AN INVESTIGATION INTO THE REASONS ITEM RESPONSE THEORY SCALES SHOW SMALLER VARIABILITY FOR HIGHER ACHIEVING GROUPS

by

Md Hafidz Omar

An Abstract
A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Education
(Educational Measurement and Statistics)
in the Graduate college of
The University of Iowa

December 1995
Thesis supervisor: Professor Hiram Dee Hoover

Unlike other scales, Item response Theory (IRT) scales often show decreasing variability in achievement as students progress across grades. This thesis was an investigation into why this effect occurs.

In order to control for extraneous factors related to the effect, we employed the test-item regression equation by using SAS/IML software to simulate dichotomously scored test items for both the 2- and 3-parameter IRT models. The varying item conditions for the simulations were made to reflect classical p-values and biserials like those occurring on actual multilevel test batteries. Data were simulated for two adjacent test levels with either 10, 20 , or 30 common items and 30,20, or 10 uniques items respectively at the two levels. Abilities were specified to be more variable at the upper grade level. LOGIST 5.0 and BILOG 3.0 were then used to estimate item and ability parameters for the simulated datasets.

Regardless of the estimation program used, theta scale shrinkage occurred in all but few of the simulated test datasets. The only datasets not showing scale shrinkage were ones in which the level-to-level differences in difficulty (p-values) and discrimination (biserials) were atypical for actual data.

The findings demonstrate that when IRT is used to scale real world data, the relationship of classical values of item difficulty and discrimination to their IRT counterparts, location and slope of either the logistic or normal ogive item characteristic curve, virtually guarantees lower

variability for higher achieving groups. This result holds even if the scaling is performed on unidimensional data simulated from populations where the true variability is greater for the higher achieving group. That is, scale shrinkage is an artifact of IRT Scaling.