A Primer on Rasch Analysis

The Rasch model is a member of the family of item-response latent-trait models. Using a set of carefully selected survey items (questions), it produces an interval scale that determines *item difficulties* and *person measures*. The items are arranged on the scale according to how likely they are to be endorsed (*item difficulty*). The scale is then used to show *person measure*, a quantitative measure of a person's attitude on a unidimensional scale. In other words, the items are used to define the measure's scale, and people are then placed on this scale based on their responses to the items in the measure. The scale units are logits (log odds units), which are linear and therefore suitable for use in simple statistical procedures.

Measures contain several related items (usually between four and eight). To create these item clusters, CCSR analysts select items that belong together according to education theory. Determinations as to which items to keep in the final measure are based on conceptual coherence as well as the statistical fit of the group of items. Unless there are strong conceptual reasons, CCSR analysts eliminate items with high misfit statistics.

Each person and item is assigned a measure score that represents where they fall on the scale. In addition, each person and item has a true standard error (the precision of the measure) and a fit statistic (the statistical coherence of the measure). The fit statistics are calculated by taking the mean squared deviations of the difference between the expected values and the observed values. The fit statistics have an expected value of 1.0; items with fit statistics substantially greater than 1.0 may belong to a construct different from the one underlying other items in the cluster and may not belong in the cluster.

The school-level means included in this data set are aggregates of individual responses that have been weighted by the inverse of the standard error. Therefore, individual responses that are less reliable or have missing data receive less weight, and more reliable responses receive greater weight in creating a school average of the measure.

In past public reporting, CCSR has often converted the logit scale to a ten-point scale for ease in interpreting results. The scale scores reported here are still on the logit scale, the formula to calculate the ten-point scale is:

$$\text{measure‡} = \frac{[10 * (\text{measure} - \text{minimum})]}{[\text{maximum} - \text{minimum}]}$$

where *measure*‡ is the item's new score on the ten-point scale, *measure* is the item's original scale score, *minimum* is the lowest scale score of any item in the measure, and *maximum* is the highest scale score of any item in the measure. For example, to recompute the item difficulty for item *prs12q01* in PRES (Academic Press) for a ten-point scale:

$$\text{PRES*} = \frac{[10 * (.81 - (-.72))]}{[.92 - (-.72)]}$$

where PRES* = 9.3.

CCSR used the Bigsteps computer program produced by MESA Press at the University of Chicago to develop these measures. For more information on Rasch

analysis, see Benjamin D. Wright and Geoffrey N. Masters, *Rating Scale Analysis: Rasch Measurement* (Chicago: MESA Press, 1982).

Creating Categories for the Rasch Measures

In order to summarize the results for each of the measures, we classified the possible range of each logit scale into three or four categories. These categories usually ranged from most often or most positive to least often or least positive. For example, the categories for Teacher-Parent Trust were very strong trust, strong trust, minimal trust, and no trust. The process typically used to develop these categories follows.

Table 2.2 of the Bigsteps output, which shows the expected responses to each item by person measure, was used in this process. The transition point between the top and second-from-the-top response for the most difficult-to-endorse item was chosen as the cutpoint between the top and second measure category. For example, for a measure that has a four-point scale of *strongly agree* to *strongly disagree*, the definition for the top measure category is "all people strongly agree with all items." However, if the most positive responses were actually the top two response categories (i.e. *nearly all* and *most*) then they were treated as the same response category. The transition point between *most* and the next response category, *about half*, was used as the cutpoint between the top and second-from-the top measure categories.

A similar procedure was used to determine the bottom category. The transition point between the bottom and second-from-the-bottom response categories (i.e., *never* and *sometimes*) for the easiest-to endorse item was the cutpoint between the bottom and second-from-the-bottom measure categories.

Dividing the middle range into two measure categories was a more difficult task. If there was enough variability it was divided into two measure categories, making four measure categories all together. Otherwise, the middle range was left as a single measure category, creating three measure categories. The goal was to create measure categories that were as homogenous as possible with respect to expected responses. In some cases, the items were grouped thematically or conceptually; then, attention was paid to keeping the thematic or conceptual groups homogeneous with respect to their response category.

There were some cases where the most difficult-to-endorse item was answered very differently from the remaining items. If the remaining items were answered in a similar fashion, we used the middle item in the item set and its transition points to determine most of the cutpoints. For example, we used the middle item to determine the cutpoint between the top and second-from-the-top measure categories and between the second and third-from-the-top measure categories. Efforts were made to maintain continuity in category meaning between measures.