An Introduction to
Three Item Testing

Kirk Becker
The Riverside Publishing Company

Using the Rasch model, the characteristics of a test or survey can be examined despite the presence of missing data, but is this also true about the characteristics of a population? In other words, is it always necessary to administer a test or survey in full in order to find out about a population of interest?

In order to compare the means of two populations on an instrument, many would say that all items on the instrument must be administered. Although this might be true for a completely untried test or survey, once the items have been calibrated only three items are needed. When items have been scaled using a population as a reference point, this reference point (the difficulty of the items in logits) can then be used to measure the ability level of individuals and the mean ability level of groups, in the same units. The Rasch model allows for a direct transformation between raw scores and logit measures. If a population mean in logits is known relative to a set of item calibrations, the population mean in raw score units can then be determined. For studies in which the population parameters are the main point of interest, this can mean huge savings in terms of time and money.

How is it possible to estimate population parameters without administering a complete measure to a large, representative sample? Data collected during the development of the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCullum, 1997) and the Stanford-Binet Intelligence Test: Fourth Edition (Thorndike, Hagen, & Sattler, 1996) was used to investigate this question.

Any pair of variables contains a great deal of information about a population that answers them. Consider the performance of 9-year-olds on a pair of items from the UNIT:

<table>
<thead>
<tr>
<th>Item</th>
<th>Right:</th>
<th>Wrong:</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>178=S_{19}</td>
<td>35=S_{19}</td>
</tr>
<tr>
<td></td>
<td>76=S_{01}</td>
<td>68=S_{01}</td>
</tr>
</tbody>
</table>

If most individuals in a population fail the pair of items ($S_{19}$), then the population mean should logically be lower than the difficulty of the two items. Likewise, if the majority of a population pass a pair of items ($S_{19}$), then the population mean should logically be higher than the difficulty of the items. The ratio of $S_{19}$ to $S_{01}$ is therefore related to the mean of the population on the entire test, however it is also
a function of the item difficulty difference. The other two cells in the cross-tabulation (Table 1) are highly related to the difference in difficulty between the items. If item 19 had been very easy and item 16 very difficult, most of the population would have fallen into cell $S_{91}$. Likewise, if item 19 were difficult and item 16 easy, most of the population would have been in cell $S_{90}$. In order to examine how these relate to item difficulty and population mean, the following ratios will be used:

$$\log \left( \frac{S_{19}}{S_{90}} \right) \quad \log \left( \frac{S_{91}}{S_{90}} \right)$$

To examine the effect item difficulty difference has on the first relationship, the cross-tabs of several item pairs were examined. For cross-tabs between one item (item 19) and a set of other items, $\log \left( \frac{S_{19}}{S_{90}} \right)$ and $\log \left( \frac{S_{91}}{S_{90}} \right)$ are both directly related to the difference in difficulty between the items. Conceptually, the ratio $\log \left( \frac{S_{19}}{S_{90}} \right)$ should reveal the difference in item difficulty for a pair of items, and as Graph 1 shows, this relationship is born out. Because the mean item difficulty is set to 0, the scale of the item calibrations differs from that of the ratio, however a simple linear transformation allows us to place these sets of values on an identity line (Graph 2).

The formula for scaling the $y$-intercept of $\log \left( \frac{S_{19}}{S_{90}} \right)$ versus $\log \left( \frac{S_{91}}{S_{90}} \right)$ to the population mean is known in this case because the means are known. The slope of this line appears to be constant ($m=-0.5$) across multiple tests and populations. As Graph 5 shows, the intercept is the difficulty of the constant item in the cross-tabs.

UNIT
Analogic Reasoning subtest: population mean = $-51x + 1.4$

Symbolic Memory subtest: population mean = $-4x + .41$

Spatial Memory subtest: population mean = $-4x + .06$

Stanford-Binet

Vocabulary subtest: population mean = $-42x$

Comprehension subtest: population mean = $-54x + 2.8$
To summarize, the steps for estimating a population mean from 3 items are as follows:

1. Administer three items from a test that has been calibrated.
2. For the two pairs of items (AB and AC) calculate the ratios log (S11/S00) and log (S10/S01) for the population of interest.
3. Perform a linear transformation on log (S10/S01) so that the plot of log (S10/S01) versus A-B and A-C is an identity.
4. Using the same scaling factor, perform the same linear transformation on the two log (S11/S00) values.
5. Determine the y-intercept of the rescaled log (S11/S00) versus log (S10/S01) plot for the two item pairs.
6. The y-intercept should be related to the population mean according to the following formula

\[
\text{Population mean} = -1/2 \times (y\text{-intercept}) + (\text{difficulty of A})
\]

References