DOCUMENT RESUME

ED 262 062 TM 850 537

AUTHOR	Phillips, Gary W.; Huynh, Huynh
TITLE	Distributional Projections: A Practical Application of the Rasch Model.
PUB DATE	Apr 85
NOTE	12p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, April 1-4, 1985).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Achievement Tests; Elementary Secondary Education; Item Analysis; Item Banks; *Latent Trait Theory; Multiple Choice Tests; Predictive Measurement; Raw Scores; Scores; *Statistical Distributions; *Test Construction; *Test Format; Test Reliability
IDENTIFIERS	*Rasch Model

ABSTRACT

A procedure which may be used to project the frequency distribution of one test onto that of another test is described and illustrated. The procedure is useful when a test developer wishes to construct an alternate form with preferred distributional characteristics. For example, the test developer may wish to construct a new test form with a completely different set of items but which yield comparable measures of central tendency, dispersion, skewness and kurtosis, reliability (KR-20, KR-21, Kappa), norms, pass-fail rates, etc. Rather than constructing alternate parallel tests, the test developer may wish to construct alternate forms which are substantially easier or harder. The procedure is also applicable in that situation. The technique assumes that items on both tests have been previously calibrated and equated with the one-parameter Rasch model. The three-parameter logistic model may also be used when number-right rather than pattern scoring is used. A basic skills multiple choice test consisting of 21 items or 32 items provides the illustration. (Author/GDC)

DISTRIBUTIONAL PROJECTIONS:

A PRACTICAL APPLICATION OF THE RASCH MODEL

GARY W. PHILLIPS
MARYLAND STATE DEPARTMENT OF EDUCATION

HUYNH HUYNH
UNIVERSITY OF SOUTH CAROLINA

1985

Paper presented at the annual convention of the National Council for Measurement in Education, Chicago, 1985.

Distributional Projections: A Practical Application of the Rasch Model

ABSTRACT

This paper describes and provides an example of a procedure which may be used to project the frequency distribution of one test on to that of another test. The procedure is useful when a test developer wishes to construct an alternate test form with preferred distributional characteristics. For example the test developer may wish to construct a new test form with a completely different set of items but which yield comparable measures of central tendency, dispersion, skewness and kurtosis, reliability (KR-20, KR-21, kappa), norms, pass-fail rates, etc. Rather than constructing alternate parallel tests the test developer may wish to construct alternate forms which are substantially easier or harder. The procedure is also applicable in this situation. The technique assumes that items on both tests have been previously calibrated and equated with the one-parameter Rasch model (the procedure may be easily extended to the three-parameter logistic model when number-right rather than pattern scoring is used).

INTRODUCTION

In this paper we address the problem of estimating the observed score distribution of an alternate form of a test. The alternate form might be a lengthened form of the original test, or be composed of a completely different set of items.

There are two common approaches to estimating observed score distributions. The first is with the negative-hypergeometric distribution within the contest of the binomial error model and multiple-matrix sampling (Lord, 1962). The second common method is to use item response theory models and assume that the conditional distribution of raw scores given Θ is a compound binomial distribution [usually estimated by a two-term approximation] (Lord, 1968, p. 524-526). The technique used in this paper also uses item response theory models. However, in the approach used here the conditional distribution of raw scores is calculated directly from the model.

Determining the Conditional Frequency Distribution

Given a test composed of m items (each with known item parameters) the conditional distribution of the raw score X can be determined This is done by letting the vector $\underline{U}$ denote the responses of an examinee with ability θ to an m-item test. The response on the ith item, Ui, is either 0 or 1. For the ith item the probability of Ui of an examinee with ability θ is

$$P(U_i|\theta) = P_i^{U_i} Q_i^{1-U_i} \qquad (1)$$

where $Q_i = 1 - P_i$.

Since it is assumed that item responses are independent the probability of any one of the 2^m observed vector of responses is equal to the product of the probabilities for each item reponse. Hence, for an examinee with ability θ the probability of the vector $\underline{U}$ is

$$P(\underline{U}|\theta) = \prod_{i=1}^{m} P_i^{U_i} Q_i^{1-U_i} \qquad (2)$$

The conditional distribution of raw scores at each ability θ, $f(X|\theta)$, can now be determined. Letting the raw score X equal $\sum Ui$ the conditional distribution of X given θ is

$$f(X|\theta) = \sum_{k=1}^{\binom{m}{X}} \left[\prod_{i=1}^{m} P_i^{u_i} Q_i^{1-u_i} \right]. \qquad (3)$$

The subscript k is used to sum each of the $\binom{m}{X}$ vectors for each θ such that k = 1, 2, ..., $\binom{m}{X}$. The upper limit of the subscript is always $\binom{m}{X}$ because there are $\binom{m}{X}$ vectors of $\underline{U}$ which result in X.

Determining the Marginal Frequency Distribution

Equations 2 and 3 may be used to make distributional projections from an old m-item test to a new n-item test. For an old m-item test let the frequencies at ability values of $\theta 0$, $\theta 1$, .., θm, be NO, N1,.., Nm. We are interested in estimating the frequencies NO, N1,.., Nn on the new n-item test at ability values $\theta 0$, $\theta 1$,.., θn. Letting Y indicate the raw score on the new n-item test equation 3 can be used to obtain

BEST COPY AVAILABLE

$f(Y|\theta)$ for values of θ associated with the raw scores on the old m-item tset. Taking the weighted mean of $f(Y|\theta)$ across the ability range from $\theta 0$ to θm will yield the projected marginal frequency distribution of raw scores $f(Y)$. That is

$$f(Y) = \sum_{j=0}^{m} f(Y|\theta j)Nj/N. \quad (4)$$

METHOD

The data for the example were obtained from a bank of calibrated items on a basic skills multiple-choice test. During the field testing of the bank a 21 item form was field tested along with a 32 item form. Both forms were administered simultaneously to two random groups of examinees. There were no common items among the two forms nor common people between the random groups. The examinee performance on the 21 item test was used to predict the frequency distribution on the 32 item test. Since the 32 item test was actually administered during the field testing it was possible to compare its predicted frequency distribution to its actual frequency distribution.

RESULTS

Table 1 shows the raw scores, ability estimates, and the predicted and actual frequency distributions on the 32 item test. A comparison of the predicted and actual frequency distributions reveal that they are very similar. In fact much of the discrepancy between the two distributions would be eliminated if the actual distribution were smoothed (note: the procedure described in this paper can also be used for smoothing distributions).

Table 2 further indicates that the predicted and actual frequency distributions are similar. The means, standard deviations, and KR-20,s are virtually identical.

DISCUSSION

THis paper described the mathematical basis and provided an empirical example of a method of estimating the frequency distribution on an alternate form of a test. The empirical example was one in which performance on a 21-item test was used to estimate derformance on a 32-item test. The two tests contained two different sets of items (from a calibrated bank) and were administered to two different groups of examinees (from the same population). The results indicated that the estimated distribution was sufficiently close to the actual distribution for practical work.

REFERENCES

Gustafson, J. E. A solution of the conditional estimation problem on long tests in the Rasch model. Educational and Psychological Measurement, 1980, 40, 337-385.

Lord, F. M. and Keats, J. A. A theoretical distribution for mental test scores. Psychometrika, 1962, 27, 59-72.

Lord, F. M. and Novick, M. R. Statistical Theories of Mental Test Scores. Reading, Mass.:Addison-Wesley, 1968.

Lord, F. M. Estimating norms by item-sampling. Educational and Psychological Measurement, 1962, 22, 259-267.

Table 1

Predicted And Actual Distribution For 32 Item Test

Raw Score	Logit Ability	Predicted Frequency	Actual Frequency
0	-5.148	.21	.00
1	-4.074	.50	.00
2	-3.299	.71	.50
3	-2.814	.80	.99
4	-2.449	.86	.50
5	-2.149	1.00	1.49
6	-1.892	1.26	1.99
7	-1.663	1.60	3.48
8	-1.454	1.95	1.49
9	-1.261	2.26	1.99
10	-1.079	2.53	1.99
11	-.906	2.79	5.47
12	-.740	3.06	4.97
13	-.579	3.33	4.97
14	-.421	3.58	4.47
15	-.266	3.80	4.47
16	-.112	3.98	2.98
17	.042	4.14	4.47
18	.196	4.28	1.99
19	.352	4.39	2.98
20	.511	4.45	2.48
21	.674	4.46	4.47
22	.843	4.44	3.48
23	1.020	4.42	2.98
24	1.207	4.38	2.98
25	1.408	4.30	2.98
26	1.628	4.15	1.99
27	1.874	3.91	4.47
28	2.154	3.62	1.99
29	2.508	3.37	6.47
30	2.973	3.37	3.98
31	3.725	3.88	3.98
32	4.776	4.19	6.47
		100.00	100.00

Table 2

Descriptive Statistics For the Predicted and Actual Distribution of the 32 Item Test

Descriptive Statistic	Predicted	Actual
Mean Raw Score	19.72	19.31
Standard Deviation Raw Score	7.62	8.26
Mean Logit Ability	.68	.69
Standard Deviation Logit Ability	1.66	1.84
KR-21	.90	.92

12