Item Analysis for Multiple-Choice Achievement Items Using R

Companion Reading: Bandalos, p. 121-131

 # The dataset <PIRLS 2011_Morocco_MC reading items.csv> shows responses of 1,512 Moroccan fourth-graders to the 7 multiple-choice reading comprehension items following one reading passage on the 2011 Progress in International Reading Literacy (PIRLS) 2011 test. Item responses are labeled as "A", "B", "C", "D", or "" = omitted/missing.

The document "PIRLS 2011 released items_Passage 1" contains the complete English-translated text of the reading passage and test items, which were administered in Arabic. The keyed correct answer for each item is indicated.

######## **Annotated Output/Results** #########

 # Distractor analysis results: The values in the "lower" column represent proportions of test-takers who chose each of the four response options (or omitted the item) among those who had scores in the lowest 1/3 of the 7-item total score distribution. The values in the "mid66" and "upper" columns represent test-takers who had scores in the middle 1/3, and upper 1/3 of the total score distribution. Values in a column will sum to 1, representing 100% of test-takers in each total score group.

On Item 1, nearly all students in the upper score group (94%), and most students in the middle group (72%), selected the correct answer, A. Among the incorrect options, distractor D seems to have been the most attractive to test-takers. In the lowest score group, more students selected distractor D (44%) than the correct answer (23%). [Examining the item's text might reveal the reason.] None of the distractor options appear to be non-functional in this population -- all were selected by non-trivial proportions of test-takers. Only 17 students skipped Item 1, and all were in the lowest group on total test scores.

On Item 3, a sizable proportion of students in the upper (35%) and middle (55%) total score groups selected distractor B rather than the correct answer, A. This distractor should be re-evaluated by language and literacy education experts from this country to ensure A is an unambiguous correct response.

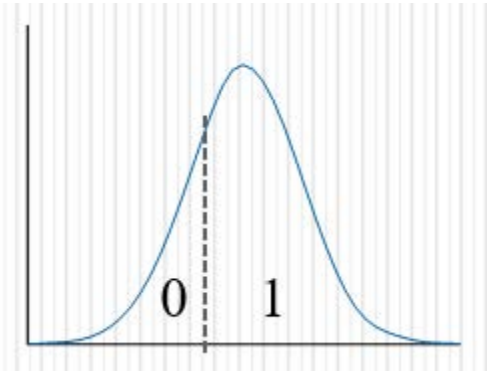
More than 200 students skipped Item 11. If this item appeared toward the end of the test, possibly the allowed testing time should be reconsidered. This item was difficult. In the upper total score group, more students chose the correct answer A (38%) than any distractor, but sizable proportions of students chose each distractor, particularly distractor C (30%), and 10% of students omitted the item. A similar pattern was evident in the middle score group, and in the lower score group, all three distractors were chosen more often than the correct answer.

```
> distractorAnalysis(PIRLS_Mor, PIRLSkey, pTable=TRUE, digits = 3, nGroups = 3,
csvReport="C:/Users/username/Desktop/Mor_distractor analysis.csv")


$Item01
  correct key   n lower mid66 upper
               17 0.022 0.000 0.000
A       *    A 808 0.225 0.717 0.937
B            B 151 0.169 0.053 0.012
C            C 121 0.144 0.029 0.004
D            D 415 0.440 0.201 0.047
```

```
$Item03
  correct key   n lower mid66 upper
              59 0.047 0.037 0.027
A       *   A 451 0.128 0.287 0.575
B           B 614 0.395 0.545 0.354
C           C 240 0.272 0.061 0.027
D           D 148 0.157 0.070 0.018


$Item11
  correct key   n lower mid66 upper
             204 0.145 0.176 0.098
A       *   A 364 0.146 0.270 0.376
B           B 278 0.241 0.139 0.115
C           C 445 0.299 0.262 0.303
D           D 221 0.168 0.152 0.108
```

 # Item analysis results: ***Proportion-correct item difficulty statistics***, displayed in the itemMean column, suggest these items vary in difficulty, which is desirable for an instrument meant to characterize group academic achievement, but that overall these items are fairly difficult for test-takers. The proportion-correct difficulty of Item 8 is near the minimum value that might be acceptable on a cross-national assessment.

Biserial item-total score correlations ('bis') assume that a continuum of possible relevant knowledge underlies observed 0/1 responses on each item, so that some test-takers are, for instance, just barely above the threshold for the "correct" category while other test-takers are far beyond the threshold. That is, this correlation assumes that test-takers' knowledge of the correct answer is not categorically 'none' (0) or 'complete' (1), but rather ranges across various levels of partial knowledge. Pearson item-total correlations (which are often labeled 'point-biserial' in the item analysis context) instead assume that both items involved in the correlation have approximately-continuous observed response distributions, an assumption which is violated to a substantial degree by binary items. Biserial correlations will generally be larger than point-biserial correlation values, and also are theoretically a better estimate of the population item-total correlation.



Biserial correlations between each item's score and the 6-item total score composed from the other items suggest problems with this item set in this test-taker population. Items 8 and 11 have weak correlations with the total score, and Item 3 also does not look good. [The item analysis report should then proceed to evaluate the text of these items. For instance, Items 8 and 11 require inference beyond literal text comprehension. Item 3 has one distractor that might be problematic as a 'wrong' answer in collectivist societies. There could be translation-related difficulties, although this testing program invests a relatively large amount of resources in test development for cross-national assessment, and this conclusion would not be possible to evaluate directly without expert working knowledge of both languages.]

```
> results <- itemAnalysis(scoredMordata2, itemReport=TRUE, bisFlag = .30)
> results$itemReport

       itemName  itemMean       pBis        bis lowBis
1 scored.Item01 0.5343915 0.46723897 0.58639056
```

2	scored.Item02	0.4470899	0.37368825	0.46985751	
3	scored.Item03	0.2982804	0.22667984	0.29905065	X
4	scored.Item04	0.5767196	0.38825138	0.48992609	
5	scored.Item06	0.3776455	0.39429587	0.50299838	
6	scored.Item08	0.1157407	0.12063429	0.19792058	X
7	scored.Item11	0.2407407	0.04837777	0.06642069	X