

MEASUREMENT OF THE 'GOODNESS OF FIT' OF RASCH'S  
PROBABILISTIC MODEL OF ITEM ANALYSIS TO OBJECTIVE  
ACHIEVEMENT TESTS OF THE WEST AFRICAN SCHOOL  
CERTIFICATE EXAMINATION

Dr. Moses Adedoyin Soriyan  
West African Examinations Council, Lagos, Nigeria

Summary

Reference: Not listed (see \*)

Rasch's Model, based on the assumption of equal item discriminating powers, was measured by the probability that the deviation of the data from the expectation was more extreme than the observed deviation. The model was found to fit a number of test items of each test - Maths, Geography, Biology, English. The index of item discrimination was taken as the slope of the item characteristic curve at the point where the probability of correct response was 0.50, and the discrimination indices of items that fit the model were found to be unequal lying in the range of 0.50 - 1.25. Person free item calibration was investigated by sampling, and person free item per se (without reference to the item parameters) was found feasible.

Person free calibration of item parameters was not attained.

Item free person measurement was not achieved.

Report\*

The Rasch model of item analysis assumes that the probability that a person answers an item correctly depends only on the easiness of the item and the ability of the person with respect to the kind of items, and measures item easiness and person ability simultaneously on a ratio scale. It is equivalent to the logistic test model under the assumption of equal item discriminating powers. The Rasch model claims person-free item calibration and item-free person measurement.

This study sought to (a) measure the fit of the model to high school achievement test items, (b) find to what extent the items that fit the model have equal item discriminations, (c) investigate the invariance claims, and (d) find to what extent each achievement test measures a unidimensional trait.

The fit of the model to each item was measured by the probability that the deviation of the data from the expectation of the model was as extreme or more extreme than the observed deviation. Items with probability greater than 0.01 were selected. The model was found to fit a fairly good number of the items of each test - mathematics, geography biology, and English.

The discrimination indices of items that fit the model were found to be quite unequal, many of them lying in the range 0.50 - 1.25. The index of item discrimination was taken as the slope of the item characteristic curve at the point where the probability of correct response was 0.50.

The claim of person-free item calibration was investigated in two steps. First, for each test, the number of selected items common to three samples (whole sample, top sample, bottom sample) was expressed as a percentage of the number of items selected for the whole sample. These percentages were quite high. Secondly, top sample item easiness was plotted against bottom sample item easiness. The points tended to form a "fan pattern" having its apex near the origin.

The claim of item-free person measurement was also investigated. The abilities of the same group of persons were measured by a hard test and an easy test composed of items selected by the model. A plot of hard test ability against easy test ability showed that the points tended to form "fan pattern" also.

Each of the achievement tests was found to have high internal consistency coefficient. Also, for each achievement test, a new test composed of items selected from the original test by the Rasch model had nearly the same internal consistency coefficient (corrected for length of original test) as the original test.

The conclusions of this research were as follows:

(a) The Rasch model could fit achievement test items in the usual academic disciplines of the high school; it could select the items without altering appreciably the internal consistency of the tests.

(b) The Rasch model, despite the assumption of equal item discriminations, would select items whose discrimination indices might be unequal.

(c) Person-free item selection per se (without reference to the item parameters) was feasible, using the Rasch model.

(d) Person free calibration of item parameters was not attained.

(e) Item-free person measurement was not achieved.

---

\* A reprint from Dissertation Abstracts International Vol. XXXII, Number 8, 1972, submitted by the author.