

Ina V.S. Mullis
Michael O. Martin
Boston College

6.1 CROSS-COUNTRY ITEM STATISTICS

In order to assess the statistical properties of the items before proceeding with item response theory (IRT) scaling (see Chapter 7), TIMSS computed a series of statistics for every item in every country. These basic item statistics (see Figure 6.1 for an example item) were produced by the IEA Data Processing Center. For each item, the basic display presents the number of students that responded in each country, the difficulty level (the percentage of students that answered the item correctly), and the discrimination index (the point-biserial correlation between success on the item and a total score).¹ For multiple-choice items the display presents the percentage of students that chose each option, including the percentage that omitted or did not reach the item, and the point-biserial correlation between each option and the total score. For free-response items (which could have more than one score level), the display presents the difficulty and discrimination of each score level.

As a prelude to the main IRT scaling, the display presents some statistics from a preliminary Rasch analysis, including the Rasch item difficulty for each item, the standard error of this difficulty estimate, and an index of the goodness-of-fit of the item to the Rasch model (Wu, 1997).

The item-analysis display presents the difficulty level of each item separately for male and female students, and, because the TIMSS IRT scaling spans two grades at Population 1 and Population 2, separately for lower- and upper-grade students. As a guide to the overall statistical properties of the item, it also presents the international item difficulty (the mean of the item difficulties across countries) and the international item discrimination (the mean of the item discriminations).

As an aid to reviewers, the item-analysis display includes a series of “flags” signaling the presence of one or more conditions that might indicate a problem with an item. The following conditions are flagged:

- Item difficulty exceeds 95 percent in the sample as a whole
- Item difficulty is less than 25 percent for 4-option multiple-choice items in the sample as a whole (20 percent for 5-option items)

¹ For the purpose of computing the discrimination index, the total score was the percentage of items a student answered correctly in mathematics or science.

Figure 6.1 Examples of Cross-Country Item Analysis

| Country | Correct Answer | | Flags | Percentages for each alternative | | | | | | | | | | Point biserials for each alternative | | | | | | | | | | Rasch | | | | | Group Difficulties | | | | | International Mean | |
|---------|----------------|------|-------|----------------------------------|------|------|-------|-----|------|-----|------|-------|-------|--------------------------------------|-------|-------|-------|-------|-------|-------|-------|------|------|-------|------|------|------|-------|--------------------|--|--|--|--|--------------------|--|
| | N | DIFF | | DISCR | A | B | C | D | E | W | OMIT | NR | A | B | C | D | E | W | OMIT | NR | RDIFF | SE | FIT | MAL | FEM | LOW | UPP | IDIFF | IDISCR | | | | | | |
| AUS | 5112 | 78.5 | 0.36 | qG | 3.3 | 9.8 | 78.5* | 2.4 | 5.2 | 0.6 | 0.1 | -0.21 | -0.14 | .36* | -0.10 | -0.22 | -0.07 | -0.04 | -0.04 | -1.14 | 0.04 | 1.07 | 77.9 | 79.1 | 70.3 | 76.7 | 68.1 | 0.38 | | | | | | | |
| AUT | 2241 | 70.3 | 0.41 | q | 4.6 | 12.4 | 70.3* | 3.4 | 6.6 | 2.2 | 0.1 | -0.19 | -0.16 | .41* | -0.11 | -0.19 | -0.14 | -0.06 | -0.34 | 0.05 | 1.06 | 71.5 | 69.2 | 70.3 | 70.3 | 68.1 | 0.38 | | | | | | | | |
| BFL | 2123 | 82.9 | 0.37 | qG | 2.7 | 8.4 | 82.9* | 2.5 | 2.9 | 0.5 | 0.0 | -0.07 | -0.13 | .37* | -0.13 | -0.17 | -0.07 | -0.01 | -0.79 | 0.06 | 1.00 | 83.3 | 82.6 | 82.1 | 83.8 | 68.1 | 0.38 | | | | | | | | |
| BFR | 1841 | 72.5 | 0.39 | q | 8.0 | 12.2 | 72.5* | 2.2 | 3.3 | 1.3 | 0.1 | -0.24 | -0.14 | .39* | -0.09 | -0.17 | -0.11 | -0.02 | -0.60 | 0.06 | 1.05 | 73.6 | 71.7 | 70.1 | 74.6 | 68.1 | 0.38 | | | | | | | | |
| BGR | 1285 | 59.2 | 0.46 | qG | 8.2 | 16.2 | 59.2* | 4.4 | 6.8 | 3.1 | 0.0 | -0.17 | -0.15 | .46* | -0.12 | -0.21 | -0.19 | -0.00 | 0.06 | 0.06 | 1.04 | 59.2 | 59.2 | 55.5 | 62.9 | 68.1 | 0.38 | | | | | | | | |
| CAN | 6180 | 71.5 | 0.41 | qG | 5.5 | 10.9 | 71.5* | 4.1 | 7.3 | 0.4 | 0.1 | -0.21 | -0.18 | .41* | -0.11 | -0.20 | -0.08 | -0.03 | -0.83 | 0.03 | 1.00 | 72.0 | 71.2 | 69.3 | 73.7 | 68.1 | 0.38 | | | | | | | | |
| CHE | 3718 | 77.4 | 0.42 | qG | 4.2 | 9.7 | 77.4* | 3.1 | 3.6 | 0.7 | 0.2 | -0.25 | -0.19 | .42* | -0.12 | -0.15 | -0.11 | -0.04 | -0.67 | 0.04 | 1.02 | 77.8 | 77.1 | 78.4 | 82.6 | 68.1 | 0.38 | | | | | | | | |
| CHE | 588 | 79.8 | 0.38 | qG | 3.9 | 8.3 | 79.8* | 2.2 | 4.3 | 0.5 | 0.0 | -0.21 | -0.17 | .38* | -0.11 | -0.19 | -0.08 | 0.00 | -0.73 | 0.11 | 0.98 | 80.9 | 78.5 | 78.9 | 80.6 | 68.1 | 0.38 | | | | | | | | |
| CHE | 570 | 75.8 | 0.36 | qS | 3.3 | 11.8 | 75.8* | 3.2 | 3.9 | 0.4 | 0.4 | -0.16 | -0.15 | .38* | -0.10 | -0.20 | -0.10 | -0.10 | -0.50 | 0.11 | 1.03 | 74.2 | 77.5 | 73.0 | 78.6 | 68.1 | 0.38 | | | | | | | | |
| COL | 1978 | 35.9 | 0.25 | qS | 14.0 | 19.2 | 35.9* | 7.7 | 14.8 | 5.8 | 1.6 | 0.00 | -0.02 | .25* | -0.05 | -0.15 | -0.15 | -0.09 | -0.58 | 0.05 | 1.07 | 36.3 | 35.7 | 35.4 | 36.5 | 68.1 | 0.38 | | | | | | | | |
| CSK | 2485 | 64.8 | 0.38 | qSG | 5.5 | 19.7 | 64.8* | 3.3 | 3.8 | 1.1 | 0.2 | -0.21 | -0.18 | .38* | -0.10 | -0.14 | -0.08 | -0.05 | -0.02 | 0.05 | 1.11 | 67.3 | 62.3 | 62.7 | 66.9 | 68.1 | 0.38 | | | | | | | | |
| CYP | 2129 | 53.9 | 0.35 | q | 9.1 | 16.6 | 53.9* | 6.0 | 10.1 | 2.5 | 0.0 | -0.12 | -0.10 | .35* | -0.08 | -0.19 | -0.12 | -0.03 | -0.43 | 0.05 | 1.10 | 54.7 | 53.0 | 52.1 | 55.6 | 68.1 | 0.38 | | | | | | | | |
| DEU | 2258 | 70.2 | 0.41 | q | 5.3 | 9.8 | 70.2* | 3.7 | 6.2 | 2.3 | 0.4 | -0.22 | -0.11 | .41* | -0.09 | -0.16 | -0.19 | -0.09 | -0.86 | 0.05 | 1.03 | 70.4 | 70.7 | 66.8 | 73.6 | 68.1 | 0.38 | | | | | | | | |
| DNK | 2372 | 77.4 | 0.43 | q | 3.7 | 9.6 | 77.4* | 3.1 | 3.7 | 1.4 | 0.4 | -0.20 | -0.20 | .43* | -0.13 | -0.20 | -0.14 | -0.04 | -1.28 | 0.05 | 0.97 | 78.4 | 76.3 | 78.0 | 86.4 | 68.1 | 0.38 | | | | | | | | |
| ESP | 2813 | 54.7 | 0.40 | qS | 8.7 | 19.0 | 54.7* | 4.5 | 8.7 | 4.4 | 0.1 | -0.19 | -0.10 | .40* | -0.09 | -0.19 | -0.15 | 0.00 | -0.26 | 0.04 | 1.03 | 56.8 | 52.6 | 50.5 | 58.8 | 68.1 | 0.38 | | | | | | | | |
| FRA | 2245 | 75.6 | 0.35 | qS | 6.1 | 11.4 | 75.6* | 1.7 | 3.4 | 0.4 | 0.3 | -0.20 | -0.14 | .35* | -0.06 | -0.18 | -0.05 | -0.07 | -0.90 | 0.05 | 1.04 | 78.1 | 73.2 | 70.7 | 80.7 | 68.1 | 0.38 | | | | | | | | |
| GBR | 1322 | 77.5 | 0.42 | qG | 2.6 | 11.6 | 77.5* | 2.7 | 5.2 | 0.4 | 0.0 | -0.20 | -0.20 | .42* | -0.12 | -0.23 | -0.10 | 0.00 | -1.48 | 0.07 | 0.99 | 78.6 | 76.2 | 77.5 | 77.4 | 68.1 | 0.38 | | | | | | | | |
| GRC | 3017 | 45.4 | 0.42 | G | 17.8 | 14.5 | 45.4* | 6.5 | 11.2 | 2.8 | 0.1 | -0.17 | -0.07 | .42* | -0.07 | -0.22 | -0.12 | 0.00 | 0.02 | 0.04 | 1.06 | 46.2 | 44.5 | 38.1 | 52.6 | 68.1 | 0.38 | | | | | | | | |
| HKG | 2537 | 85.6 | 0.42 | qS | 1.7 | 5.5 | 85.6* | 2.8 | 3.9 | 0.4 | 0.0 | -0.23 | -0.24 | .42* | -0.09 | -0.22 | -0.05 | 0.00 | -1.09 | 0.06 | 1.00 | 86.8 | 84.0 | 84.3 | 86.8 | 68.1 | 0.38 | | | | | | | | |
| HUN | 2224 | 61.5 | 0.46 | qSG | 9.2 | 15.5 | 61.5* | 5.0 | 1.9 | 6.6 | 0.4 | -0.24 | -0.16 | .46* | -0.18 | -0.10 | -0.17 | -0.06 | -0.06 | 0.05 | 1.02 | 62.9 | 60.4 | 55.0 | 68.3 | 68.1 | 0.38 | | | | | | | | |
| IRL | 2332 | 77.7 | 0.41 | qS | 3.7 | 11.2 | 77.7* | 3.1 | 3.6 | 0.1 | 0.0 | -0.23 | -0.20 | .41* | -0.11 | -0.19 | -0.05 | -0.03 | -1.07 | 0.05 | 1.01 | 77.7 | 77.8 | 75.3 | 80.3 | 68.1 | 0.38 | | | | | | | | |
| IRN | 2755 | 42.1 | 0.30 | SG | 17.5 | 20.3 | 42.1* | 6.0 | 11.7 | 2.1 | 0.0 | -0.13 | -0.05 | .30* | -0.04 | -0.20 | -0.01 | 0.00 | -0.37 | 0.04 | 1.05 | 41.3 | 43.1 | 35.5 | 48.8 | 68.1 | 0.38 | | | | | | | | |
| ISL | 1388 | 73.1 | 0.40 | q | 3.6 | 13.4 | 73.1* | 4.0 | 5.0 | 0.4 | 0.1 | -0.18 | -0.20 | .40* | -0.15 | -0.15 | -0.06 | -0.04 | -1.37 | 0.07 | 0.98 | 73.5 | 72.7 | 69.3 | 77.5 | 68.1 | 0.38 | | | | | | | | |
| ISR | 518 | 63.3 | 0.38 | q | 7.9 | 16.6 | 63.3* | 5.0 | 5.6 | 1.2 | 0.4 | -0.16 | -0.15 | .38* | -0.18 | -0.10 | -0.10 | -0.16 | -0.18 | 0.10 | 1.10 | 65.0 | 64.0 | 63.3 | 65.3 | 68.1 | 0.38 | | | | | | | | |
| JPN | 3913 | 83.3 | 0.33 | qSG | 1.9 | 8.6 | 83.3* | 2.4 | 3.7 | 0.0 | 0.0 | -0.17 | -0.14 | .33* | -0.13 | -0.22 | 0.00 | 0.00 | -0.73 | 0.05 | 1.09 | 82.4 | 84.2 | 82.1 | 84.4 | 68.1 | 0.38 | | | | | | | | |
| KOR | 2160 | 91.3 | 0.45 | qSG | 1.7 | 3.1 | 91.3* | 2.0 | 1.8 | 0.1 | 0.0 | -0.19 | -0.20 | .45* | -0.21 | -0.28 | -0.04 | 0.00 | -1.84 | 0.08 | 0.91 | 92.9 | 89.3 | 90.7 | 91.9 | 68.1 | 0.38 | | | | | | | | |
| KWT | 695 | 42.8 | 0.27 | q | 14.8 | 16.4 | 42.8* | 7.1 | 14.5 | 4.1 | 0.3 | -0.10 | -0.03 | .27* | -0.09 | -0.14 | -0.06 | -0.10 | -0.66 | 0.09 | 1.06 | 41.4 | 43.5 | 43.5 | 43.5 | 68.1 | 0.38 | | | | | | | | |
| LTU | 1882 | 45.2 | 0.39 | sS | 10.2 | 24.0 | 45.2* | 5.6 | 6.9 | 6.5 | 0.4 | -0.15 | -0.09 | .39* | -0.04 | -0.18 | -0.22 | -0.06 | 0.02 | 0.05 | 1.06 | 42.3 | 47.8 | 39.7 | 50.7 | 68.1 | 0.38 | | | | | | | | |
| LVA | 1887 | 47.3 | 0.35 | q | 8.2 | 26.0 | 47.3* | 4.6 | 9.4 | 3.9 | 0.6 | -0.16 | -0.06 | .35* | -0.07 | -0.21 | -0.12 | -0.05 | 0.06 | 0.05 | 1.10 | 46.7 | 47.9 | 43.5 | 51.2 | 68.1 | 0.38 | | | | | | | | |
| MEX | 4371 | 48.2 | 0.33 | q | 16.4 | 18.6 | 48.2* | 4.0 | 10.8 | 1.2 | 0.4 | -0.17 | -0.06 | .33* | -0.08 | -0.19 | -0.07 | -0.04 | -0.88 | 0.03 | 1.02 | 48.3 | 48.0 | 45.8 | 50.7 | 68.1 | 0.38 | | | | | | | | |
| NLD | 1546 | 78.4 | 0.36 | q | 3.2 | 13.8 | 78.4* | 1.6 | 2.7 | 0.3 | 0.2 | -0.27 | -0.17 | .36* | -0.08 | -0.14 | -0.11 | -0.08 | -0.98 | 0.07 | 1.06 | 78.4 | 78.2 | 74.9 | 82.0 | 68.1 | 0.38 | | | | | | | | |
| NOR | 2144 | 72.8 | 0.37 | qG | 4.2 | 12.3 | 72.8* | 3.3 | 6.5 | 0.9 | 0.1 | -0.21 | -0.15 | .37* | -0.08 | -0.18 | -0.12 | -0.03 | -1.12 | 0.05 | 1.02 | 72.0 | 73.3 | 70.2 | 74.8 | 68.1 | 0.38 | | | | | | | | |
| NZL | 2543 | 77.1 | 0.41 | qS | 4.3 | 10.5 | 77.1* | 3.1 | 4.7 | 0.1 | 0.0 | -0.20 | -0.20 | .41* | -0.14 | -0.19 | -0.02 | -0.04 | -1.41 | 0.05 | 1.00 | 79.0 | 75.2 | 73.9 | 79.8 | 68.1 | 0.38 | | | | | | | | |
| PHL | 4478 | 49.3 | 0.15 | qbBFSG | 6.5 | 25.8 | 49.3* | 4.7 | 13.2 | 0.4 | 0.1 | -0.06 | 0.03 | .15* | -0.07 | -0.16 | -0.03 | -0.02 | -0.92 | 0.03 | 1.20 | 50.2 | 48.7 | 49.7 | 49.0 | 68.1 | 0.38 | | | | | | | | |
| PRT | 2486 | 64.0 | 0.35 | qS | 6.9 | 17.7 | 64.0* | 3.0 | 6.1 | 2.2 | 0.2 | -0.17 | -0.13 | .35* | -0.11 | -0.15 | -0.12 | 0.01 | -1.14 | 0.05 | 1.01 | 67.5 | 60.4 | 60.9 | 67.0 | 68.1 | 0.38 | | | | | | | | |
| ROM | 2789 | 43.5 | 0.35 | sS | 16.6 | 16.9 | 43.5* | 6.9 | 9.3 | 4.0 | 0.5 | -0.12 | -0.08 | .35* | -0.07 | -0.18 | -0.12 | -0.04 | 0.19 | 0.04 | 1.09 | 45.9 | 41.1 | 39.9 | 47.1 | 68.1 | 0.38 | | | | | | | | |
| RUS | 3036 | 54.4 | 0.31 | qF | 8.5 | 24.4 | 54.4* | 2.6 | 5.6 | 3.3 | 0.4 | -0.17 | -0.04 | .31* | -0.09 | -0.20 | -0.15 | -0.02 | 0.20 | 0.04 | 1.19 | 54.9 | 53.9 | 51.5 | 57.3 | 68.1 | 0.38 | | | | | | | | |
| SCO | 2127 | 76.4 | 0.39 | q | 2.8 | 12.4 | 76.4* | 2.9 | 4.8 | 0.3 | 0.0 | -0.22 | -0.17 | .39* | -0.11 | -0.23 | -0.07 | -0.02 | -1.48 | 0.06 | 1.01 | 76.1 | 76.6 | 73.9 | 76.9 | 68.1 | 0.38 | | | | | | | | |
| SGP | 3096 | 85.5 | 0.36 | qs | 1.8 | 7.2 | 85.5* | 2.2 | 3.2 | 0.1 | 0.0 | -0.12 | -0.24 | .36* | -0.14 | -0.17 | -0.04 | 0.01 | -0.56 | 0.05 | 1.02 | 84.5 | 86.4 | 82.0 | 86.2 | 68.1 | 0.38 | | | | | | | | |
| SLV | 2650 | 61.1 | 0.37 | qFG | 5.5 | 20.8 | 61.1* | 4.0 | 5.8 | 1.4 | 0.2 | -0.17 | -0.16 | .37* | -0.08 | -0.18 | -0.09 | -0.05 | 0.03 | 0.05 | 1.13 | 60.3 | 61.8 | 58.6 | 63.6 | 68.1 | 0.38 | | | | | | | | |
| SVN | 2086 | 68.2 | 0.38 | qs | 5.5 | 17.3 | 68.2* | 3.5 | 3.2 | 0.8 | 0.0 | -0.21 | -0.19 | .38* | -0.09 | -0.13 | -0.06 | 0.00 | -0.46 | 0.05 | 1.06 | 69.9 | 66.7 | 62.8 | 73.7 | 68.1 | 0.38 | | | | | | | | |
| SWE | 3277 | 73.9 | 0.38 | q | 3.3 | 13.5 | 73.9* | 4.0 | 4.7 | 0.3 | 0.1 | -0.20 | -0.16 | .38* | -0.13 | -0.20 | -0.08 | -0.04 | -0.96 | 0.04 | 1.05 | 73.4 | 74.4 | 74.6 | 80.0 | 68.1 | 0.38 | | | | | | | | |
| USA | 4108 | 68.7 | 0.37 | qS | 5.4 | 14.5 | 68.7* | 3.8 | 5.8 | 0.7 | 0.0 | -0.21 | -0.19 | .37* | -0.11 | -0.13 | -0.07 | 0.01 | -1.04 | 0.04 | 1.06 | 67.6 | 72.0 | 66.4 | 71.6 | 68.1 | 0.38 | | | | | | | | |

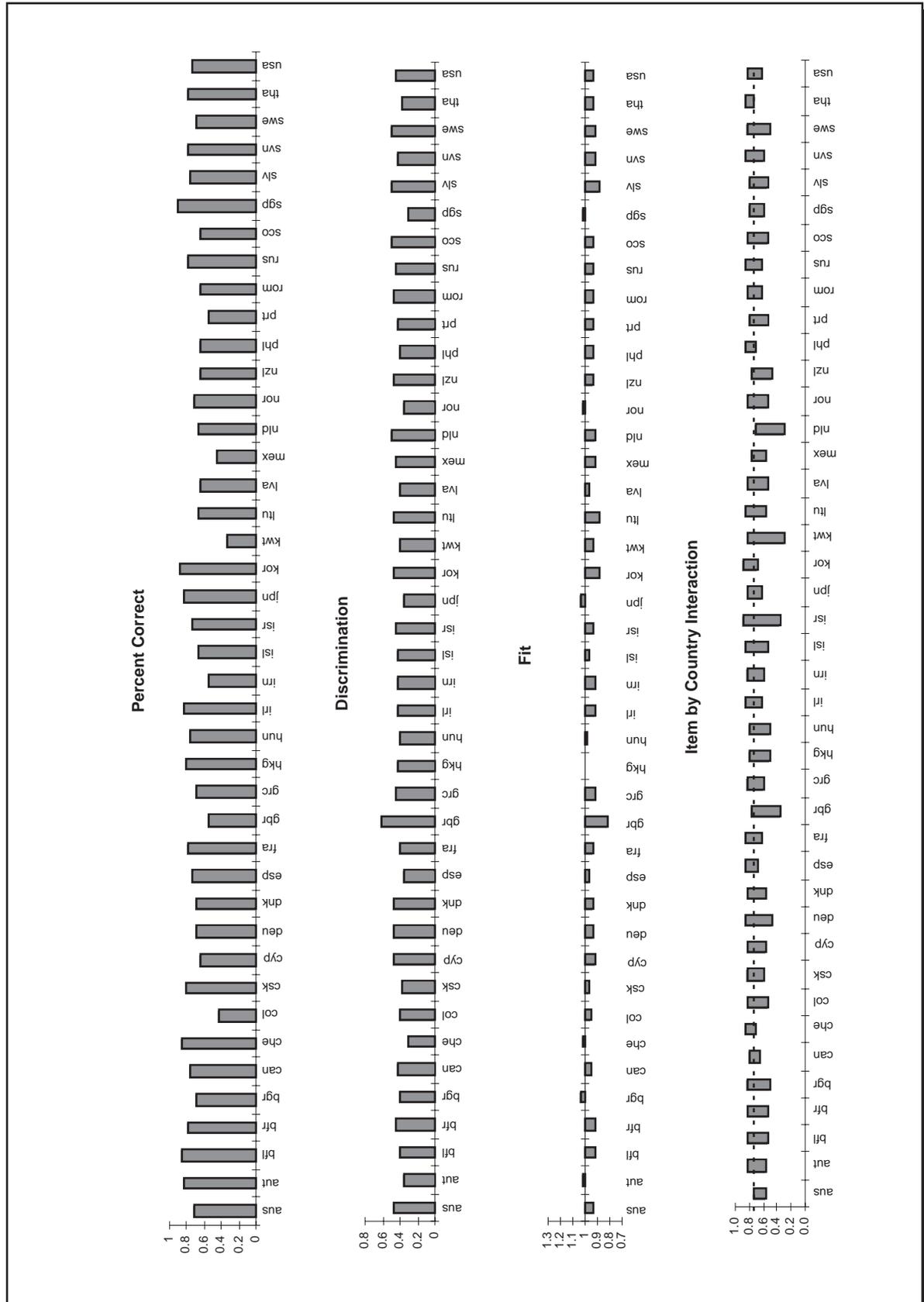
- Item difficulty exceeds 95 percent or is less than 25 percent (20 percent for 5-option items) for students in the lower grade
- Item difficulty exceeds 95 percent or is less than 25 percent (20 percent for 5-option items) for students in the upper grade
- One or more of the distracter percentages is less than 5 percent
- One or more of the distracter percentages is greater than the percentage for the correct answer
- Point-biserial correlation for one or more of the distracters exceeds zero
- Item discrimination (i.e., the point-biserial for the correct answer) is less than 0.2
- Item discrimination does not increase with each score level (for an item with more than one score level)
- Rasch goodness-of-fit index is less than 0.88 or greater than 1.12
- Difficulty levels on the item are significantly different for males and females
- Difference in item difficulty levels between males and females diverge significantly from the average difference between males and females across all the items making up the total score
- Difference in item difficulty levels between lower and upper grades diverge significantly from the average difference between lower and upper grades on all the items making up the total score.

Although not all of these conditions necessarily indicate a problem, the flags are a useful way to draw a reviewer's attention to potential sources of concern. The IEA Data Processing Center also produced information about the inter-rater agreement for the free-response items.

6.2 GRAPHICAL DISPLAYS

As a further aid to reviewing the psychometric characteristics of the items, the Australian Council for Educational Research (ACER) produced graphical representations of selected item statistics for each participating country (see Figure 6.2). This display presents, for each item, the difficulty level and discrimination for every country, together with the Rasch goodness-of-fit statistic and an indication of the item-by-country interaction. The item-by-country interaction chart plots a confidence interval for the probability of success on the item in each country against the average probability of success across all countries. The graphical representations allow comparisons across countries on these statistics at a glance.

Figure 6.2 Example of Graphical Displays of Cross-Country Item Statistics - Mathematics - Population 2



6.3 SUMMARY INFORMATION FOR POTENTIALLY PROBLEMATIC ITEMS

Although the system of flagging potentially problematic conditions and the graphical summaries were both very helpful in identifying items with possible problems, the task of reviewing the characteristics of each item in each country was still considerable. To ensure that no serious item problem would go unnoticed, ACER also provided, for each item, a list of countries that exhibited one or more potentially serious characteristics (see Figure 6.3). Countries were listed in this display if the item had a significant item-by-country interaction (i.e., students in the country found the item easier or more difficult than items in general), or if they exhibited problematic discrimination (i.e., the point-biserial for a distracter was greater than .05, the point-biserial for the correct answer was negative, or, for items with more than one score point, the point-biserial did not increase with each score level). Countries were also listed if their data showed poor fit to the Rasch model for that item.

6.4 ITEM CHECKING PROCEDURES

Prior to the international scaling of the Population 1 and 2 achievement data by ACER, the International Study Center conducted a thorough review of the item statistics for all participating countries to ensure that items were performing comparably across countries. Although only a small number of items were found to be inappropriate for international comparisons, throughout the series of item-checking steps a number of reasons were discovered for differences in items across countries. Most of these were inadvertent changes in the items during the printing process, including omitting an item option or misprinting the graphics associated with an item. However, differences attributable to translation problems were found for an item or two in several countries.

In particular, items with the following problems were considered for possible deletion from the international database:

- Errors were detected during translation verification but were not corrected before test administration
- Data cleaning revealed more or fewer options than in the original version of the item
- The item analysis information showed the item to have a negative biserial
- The item-by-country interaction results showed a very large negative interaction for a given country
- The item-fit statistic indicated the item was not fitting the model
- For free-response items, the within-country scoring reliability data showed an agreement of less than 70% for the score level. Also, performance in items with more than one score level was not ordered by score, or correct levels were associated with negative point-biserials.

Figure 6.3 Example Summary Information for Items with Poor Statistics for Some Countries

| Country | Item-by-Country Interactions | | Non-key PB is Positive | Discrimination | | Fit |
|-----------------|------------------------------|--|------------------------------|-----------------------|------------------------|--------------|
| | Easier than Expected | Harder than Expected | | Key PB is Negative | Ability not Ordered | Fit Large |
| Table=#Name | | | | | | |
| <i>Item 119</i> | <i>BSMSQ15</i> | <i>BSMS/WHICH IS NOT A CHEMICAL CHANGE (A)</i> | | | | |
| DEU | | | X | | | |
| HKG | | | X | | | |
| ISL | | | X | | | |
| ISR | | | X | | | |
| NOR | | | X | | | |
| PHL | | | | | | |
| <i>Item 120</i> | <i>BSMSQ16</i> | <i>BSMS/HOW LONG TAKE LIGHT FROM STAR (D)</i> | | | | |
| COL | | | X | | | |
| CYP | | | X | | | |
| DEU | | | X | | | |
| GRC | | | X | | | |
| HKG | | | X | | | |
| ISR | | | X | | | |
| KOR | | | X | | | |
| MEX | | | X | | | |
| ROM | | | X | | | |
| THA | | | X | X | | |

The statistics and translation verification documentation were used as pointers towards checking actual booklets and contacting National Research Coordinators. If a problem could be detected by the International Study Center (such as a negative point-biserial for a correct answer or too few options for the multiple-choice questions), the item was deleted from the international scaling. However, if there was a question about potential translation or cultural issues, then the NRC was queried, and the International Study Center abided by the decision made by the NRC. In several cases, NRCs consulted mathematics or science experts before making a decision.

Considering that the checking involved approximately 500 items for each of more than 40 countries, very few deviations from the international format were revealed. Table 6.1 contains a list of the changes made in the international database for Populations 1 and 2.

Table 6.1 Recodes Made to Free-Response Item Codes in the Written Assessment and Performance Assessment Items

| | Item | Variable | Recode | Comment |
|--|-----------|----------|---|---|
| General | All Items | | 37, 38 → 39 27, 28 → 29 17, 18 → 19 77, 78 → 79 | Country-specific diagnostic codes recoded to 'other' categories within the score level. |
| | K10 | BSMMK08 | 71 → 70 | Training team found it difficult to distinguish between the 70 and 71 codes; both codes combined in 70. |
| Population 2 - Written Assessment Items | L04 | BSESL04 | 20 → 10 21 → 11 29 → 19 10 → 74 11 → 75 12 → 76 19 → 79 | Only 20s have positive point-biserial correlation; change to 1-point item codes. |
| | M11 | BSESM11 | 10, 11, 12, 13 → 71 20, 21, 22, 23, 24, 25 → 72 30 → 10 31 → 11 | Only 30s have positive point-biserial correlation; change to 1-point item codes. |
| | Y01 | BSESY01 | 20 → 10 21 → 11 22 → 12 29 → 19 10 → 73 11 → 74 19 → 75 | Only 20s have positive point-biserial correlation; change to 1-point item codes. |
| | Y02 | BSESY02 | 21 → 19 | Typographical error in category 21 in coding guide. |
| | J03 | BSSSJ03 | 19 → 10 | Typographical error in coding guide. |
| | M12 | BSSSM12 | 19 → 10 | Typographical error in coding guide. |
| | O14 | BSES014 | 20 → 10 29 → 19 10 → 72 11 → 73 19 → 74 | Only 20s have positive point-biserial correlation. |
| | Q18 | BSSSQ18 | 19 → 10 29 → 20 | Typographical error in coding guide. |
| | L16 | BSSML16 | 19 → 10 | Typographical error in coding guide. |
| | M06 | BSSMM06 | 19 → 10 | Typographical error in coding guide. |
| | M08 | BSSMM08 | 19 → 10 | Typographical error in coding guide. |
| | Q10 | BSSMQ10 | 19 → 10 | Typographical error in coding guide. |
| | R13 | BSSMR13 | 74 → 79 | Typographical error in code 74 (28 instead of 280); leaves gap in 7* diagnostic codes. |
| | S01A | BSEMS01A | 19 → 10 | Typographical error in coding guide. |
| | S02A | BSEMS02A | 19 → 10 | Typographical error in coding guide. |
| | T01A | BSEMT01A | 29 → 20 | Typographical error in coding guide. |
| | T02A | BSEMT02A | 19 → 10 | Typographical error in coding guide. |
| | U01A | BSEMU01A | 19 → 10 | Typographical error in coding guide. |
| | U02A | BSEMU02A | 19 → 10 29 → 20 | Typographical error in coding guide. |
| | U02B | BSEMU02B | 19 → 10 29 → 20 | Typographical error in coding guide. |

Table 6.1 Recodes Made to Free-Response Item Codes in the Written Assessment and Performance Assessment Items (Continued)

| | Item | Variable | Recode | Comment |
|--|--|------------------|---|--|
| Population 1 - Written Assessment Items | T04A | ASEMT04A | 20 → 10 29 → 19 10 → 72 11 → 73 | Only 20s have positive point-biserial correlation. |
| | T04B | ASEMT04B | 20 → 10 29 → 19 10 → 72 11 → 73 | Only 20s have positive point-biserial correlation. |
| | V04A | ASEMV04A | 30 → 20 20 → 12 21 → 13 | Differentiation between 30s, 20s, and 10s not clear. |
| | Y01 | ASESY01 | 20 → 10 29 → 19 10 → 72 11 → 73 19 → 74 | Only 20s have positive point-biserial correlation. |
| | Z02 | ASESZ02 | 30 → 10 31 → 11 20 → 71 29 → 72 10, 11, 12, 13 → 73 | Only 30s have positive point-biserial correlation. |
| Performance Assessment Items | Task M2 (Calculator) Item 5 (Population 2) | BSPM25 | 11 → 12 | Error in coding guide: valid codes listed as 10,12, 19 (no code 11). Recoded 11 codes used in some countries. |
| | Task M5 (Packaging) Item 1 (Populations 2 & 1) | BSPM51 ASPM51 | 30 → 22 31 → 23 | Two versions of task used across countries: original asked for 2 OR 3 boxes; revised asked for 3. Item changed to 2-point value for report tables; changed codes for 3 correct boxes (30,31) to 2-point codes (22,23). |
| | Task S5 (Solutions) Item 2A (Population 2) | BSPS52A | 99 → 98 | Administrator notes not coded consistently across countries; invalid 99 codes (blank) used in several countries recoded to not administered. Item omitted from report table but kept in data file. |
| | Task S5 (Solutions) Item 4 (Population 2) | BSPS54 | 10 → 21 | Coding guide revised based on reports of problematic scoring during training development. |
| | Task S6 (Containers) Item 1A (Population 1) | ASPS61A | 99 → 98 | Administrator notes not coded consistently across countries; invalid 99 codes (blank) used in several countries recoded to not administered. |

REFERENCES

Wu, M.L. (1997). *The development and application of a fit test for use with marginal maximum likelihood estimation and generalised item response models*. Unpublished master's dissertation, University of Melbourne.