10 TIMSS Test-Curriculum Matching Analysis

Albert E. Beaton
Eugenio J. Gonzalez
Boston College

10.1 INTRODUCTION

TIMSS developed international tests of mathematics and science that reflect as far as possible the various curricula of the participating countries. The subject matter coverage of these tests was reviewed by the TIMSS Subject Matter Advisory Committee (SMAC), which consists of mathematics and science educators and practitioners from around the world, and the test was approved for use by the National Research Coordinators (NRCs) of the participating countries. Although every effort was made in TIMSS to ensure the widest possible subject matter coverage, no test can measure all that is taught or learned in every participating country.

Given that no test can cover the curriculum in every country completely, the question arises as to how well the items on the tests match the curricula of each of the participating countries. To address this issue, TIMSS asked each country to indicate which items on the tests, if any, were inappropriate to its curriculum. For each country in turn, TIMSS took the list of remaining items, and computed the average percentage correct on these items for that country and all other countries. This allowed each country to select only those items on the tests that they would like included, and to compare the performance of their students on those items with the performance of the students in each of the other participating countries on that set of items. However, in addition to comparing the performance of all countries on the set of items chosen by each country, the Test-Curriculum Matching Analysis (TCMA) also shows each country's performance on the items chosen by each of the other countries. In these analyses, each country was able to see the performance of all countries on the items appropriate for its curriculum, but to see also the performance of its students on items judged appropriate for the curriculum in other countries.

Each NRC was given a questionnaire with all the items included in the TIMSS tests and was asked to indicate, for each item, whether it was considered an appropriate item for their curriculum. The questionnaire sought the information separately for each item at each grade level at which the items were administered. The results from these questionnaires were then used to assess the curricular coverage of the items in the tests, and the effect on the test results of all countries of omitting those items identified by each NRC or their representative. It must be stressed that this analysis was not intended to replace the carefully constructed and agreed-upon tests that TIMSS used for its international comparisons and research analyses. The IRT scaling and research analyses used all items that were included in the tests and that met psychometric standards. In

the TCMA analysis, items identified by NRCs were omitted from test results only in the series of analyses designed to illuminate and explain the international comparisons based on the entire test.

10.2 THE ANALYTICAL METHOD OF THE TCMA

The TCMA makes use of the average proportion-correct technology described in Chapter 9. The basic item-level data for a participating country at a grade level were represented by the matrix D_{ikj}. This matrix contains elements d_{ikj}, which represent the scored response of student i in country k to item j. The possible values for item j are 0 or 1 for multiple-choice items, and between 0 and 3 for multiple-score items. Most of the elements of D are missing since each student took only one of eight possible booklets administered at a grade level. Depending on the booklet, each student took between one-fifth and two-fifths of the total item pool (Adams and Gonzalez, 1996).

The information provided by the NRC as to whether or not an item should be omitted from these analyses for the particular grade were summarized in a matrix T_{kj} where the elements t_{kj} represent the information that the NRC in country k submitted about item j (for a particular grade). The actual responses of the NRCs for an item were 0 (meaning omit this item for my country) or 1 (meaning include it). Given that multiple-score items were included in the TIMSS tests, both matrices D_{ikj} and T_{kj} were then converted to $D_{ikj'}$ and $T_{kj'}$ matrices as described in the previous chapter. In that conversion, the score points on each item in the matrix $D_{ikj'}$ were transformed into their binary representation, and the item selection by the NRC, contained in the matrix T_{kj}, was transformed into a matrix that matched the $D_{ikj'}$.

Although the procedure described here will work generally for any item selection proportion from 0 to 1, the TCMA analysis in TIMSS was limited to a binary choice of either including or excluding the item at the specific grade level. The computational procedure used for the TCMA analysis was as follows. First form the $P'_{kj'}$ matrix. The elements in matrix $P'_{kj'}$ are computed from the D_{ikj} matrix after the transformations and estimation outlined in the Chapter 9 are applied to the data. The elements of $P'_{kj'}$ are the weighted averages of the student responses in country k to item j', that is, the average of the student responses $d_{ikj'}$, estimated for some elements. Under the TIMSS design, students not administered particular items may be considered missing at random and treated as not having taken the item. Item responses coded as not reached or omitted are treated as incorrect responses.

The next step is to compute a Test Coverage Index. A reasonable index is the percentage of the total possible test points that were deemed appropriate by each country. The total possible test points in a TIMSS test are equal to C_t, and the total possible score on the items deemed appropriate in country k is computed as

$$C_k = \sum_{j'} t_{kj'} \quad .$$

The Test Coverage Index can then be computed as the ratio of the total possible score on the items deemed appropriate in country *k* to the total possible test points in the TIMSS test:

$$\text{Test Coverage Index} = \frac{C_k}{C_t} \quad .$$

The Test Coverage Index indicates the proportion of score points of the test that was considered appropriate to the curriculum in the country. The TCI for each country is presented in Tables 10.1 and 10.2.

Table 10.1 Test Coverage Index for the TIMSS Mathematics Tests

Country	3rd grade	4th Grade	7th Grade	8th Grade
Australia	0.56	0.98	0.70	0.95
Austria	-	-	0.81	0.91
Belgium (Fl)	-	-	0.64	0.86
Belgium (Fr)	-	-	0.64	0.85
Bulgaria	-	-	0.73	0.73
Canada	0.56	0.88	0.54	0.91
Colombia	-	-	0.55	0.82
Cyprus	0.68	0.88	0.62	0.77
Czech Republic	0.54	0.74	0.90	0.93
Denmark	-	-	0.36	0.83
England	0.48	0.76	0.57	0.80
France	-	-	0.79	0.86
Germany	-	-	0.81	0.96
Greece	0.45	0.81	0.67	0.47
Hong Kong	0.40	0.81	0.86	0.93
Hungary	0.65	0.85	0.98	1.00
Iceland	0.30	0.65	0.63	0.82
Iran, Islamic Rep.	0.58	0.86	0.79	0.91
Ireland	0.36	0.76	0.70	0.90
Israel	-	0.74	0.00	0.98
Japan	0.74	0.89	0.90	0.94
Korea	0.61	0.43	0.89	0.91
Kuwait	-	0.58	0.00	0.86
Latvia	0.45	0.93	0.93	0.99
Lithuania	-	-	0.90	0.96
Netherlands	0.23	0.52	0.50	0.72
New Zealand	0.63	0.87	0.71	0.90
Norway	0.72	0.88	0.73	0.93
Portugal	0.90	0.90	0.91	0.94
Romania	-	-	0.54	0.88
Russian Federation	-	-	0.75	0.78
Scotland	0.41	0.81	0.47	0.77
Singapore	0.51	0.74	0.78	0.89
Slovak Republic	-	-	0.94	0.94
Slovenia	0.71	0.79	0.89	0.93
South Africa	-	-	0.50	0.80
Spain	-	-	0.93	0.98
Sweden	-	-	0.62	0.78
Switzerland	-	-	0.56	0.82
Thailand	-	-	-	-
United States	1.00	1.00	1.00	1.00

Table 10.2 Test Coverage Index for the TIMSS Science Tests

Country	3rd Grade	4th Grade	7th Grade	8th Grade
Australia	0.47	0.76	0.64	0.91
Austria	-	-	0.36	0.90
Belgium (Fl)	-	-	0.32	0.67
Belgium (Fr)	-	-	0.16	0.40
Bulgaria	-	-	0.72	0.77
Canada	0.61	0.89	0.53	0.83
Colombia	-	-	0.74	0.77
Cyprus	0.42	0.58	0.20	0.53
Czech Republic	0.38	0.90	0.74	0.93
Denmark	-	-	0.14	0.48
England	0.30	0.50	0.71	0.85
France	-	-	0.18	0.50
Germany	-	-	0.60	0.88
Greece	0.29	0.68	0.49	0.76
Hong Kong	0.32	0.40	0.22	0.47
Hungary	0.39	0.47	0.67	0.88
Iceland	0.89	0.90	1.00	1.00
Iran, Islamic Rep.	0.37	0.81	0.33	0.60
Ireland	0.13	0.26	0.41	0.62
Israel	0.23	0.32	-	0.70
Italy	0.34	0.93	0.64	0.88
Japan	0.16	0.28	0.31	0.59
Korea	0.10	0.24	0.29	0.40
Kuwait	-	0.81	-	0.90
Latvia (LSS)	0.70	0.99	0.32	0.77
Lithuania	-	-	0.54	0.82
Mexico	-	-	0.36	0.39
Netherlands	0.32	0.65	0.23	0.70
New Zealand	0.65	0.86	0.75	0.86
Norway	0.47	0.59	0.63	0.76
Portugal	0.80	0.80	0.55	0.91
Romania	-	-	0.62	0.68
Russian Federation	-	-	0.34	0.66
Scotland	0.31	0.43	0.34	0.66
Singapore	0.30	0.50	0.57	0.75
Slovak Republic	-	-	0.76	0.88
Slovenia	0.88	0.93	0.90	0.96
South Africa	-	-	0.18	0.51
Spain	-	-	0.90	1.00
Sweden	-	-	0.59	0.86
Switzerland	-	-	0.25	0.72
Thailand	-	-	-	-
United States	1.00	1.00	1.00	1.00

After computing the TCI, the next step was to compute the normalized weight matrix. To facilitate cross-national comparisons, it is useful to anchor the various national proficiency estimates in a common manner. The national proficiency estimates described in the next section have the property that, if the students in a country correctly answer

all of the items deemed appropriate for that country, then the country will receive a value of 100; if the students answer all of those items incorrectly, then the country will receive a value of zero. Items not deemed appropriate to the curriculum of a country are not used in computing these values. In situations where the information in T is either 1 (include) or 0 (omit), the country values may be considered percentages of possible points attained on included items. If T contains proportions other than 0 and 1, then the country values may be greater than 100, in which case the students answered more items correctly than was expected from the values in T.

To compute such country estimates, it is necessary to compute the matrix $W_{kj'}$, with the elements $w_{kj'}$, where the matrix elements are computed as follows:

$$w_{kj} = \frac{t_{kj'}}{\sum_{j'} t_{kj'}^2}$$

where the denominator of this equation is the sum of the squares of the NRCs' judgments to the items.

The Country Comparison Matrix can be computed from $P_{kj'}$ and $W_{kj'}$ by the matrix multiplication

$$C_{kk'} = 100 * (W_{kj'} * P'_{kj'})$$

where the elements of $C_{kk'}$ indicate how the students in country k' scored on the items deemed appropriate in country k.

Another way to directly estimate the $C_{kk'}$ matrix without going through the intermediate step of computing the w_{kj} matrix is as follows:

$$C_{kk'} = \frac{\sum_{j'} t_{kj'} * p_{kj'}}{\sum_{j'} t_{kj'}^2} * 100$$

The estimates in the resulting Country Comparison Matrix are unbiased estimators of average student performance based on the items selected by each country for inclusion in the TCMA. The precision of estimates varies as a result of the test booklet rotation as well as the different school and student sampling plans.

10.3 COMPUTING STANDARD ERRORS

The computation of the standard error for the TCMA is a continuation of the procedure described for computing the standard error for the average percent correct. Once the $P_{kj}'^{h'}$ matrices are obtained, we then continue to compute each of the $C_{kk'}^{h'}$ matrices, which can be computed with each of the different $P_{kj}'^{h'}$ replicate matrices. This is accomplished in a straightforward manner by use of the following multiplication:

$$C_{kk'}^{h'} = \frac{\sum_{j'} t_{kj'} * p_{kj'}^{h'}}{\sum_{j'} t_{kj'}^2} * 100$$

The jackknifed standard errors for each of the elements in the $C_{kk'}$ matrix are then computed by applying the following formula

$$jse_{C_{kk'}} = \sqrt{\sum_{h'} (c_{kk'} - c'^{h'}_{kk'})^2}$$

REFERENCES

Adams, R. J. and Gonzalez, E. J. (1996). The TIMSS test design. In Martin, M.O. and Kelly, D.L. (Eds.), *TIMSS technical report, volume I: Design and development*. Chestnut Hill, MA: Boston College.