CHAPTER 10

Reviewing the PIRLS 2016 Achievement Item Statistics

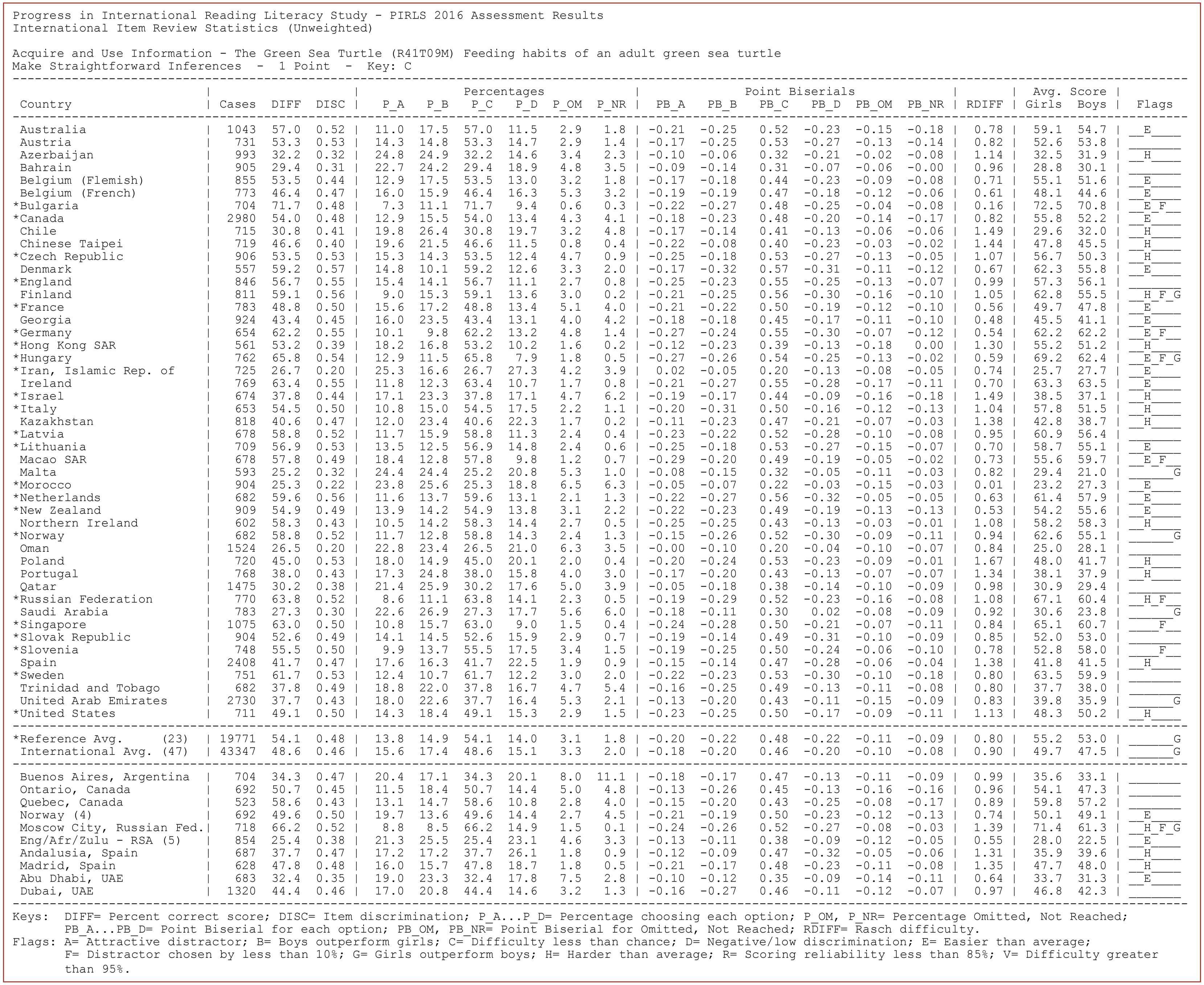
Pierre Foy
Michael O. Martin
Ina V.S. Mullis
Liqun Yin

The TIMSS & PIRLS International Study Center conducted an in depth review of a range of diagnostic statistics to examine and evaluate the psychometric characteristics of each achievement item across the countries that participated in the PIRLS 2016 assessments. This review of item statistics is essential to the successful application of item response theory (IRT) scaling to derive student achievement scores for analysis and reporting. The review played a crucial role in the quality assurance of the PIRLS 2016 achievement data prior to scaling, making it possible to detect unusual item properties that could signal a problem or error for a particular country. For example, an item that was uncharacteristically easy or difficult, or had an unusually low discriminating power, could indicate a potential problem with either translation or printing. Similarly, a constructed response item with unusually low scoring reliability could indicate a problem with a scoring guide in a particular country. In the rare instances where such items were found, the country's translation verification documents and printed booklets were examined for flaws or inaccuracies and, if necessary, the item was removed from the international database for that country.

Statistics for Item Review

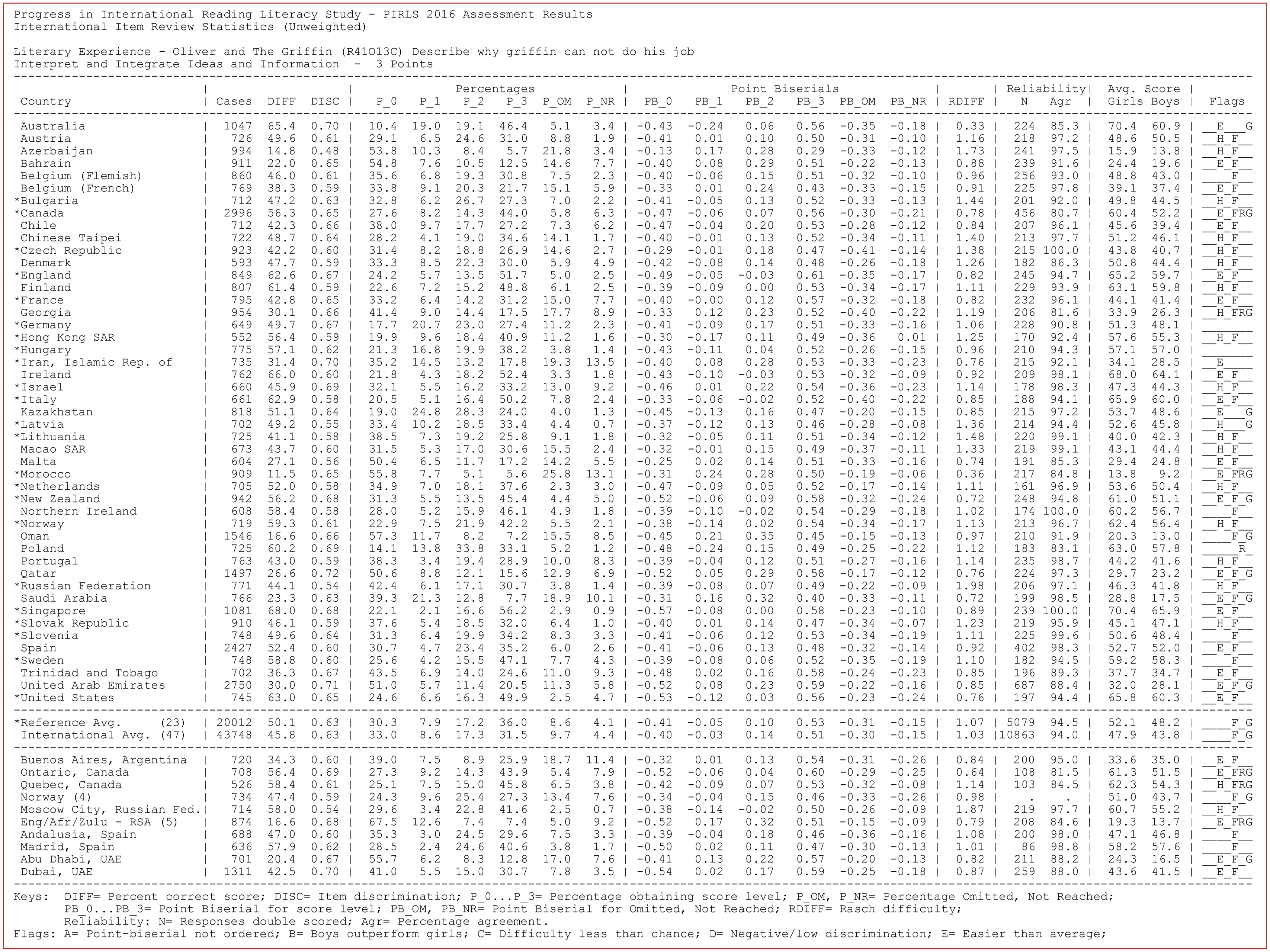
The TIMSS & PIRLS International Study Center computed item statistics for all achievement items in the 2016 assessments, including PIRLS (175 items), PIRLS Literacy (183 items), and ePIRLS (91 items). The item statistics for each of the participating countries were then carefully reviewed. Exhibits 10.1 and 10.2 show actual samples of the statistics calculated for a multiple-choice and a constructed response item, respectively.

Exhibit 10.1: Example International Item Statistics for a PIRLS 2016 Multiple-Choice Item



```
Progress in International Reading Literacy Study - PIRLS 2016 Assessment Results
International Item Review Statistics (Unweighted)

Acquire and Use Information - The Green Sea Turtle (R41T09M) Feeding habits of an adult green sea turtle
Make Straightforward Inferences  -  1 Point  -  Key: C
------------------------------------------------------------------------------------------------------------------------------------
                               |                  |           Percentages                  |           Point Biserials                   |       | Avg. Score  |
 Country                       | Cases  DIFF  DISC |  P_A   P_B   P_C   P_D  P_OM  P_NR |  PB_A   PB_B   PB_C   PB_D  PB_OM  PB_NR | RDIFF | Girls  Boys |  Flags
------------------------------------------------------------------------------------------------------------------------------------
 Australia                     |  1043  57.0  0.52 | 11.0  17.5  57.0  11.5   2.9   1.8 | -0.21  -0.25   0.52  -0.23  -0.15  -0.18 |  0.78 |  59.1  54.7 | __E____
 Austria                       |   731  53.3  0.53 | 14.3  14.8  53.3  14.7   2.9   1.4 | -0.17  -0.25   0.53  -0.27  -0.13  -0.14 |  0.82 |  52.6  53.8 | _______
 Azerbaijan                    |   993  32.2  0.32 | 24.8  24.9  32.2  14.6   3.4   2.3 | -0.10  -0.06   0.32  -0.21  -0.02  -0.08 |  1.14 |  32.5  31.9 | __H____
 Bahrain                       |   905  29.4  0.31 | 22.7  24.2  29.4  18.9   4.8   3.5 | -0.09  -0.14   0.31  -0.07  -0.06  -0.00 |  0.96 |  28.8  30.1 | _______
 Belgium (Flemish)             |   855  53.5  0.44 | 12.9  17.5  53.5  13.0   3.2   1.8 | -0.17  -0.18   0.44  -0.23  -0.09  -0.08 |  0.71 |  55.1  51.6 | __E____
 Belgium (French)              |   773  46.4  0.47 | 16.0  15.9  46.4  16.3   5.3   3.2 | -0.19  -0.19   0.47  -0.18  -0.12  -0.06 |  0.61 |  48.1  44.6 | __E____
*Bulgaria                      |   704  71.7  0.48 |  7.3  11.1  71.7   9.4   0.6   0.3 | -0.22  -0.27   0.48  -0.25  -0.04  -0.08 |  0.16 |  72.5  70.8 | __E_F__
*Canada                        |  2980  54.0  0.48 | 12.9  15.5  54.0  13.4   4.3   4.1 | -0.18  -0.23   0.48  -0.20  -0.14  -0.17 |  0.82 |  55.8  52.2 | __E____
 Chile                         |   715  30.8  0.41 | 19.8  26.4  30.8  19.7   3.2   4.8 | -0.17  -0.14   0.41  -0.13  -0.06  -0.06 |  1.49 |  29.6  32.0 | __H____
 Chinese Taipei                |   719  46.6  0.40 | 19.6  21.5  46.6  11.5   0.8   0.4 | -0.22  -0.08   0.40  -0.23  -0.03  -0.02 |  1.44 |  47.8  45.5 | __H____
*Czech Republic                |   906  53.5  0.53 | 15.3  14.3  53.5  12.4   4.7   0.9 | -0.25  -0.18   0.53  -0.27  -0.13  -0.05 |  1.07 |  56.7  50.3 | __H____
 Denmark                       |   557  59.2  0.57 | 14.8  10.1  59.2  12.6   3.3   2.0 | -0.17  -0.32   0.57  -0.31  -0.11  -0.12 |  0.67 |  62.3  55.8 | __E____
*England                       |   846  56.7  0.55 | 15.4  14.1  56.7  11.1   2.7   0.8 | -0.25  -0.23   0.55  -0.25  -0.13  -0.07 |  0.99 |  57.3  56.1 | _______
 Finland                       |   811  59.1  0.56 |  9.0  15.3  59.1  13.6   3.0   0.2 | -0.21  -0.25   0.56  -0.30  -0.16  -0.10 |  1.05 |  62.8  55.5 | __H_F_G
*France                        |   783  48.8  0.50 | 15.6  17.2  48.8  13.4   5.1   4.0 | -0.21  -0.22   0.50  -0.19  -0.12  -0.10 |  0.56 |  49.7  47.8 | __E____
 Georgia                       |   924  43.4  0.45 | 16.0  23.5  43.4  13.1   4.0   4.2 | -0.18  -0.18   0.45  -0.17  -0.11  -0.10 |  0.48 |  45.5  41.1 | __E____
*Germany                       |   654  62.2  0.55 | 10.1   9.8  62.2  13.2   4.8   1.4 | -0.27  -0.24   0.55  -0.30  -0.07  -0.12 |  0.54 |  62.2  62.2 | __E_F__
*Hong Kong SAR                 |   561  53.2  0.39 | 18.2  16.8  53.2  10.2   1.6   0.2 | -0.12  -0.23   0.39  -0.13  -0.18   0.00 |  1.30 |  55.2  51.2 | __H____
*Hungary                       |   762  65.8  0.54 | 12.9  11.5  65.8   7.9   1.8   0.5 | -0.27  -0.26   0.54  -0.25  -0.13  -0.02 |  0.59 |  69.2  62.4 | __E_F_G
*Iran, Islamic Rep. of         |   725  26.7  0.20 | 25.3  16.6  26.7  27.3   4.2   3.9 |  0.02  -0.05   0.20  -0.13  -0.08  -0.05 |  0.74 |  25.7  27.7 | __E____
 Ireland                       |   769  63.4  0.55 | 11.8  12.3  63.4  10.7   1.7   0.8 | -0.21  -0.27   0.55  -0.28  -0.17  -0.11 |  0.70 |  63.3  63.5 | __E____
*Israel                        |   674  37.8  0.44 | 17.1  23.3  37.8  17.1   4.7   6.2 | -0.19  -0.17   0.44  -0.09  -0.16  -0.18 |  1.49 |  38.5  37.1 | __H____
*Italy                         |   653  54.5  0.50 | 10.8  15.0  54.5  17.5   2.2   1.1 | -0.20  -0.31   0.50  -0.16  -0.12  -0.13 |  1.04 |  57.8  51.5 | __H____
 Kazakhstan                    |   818  40.6  0.47 | 12.0  23.4  40.6  22.3   1.7   0.2 | -0.11  -0.23   0.47  -0.21  -0.07  -0.03 |  1.38 |  42.8  38.7 | __H____
*Latvia                        |   678  58.8  0.52 | 11.7  15.9  58.8  11.3   2.4   0.4 | -0.23  -0.22   0.52  -0.28  -0.10  -0.08 |  0.95 |  60.9  56.4 | _______
*Lithuania                     |   709  56.9  0.53 | 13.5  12.5  56.9  14.8   2.4   0.6 | -0.25  -0.18   0.53  -0.27  -0.15  -0.07 |  0.70 |  58.7  55.1 | __E____
 Macao SAR                     |   678  57.8  0.49 | 18.4  12.8  57.8   9.8   1.2   0.7 | -0.29  -0.20   0.49  -0.19  -0.05  -0.02 |  0.73 |  55.6  59.7 | __E_F__
 Malta                         |   593  25.2  0.32 | 24.4  24.4  25.2  20.8   5.3   1.0 | -0.08  -0.15   0.32  -0.05  -0.11  -0.03 |  0.82 |  29.4  21.0 | ______G
*Morocco                       |   904  25.3  0.22 | 23.8  25.6  25.3  18.8   6.5   6.3 | -0.05  -0.07   0.22  -0.03  -0.15  -0.03 |  0.01 |  23.2  27.3 | __E____
*Netherlands                   |   682  59.6  0.56 | 11.6  13.7  59.6  13.1   2.1   1.3 | -0.22  -0.27   0.56  -0.32  -0.05  -0.05 |  0.63 |  61.4  57.9 | __E____
*New Zealand                   |   909  54.9  0.49 | 13.9  14.2  54.9  13.8   3.1   2.2 | -0.22  -0.23   0.49  -0.19  -0.13  -0.13 |  0.53 |  54.2  55.6 | __E____
 Northern Ireland              |   602  58.3  0.43 | 10.5  14.2  58.3  14.4   2.7   0.5 | -0.25  -0.25   0.43  -0.13  -0.03  -0.01 |  1.08 |  58.2  58.3 | __H____
*Norway                        |   682  58.8  0.52 | 11.7  12.8  58.8  14.3   2.4   1.3 | -0.15  -0.26   0.52  -0.30  -0.09  -0.11 |  0.94 |  62.6  55.1 | ______G
 Oman                          |  1524  26.5  0.20 | 22.8  23.4  26.5  21.0   6.3   3.5 | -0.00  -0.10   0.20  -0.04  -0.10  -0.07 |  0.84 |  25.0  28.1 | _______
 Poland                        |   720  45.0  0.53 | 18.0  14.9  45.0  20.1   2.0   0.4 | -0.20  -0.24   0.53  -0.23  -0.09  -0.01 |  1.67 |  48.0  41.7 | __H____
 Portugal                      |   768  38.0  0.43 | 17.3  24.8  38.0  15.8   4.0   3.0 | -0.17  -0.20   0.43  -0.13  -0.07  -0.07 |  1.34 |  38.1  37.9 | __H____
 Qatar                         |  1475  30.2  0.38 | 21.4  25.9  30.2  17.6   5.0   3.9 | -0.05  -0.18   0.38  -0.14  -0.10  -0.09 |  0.98 |  30.9  29.4 | _______
*Russian Federation            |   770  63.8  0.52 |  8.6  11.1  63.8  14.1   2.3   0.5 | -0.19  -0.29   0.52  -0.23  -0.16  -0.08 |  1.08 |  67.1  60.4 | __H_F__
 Saudi Arabia                  |   783  27.3  0.30 | 22.6  26.9  27.3  17.7   5.6   6.0 | -0.18  -0.11   0.30   0.02  -0.08  -0.09 |  0.92 |  30.6  23.8 | ______G
*Singapore                     |  1075  63.0  0.50 | 10.8  15.7  63.0   9.0   1.5   0.4 | -0.24  -0.28   0.50  -0.21  -0.07  -0.11 |  0.84 |  65.1  60.7 | ____F__
*Slovak Republic               |   904  52.6  0.49 | 14.1  14.5  52.6  15.9   2.9   0.7 | -0.19  -0.14   0.49  -0.31  -0.10  -0.09 |  0.85 |  52.0  53.0 | _______
*Slovenia                      |   748  55.5  0.50 |  9.9  13.7  55.5  17.5   3.4   1.5 | -0.19  -0.25   0.50  -0.24  -0.06  -0.10 |  0.78 |  52.8  58.0 | ____F__
 Spain                         |  2408  41.7  0.47 | 17.6  16.3  41.7  22.5   1.9   0.9 | -0.15  -0.14   0.47  -0.28  -0.06  -0.04 |  1.38 |  41.8  41.5 | __H____
*Sweden                        |   751  61.7  0.53 | 12.4  10.7  61.7  12.2   3.0   2.0 | -0.22  -0.23   0.53  -0.30  -0.10  -0.18 |  0.80 |  63.5  59.9 | _______
 Trinidad and Tobago           |   682  37.8  0.49 | 18.8  22.0  37.8  16.7   4.7   5.4 | -0.16  -0.25   0.49  -0.13  -0.11  -0.08 |  0.80 |  37.7  38.0 | _______
 United Arab Emirates          |  2730  37.7  0.43 | 18.0  22.6  37.7  16.4   5.3   2.1 | -0.13  -0.20   0.43  -0.11  -0.15  -0.09 |  0.83 |  39.8  35.9 | ______G
*United States                 |   711  49.1  0.50 | 14.3  18.4  49.1  15.3   2.9   1.5 | -0.23  -0.25   0.50  -0.17  -0.09  -0.11 |  1.13 |  48.3  50.2 | __H____
------------------------------------------------------------------------------------------------------------------------------------
*Reference Avg.      (23)      | 19771  54.1  0.48 | 13.8  14.9  54.1  14.0   3.1   1.8 | -0.20  -0.22   0.48  -0.22  -0.11  -0.09 |  0.80 |  55.2  53.0 | ______G
 International Avg.  (47)      | 43347  48.6  0.46 | 15.6  17.4  48.6  15.1   3.3   2.0 | -0.18  -0.20   0.46  -0.20  -0.10  -0.08 |  0.90 |  49.7  47.5 | ______G
------------------------------------------------------------------------------------------------------------------------------------
 Buenos Aires, Argentina       |   704  34.3  0.47 | 20.4  17.1  34.3  20.1   8.0  11.1 | -0.18  -0.17   0.47  -0.13  -0.11  -0.09 |  0.99 |  35.6  33.1 | _______
 Ontario, Canada               |   692  50.7  0.45 | 11.5  18.4  50.7  14.4   5.0   4.8 | -0.13  -0.26   0.45  -0.13  -0.16  -0.16 |  0.96 |  54.1  47.3 | _______
 Quebec, Canada                |   523  58.6  0.43 | 13.1  14.7  58.6  10.8   2.8   4.0 | -0.15  -0.20   0.43  -0.25  -0.08  -0.17 |  0.89 |  59.8  57.2 | _______
 Norway (4)                    |   692  49.6  0.50 | 19.7  13.6  49.6  14.4   2.7   4.5 | -0.21  -0.19   0.50  -0.23  -0.12  -0.13 |  0.74 |  50.1  49.1 | __E____
 Moscow City, Russian Fed.     |   718  66.2  0.52 |  8.8   8.5  66.2  14.9   1.5   0.1 | -0.24  -0.26   0.52  -0.27  -0.08  -0.03 |  1.39 |  71.4  61.3 | __H_F_G
 Eng/Afr/Zulu - RSA (5)        |   854  25.4  0.38 | 21.3  25.5  25.4  23.1   4.6   3.3 | -0.13  -0.11   0.38  -0.09  -0.12  -0.05 |  0.55 |  28.0  22.5 | __E____
 Andalusia, Spain              |   687  37.7  0.47 | 17.2  17.2  37.7  26.1   1.8   0.9 | -0.12  -0.09   0.47  -0.32  -0.05  -0.06 |  1.31 |  35.9  39.6 | __H____
 Madrid, Spain                 |   628  47.8  0.48 | 16.0  15.7  47.8  18.7   1.8   0.5 | -0.21  -0.17   0.48  -0.23  -0.11  -0.08 |  1.35 |  47.7  48.0 | __H____
 Abu Dhabi, UAE                |   683  32.4  0.35 | 19.0  23.3  32.4  17.8   7.5   2.8 | -0.10  -0.12   0.35  -0.09  -0.14  -0.11 |  0.64 |  33.7  31.3 | __E____
 Dubai, UAE                    |  1320  44.4  0.46 | 17.0  20.8  44.4  14.6   3.2   1.3 | -0.16  -0.27   0.46  -0.11  -0.12  -0.07 |  0.97 |  46.8  42.3 | _______
------------------------------------------------------------------------------------------------------------------------------------
Keys:  DIFF= Percent correct score; DISC= Item discrimination; P_A...P_D= Percentage choosing each option; P_OM, P_NR= Percentage Omitted, Not Reached;
       PB_A...PB_D= Point Biserial for each option; PB_OM, PB_NR= Point Biserial for Omitted, Not Reached; RDIFF= Rasch difficulty.
Flags: A= Attractive distractor; B= Boys outperform girls; C= Difficulty less than chance; D= Negative/low discrimination; E= Easier than average;
       F= Distractor chosen by less than 10%; G= Girls outperform boys; H= Harder than average; R= Scoring reliability less than 85%; V= Difficulty greater
       than 95%.
```

Exhibit 10.2: Example International Item Statistics for a PIRLS 2016 Constructed Response Item



```
Progress in International Reading Literacy Study - PIRLS 2016 Assessment Results
International Item Review Statistics (Unweighted)

Literary Experience - Oliver and The Griffin (R41O13C) Describe why griffin can not do his job
Interpret and Integrate Ideas and Information  -  3 Points
-------------------------------------------------------------------------------------------------------------------------------------------------------------
                           |                   |            Percentages              |              Point Biserials                 |       | Reliability|  Avg. Score |
 Country                   | Cases  DIFF  DISC |  P_0   P_1   P_2   P_3  P_OM  P_NR  |  PB_0   PB_1   PB_2   PB_3  PB_OM  PB_NR | RDIFF |   N    Agr  | Girls Boys  |  Flags
-------------------------------------------------------------------------------------------------------------------------------------------------------------
 Australia                 |  1047  65.4  0.70 | 10.4  19.0  19.1  46.4   5.1   3.4 | -0.43  -0.24   0.06   0.56  -0.35  -0.18 |  0.33 |  224  85.3 |  70.4  60.9 | __E___G
 Austria                   |   726  49.6  0.61 | 29.1   6.5  24.6  31.0   8.8   1.9 | -0.41   0.01   0.10   0.50  -0.31  -0.10 |  1.16 |  218  97.2 |  48.6  50.5 | __H_F__
 Azerbaijan                |   994  14.8  0.48 | 53.8  10.3   8.4   5.7  21.8   3.4 | -0.13   0.17   0.28   0.29  -0.33  -0.12 |  1.73 |  241  97.5 |  15.9  13.8 | __H_F__
 Bahrain                   |   911  22.0  0.65 | 54.8   7.6  10.5  12.5  14.6   7.7 | -0.40   0.08   0.29   0.51  -0.22  -0.13 |  0.88 |  239  91.6 |  24.4  19.6 | __E_F__
 Belgium (Flemish)         |   860  46.0  0.61 | 35.6   6.8  19.3  30.8   7.5   2.3 | -0.40  -0.06   0.15   0.51  -0.32  -0.10 |  0.96 |  256  93.0 |  48.8  43.0 | ____F__
 Belgium (French)          |   769  38.3  0.59 | 33.8   9.1  20.3  21.7  15.1   5.9 | -0.33   0.01   0.24   0.43  -0.33  -0.15 |  0.91 |  225  97.8 |  39.1  37.4 | __E_F__
*Bulgaria                  |   712  47.2  0.63 | 32.8   6.2  26.7  27.3   7.0   2.2 | -0.41  -0.05   0.13   0.52  -0.33  -0.13 |  1.44 |  201  92.0 |  49.8  44.5 | __H_F__
*Canada                    |  2996  56.3  0.65 | 27.6   8.2  14.3  44.0   5.8   6.3 | -0.47  -0.06   0.07   0.56  -0.30  -0.21 |  0.78 |  456  80.7 |  60.4  52.2 | __E_FRG
 Chile                     |   712  42.3  0.66 | 38.0   9.7  17.7  27.2   7.3   6.2 | -0.47  -0.04   0.20   0.53  -0.28  -0.12 |  0.84 |  207  96.1 |  45.6  39.4 | __E_F__
 Chinese Taipei            |   722  48.7  0.64 | 28.2   4.1  19.0  34.6  14.1   1.7 | -0.40  -0.01   0.13   0.52  -0.34  -0.11 |  1.40 |  213  97.7 |  51.2  46.1 | __H_F__
*Czech Republic            |   923  42.2  0.60 | 31.4   8.2  18.8  26.9  14.6   2.7 | -0.29  -0.01   0.18   0.47  -0.41  -0.14 |  1.38 |  215 100.0 |  43.8  40.7 | __H_F__
 Denmark                   |   593  47.7  0.59 | 33.3   8.5  22.3  30.0   5.9   4.9 | -0.42  -0.08   0.14   0.48  -0.26  -0.18 |  1.26 |  182  86.3 |  50.8  44.4 | __H_F__
*England                   |   849  62.6  0.67 | 24.2   5.7  13.5  51.7   5.0   2.5 | -0.49  -0.05  -0.03   0.61  -0.35  -0.17 |  0.82 |  245  94.7 |  65.2  59.7 | __E_F__
 Finland                   |   807  61.4  0.59 | 22.6   7.2  15.2  48.8   6.1   2.5 | -0.39  -0.09   0.00   0.53  -0.34  -0.17 |  1.11 |  229  93.9 |  63.1  59.8 | __H_F__
*France                    |   795  42.8  0.65 | 33.2   6.4  14.2  31.2  15.0   7.7 | -0.40  -0.00   0.12   0.57  -0.32  -0.18 |  0.82 |  232  96.1 |  44.1  41.4 | __E_F__
 Georgia                   |   954  30.1  0.66 | 41.4   9.0  14.4  17.5  17.7   8.9 | -0.33   0.12   0.23   0.52  -0.40  -0.22 |  1.19 |  206  81.6 |  33.9  26.3 | __H_FRG
*Germany                   |   649  49.7  0.67 | 17.7  20.7  23.0  27.4  11.2   2.3 | -0.41  -0.09   0.17   0.51  -0.33  -0.16 |  1.06 |  228  90.8 |  51.3  48.1 | _______
*Hong Kong SAR             |   552  56.4  0.59 | 19.9   9.6  18.4  40.9  11.2   1.6 | -0.30  -0.17   0.11   0.49  -0.36   0.01 |  1.25 |  170  92.4 |  57.6  55.3 | __H_F__
*Hungary                   |   775  57.1  0.62 | 21.3  16.8  19.9  38.2   3.8   1.4 | -0.43  -0.11   0.04   0.52  -0.26  -0.15 |  0.96 |  210  94.3 |  57.1  57.0 | _______
*Iran, Islamic Rep. of     |   735  31.4  0.70 | 35.2  14.5  13.2  17.8  19.3  13.5 | -0.40   0.08   0.28   0.53  -0.33  -0.23 |  0.76 |  215  92.1 |  34.1  28.5 | __E____
 Ireland                   |   762  66.0  0.60 | 21.8   4.3  18.2  52.4   3.3   1.8 | -0.43  -0.10  -0.03   0.53  -0.32  -0.09 |  0.92 |  209  98.1 |  68.0  64.1 | __E_F__
*Israel                    |   660  45.9  0.69 | 32.1   5.5  16.2  33.2  13.0   9.2 | -0.46   0.01   0.22   0.54  -0.36  -0.23 |  1.14 |  178  98.3 |  47.3  44.3 | __H_F__
*Italy                     |   661  62.9  0.58 | 20.5   5.1  16.4  50.2   7.8   2.4 | -0.33  -0.06  -0.02   0.52  -0.40  -0.22 |  0.85 |  188  94.1 |  65.9  60.0 | __E_F__
 Kazakhstan                |   818  51.1  0.64 | 19.0  24.8  28.3  24.0   4.0   1.3 | -0.45  -0.13   0.16   0.47  -0.20  -0.15 |  0.85 |  215  97.2 |  53.7  48.6 | __E___G
*Latvia                    |   702  49.2  0.55 | 33.4  10.2  18.5  33.4   4.4   0.7 | -0.37  -0.12   0.13   0.46  -0.28  -0.08 |  1.36 |  214  94.4 |  52.6  45.8 | __H___G
*Lithuania                 |   725  41.1  0.58 | 38.5   7.3  19.2  25.8   9.1   1.8 | -0.32  -0.05   0.11   0.51  -0.34  -0.12 |  1.48 |  220  99.1 |  40.0  42.3 | __H_F__
 Macao SAR                 |   673  43.7  0.60 | 31.5   5.3  17.0  30.6  15.5   2.4 | -0.32  -0.01   0.15   0.49  -0.37  -0.11 |  1.33 |  219  99.1 |  43.1  44.4 | __H_F__
 Malta                     |   604  27.1  0.56 | 50.4   6.5  11.7  17.2  14.2   5.5 | -0.25   0.02   0.14   0.51  -0.33  -0.16 |  0.74 |  191  85.3 |  29.4  24.8 | __E_F__
*Morocco                   |   909  11.5  0.65 | 55.8   7.7   5.1   5.6  25.8  13.1 | -0.31   0.24   0.28   0.50  -0.19  -0.06 |  0.36 |  217  84.8 |  13.8   9.2 | __E_FRG
*Netherlands               |   705  52.0  0.58 | 34.9   7.0  18.1  37.6   2.3   3.0 | -0.47  -0.09   0.05   0.52  -0.17  -0.14 |  1.11 |  161  96.9 |  53.6  50.4 | __H_F__
*New Zealand               |   942  56.2  0.68 | 31.3   5.5  13.5  45.4   4.4   5.0 | -0.52  -0.06   0.09   0.58  -0.32  -0.24 |  0.72 |  248  94.8 |  61.0  51.1 | __E_F_G
 Northern Ireland          |   608  58.4  0.58 | 28.0   5.2  15.9  46.1   4.9   1.8 | -0.39  -0.10  -0.02   0.54  -0.29  -0.18 |  1.02 |  174 100.0 |  60.2  56.7 | ____F__
*Norway                    |   719  59.3  0.61 | 22.9   7.5  21.9  42.2   5.5   2.1 | -0.38  -0.14   0.02   0.54  -0.34  -0.17 |  1.13 |  213  96.7 |  62.4  56.4 | __H_F__
 Oman                      |  1546  16.6  0.66 | 57.3  11.7   8.2   7.2  15.5   8.5 | -0.45   0.21   0.35   0.45  -0.15  -0.13 |  0.97 |  210  91.9 |  20.3  13.0 | ____F_G
 Poland                    |   725  60.2  0.69 | 14.1  13.8  33.8  33.1   5.2   1.2 | -0.48  -0.24   0.15   0.49  -0.25  -0.22 |  1.12 |  183  83.1 |  63.0  57.8 | _____R_
 Portugal                  |   763  43.0  0.59 | 38.3   3.4  19.4  28.9  10.0   8.3 | -0.39  -0.04   0.12   0.51  -0.27  -0.16 |  1.14 |  235  98.7 |  44.2  41.6 | __H_F__
 Qatar                     |  1497  26.6  0.72 | 50.6   8.8  12.1  15.6  12.9   6.9 | -0.52   0.05   0.29   0.58  -0.17  -0.12 |  0.76 |  224  97.3 |  29.7  23.2 | __E_F_G
*Russian Federation        |   771  44.1  0.54 | 42.4   6.1  17.1  30.7   3.8   1.4 | -0.39  -0.08   0.07   0.49  -0.22  -0.09 |  1.98 |  206  97.1 |  46.3  41.8 | __H_F__
 Saudi Arabia              |   766  23.3  0.63 | 39.3  21.3  12.8   7.7  18.9  10.1 | -0.31   0.16   0.32   0.40  -0.33  -0.11 |  0.72 |  199  98.5 |  28.8  17.5 | __E_F_G
*Singapore                 |  1081  68.0  0.68 | 22.1   2.1  16.6  56.2   2.9   0.9 | -0.57  -0.08   0.00   0.58  -0.23  -0.10 |  0.89 |  239 100.0 |  70.4  65.9 | __E_F__
*Slovak Republic           |   910  46.1  0.59 | 37.6   5.4  18.5  32.0   6.4   1.0 | -0.40   0.01   0.14   0.47  -0.34  -0.07 |  1.23 |  219  95.9 |  45.1  47.1 | __H_F__
*Slovenia                  |   748  49.6  0.64 | 31.3   6.4  19.9  34.2   8.3   3.3 | -0.41  -0.06   0.12   0.53  -0.34  -0.19 |  1.11 |  225  99.6 |  50.6  48.4 | ____F__
 Spain                     |  2427  52.4  0.60 | 30.7   4.7  23.4  35.2   6.0   2.6 | -0.41  -0.06   0.13   0.48  -0.32  -0.14 |  0.92 |  402  98.3 |  52.7  52.0 | __E_F__
*Sweden                    |   748  58.8  0.60 | 25.6   4.2  15.5  47.1   7.7   4.3 | -0.39  -0.08   0.06   0.52  -0.35  -0.19 |  1.10 |  182  94.5 |  59.2  58.3 | ____F__
 Trinidad and Tobago       |   702  36.3  0.67 | 43.5   6.9  14.0  24.6  11.0   9.3 | -0.48   0.02   0.16   0.58  -0.24  -0.23 |  0.85 |  196  89.3 |  37.7  34.7 | __E_F__
 United Arab Emirates      |  2750  30.0  0.71 | 51.0   5.7  11.4  20.5  11.3   5.8 | -0.52   0.08   0.23   0.59  -0.22  -0.16 |  0.85 |  687  88.4 |  32.0  28.1 | __E_F_G
*United States             |   745  63.0  0.65 | 24.6   6.6  16.3  49.9   2.5   4.7 | -0.53  -0.12   0.03   0.56  -0.23  -0.24 |  0.76 |  197  94.4 |  65.8  60.3 | __E_F__
-------------------------------------------------------------------------------------------------------------------------------------------------------------
*Reference Avg.      (23)  | 20012  50.1  0.63 | 30.3   7.9  17.2  36.0   8.6   4.1 | -0.41  -0.05   0.10   0.53  -0.31  -0.15 |  1.07 | 5079  94.5 |  52.1  48.2 | ____F_G
 International Avg.  (47)  | 43748  45.8  0.63 | 33.0   8.6  17.3  31.5   9.7   4.4 | -0.40  -0.03   0.14   0.51  -0.30  -0.15 |  1.03 |10863  94.0 |  47.9  43.8 | ____F_G
-------------------------------------------------------------------------------------------------------------------------------------------------------------
 Buenos Aires, Argentina   |   720  34.3  0.60 | 39.0   7.5   8.9  25.9  18.7  11.4 | -0.32   0.01   0.13   0.54  -0.31  -0.26 |  0.84 |  200  95.0 |  33.6  35.0 | __E_F__
 Ontario, Canada           |   708  56.4  0.69 | 27.3   9.2  14.3  43.9   5.4   7.9 | -0.52  -0.06   0.04   0.60  -0.29  -0.25 |  0.64 |  108  81.5 |  61.3  51.5 | __E_FRG
 Quebec, Canada            |   526  58.4  0.61 | 25.1   7.5  15.0  45.8   6.5   3.8 | -0.42  -0.09   0.07   0.53  -0.32  -0.08 |  1.14 |  103  84.5 |  62.3  54.3 | __H_FRG
 Norway (4)                |   734  47.4  0.59 | 24.3   9.6  25.4  27.3  13.4   7.6 | -0.34  -0.04   0.15   0.46  -0.33  -0.26 |  0.98 |    .     . |  51.0  43.7 | ____F_G
 Moscow City, Russian Fed. |   714  58.0  0.54 | 29.6   3.4  22.8  41.6   2.5   0.7 | -0.38  -0.14  -0.02   0.50  -0.26  -0.09 |  1.87 |  219  97.7 |  60.7  55.2 | __H_F__
 Eng/Afr/Zulu - RSA (5)    |   874  16.6  0.68 | 67.5  12.6   7.4   7.4   5.0   9.2 | -0.52   0.17   0.32   0.51  -0.15  -0.09 |  0.79 |  208  84.6 |  19.3  13.7 | __E_FRG
 Andalusia, Spain          |   688  47.0  0.60 | 35.3   3.0  24.5  29.6   7.5   3.3 | -0.39  -0.04   0.18   0.46  -0.36  -0.16 |  1.08 |  200  98.0 |  47.1  46.8 | ____F__
 Madrid, Spain             |   636  57.9  0.62 | 28.5   2.4  24.6  40.6   3.8   1.7 | -0.50   0.02   0.11   0.47  -0.30  -0.13 |  1.01 |   86  98.8 |  58.2  57.6 | ____F__
 Abu Dhabi, UAE            |   701  20.4  0.67 | 55.7   6.2   8.3  12.8  17.0   7.6 | -0.41   0.13   0.22   0.57  -0.20  -0.13 |  0.82 |  211  88.2 |  24.3  16.5 | __E_F_G
 Dubai, UAE                |  1311  42.5  0.70 | 41.0   5.5  15.0  30.7   7.8   3.5 | -0.54   0.02   0.17   0.59  -0.25  -0.18 |  0.87 |  259  88.0 |  43.6  41.5 | __E_F__
-------------------------------------------------------------------------------------------------------------------------------------------------------------
Keys:  DIFF= Percent correct score; DISC= Item discrimination; P_0...P_3= Percentage obtaining score level; P_OM, P_NR= Percentage Omitted, Not Reached;
       PB_0...PB_3= Point Biserial for score level; PB_OM, PB_NR= Point Biserial for Omitted, Not Reached; RDIFF= Rasch difficulty;
       Reliability: N= Responses double scored; Agr= Percentage agreement.
Flags: A= Point-biserial not ordered; B= Boys outperform girls; C= Difficulty less than chance; D= Negative/low discrimination; E= Easier than average;
```

For all items, regardless of format (i.e., multiple-choice or constructed response), statistics included 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 total score).[1] Also provided was an estimate of the difficulty of the item using a Rasch one-parameter IRT model. Statistics for each item were displayed alphabetically by country, together with an international average—i.e., based on all participating countries listed above the international average—and a reference average—based on a pool of countries that have participated regularly in the PIRLS assessments—for each statistic. The reference countries are shown with an asterisk next to their names. The international and reference averages of the item difficulties and item discriminations served as guides to the overall statistical properties of the items. The item review outputs also listed the benchmarking participants.

Statistics displayed for multiple-choice items included the percentage of students that chose each response option—as well as the percentage of students that omitted or did not reach the item—and the point-biserial correlations for each response option. Statistics displayed for constructed response items (which could have 1, 2, or 3 score points) included the percent correct and point-biserial of each score level. Constructed response item tables also provided information about the reliability with which each item was scored in each country, showing the total number of double-scored responses and the percentage of score agreement between the scorers.

During item review, "not reached" responses (i.e., items toward the end of the booklet that the student did not attempt)[2] were treated as "not administered" and thus did not contribute to the calculation of the item statistics. However, the percentage of students not reaching each item was reported. Omitted responses, although treated as incorrect, were tabulated separately from incorrect responses for the sake of distinguishing students who provided no form of response from students who attempted a response.

The definitions and detailed descriptions of the statistics that were calculated are given below. The statistics are listed in order of their appearance in the item review outputs:

CASES: This is the number of students to whom the item was administered. Not-reached responses were not included in this count.

DIFF: The item difficulty is the average percent correct on an item. For a 1-point item, including all multiple-choice items, it is the percentage of students providing a fully correct response to the item. For 2-point and 3-point items, it is the average percentage of points. For example, if 25 percent of students scored 2 points, 50 percent scored 1 point on a 2-point item, and the other 25 percent score 0 points, then the average percent correct for such an item would be 50 percent. For this statistic, not reached responses were not included.

[1] For computing point-biserial correlations, the total score is the percentage of points a student has scored on the items (s)he was administered. Not reached responses are not included in the total score.

[2] An item was considered "not reached" if the item itself and the item immediately preceding it were not answered and no subsequent items had been attempted. The decision as to whether an item was not reached was made separately for part 1 and part 2 of each assessment booklet.

DISC: The item discrimination is computed as the correlation between the response to an item and the total score on all items administered to a student. Items exhibiting good measurement properties should have a moderately positive correlation, indicating that the more able students get the item right, the less able get it wrong. For this statistic, not reached items were not included.

PCT_A, PCT_B, PCT_C, and PCT_D: Available for multiple-choice items. Each column indicates the percentage of students choosing the particular response option for the item (A, B, C, or D). [3] Not reached responses were excluded from the denominator.

PCT_0, PCT_1, PCT_2, and PCT_3: Available for constructed response items. Each column indicates the percentage of students responding at that particular score level, up to and including the maximum score level for the item. Not reached items were excluded from the denominator.

PCT_OM: Percentage of students who, having reached the item, did not provide a response. Not reached responses were excluded from the denominator.

PCT_NR: Percentage of students who did not reach the item. This statistic is the number of students who did not reach an item as a percentage of all students who were administered that item, including those who omitted or did not reach that item.

PB_A, PB_B, PB_C, and PB_D: Available for multiple-choice items. These columns show the point-biserial correlations between choosing each of the response options (A, B, C, or D) and the total score on all of the items administered to a student. Items with good psychometric properties have moderately positive correlations for the correct option and negative correlations for the distracters (the incorrect options). Not reached responses were not included in these calculations.

PB_0, PB_1, PB_2, and PB_3: Available for constructed response items. These columns present the point-biserial correlations between the score levels on the item and the overall score on all of the items the student was administered. For items with good measurement properties, the correlation coefficients should monotonically increase from negative to positive as the score on the item increases. Not reached responses were not included in these calculations.

PB_OM: The point-biserial correlation between a binary variable indicating an omitted response to the item, and the total score on all items administered to a student. This correlation should be negative or near zero. Not reached responses were not included in this statistic.

[3] ePIRLS included multiple-choice items with as many as six response options, thus adding options E and F.

PB_NR: The point-biserial correlation between a binary variable indicating a not-reached response to the item, and the total score on all items administered to a student. This correlation should be negative or near zero.

RDIFF: An estimate of the difficulty of an item based on a Rasch one-parameter IRT model applied to the achievement data of a given country. The difficulty estimate is expressed in the logit metric (with a positive logit indicating a difficult item) and was scaled so that the average Rasch item difficulty across all items within each country was zero.

Reliability (N): To provide a measure of the reliability of the scoring of the constructed response items, items in approximately 25 percent of the test booklets in each country were independently scored by two scorers. This column indicates the number of responses that were double-scored for a given item in a country.

Reliability (Agr): This column contains the percentage of agreement on the scores assigned by the two independent PIRLS scorers.

As an aid to the reviewers, the item review displays included a series of flags signaling the presence of one or more conditions that might indicate a problem with an item. The following conditions were flagged:

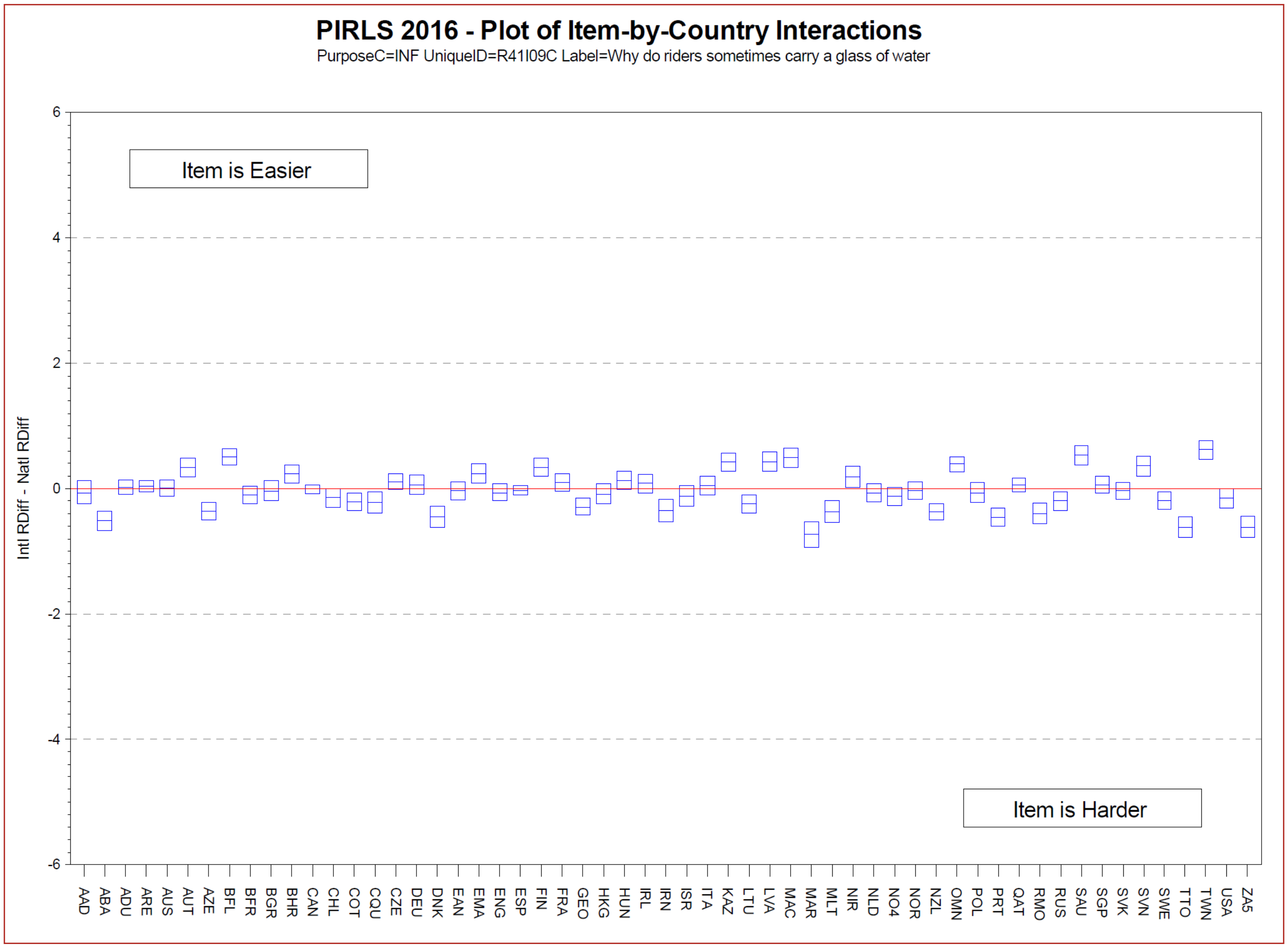
- The item discrimination (DISC) was less than 0.10 (flag D)
- The item difficulty (DIFF) was less than 25% for multiple-choice items (flag C)
- The item difficulty (DIFF) exceeded 95% (flag V)
- The Rasch difficulty estimate (RDIFF) for a given country made the item either easier (flag E) or more difficult (flag H) relative to the international average for that item
- The point-biserial correlation for at least one distracter in a multiple-choice item was positive, or the point-biserial correlations across the score levels of a constructed response item were not ordered (flag A)
- The percentage of students selecting one of the response options for a multiple-choice item, or one of the score values for a constructed response item, was less than 10% (flag F)
- Scoring reliability for agreement on the score value of a constructed response item was less than 85% (flag R)

Although not all of these conditions necessarily indicated a problem, the flags were a useful tool to draw attention to potential sources of concern.

Item–by–Country Interaction

Although countries are expected to exhibit some variation in performance across items, in general countries with high average performance on the assessment should perform relatively well on each of the items, and low-scoring countries should do less well on each of the items. When this does not occur (e.g., when a high-performing country has low performance on an item on which other countries are doing well), there is said to be an item-by-country interaction. When large, such item-by-country interactions may be a sign that an item is flawed in some way and that steps should be taken to address the problem. To assist in detecting sizeable item-by-country interactions, the TIMSS & PIRLS International Study Center produced a graphical display for each item showing the difference between each country's Rasch item difficulty and the international average Rasch item difficulty across all countries. An example of the graphical displays is provided in Exhibit 10.3.

Exhibit 10.3: Example Plot of Item-by-Country Interaction for a PIRLS 2016 Item



In each of these item-by-country interaction displays, the difference in Rasch item difficulty for each country is presented as a 95 percent confidence interval, which includes a built-in Bonferroni correction for multiple comparisons across the participating countries. The limits for this confidence interval were computed as follows:

$$\text{Upper Limit} = RDIFF_{i.} - RDIFF_{ik} + SE(RDIFF_{ik}) \cdot Z_b \qquad (10.1)$$

$$\text{Lower Limit} = RDIFF_{i.} - RDIFF_{ik} - SE(RDIFF_{ik}) \cdot Z_b \qquad (10.2)$$

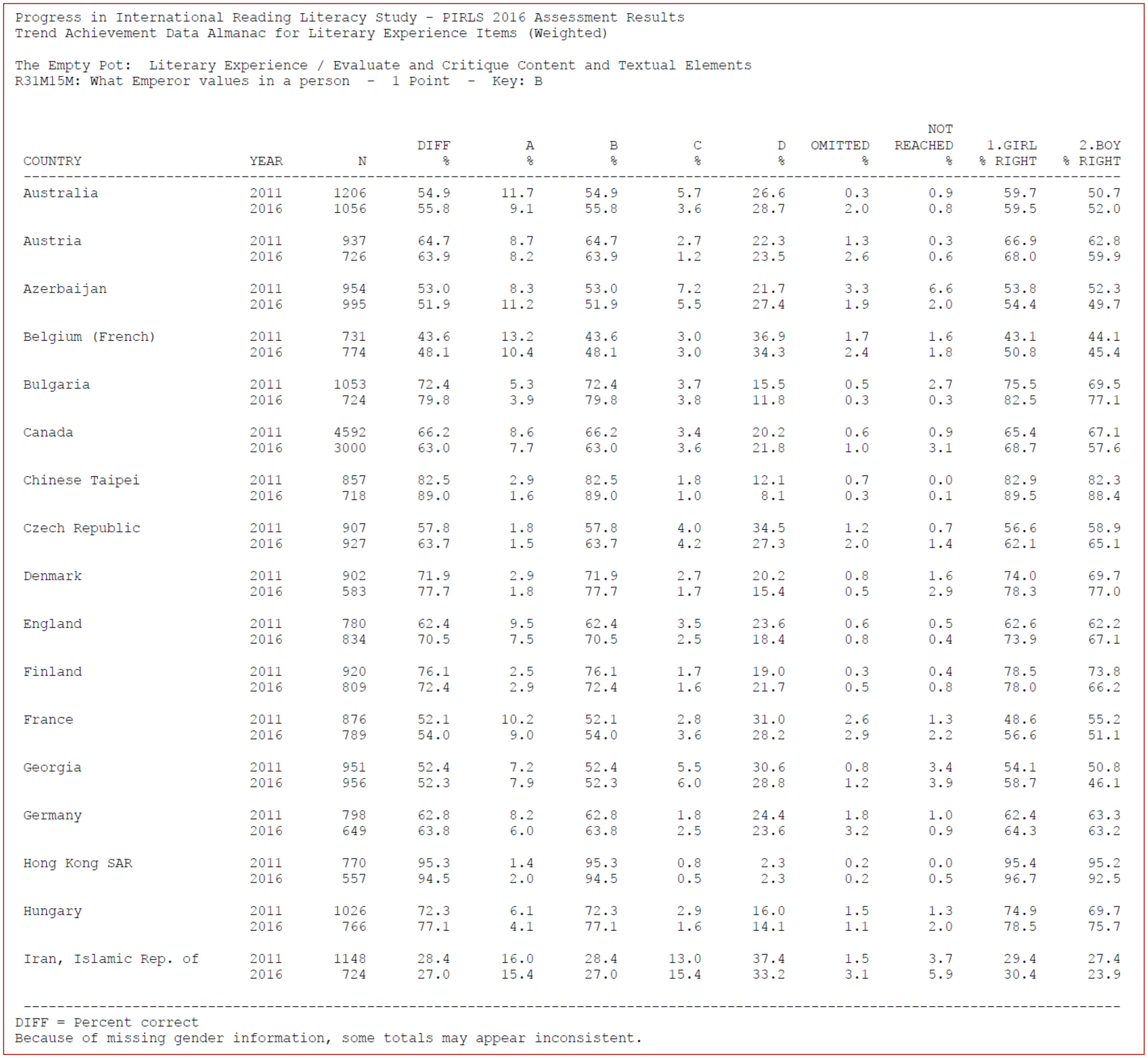
where $RDIFF_{ik}$ is the Rasch difficulty of item i in country k, $RDIFF_{i.}$ is the international average Rasch difficulty of item i, $SE(RDIFF_{ik})$ is the standard error of the Rasch difficulty of item i in country k, and Z_b is the 95% critical value from the Z distribution corrected for multiple comparisons using the Bonferroni procedure.

Trend Item Review

In order to measure trends, PIRLS 2016 included achievement items from previous assessments as well as items developed for use for the first time in 2016. Accordingly, the PIRLS 2016 assessments included items from 2001, 2006, 2011, and 2016. An important review step, therefore, was to check that these “trend items” had statistical properties in 2016 similar to those they had in the previous assessments (e.g., a PIRLS item that was relatively easy in 2011 should still be relatively easy in 2016).

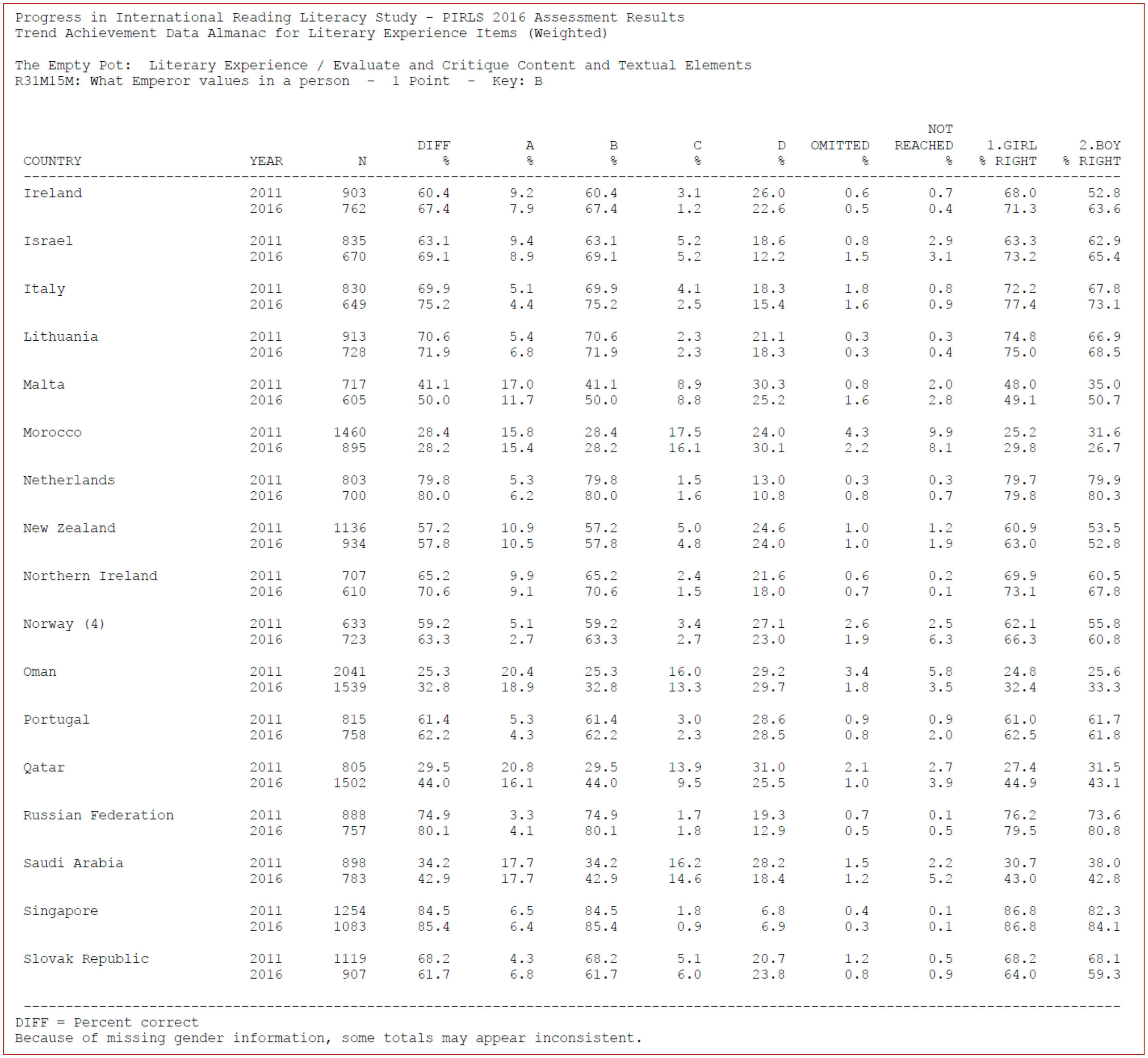
As can be seen in the example in Exhibit 10.4, the trend item review focused on statistics for trend items from the current and previous assessments (2016 and 2011) for countries that participated in both. For each country, trend item statistics included the percentage of students in each score category (or response option for multiple-choice items) for each assessment, as well as the difficulty of the item and the percent correct by gender. In reviewing these item statistics, the aim was to detect any unusual changes in item difficulties between administrations, which might indicate a problem in using the item to measure trends.

Exhibit 10.4: Example Item Statistics for a PIRLS 2016 Trend Item



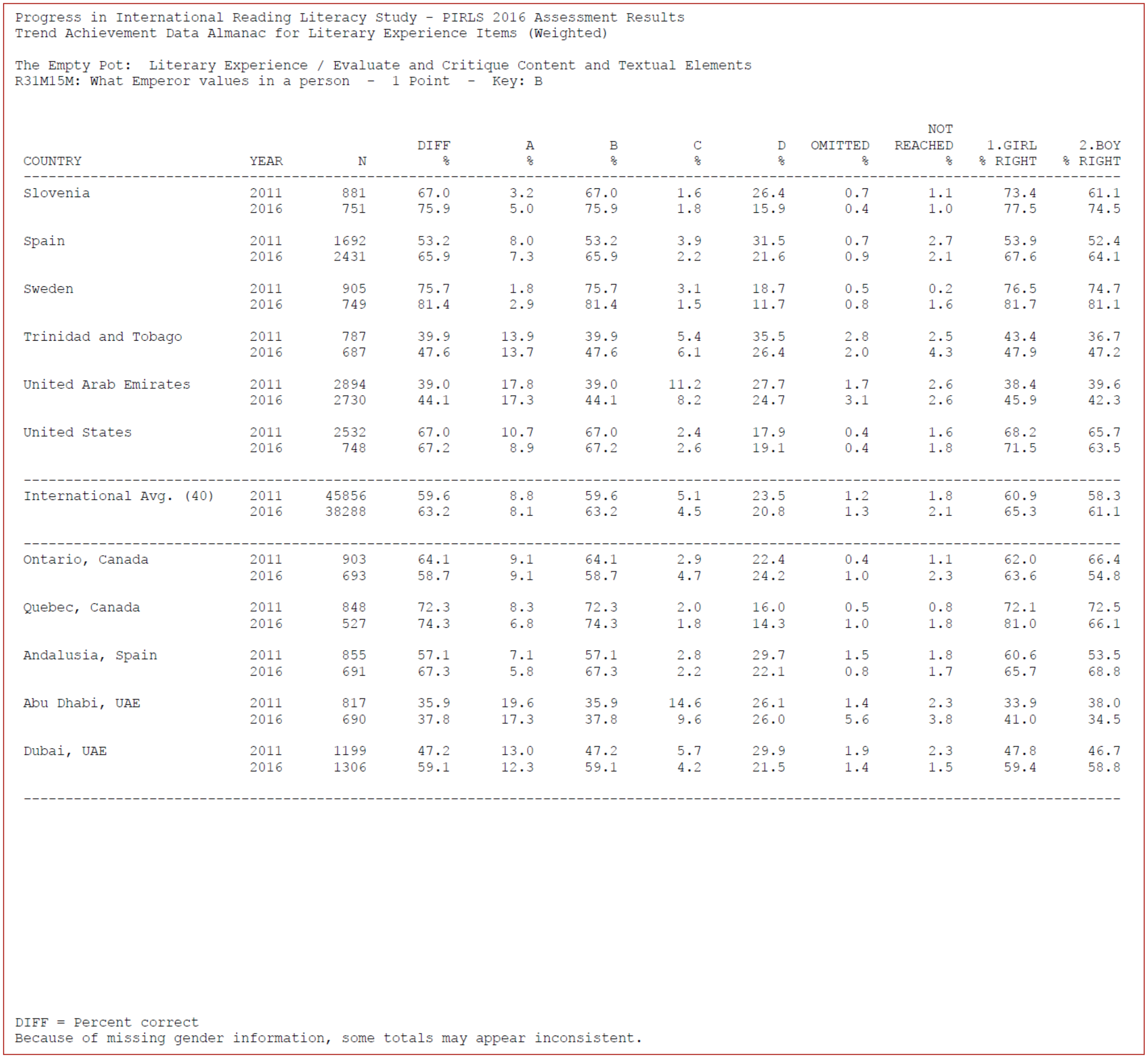
```
Progress in International Reading Literacy Study - PIRLS 2016 Assessment Results
Trend Achievement Data Almanac for Literary Experience Items (Weighted)

The Empty Pot:  Literary Experience / Evaluate and Critique Content and Textual Elements
R31M15M: What Emperor values in a person  -  1 Point  -  Key: B


                                                                                    NOT
                                            DIFF       A       B       C       D  OMITTED  REACHED   1.GIRL    2.BOY
 COUNTRY                    YEAR        N      %       %       %       %       %        %        %  % RIGHT  % RIGHT
 ----------------------------------------------------------------------------------------------------------------------
 Australia                  2011     1206   54.9    11.7    54.9     5.7    26.6      0.3      0.9     59.7     50.7
                            2016     1056   55.8     9.1    55.8     3.6    28.7      2.0      0.8     59.5     52.0

 Austria                    2011      937   64.7     8.7    64.7     2.7    22.3      1.3      0.3     66.9     62.8
                            2016      726   63.9     8.2    63.9     1.2    23.5      2.6      0.6     68.0     59.9

 Azerbaijan                 2011      954   53.0     8.3    53.0     7.2    21.7      3.3      6.6     53.8     52.3
                            2016      995   51.9    11.2    51.9     5.5    27.4      1.9      2.0     54.4     49.7

 Belgium (French)           2011      731   43.6    13.2    43.6     3.0    36.9      1.7      1.6     43.1     44.1
                            2016      774   48.1    10.4    48.1     3.0    34.3      2.4      1.8     50.8     45.4

 Bulgaria                   2011     1053   72.4     5.3    72.4     3.7    15.5      0.5      2.7     75.5     69.5
                            2016      724   79.8     3.9    79.8     3.8    11.8      0.3      0.3     82.5     77.1

 Canada                     2011     4592   66.2     8.6    66.2     3.4    20.2      0.6      0.9     65.4     67.1
                            2016     3000   63.0     7.7    63.0     3.6    21.8      1.0      3.1     68.7     57.6

 Chinese Taipei             2011      857   82.5     2.9    82.5     1.8    12.1      0.7      0.0     82.9     82.3
                            2016      718   89.0     1.6    89.0     1.0     8.1      0.3      0.1     89.5     88.4

 Czech Republic             2011      907   57.8     1.8    57.8     4.0    34.5      1.2      0.7     56.6     58.9
                            2016      927   63.7     1.5    63.7     4.2    27.3      2.0      1.4     62.1     65.1

 Denmark                    2011      902   71.9     2.9    71.9     2.7    20.2      0.8      1.6     74.0     69.7
                            2016      583   77.7     1.8    77.7     1.7    15.4      0.5      2.9     78.3     77.0

 England                    2011      780   62.4     9.5    62.4     3.5    23.6      0.6      0.5     62.6     62.2
                            2016      834   70.5     7.5    70.5     2.5    18.4      0.8      0.4     73.9     67.1

 Finland                    2011      920   76.1     2.5    76.1     1.7    19.0      0.3      0.4     78.5     73.8
                            2016      809   72.4     2.9    72.4     1.6    21.7      0.5      0.8     78.0     66.2

 France                     2011      876   52.1    10.2    52.1     2.8    31.0      2.6      1.3     48.6     55.2
                            2016      789   54.0     9.0    54.0     3.6    28.2      2.9      2.2     56.6     51.1

 Georgia                    2011      951   52.4     7.2    52.4     5.5    30.6      0.8      3.4     54.1     50.8
                            2016      956   52.3     7.9    52.3     6.0    28.8      1.2      3.9     58.7     46.1

 Germany                    2011      798   62.8     8.2    62.8     1.8    24.4      1.8      1.0     62.4     63.3
                            2016      649   63.8     6.0    63.8     2.5    23.6      3.2      0.9     64.3     63.2

 Hong Kong SAR              2011      770   95.3     1.4    95.3     0.8     2.3      0.2      0.0     95.4     95.2
                            2016      557   94.5     2.0    94.5     0.5     2.3      0.2      0.5     96.7     92.5

 Hungary                    2011     1026   72.3     6.1    72.3     2.9    16.0      1.5      1.3     74.9     69.7
                            2016      766   77.1     4.1    77.1     1.6    14.1      1.1      2.0     78.5     75.7

 Iran, Islamic Rep. of      2011     1148   28.4    16.0    28.4    13.0    37.4      1.5      3.7     29.4     27.4
                            2016      724   27.0    15.4    27.0    15.4    33.2      3.1      5.9     30.4     23.9

 ----------------------------------------------------------------------------------------------------------------------
DIFF = Percent correct
Because of missing gender information, some totals may appear inconsistent.
```

Exhibit 10.4: Example Item Statistics for a PIRLS 2016 Trend Item (Continued)



```
Progress in International Reading Literacy Study - PIRLS 2016 Assessment Results
Trend Achievement Data Almanac for Literary Experience Items (Weighted)

The Empty Pot:  Literary Experience / Evaluate and Critique Content and Textual Elements
R31M15M: What Emperor values in a person  -  1 Point  -  Key: B


                                                                                          NOT
                                          DIFF         A         B         C         D  OMITTED   REACHED    1.GIRL     2.BOY
 COUNTRY                   YEAR       N      %         %         %         %         %        %         %   % RIGHT   % RIGHT
 ---------------------------------------------------------------------------------------------------------------------------------
 Ireland                   2011     903   60.4       9.2      60.4       3.1      26.0      0.6       0.7      68.0      52.8
                           2016     762   67.4       7.9      67.4       1.2      22.6      0.5       0.4      71.3      63.6

 Israel                    2011     835   63.1       9.4      63.1       5.2      18.6      0.8       2.9      63.3      62.9
                           2016     670   69.1       8.9      69.1       5.2      12.2      1.5       3.1      73.2      65.4

 Italy                     2011     830   69.9       5.1      69.9       4.1      18.3      1.8       0.8      72.2      67.8
                           2016     649   75.2       4.4      75.2       2.5      15.4      1.6       0.9      77.4      73.1

 Lithuania                 2011     913   70.6       5.4      70.6       2.3      21.1      0.3       0.3      74.8      66.9
                           2016     728   71.9       6.8      71.9       2.3      18.3      0.3       0.4      75.0      68.5

 Malta                     2011     717   41.1      17.0      41.1       8.9      30.3      0.8       2.0      48.0      35.0
                           2016     605   50.0      11.7      50.0       8.8      25.2      1.6       2.8      49.1      50.7

 Morocco                   2011    1460   28.4      15.8      28.4      17.5      24.0      4.3       9.9      25.2      31.6
                           2016     895   28.2      15.4      28.2      16.1      30.1      2.2       8.1      29.8      26.7

 Netherlands               2011     803   79.8       5.3      79.8       1.5      13.0      0.3       0.3      79.7      79.9
                           2016     700   80.0       6.2      80.0       1.6      10.8      0.8       0.7      79.8      80.3

 New Zealand               2011    1136   57.2      10.9      57.2       5.0      24.6      1.0       1.2      60.9      53.5
                           2016     934   57.8      10.5      57.8       4.8      24.0      1.0       1.9      63.0      52.8

 Northern Ireland          2011     707   65.2       9.9      65.2       2.4      21.6      0.6       0.2      69.9      60.5
                           2016     610   70.6       9.1      70.6       1.5      18.0      0.7       0.1      73.1      67.8

 Norway (4)                2011     633   59.2       5.1      59.2       3.4      27.1      2.6       2.5      62.1      55.8
                           2016     723   63.3       2.7      63.3       2.7      23.0      1.9       6.3      66.3      60.8

 Oman                      2011    2041   25.3      20.4      25.3      16.0      29.2      3.4       5.8      24.8      25.6
                           2016    1539   32.8      18.9      32.8      13.3      29.7      1.8       3.5      32.4      33.3

 Portugal                  2011     815   61.4       5.3      61.4       3.0      28.6      0.9       0.9      61.0      61.7
                           2016     758   62.2       4.3      62.2       2.3      28.5      0.8       2.0      62.5      61.8

 Qatar                     2011     805   29.5      20.8      29.5      13.9      31.0      2.1       2.7      27.4      31.5
                           2016    1502   44.0      16.1      44.0       9.5      25.5      1.0       3.9      44.9      43.1

 Russian Federation        2011     888   74.9       3.3      74.9       1.7      19.3      0.7       0.1      76.2      73.6
                           2016     757   80.1       4.1      80.1       1.8      12.9      0.5       0.5      79.5      80.8

 Saudi Arabia              2011     898   34.2      17.7      34.2      16.2      28.2      1.5       2.2      30.7      38.0
                           2016     783   42.9      17.7      42.9      14.6      18.4      1.2       5.2      43.0      42.8

 Singapore                 2011    1254   84.5       6.5      84.5       1.8       6.8      0.4       0.1      86.8      82.3
                           2016    1083   85.4       6.4      85.4       0.9       6.9      0.3       0.1      86.8      84.1

 Slovak Republic           2011    1119   68.2       4.3      68.2       5.1      20.7      1.2       0.5      68.2      68.1
                           2016     907   61.7       6.8      61.7       6.0      23.8      0.8       0.9      64.0      59.3

 ---------------------------------------------------------------------------------------------------------------------------------
DIFF = Percent correct
Because of missing gender information, some totals may appear inconsistent.
```

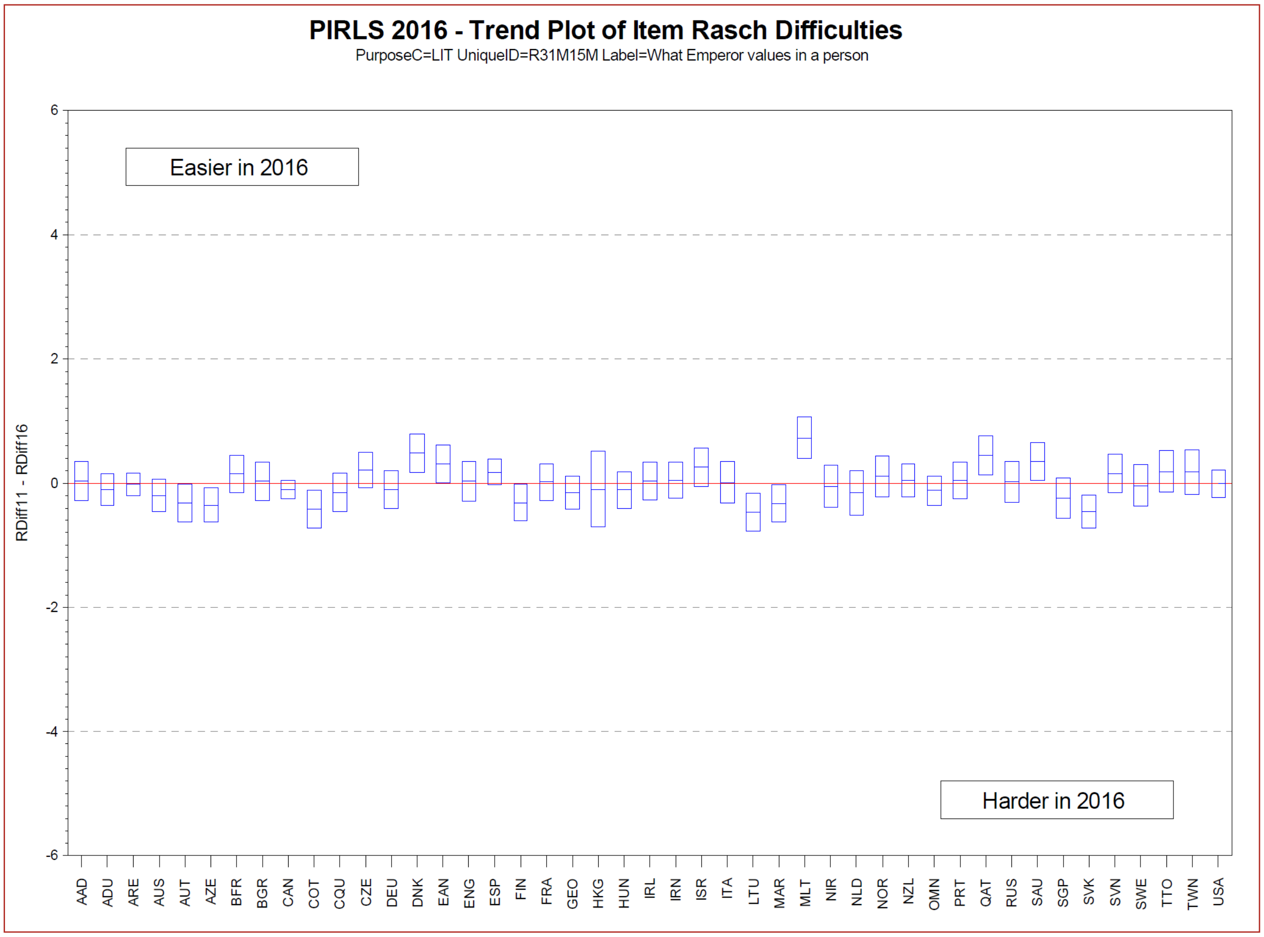
Exhibit 10.4: Example Item Statistics for a PIRLS 2016 Trend Item (Continued)



```
Progress in International Reading Literacy Study - PIRLS 2016 Assessment Results
Trend Achievement Data Almanac for Literary Experience Items (Weighted)

The Empty Pot:  Literary Experience / Evaluate and Critique Content and Textual Elements
R31M15M: What Emperor values in a person  -  1 Point  -  Key: B


                                                                                       NOT
                                          DIFF       A       B       C        D  OMITTED  REACHED   1.GIRL   2.BOY
 COUNTRY                   YEAR       N      %       %       %       %        %        %        %  % RIGHT  % RIGHT
 ---------------------------------------------------------------------------------------------------------------------
 Slovenia                  2011     881   67.0     3.2    67.0     1.6     26.4      0.7      1.1     73.4     61.1
                           2016     751   75.9     5.0    75.9     1.8     15.9      0.4      1.0     77.5     74.5

 Spain                     2011    1692   53.2     8.0    53.2     3.9     31.5      0.7      2.7     53.9     52.4
                           2016    2431   65.9     7.3    65.9     2.2     21.6      0.9      2.1     67.6     64.1

 Sweden                    2011     905   75.7     1.8    75.7     3.1     18.7      0.5      0.2     76.5     74.7
                           2016     749   81.4     2.9    81.4     1.5     11.7      0.8      1.6     81.7     81.1

 Trinidad and Tobago       2011     787   39.9    13.9    39.9     5.4     35.5      2.8      2.5     43.4     36.7
                           2016     687   47.6    13.7    47.6     6.1     26.4      2.0      4.3     47.9     47.2

 United Arab Emirates      2011    2894   39.0    17.8    39.0    11.2     27.7      1.7      2.6     38.4     39.6
                           2016    2730   44.1    17.3    44.1     8.2     24.7      3.1      2.6     45.9     42.3

 United States             2011    2532   67.0    10.7    67.0     2.4     17.9      0.4      1.6     68.2     65.7
                           2016     748   67.2     8.9    67.2     2.6     19.1      0.4      1.8     71.5     63.5

 ---------------------------------------------------------------------------------------------------------------------
 International Avg. (40)   2011   45856   59.6     8.8    59.6     5.1     23.5      1.2      1.8     60.9     58.3
                           2016   38288   63.2     8.1    63.2     4.5     20.8      1.3      2.1     65.3     61.1

 ---------------------------------------------------------------------------------------------------------------------
 Ontario, Canada           2011     903   64.1     9.1    64.1     2.9     22.4      0.4      1.1     62.0     66.4
                           2016     693   58.7     9.1    58.7     4.7     24.2      1.0      2.3     63.6     54.8

 Quebec, Canada            2011     848   72.3     8.3    72.3     2.0     16.0      0.5      0.8     72.1     72.5
                           2016     527   74.3     6.8    74.3     1.8     14.3      1.0      1.8     81.0     66.1

 Andalusia, Spain          2011     855   57.1     7.1    57.1     2.8     29.7      1.5      1.8     60.6     53.5
                           2016     691   67.3     5.8    67.3     2.2     22.1      0.8      1.7     65.7     68.8

 Abu Dhabi, UAE            2011     817   35.9    19.6    35.9    14.6     26.1      1.4      2.3     33.9     38.0
                           2016     690   37.8    17.3    37.8     9.6     26.0      5.6      3.8     41.0     34.5

 Dubai, UAE                2011    1199   47.2    13.0    47.2     5.7     29.9      1.9      2.3     47.8     46.7
                           2016    1306   59.1    12.3    59.1     4.2     21.5      1.4      1.5     59.4     58.8

 ---------------------------------------------------------------------------------------------------------------------









DIFF = Percent correct
Because of missing gender information, some totals may appear inconsistent.
```

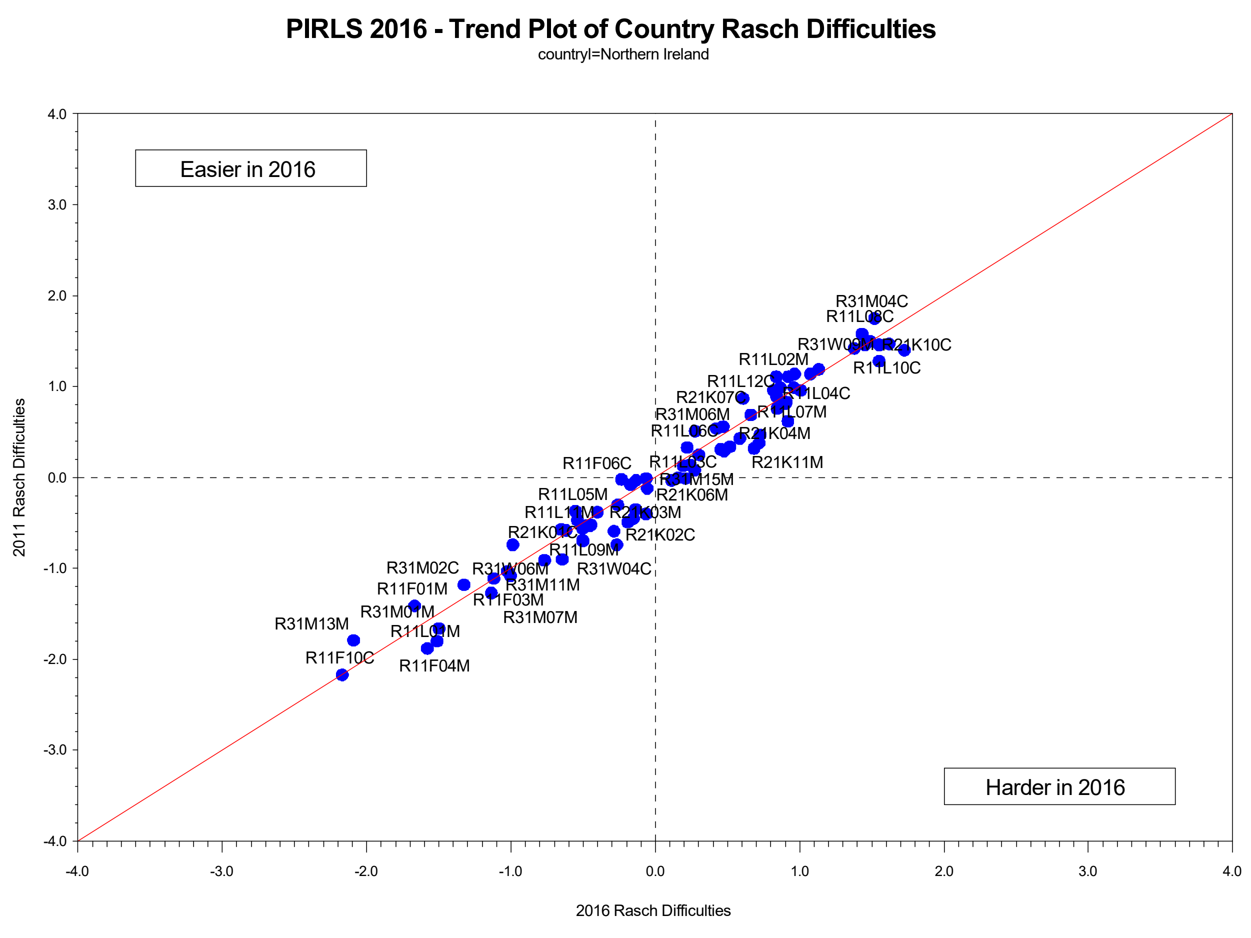
Although some changes in item difficulties were anticipated as countries' overall achievement may have improved or declined, items were noted if the difference between the Rasch difficulties across the two assessments for a particular country was greater than 2 logits. The TIMSS & PIRLS International Study Center used two different graphical displays to examine the differences in item difficulties. The first of these, shown for an example item in Exhibit 10.5, displays the difference in Rasch item difficulty of the item between 2016 and 2011 for each country. A positive difference for a country indicates that the item was relatively easier in 2016, and a negative difference indicates that the item was relatively more difficult.

Exhibit 10.5: Example Plot of Differences in Rasch Item Difficulties Between 2016 and 2011 for a PIRLS 2016 Trend Item



The second graphical display, presented in Exhibit 10.6, shows the performance of a given country on all trend items simultaneously. For each country, the graph plots the 2016 Rasch difficulty of every trend item against its Rasch difficulty in 2011. Where there were no differences between the difficulties in the two successive administrations, the data points aligned on or near the diagonal.

Exhibit 10.6: Example Plot of Rasch Item Difficulties Across PIRLS 2016 Trend Items by Country



Reliability

Documenting the reliability of the PIRLS 2016 assessments was a critical quality control step in reviewing the items. As one indicator of reliability, the review considered Cronbach's Alpha coefficient of reliability calculated at the assessment booklet level. Secondly, the scoring of the constructed response items had to meet specific reliability criteria in terms of consistent within-country scoring, cross-country scoring, and across assessment or trend scoring.

Test Reliability

Exhibit 10.7 displays the PIRLS, PIRLS Literacy, and ePIRLS test reliability coefficients for every country, respectively. These coefficients are the median Cronbach's alpha reliability across all PIRLS 2016 assessment booklets. In general, reliabilities were relatively high. For PIRLS, the international median reliability (the median of the reliability coefficients for all countries) was 0.83. The international median reliability for PIRLS Numeracy was 0.92, whereas the international median reliability for ePIRLS was 0.92.

Exhibit 10.7: Cronbach's Alpha Reliability Coefficient – PIRLS 2016

Country	Reliability Coefficient		
	PIRLS	PIRLS Literacy	ePIRLS
Australia	0.91	—	—
Austria	0.86	—	—
Azerbaijan	0.89	—	—
Bahrain	0.91	—	—
Belgium (Flemish)	0.86	—	—
Belgium (French)	0.87	—	—
Bulgaria	0.91	—	—
Canada	0.89	—	0.90
Chile	0.90	—	—
Chinese Taipei	0.87	—	0.90
Czech Republic	0.88	—	—
Denmark	0.88	—	0.90
Egypt	—	0.92	—
England	0.90	—	—
Finland	0.88	—	—
France	0.88	—	—
Georgia	0.89	—	0.90
Germany	0.90	—	—
Hong Kong SAR	0.85	—	—
Hungary	0.89	—	—
Iran, Islamic Rep. of	0.90	0.92	—
Ireland	0.89	—	0.90
Israel	0.92	—	0.92
Italy	0.87	—	0.89
Kazakhstan	0.86	—	—
Kuwait	—	0.90	—
Latvia	0.86	—	—
Lithuania	0.88	—	—
Macao SAR	0.87	—	—
Malta	0.89	—	—

Exhibit 10.7: Cronbach's Alpha Reliability Coefficient – PIRLS 2016 (Continued)

Country	Reliability Coefficient		
	PIRLS	PIRLS Literacy	ePIRLS
Morocco	0.86	0.91	—
Netherlands	0.86	—	—
New Zealand	0.92	—	—
Northern Ireland	0.90	—	—
Norway	0.87	—	0.89
Oman	0.91	—	—
Poland	0.88	—	—
Portugal	0.87	—	0.89
Qatar	0.92	—	—
Russian Federation	0.87	—	—
Saudi Arabia	0.90	—	—
Singapore	0.91	—	0.92
Slovak Republic	0.90	—	—
Slovenia	0.89	—	0.90
South Africa	—	0.90	—
Spain	0.87	—	—
Sweden	0.88	—	0.90
Trinidad and Tobago	0.92	—	—
United Arab Emirates	0.93	—	0.93
United States	0.90	—	0.91
International Median	**0.89**	**0.91**	**0.90**
Benchmarking Participants			
Buenos Aires, Argentina	0.90	—	—
Ontario, Canada	0.90	—	—
Quebec, Canada	0.85	—	—
Denmark (3)	—	0.88	—
Norway (4)	0.88	—	—
Moscow City, Russian Fed.	0.83	—	—
Eng/Afr/Zulu - RSA (5)	0.91	—	—
Andalusia, Spain	0.86	—	—
Madrid, Spain	0.85	—	—
Abu Dhabi, UAE	0.92	—	0.93
Dubai, UAE	0.92	—	0.93

Scoring Reliability for Constructed Response Items

A sizeable proportion of the items in the PIRLS 2016 assessments were constructed response items, comprising about half of the assessment score points. An essential requirement for use of such items is that they be reliably scored by all participants. That is, a particular student response should receive the same score, regardless of the scorer. In conducting PIRLS 2016, measures taken to ensure that the constructed response items were scored reliably in all countries included developing scoring guides for each constructed response question (that provided descriptions of acceptable responses for each score point value) and providing extensive training in the application of the scoring guides. See Chapter 1: Developing the PIRLS 2016 Achievement Items for more information on the scoring guides and see Chapter 6: Survey Operations for PIRLS 2016 for information on the scoring process.

Within-Country Scoring Reliability

To gather and document information about the within-country agreement among scorers for PIRLS 2016, a random sample of approximately 25 percent of the assessment booklets was selected to be scored independently by two scorers. The inter-scorer agreement for each item in each country was examined as part of the item review process. Exact percent agreement across items was high on average across countries—96 percent or above, on average internationally. See Appendix 10A for the average and range of the within-country percentage of correctness score agreement across all items. The PIRLS Within-Country Scoring Reliability documents also provide the average and range of the within-country percentage of diagnostic score agreement.

Trend Item Scoring Reliability

The TIMSS & PIRLS International Study Center also took steps to show that the 2016 constructed response items used in PIRLS 2011 were scored in the same way in both assessments. In anticipation of this, countries that participated in PIRLS 2011 sent samples of scored student booklets from the 2011 data collections to IEA Hamburg, where they were digitally scanned and stored for later use. As a check on scoring consistency from one administration to the next, staff members working in each country on scoring the 2016 data were asked also to score these 2011 responses using the Trend Reliability Scoring Software developed by IEA Hamburg. Each country scored 200 responses for 22 PIRLS reading items (South Africa scored 24 PIRLS Literacy reading items for their fourth grade sample).

There was a very high degree of scoring consistency in PIRLS 2016. The exact agreement between the scores awarded in 2011 and those given by the 2016 scorers was 95 percent on average internationally. The average and range of scoring consistency over time can be found in Appendix 10B.

Cross-Country Scoring Reliability Study

It also was important to document the consistency of scoring across countries. Because of the many different languages in use in PIRLS 2016, establishing the reliability of constructed response scoring across all countries was not feasible. However, the TIMSS & PIRLS International Study Center did conduct a cross-country study of scoring reliability among Northern Hemisphere countries that had scorers who were proficient in English. A sample of student responses was provided by the English-speaking Southern Hemisphere countries. Cross-country scoring included 200 student responses for 22 PIRLS reading items. This set of student responses in English was then scored independently in each country that had two scorers proficient in English, using the Cross-Country Scoring Reliability Software provided by IEA Hamburg. In all, scorers from 44 countries and four benchmarking participants took part in the process. Making all possible comparisons among scorers gave a total of 1,128 possible comparisons for each student response to each item, and resulted in more than 225,600 total comparisons when aggregated across all 200 student responses to any given item.

Agreement across countries was defined in terms of the percentage of these comparisons that were in exact agreement. On average internationally, scorer reliability across countries in PIRLS 2016 was high, with an exact agreement in the scores awarded of 85 percent on average internationally. See Appendix 10C for the results of the cross-country scoring reliability study.

Item Review Procedures

Using the information from the comprehensive collection of item analyses and reliability data that were computed and summarized for PIRLS 2016, the TIMSS & PIRLS International Study Center thoroughly reviewed all item statistics for every participating country and benchmarking participant to ensure that the items were performing comparably across countries. In particular, items with the following problems were considered for possible deletion from the international database:

- An error was detected during translation verification but was not corrected before test administration
- Data checking revealed a multiple-choice item with more or fewer options than in the international version
- The item analysis showed the item to have a negative biserial, or, for an item with more than 1 score point, point biserials that did not increase with each score level
- The item-by-country interaction results showed a very large negative interaction for a particular country

- For constructed response items, the within-country scoring reliability data showed an agreement of less than 70 percent
- For trend items, an item performed substantially differently in 2016 compared to the PIRLS 2011 administration, or an item was not included in the previous assessment for a particular country

When item statistics indicated a problem with an item, translation verification documentation was used as an aid in checking the test booklets. If a question remained about potential translation or cultural issues, however, then the National Research Coordinator was consulted before deciding how the item should be treated.

The checking of the PIRLS 2016 achievement data involved review of almost 400 items and resulted in the detection of very few items that were inappropriate for international comparisons. The items found to be problematic during the review process primarily had issues related to translation or printing problems. See Appendix 10D: Country Adaptations to Items and Item Scoring for a list of deleted items, as well as a list of recodes made to constructed response item codes. There also were a number of items in each study that were combined, or derived, for scoring purposes. See Appendix 10E for details about how score points were awarded for each derived item.

Appendix 10A: PIRLS 2016 Within-Country Scoring Reliability for the Constructed Response Items

PIRLS 2016 Within-Country Scoring Reliability for the Constructed Response Items

Country	PIRLS		
	Average of Percent Agreement Across Items	Range of Percent Agreement	
		Minimum	Maximum
Australia	92	68	100
Austria	96	88	100
Azerbaijan	98	93	100
Bahrain	96	89	100
Belgium (Flemish)	95	84	100
Belgium (French)	99	95	100
Bulgaria	96	84	100
Canada	89	68	100
Chile	98	94	100
Chinese Taipei	98	91	100
Czech Republic	100	98	100
Denmark	90	62	100
England	95	75	100
Finland	96	79	100
France	94	83	100
Georgia	91	72	100
Germany	93	72	100
Hong Kong SAR	97	84	100
Hungary	97	91	100
Iran, Islamic Rep. of	95	85	100
Ireland	99	94	100
Israel	96	87	100
Italy	95	86	100
Kazakhstan	99	96	100
Latvia	96	83	100
Lithuania	99	97	100
Macao SAR	99	96	100
Malta	91	76	100
Morocco	89	68	99
Netherlands	96	82	100

PIRLS 2016 Within-Country Scoring Reliability for the Constructed Response Items (Continued)

Country	PIRLS		
	Average of Percent Agreement Across Items	Range of Percent Agreement	
		Minimum	Maximum
New Zealand	95	79	100
Northern Ireland	100	100	100
Norway	97	91	100
Oman	94	85	100
Poland	94	80	100
Portugal	98	90	100
Qatar	98	93	100
Russian Federation	99	94	100
Saudi Arabia	98	92	100
Singapore	100	99	100
Slovak Republic	98	91	100
Slovenia	97	82	100
Spain	98	90	100
Sweden	96	88	100
Trinidad and Tobago	89	66	100
United Arab Emirates	93	86	100
United States	97	91	100
International Avg.	**96**	**86**	**100**
Benchmarking Participants			
Buenos Aires, Argentina	93	79	100
Ontario, Canada	88	70	100
Quebec, Canada	89	59	100
Moscow City, Russian Fed.	98	90	100
Eng/Afr/Zulu - RSA (5)	92	76	100
Andalusia, Spain	98	91	100
Madrid, Spain	98	87	100
Abu Dhabi, UAE	93	84	100
Dubai, UAE	93	86	100

PIRLS Literacy 2016 Within-Country Scoring Reliability for the PIRLS Literacy Constructed Response Items

Country	PIRLS Literacy		
	Average of Percent Agreement Across Items	Range of Percent Agreement	
		Minimum	Maximum
Egypt	97	88	100
Iran, Islamic Rep. of	96	76	100
Kuwait	90	61	100
Morocco	89	33	100
South Africa	94	83	100
International Avg.	**93**	**68**	**100**
Benchmarking Participant			
Denmark (3)	95	68	100

ePIRLS 2016 Within-Country Scoring Reliability for the ePIRLS Constructed Response Items

Country	ePIRLS		
	Average of Percent Agreement Across Items	Range of Percent Agreement	
		Minimum	Maximum
Canada	92	79	99
Chinese Taipei	96	90	100
Denmark	91	72	99
Georgia	94	84	100
Ireland	95	90	100
Israel	95	90	100
Italy	95	88	100
Norway	98	96	100
Portugal	95	87	100
Singapore	100	100	100
Slovenia	90	70	100
Sweden	95	86	100
United Arab Emirates	94	87	100
United States	94	86	100
International Avg.	**95**	**86**	**100**
Benchmarking Participants			
Abu Dhabi, UAE	94	86	100
Dubai, UAE	93	85	100

Appendix 10B: PIRLS 2016 Trend Scoring Reliability for the Constructed Response Items

PIRLS 2016 Trend Scoring Reliability for the Constructed Response Items

Country	Average of Percent Agreement Across Items	Range of Percent Agreement	
		Minimum	Maximum
Australia	95	80	100
Austria	96	82	100
Azerbaijan	92	66	100
Belgium (French)	97	87	100
Bulgaria	96	81	100
Canada	94	79	100
Chinese Taipei	95	81	100
Czech Republic	96	80	100
Denmark	95	78	100
England	96	80	100
Finland	95	78	100
France	93	69	100
Georgia	93	76	100
Germany	96	85	100
Hong Kong SAR	98	88	100
Hungary	95	78	100
Iran, Islamic Rep. of	95	82	100
Ireland	96	86	100
Israel	95	78	100
Italy	94	82	100
Lithuania	97	90	100
Netherlands	94	69	100
New Zealand	96	80	100
Northern Ireland	96	83	100
Norway	96	87	100
Oman	95	82	100
Poland	96	82	100
Portugal	92	66	100
Qatar	91	62	100
Russian Federation	96	83	100
Singapore	96	83	100
Slovak Republic	94	82	100

PIRLS 2016 Trend Scoring Reliability for the Constructed Response Items (Continued)

Country	Average of Percent Agreement Across Items	Range of Percent Agreement	
		Minimum	Maximum
South Africa	93	72	100
Spain	92	72	100
Sweden	95	78	100
Trinidad and Tobago	92	73	100
United Arab Emirates	93	56	100
United States	94	74	100
International Avg.	**95**	**78**	**100**
Benchmarking Participants			
Eng/Afr/Zulu - RSA (5)	91	65	100
Dubai, UAE	90	51	100

Appendix 10C: PIRLS 2016 Cross-Country Scoring Reliability for the Constructed Response Items

PIRLS 2016 Cross-Country Scoring Reliability for the Constructed Response Items

Item Label	Total Valid Comparisons	Percent Exact Agreement
Empty Pot R31M02C	214,879	97
Empty Pot R31M04C	204,588	88
Empty Pot R31M09C	212,582	86
Empty Pot R31M10C	216,460	92
Empty Pot R31M16C	216,989	92
Honey R31W01C	221,321	94
Honey R31W02C	211,896	78
Honey R31W04C	213,069	96
Honey R31W11C	217,978	97
Honey R31W13C	217,192	84
Sharks R21K01C	214,490	81
Sharks R21K02C	216,596	93
Sharks R21K05C	212,590	87
Sharks R21K07C	208,487	81
Sharks R21K10C	213,352	82
Sharks R21K12C	214,311	77
Shiny Straw R21Y03C	210,586	89
Shiny Straw R21Y09C	215,727	82
Shiny Straw R21Y10C	212,668	78
Shiny Straw R21Y12C	214,658	80
Shiny Straw R21Y13C	215,811	65
Shiny Straw R21Y14C	209,761	73
Average Percent Agreement		**85**

Appendix 10D: Country Adaptations to PIRLS 2016 Items and Item Scoring

Country Adaptations to PIRLS 2016 Items and Item Scoring

PIRLS and PIRLS Literacy
Deleted Items
MALTA
The Green Sea Turtle's Journey of a Lifetime Item 15, R41T15M (Negative discrimination)
How Did We Learn to Fly? Item 2, L21E02C (Translation error)
NETHERLANDS
Sharks Item 4, R21K04M (Negative discrimination)
NORWAY
Sharks Item 2, R21K02C (Printing error)
Shiny Straw Item 3, R21Y03C (Printing error)
Empty Pot Item 7, R31M07M (Translation error)
SAUDI ARABIA
Oliver and the Griffin Item 6, R41O06M (Low discrimination)
SINGAPORE
The Green Sea Turtle's Journey of a Lifetime Item 11, R41T11C (Scoring error)
SLOVAK REPUBLIC
Empty Pot Item 4, R31M04C (Translation error)
Constructed Response Items with Category Recoding
Icelandic Horses Item 15, R41I15C (Recoded from 2 into 1)
African Rhinos and Oxpecker Birds Item 17, L21C17C (Recoded from 2 into 1)
Flowers on the Roof Item 12, R11F12C (Recoded from 3 into 2)
ePIRLS
Deleted Items
GEORGIA
Rainforests Item 6, E11R06C (Missing data)
Constructed Response Items with Category Recoding
Zebra and Wildebeest Migration Item 12, E11Z12C (Recoded 2 to 1)

Appendix 10E: Derived Items in PIRLS 2016

Derived Items in PIRLS 2016

PIRLS and PIRLS Literacy

Where’s the Honey? Item 7, R31W07C – Item parts A, B, and C are combined to create a 3-point item, where 3 score points are awarded if all parts are correct, 2 score points are awarded if two parts are correct, and 1 score point is awarded if only one part is correct

Empty Pot Item 17, R31M17C – Item parts A, B, and C are combined to create a 3-point item, where 3 score points are awarded if all parts are correct, 2 score points are awarded if two parts are correct, and 1 score point is awarded if only one part is correct

Ants Item 12, L11A12CZ – Item parts A, B, and C are summed to create a 3-point item

Ants Item 13, L11A13CZ – Item parts B–E are combined to create a 2-point item, where 2 points are awarded if all 4 parts are correct, 1 point is awarded if 3 parts are correct, and 0 points are awarded if 2 or fewer parts are correct

ePIRLS

Mars Item 16, E11M16C – Item parts A through D are combined to create a 2-point item, where 2 points are awarded if all 4 parts are correct, 1 point is awarded if 3 parts are correct, and 0 points are awarded if 2 or fewer parts are correct

Rainforests Item 3, E11R03C – Item parts A through D are combined to create a 2-point item, where 2 points are awarded if all 4 parts are correct, 1 point is awarded if 3 parts are correct, and 0 points are awarded if 2 or fewer parts are correct

Rainforests Item 7, E11R07C – Item parts A through D are combined to create a 2-point item, where 2 points are awarded if all 4 parts are correct, 1 point is awarded if 3 parts are correct, and 0 points are awarded if 2 or fewer parts are correct

Zebra and Wildebeest Migration item 20, E11Z20C – Item parts A through D are combined to create a 2-point item, where 2 points are awarded if all 4 parts are correct, 1 point is awarded if 3 parts are correct, and 0 points are awarded if 2 or fewer parts are correct

The Legend of Troy Item 18, E11T18C – Item parts A, B, and D are combined to create a 1-point item, where 1 point is awarded if all 3 parts are correct and 0 points are awarded if 2 or fewer parts are correct