DOCUMENT RESUME

ED 074 075 TM 002 437

AUTHOR Cypress, Beulah K.

TITLE The Effects of Diverse Test Score Distribution Characteristics on the Estimation of the Rasch Measurement Model.

PUB DATE 73

NOTE 15p.; Paper presented at the annual meeting of the American Educational Research Association, New Orleans, 1973

EDRS PRICE MF-$0.65 HC-$3.29

DESCRIPTORS Achievement Tests; Data Analysis; *Measurement Instruments; *Models; *Raw Scores; Research Methodology; *Standard Error of Measurement; State Surveys; Tables (Data); Technical Reports; *Test Results

IDENTIFIERS Florida; *Rasch Model

ABSTRACT

The potential of the Rasch model to develop scores, on a ratio scale, suitable for interindividual comparisons, from intact groups with disparate distribution characteristics was investigated. The specific problems studied were: (1) the effects of skewed test score distributions on the ability parameter of the Rasch measurement model; (2) the effects of group size on the ability parameter of the Rasch measurement model; (3) the interactive effects of skewed test score distributions and group size on the ability parameter of the Rasch measurement model; and (4) the effects of skew and total group size on the standard errors of estimate of item log easiness estimates. The data for the study in the form of item responses were randomly selected from 120,000 students who participated in the Florida State-Wide Testing Program in September 1971. The 90-item mathematics test was selected. Thirty-five raw score distributions characterized by seven levels of skew and five group sizes were constructed. Group responses were submitted to a computer program which estimated the model's parameters according to a maximum likelihood procedure. Results of the study indicated that the estimates derived from the Rasch measurement model were not independent of the group used to produce them. Differences were minimal in the middle score range, but large in low and high score range. Eleven tables present the study data. (Author/DB)

THE EFFECTS OF DIVERSE TEST SCORE DISTRIBUTION CHARACTERISTICS ON THE ESTIMATION OF THE ABILITY PARAMETER OF THE RASCH MEASUREMENT MODEL

Beulah K. Cypress
The Florida State University*

The Rasch measurement model has been suggested as a method of developing achievement test norms which does not depend on the precision of sampling techniques. Due to the property of measurement objectivity, the raw ability scores estimated according to the model have been claimed to be relatively free of the group used to produce them. The potential of the Rasch model to develop scores, on a ratio scale, suitable for interindividual comparisons, from intact groups with disparate distribution characteristics was investigated.

The specific problems probed by the study were:

(1) The effects of skewed test score distributions on the ability parameter of the Rasch measurement model.

(2) The effects of group size on the ability parameter of the Rasch measurement model.

(3) The interactive effects of skewed test score distributions and group size on the ability parameter of the Rasch measurement model.

(4) The effects of skew and total group size on the standard errors of estimate of item log easiness estimates.

A paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, 1973.

*Now at the University of Illinois College of Medicine, Center for Educational Development

The data for the study in the form of item responses were randomly selected from 120,000 students who participated in the Florida State-Wide Testing Program in September 1971. The 90 item mathematics test was selected since it met the criteria specified by the model for a well constructed achievement test consisting of many items with a range of easy and difficult items. This was necessary to minimize standard errors of estimate of ability estimates. Thirty-five raw score distributions characterized by 7 levels of skew and 5 group sizes were constructed. Skew was introduced by manipulating the percentages within score intervals of a normal distribution of 1200 frequencies. This resulted in skews designated as low positive, medium positive, high positive, low negative, medium negative, and high negative. Group sizes were 1200, 600, 300, 150 and 75. Group responses were submitted to a computer program developed by Wright and Panchapakesan (1970) which estimated the model's parameters according to a maximum likelihood procedure.

The raw ability scores generated from the data of the 35 constructed groups were compared by means of a generalized distance function,

$$d = \sum \left| (C_i - G_i') \right| ,$$

where C represents the Rasch score in the criterion group, G is the Rasch score in another group, and i is the index of Rasch score estimates for raw scores 1 to 89. The magnitude of d for each comparison was expected to reveal consistent skew, size and interaction effects.

The normal distribution within each group size was the criterion for the skew effect, and the largest group size within each level of skew was the standard for the size effect. The criterion for the interactive effect of skew and group size was the group of 1200, normally distributed test scores.

To investigate the effect of group size and skew on the standard errors of estimate of the items, 10 easy and 10 difficult items common to all groups were selected and compared across normally distributed groups of all sizes for the size effect, and across the 7 levels of skew in the groups of 1200 for the skew effect.

Estimation of the size of the distance function for each comparison revealed the following:

(1) When the underlying distribution was negatively skewed, high Rasch scores were greater than the criterion scores and low Rasch scores were less than criterion scores.

(2) When the underlying distribution was positively skewed, high Rasch scores were less than criterion scores and low Rasch scores were greater than criterion scores.

(3) Differences between Rasch ability scores and criterion scores increased as skew became increasingly more positive or negative.

(4) Differences between scores generated from negatively skewed distributions and the criterion scores were larger than differences between scores generated from positively skewed distributions and criterion scores.

(5) As a group size decreased, high Rasch ability scores were greater than criterion scores and low Rasch ability scores were less than criterion scores.

(6) The size effect was not as pronounced as the skew effect.

(7) Standard errors of estimate for easy items increased when the underlying distribution was negatively skewed and decreased when the underlying distribution was positively skewed.

(8) Standard errors of estimate for difficult items decreased when the underlying distribution was negatively skewed, and increased when the underlying distribution was positively skewed.

Results indicated that the estimates derived from the Rasch measurement model were not independent of the group used to produce them. Differences were minimal in the middle score range, but large in low and high score ranges. The need to calibrate item estimates from suitable groups was shown by the model, but a paradox was revealed that increased precision of calibration of difficult items was associated with negatively skewed distributions of total raw scores, but such distributions were less desirable for good ability estimates.

TABLE 1

PERCENTAGES AND FREQUENCIES OVER RAW SCORE RANGE FOR GROUP OF 1200 STRUCTURED ACCORDING TO VARIOUS LEVELS OF POSITIVE SKEW*

Standard Deviation of Normal Distribution	Score Interval	Normal Skew %	Freq.	Low Positive Skew %	Freq.	Medium Positive Skew %	Freq.	High Positive Skew %	Freq.
-1.0 to -3.0	1 - 30	16.0	192	25.0	300	43.0	516	66.0	792
0.0 to -1.0	31 - 45	34.0	408	50.0	600	41.0	492	30.0	360
0.0 to +1.0	46 - 60	34.0	408	18.0	216	14.5	174	4.0	48
+1.0 to +3.0	61 - 90	16.0	192	7.0	84	1.5	18	1.0	12
Totals		100.0	1200	100.0	1200	100.0	1200	100.0	1200

*Application of the same percentages to groups of 600, 300, 150, and 75 reduces the frequencies proportionately. Reversing the percentages of the score intervals produces negative skew.

TABLE 2

VALUES OF d YIELDED BY COMPARING RASCH SCORES GENERATED BY NORMAL DISTRIBUTION OF RAW SCORES TO RASCH SCORES GENERATED BY 6 LEVELS OF SKEWED RAW SCORE DISTRIBUTIONS FOR EACH OF 5 GROUP SIZES

Group Size	Skew Level							
	High Negative Skew	Medium Negative Skew	Low Negative Skew	Average d Each Group Size for Negative Skew	Low Positive Skew	Medium Positive Skew	High Positive Skew	Average d Each Group Size for Positive Skew
1200	201.1395	143.6725	99.0858	145.0993	53.3879	89.7010	129.9734	96.0208
600	233.6252	154.1207	90.0909	159.3822	53.4946	89.7309	124.0277	89.1043
300	238.8032	140.7319	81.5919	153.7090	63.5976	97.8240	134.9415	99.7544
150	229.2814	186.2706	110.8007	142.1176	60.7152	99.2386	130.0305	96.6614
75	334.9470	215.9886	92.1261	214.3539	61.9612	103.9548	118.8319	94.9159

TABLE 3

VALUES OF d YIELDED BY COMPARING RASCH SCORES GENERATED BY GROUP OF 1200 SUBJECTS TO RASCH SCORES GENERATED BY 4 OTHER GROUP SIZES FOR 7 LEVELS OF SKEW

Level of Skew	Group Size				Average d for Each Level of Skew
	600	300	150	75	
Normal	4.1757	18.2183	23.8060	31.7888	19.4972
Low Positive	.3251	2.9789	12.6766	18.4172	8.5994
Medium Positive	4.3417	10.3067	14.2176	16.8260	11.4230
High Positive	.7028	3.7951	14.0657	32.6403	12.8009
Low Negative	1.1439	5.8447	40.3070	28.9624	19.0645
Medium Negative	13.7577	14.2795	65.1151	102.2770	48.8573
High Negative	36.6965	56.3073	51.7456	167.8923	78.0854

TABLE 4

VALUES OF d YIELDED BY COMPARING RASCH SCORES GENERATED FROM A NORMAL DISTRIBUTION OF 1200 RAW SCORES TO RASCH SCORES GENERATED FROM RAW SCORE DISTRIBUTIONS OF VARYING SIZE AND DEGREE OF SKEW

Group Size	Skew Level						
	High Negative	Medium Negative	Low Negative	Normal	Low Positive	Medium Positive	High Positive
1200	201.1395	143.6725	93.0853	0	50.3970	89.7310	120.5734
600	237.4627	158.2574	94.3236	4.1757	49.3465	87.5604	121.1417
300	257.4353	158.8999	99.7444	18.2183	45.4250	79.6656	115.8135
150	252.7851	209.7814	134.3180	23.8060	37.2607	75.7128	106.5441
75	367.7806	246.7861	122.9866	31.7888	30.8570	73.0492	87.9276

TABLE 5

GROUPS RANKED BY MAGNITUDE OF DEVIATION
FROM CRITERION* GROUP

Rank	Group Size	Skew Level	Value of d	Direction of Differences	
				Low Scores	High Scores
1	600	Normal	4.1757	<C	>C
2	300	Normal	10.2183	<C	>C
3	150	Normal	23.8060	<C	>C
4	75	Low positive	30.8570	>C	<C
5	75	Normal	31.7888	<C	>C
6	150	Low positive	37.2607	>C	<C
7	300	Low positive	45.4250	>C	<C
8	600	Low positive	49.3465	>C	<C
9	1200	Low positive	50.3970	>C	<C
10	75	Medium positive	73.0492	>C	<C
11	150	Medium positive	75.7128	>C	<C
12	300	Medium positive	79.6656	>C	<C
13	600	Medium positive	87.5604	>C	<C
14	75	High positive	87.9276	>C	<C
15	1200	Medium positive	89.7310	>C	<C
16	1200	Low negative	93.0853	<C	>C
17	600	Low negative	94.3236	<C	>C
18	300	Low negative	99.7444	<C	>C
19	150	High positive	106.5441	>C	<C
20	300	High positive	115.8135	>C	<C
21	1200	High positive	120.5734	>C	<C
22	600	High positive	121.1417	>C	<C
23	75	Low negative	122.9866	<C	>C
24	150	Low negative	134.3180	<C	>C
25	1200	Medium negative	143.6725	<C	>C
26	600	Medium negative	158.2574	<C	>C
27	300	Medium negative	158.8999	<C	>C
28	1200	High negative	201.1395	<C	>C
29	150	Medium negative	209.7814	<C	>C
30	600	High negative	237.4627	<C	>C
31	75	Medium negative	246.7861	<C	>C
32	150	High negative	252.7851	<C	>C
33	300	High negative	257.4353	<C	>C
34	75	High negative	367.7806	<C	>C

*Criterion group = 1200, normal skew level

TABLE 6

STANDARD ERRORS OF ESTIMATE OF LOG EASINESS ESTIMATES OF EASY ITEMS FOR NORMAL SKEW IN 5 GROUP SIZES

Item	Group Size				
	1200	600	300	150	75
1	.124	.174	.234	.288	.384
2	.100	.142	.204	.280	.384
3	.094	.132	.183	.268	.489
4	.090	.134	.187	.268	.334
5	.089	.127	.183	.257	.334
6	.088	.122	.171	.222	.311
7	.085	.125	.178	.257	.318
8	.080	.111	.165	.233	.311
9	.079	.111	.154	.219	.290
10	.079	.113	.154	.222	.290

TABLE 7

STANDARD ERRORS OF ESTIMATE OF LOG EASINESS ESTIMATES OF DIFFICULT ITEMS. FOR NORMAL SKEW IN 5 GROUP SIZES

Item	Group Size				
	1200	600	300	150	75
1	.073	.105	.160	.222	.326
2	.074	.104	.146	.207	.318
3	.075	.105	.149	.211	.318
4	.076	.103	.146	.207	.293
5	.077	.108	.165	.234	.373
6	.077	.105	.161	.227	.345
7	.079	.115	.159	.222	.318
8	.079	.112	.160	.222	.335
9	.080	.112	.152	.222	.326
10	.088	.129	.193	.301	.473

TABLE 8

STANDARD ERRORS OF ESTIMATE OF LOG EASINESS ESTIMATES OF EASY ITEMS FOR GROUP SIZE 1200 IN 7 LEVELS OF SKEW

Item	Level of Skew						
	Low Positive	Medium Positive	High Positive	Normal	Low Negative	Medium Negative	High Negative
1	.100	.085	.072	.124	.175	.214	.251
2	.087	.074	.066	.100	.128	.147	.176
3	.085	.079	.072	.094	.104	.108	.120
4	.080	.071	.064	.090	.104	.112	.123
5	.077	.069	.062	.089	.116	.133	.147
6	.078	.069	.064	.088	.109	.124	.141
7	.078	.070	.064	.085	.100	.109	.118
8	.072	.066	.061	.080	.093	.104	.113
9	.069	.063	.061	.079	.107	.132	.176
10	.069	.063	.061	.079	.106	.133	.164

TABLE 9

STANDARD ERRORS OF ESTIMATE OF LOG EASINESS ESTIMATES
OF DIFFICULT ITEMS FOR GROUP SIZE 1200
IN 7 LEVELS OF SKEW

	Level of Skew						
Item	Low Positive	Medium Positive	High Positive	Normal	Low Negative	Medium Negative	High Negative
1	.075	.077	.083	.073	.069	.068	.066
2	.075	.074	.076	.074	.072	.070	.067
3	.079	.081	.083	.075	.070	.067	.064
4	.077	.078	.079	.076	.074	.073	.069
5	.085	.090	.094	.077	.070	.066	.063
6	.075	.074	.074	.077	.077	.075	.072
7	.085	.086	.089	.079	.073	.070	.066
8	.080	.079	.081	.079	.076	.073	.067
9	.085	.087	.092	.080	.073	.074	.065
10	.098	.099	.097	.088	.079	.074	.067

TABLE 10

EFFECTS OF SKEW ON STANDARD ERRORS OF ESTIMATE
OF EASY AND DIFFICULT ITEMS

	Negative Skew	Positive Skew
Easy Items	As Skew Increases Standard Errors Increase	As Skew Increases Standard Errors Decrease
Difficult Items	As Skew Increases Standard Errors Decrease	As Skew Increases Standard Errors Increase

TABLE 11

EFFECTS OF SKEW ON STANDARD ERRORS OF ESTIMATE OF EASY AND DIFFICULT ITEMS AND ON RASCH - ABILITY SCORES

	Negative Skew	Positive Skew
Easy Items	As Skew Increases: Standard Errors Increase; and Low Scores Decrease, High Scores Increase (A)	As Skew Increases: Standard Errors Decrease; and Low Scores Increase, High Scores Decrease (B)
Difficult Items	As Skew Increases: Standard Errors Decrease; and Low Scores Decrease, High Scores Increase (C)	As Skew Increases: Standard Errors Increase; and Low Scores Increase, High Scores Decrease (D)