Center for Response to Intervention in Early Childhood Michael C. Rodriguez
University of Minnesota 2011

The Rasch Model

Rasch modeling (Rasch, 1960) provides a framework for examining a particular test and student performance on that test at the item level. The probability of correct response for each item is modeled as a logistic function, where B is examinee ability level and D is the difficulty level of the item, or the ability at which a test taker has a predicted probability .5 of answering the item correctly. When calibrating the items for each of the measures, we use the Rasch model, which is a mathematical model for the relation between the probability of success (P) and the difference between an individual's ability (B) and an item's difficulty (D).

Rasch item calibrations are estimated using Winsteps 3.70 (Linacre, 2010). The typical formulations of this logistic (log-odds) model is $\log\left(\frac{P_{ni}}{1-P_{ni}}\right) = B_n - D_i$, or as seen in typical IRT format, as $P(x = 1|B) = \frac{e^{(B-D)}}{1+e^{(B-D)}}$, where P is a probability of person n responding correctly to item i, and the Rasch parameters are B_n, the ability of person n, and D_i, the difficulty of item i.

Linacre, J.M. (2010). *Winsteps* (Version 3.70.0) [Computer Software]. Beaverton, Oregon: Winsteps.com.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Chicago: The University of Chicago Press.

Power & the Rasch Model

Power is the probability of correctly rejecting a null hypothesis, which is not relevant in measurement. As a measurement model, the Rasch model is intended to estimate item attributes (and person attributes). An appropriate criterion for determining sample size is item calibration stability (Linacre, 1994). This is related to the precision of estimates of item attributes, rather than power per se. To achieve item calibration stability within +/- 0.50 logits (95% CI), 100 observations are needed.

This criterion is consistent with the measurement errors found in large-scale standardized assessments, such as TIMSS and PIRLS and other national exams (Wu, 2010), which are considerably longer than the proposed IGDIs. In all item and score analyses, we will continue to monitor precision of estimates as represented in the standard errors of item locations (item difficulty) and person locations (person scores). We will also monitor the standard error of measurement conditional on location on the score scale.

Linacre, J.M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, *7*(4), 328.

Wu, M. (2010). Measurement, sampling, and equating errors in large-scale assessments. *Educational Measurement: Issues and Practice, 29*(4), 15-27.