1.1		1.1	1:1	1	1:		1:	1.1	1	
Finland	56 (0.7)	0.7 0.7	0.7 0.7 0.	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	0.7 0.7 0	0.7 0.7 0	0.7 0.7	0.7 0.7	7 0.7 0.7	ì	0.7 0.7 0	0.7 0.7	0.6 0.7	0.7 0.7 (	0.7 0.7	0.7 0.7
Bulgaria	54 (1.4)		1.4			1.4			1.4		1.5	1.4		1.4	1.5		1.4		1.4	1.4		
Latvia (LSS)	53 (0.8)		0.8			0.8			0.8		0.9	0.8		0.8	0.9		0.8		0.8	0.8		
United States	52 (0.9)					0.9			0.9		0.9	0.9	1	0.9	1:0		0.9		0.9	0.0		
England New Zealand	50 (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 1 1 2 1 2 1 1 2 1 2 1	1.1 1.1 1 1 2 1 2 1 1 2 1 2 1	1.1 1.2 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	1.0 1.0	1.1 1.1
lithuania	(7.1) OC		10			10			10	1	10	10		1 0	<u>;</u>		10		10	10	1	10 10
Italy	49 (0.8)					0.8			0.8		6.0	0.8		0.8	0.8	1	0.8		6.0	0.8		
Cyprus	47 (0.4)					0.4			0.4	1	0.4	0.4	1	0.4	0.4	1	0.4		0.4	0.4	1	
Romania	47 (1.2)		1.2			1.2			1.2	1	1.3	1.2		1.2	1.3		1.2		1.2	1.2		
Moldova	46 (0.9)		0.9			0.9			0.9		0.9	0.9		0.9	1.0		0.9		0.9	0.9		
Israel	45 (0.8)	0.8 0.8				0.8			0.8		0.8	0.8		0.8	0.8		0.8		0.8	0.8		
Thailand	45 (1.1)		1.1			1.1			1:1		1.1	1:1		1:1	1:1		1.1		1.1	1.1		
Tunicio Tunicio	40 (0.8) 20 (0.5)	0.8 0.8	2 G	0.8 0.8	0.8 U.8	2.0	0.8 0.8	0.8 0.8	0.8 0.8	0 0 2 0 2 0	0.8 U.8 U	0.5 0.8 U	0.8	0.8 0.8	5 U.8 U.8		0.5 0.5 0.5	0.8 0.8	0.8 0.8	0.2 0.2	0.8	0.5 0.8
	(0.0) 85								0.0	1		0.0	1	0.0		1			c.0	0.0	1	
Turkev	37 (0.9)		6.0			6.0			0.8	÷.	6.0	0.9	1	0.9	0.9	1	6.0		0.9	0.9	÷.	
Iran, Islamic Rep.	36 (0.6)	0.6 0.7	0.6			0.6			0.6	1	0.6	0.6		0.6	0.7	1	0.6		0.7	0.6	1	
Indonesia	35 (0.8)	0.8 0.8	0.8			0.8	0.8 0.8		8 0.8 0.8		0.8	0.8 0.8 0	0.8 0.8	0.8 0.8	0.7	1	0.8	0.8 0.8	0.8	0.8 0.8 (	0.8 0.8	0.8 0.8
Chile	31 (0.8)		0.8			0.8			0.8		0.8	0.8		0.8	0.8		0.8		0.8	0.8		
Philippines	26 (0.7)		0.7			0.7			0.7		0.7	0.7		0.7	0.6		0.7		0.7	0.7		
Morocco	24 (0.2)		0.2			0.2			0.2		0.2	0.2	1	0.2	0.2		0.2		0.2	0.2		
South Africa	21 (0.6)	0.6 0.7	0.6	0.6 0.6	0.6 0.6	0.6	0.7 0.6	0.6 0.6	0.7	0.6 0.6	0.7 0.6 0	0.6 0.6 0	0.7 0.6	0.6 0.6	5 0.6 0.6	9.0 0.0	0.7 0.6 0	0.6 0.6	0.7 0.6		0.6	0.6 0.6
International Avg.	51 (0.2)	0.2 0.2	0.2	0.2 0.2	0.2 0.2	0.2	0.2 0.2	0.2 0.2	2 0.2 0.2	0.2	0.2 0.2 0	0.2 0.2 0	0.2 0.2	0.2 0.2	2 0.2 0.2	2 0.2	0.2 0.2 0	0.2 0.2	0.2 0.2	0.2 0.2 (	0.2 0.2	0.2 0.2
Number of Items (Score Points) Identified*	169	159 127 152		169 160	166 142	2 169	169 120 162	165 12	165 126 159 141 162		140 135 169 169 136	39 169 1:	36 156	169 161	1 99 153	169	128 147 1	134 145	149 165 150 169		98 126	150 141
*Of the 162 items in the mathematics test, some items had two parts and some extended- response items were scored on a two-point scale, resulting in 169 total score points.	test, some i vo–point scċ	tems had tr sle, resultin	wo parts 169	and some total scol	e extende re points.	-b	() Stá se	ndard erro 5. The matr	() Standard errors for the average percent of correct responses on all items appear in parenthe- ses. The matrix contains standard errors corresponding to the average percent of correct neconcect based on TCMM, subcars of items, as clicitation in Fublishin C1. Because are usuits are	erage per tandard e	cent of corr rrors corres of items a	ect respons ponding to	ses on all o the avei in Exhibi	items app age perce	the pare the pare pare result	t t are	Because populat Schools only.	pulation ( only.	Because population coverage falls below 65% Latvia is annotated LS Schools only.	s below 65	% Latvia is	annotated l
							2 2	unded to t	rounded to the nearest whole number, some totals may appear inconsistent	hole num	ber, some t	otals may a	appear in	consistent		2						

LSS for Latvian Speaking

**TIMSS**1999 B<sup>th</sup>grade Mathematics

351

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1998-1999.

