DOCUMENT RESUME

ED 111 868 TM 004 836

AUTHOR Ramaswamy, T.
TITLE Some Methodological Considerations in the Testing of Rasch Model Claims.
NOTE 28p.

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS *Ability; Cloze Procedure; Correlation; Hypothesis Testing; Individual Differences; *Item Analysis; Literature Reviews; Measurement Techniques; *Models; Prediction; Predictor Variables; Probability; Response Style (Tests); Statistical Analysis; *Testing
IDENTIFIERS *Rasch Model

ABSTRACT

The Rasch item analysis model is supposed to yield norm-free estimates of ability and easiness values, but there are several possible interpretations of the nature and extent of such norm-freeness. One such interpretation was that to involve the scores of one single experimental group of testees which were embedded in four differently skewed distributions of other scores, the testees had been administered easy and hard sets of items for the verifications of person-free and item-free ability estimations. Similarly for the verifications of person-and item-free easiness estimation by the Rasch model, the study involved the formation of four differently skewed sets of items in terms of their proportion right easiness values among which the same single set of experimental items had been embedded. These four sets of items were administered to bright and dumb groups of examinees. The no-guessing and constant discrimination power assumptions of the Rasch model were respectively made to be satisfied by using the cloze test blanks as items and by removing those blanks outside a narrow range of discrimination indices. Because of the possibility that the estimation errors may critically depend on the number of ties at each raw score level making the Rasch estimates of ability and easiness statistically different from one group of examinees or set of items at each of these score levels, a linear prediction model was used with the raw pupil scores or item easiness as predictors. (Author)

SOME METHODOLOGICAL CONSIDERATIONS IN THE TESTING OF RASCH MODEL CLAIMS

1. INTRODUCTION

Ever since Gulliksen's (1950) expression of the need for a response-trait model which would yield norm-free estimates of pupil-abilities and item-easinesses, there have been numerous attempts by Lazarfeld, Lord, Rasch, Birnbaum, and others in this direction. Except for the normal-ogive model of Lord, all the others are completely arbitrary mathematical functions purported to represent the binary responses of actual life. The simplest of all is that of Rasch (1966). The essence of his model is epitomized in the expression for

> Probability of getting an item of easiness e_j right by a testee of ability a_i
>
> $$P_{ij}(x) = (a_i e_j)^x/(1 + a_i e_j)$$
>
> with x equal to unity. In the above equation, x is the random variable taking values 0 or 1 according to whether the probability is for the item being scored wrong or right.

Thus, the very basic equation starts with two separate parameters of person ability a_i and item easiness e_j, with the assumption that they are independent and can likewise be estimated: easiness without regard to what sample of persons is used and ability without concern about what set of items we have at hand.*

---

* Rasch (1966), in the abstract to his paper, says "An approach to item analysis is described by means of which the difficulty of an item and the ability of an individual may sometimes be assessed without reference to the norms provided by some population." What the conditions are, to make the norm-free estimation possible are not explicitly mentioned, nor is indicated what is meant by "some population." The condition seems to be that of using the row and column frequencies of correct responses of the ordered item-person response matrix, as sufficient statistic for ability and easiness estimation; moreover, Rasch adds in the body of his paper, "..the parameter of the subjects in the subgroups may be evaluated without regard to the parameter of the other subjects; and, of course, it has already been shown these will all be independent of the item parameters. A similar statement holds for the latter." This seems to mean that there are four freenesses of Rasch estimation: person- and item-free ability and easiness estimation.

This last claim has been the one under test by several investigators and is also the subject of the present effort. This endeavor was to call attention to the various inadequacies of the past studies and correct them where it was possible.

2. REVIEW OF LITERATURE

Brooks (1965) investigated the invariance of Rasch-easiness-estimates (REE's) with respect to large ability variations in the sample of persons initially used to arrive at the easinesses (test of person-free easiness estimation). He used two samples of persons of differing mean ability and evaluated the invariance in terms of an "I-index" obtained by taking the square-root of the mean square deviations of the item-points of an empirical plot from the straight line prescribed by the Rasch model. He found that, in general, the item-points followed the theoretical line. His conclusion was based on the visual observation of the closeness of the observed and the theoretical points. The degree of closeness of these points allowing for chance errors (due to sampling, measurement, and estimation procedures) can be evaluated only by statistical inferential methods.

Wright (1968) used a different approach to corroborate the invariance of parameter estimation by the Rasch model. To verify item-free ability estimation, he gave easy and hard halves of a test to a group of Ss and estimated two sets of Rasch abilities for the same Ss from the two sets of items. The purpose of his data analysis was to see how the raw test scores (RTS's) compare with the Rasch ability estimates (RAE's). So, he found the differences, their means and standard deviations between the two sets of RAE's and the two sets of RTS's. The mean and standard deviation were small for the RAE's and large for the RTS's. This showed that the RAE's are about the same for any group of Ss, no matter what kind of items they are administered to. Here, Wright compared the means of the differences and not the individual differences themselves. Rasch might have meant invariance of RAE's for individual persons, and not for the group as represented by the mean. Moreover, small mean-difference is an artifact of the logarithmic scale of the RAE's.

To verify person-free ability, Wright determined the RAE's and percentile scores of a high and low ability group, corresponding to each RTS. He then plotted both the pairs of RAE's and percentiles (parts, because of the low and the high ability groups) against the RTS's. Whereas for the RAE's, the plots of the two groups were overlapping, they were not, for the percentile-pair. Wright concluded that for any RTS, the Rasch model gave the same estimate of ability.

Anderson et al (1968) tried to verify item-free easiness, person-free easiness, and person-free ability estimations by the Rasch model. The latter two verifications were performed by correlating the Rasch easinesses and the Rasch abilities from two groups of Ss and finding the correlation to be high (.996 and .992 respectively). It is not known how different the two groups were in the score distribution statistics like mean and the skew. Item-free easiness estimation was tested by correlating the easiness scale values with and without the items fitting the model. This correlation was found to be high again (.999). It is felt that correlation is not a precise measure of agreement between two sets of values even though it may serve to indicate the degree of relationship between the two sets.

Brink (1970) verified item-free and person-free ability estimations using simulated data with total scores being of varying standard deviations but satisfying normal distribution. He also used data of varying ranges and of rectangular distributions. He concluded that there was "no systematic differences in fit as well as no differences in values of the ability estimates." What statistical measure he used to gauge the fit is not clear. At any rate, his study shows that there is no influence of standard deviation on the Rasch estimates.

Cypress (1972) found while examining person-free ability and easiness estimations by the Rasch model, that different skews of the score distribution did affect ability and easiness estimation. The dependent variable to study the effect of person-score-skews on ability estimation was a distance measure called the sum of absolute differences. The differences were those between the RAE's corresponding to the same raw score but for two differently skewed distributions - one with known skew and the other with zero skew. Cypress used the expression given by Wright and Panchapekesan (1968) for the standard error of easiness in the investigation of person-free easiness estimation. In her study, the differently skewed distributions also had widely different means. It is not apparent which caused the observed effect on the Rasch estimates. Furthermore, comparing merely the raw sum of absolute differences does not take into consideration the sampling, estimation and other chance errors.

Tinsley and Dawis (1973) used correlational methods to determine whether RAE's were more invariant than percentiles and raw scores when they were estimated from tests that were different in item easiness (test of item-free Rasch ability estimation). They computed the correlation between each pair of the three types of scores, obtained from hard and easy parts of four kinds of analogy subtests (word, symbol, number, and picture). All the three types of scores were found to be item-dependent. The investigators attributed this to the failure of the tests to meet the assumptions of the Rasch theoretical model. To quote them, "it is illogical to assume that tests which do not fit the Rasch model will still have the characteristics attributed to it (Tinsley & Dawis, 1973)."

3. CRITIQUE OF LITERATURE

Some apparent flaws or inadequacies of the foregoing studies are listed below:

a) While Rasch meant independence or invariance of estimation in the four respects of person-free ability, person-free easiness, item-free abiltiy and finally item-free easiness, none of the investigators studied all of them.

b) Inspite of the knowledge that the Rasch model assumptions were not satisfied in their data, the investigators proceeded with their studies of verifying Rasch model claims. The assumptions of no-guessing and of constant-discrimination-power in the items of the test used, seem to be the major ones violated. That of the unidimensionality of the test might have been satisfied only in the analogy tests employed by Tinsley and Dawis (1973). Unfortunately, the meeting of the assumption of local independence of the item-responses cannot be evaluated.

c) In any verification of person-free ability or item-free easiness estimation by the Rasch model, the sample size of persons or items, and at least, the first three moments (mean, sd, and the skew) of the parameter distribution have to be considered. When the effect of one of these is being studied, the rest of them should be kept constant. None of the investigators took this into consideration (A generalized normal function is appropriate for this purpose).

d) Errors of measurement, estimation and sampling can be allowed for only in the framework of statistical inference. Such a route does not appear to have been thought of by any of the above-referenced workers.

Chi-square tests are less suitable than the parametric F-tests, since the Rasch measures are supposed to be ratio scale.

While it is true that Rasch's claims could be interpreted and tested in many ways, the "method of embedding" for the verifications of person-free ability and item-free easiness, and the usual "method of variant groups" or sets for the (cross) verifications of person-free easiness and item-free ability seemed to be useful to bring out the different nature of the issues involved. It would be edifying at this point to state the purpose of the present study explicitly and our interpretation of Rasch's claims, when these methods will also be outlined.

4. PURPOSE OF THE STUDY

> It is the purpose of this study to test through statistical inferential analysis the claims of the Rasch model estimation in the four aspects of person- and item-free ability and easiness estimation.

a,b) Person- and item-free Rasch ability estimation: For these cases, the Rasch claims are interpreted to mean that a testee or a group of testees whose scores are embedded in a series of differently skewed distributions of scores, should have the same Rasch ability estimates from any of the embedding or host distributions. This defines the case of person-free ability estimation. For item-free ability estimation, the Rasch ability estimates of the experimental Ss (who are common to all the embedding or host distributions) should be the same from the administration of either hard or easy item-sets. In effect, the object here is to study the effect of skew of the embedding subject-score distribution and that of the item-set easiness on the Rasch ability estimation.

c,d) Person- and item-free Rasch easiness estimation: The easiness estimates from the Rasch model of a given item-set should be the same whether they are estimated from low or high ability Ss. Item-free easiness estimation by the Rasch model might be said to be true if a particular set of items common to several distributions in easiness values have their Rasch easiness estimates unaffected by their presence in these host distributions. In summary, our aim here is to study the effect on Rasch easiness estimates by the skew of the embedding easiness distribution and by the ability of the subject-group used to arrive at the estimates.

5. METHOD OF ANALYSIS

a, b, c, d) i) Data: The subjects were 226 pupils in the 4th, 5th, and 6th grades of three different schools. There were approximately equal number of girls and boys in the sample. These pupils had taken five cloze tests* in social studies, out of which, one of moderate difficulty was chosen as the "experimental cloze test" (ECT) as this was going to be used for the primary purpose of estimating Rasch scores and in the subsequent inferential analysis. The other four were called the auxiliary cloze tests (ACT's) since these were to be used for the secondary purpose of creating the various skewed distributions. The ECT and the ACT's had all about 250 words and about 50 blanks. The blanks of the ECT were considered as "items" and the free responses to them were scored zero or one for wrong or right restoration of the deleted word.

Since the Rasch model estimates can be expected to follow Rasch's claims only if the model assumptions are met, the assumption of equal discrimination was forced by discarding items having a discrimination index outside a small range of 0.15 from the ones with the maximum set of

---

*Cloze tests are a kind of reading comprehension tests wherein every n'th word is systematically deleted from a passage and a blank of standard length is left in its place; pupils fill up the blanks using the context of the words on both sides of the blanks. The percent score of a pupil over the number of blanks on a randomly selected passage of a book can be taken as a reasonable measure of reading ability of the pupil with respect to that book.

observed values. The allowed range of 0.15 was taken as a reasonably small spread in a compromise to maintain constant discrimination power for all the items demanded by the model. As far as the assumption of guessing being small goes, this can be expected to be fairly well satisfied in the case of free responses, since the number of alternatives or choices or options per item can be taken as extremely large. The reciprocal of this number will be small, which reciprocal is usually taken as the probability of chance response to an item. The assumption of unifactoralness can be expected to be also satisfied in cloze test responses. Thus, once it had been seen to that that the assumptions of the Rasch model were all fairly well met, the claims were ready to be tested.

The scores on the four auxiliary cloze tests (ACT's) were dichotomized on a criterion score.* Their totals ranging from 0 to 4 were used as premeasures to create four distributions of four different skew magnitudes but of similar mean, standard deviation, and total sample size. The parent distribution had a total of 226 Ss. The four host** distributions were created by calculating the numbers (frequencies) to go in each of these five score groups so as to give the same reduced total of 189 Ss and a constant mean. The calculated frequencies were randomly drawn afresh 4 times from within each of the five score groups of the parent distribution. In order to alter the skew magnitude, the trend in the frequency distribution

---

*The ACT's were used as premeasures to create the differently skewed distributions. It is customary in the literature of cloze test studies to use the 36% score as the criterion to denote the demarcation between those who can "read" the passage and those who cannot. This was the criterion used here too. But such binary scores are not in anyway superior to the raw scores themselves apart from being convenient to calculate the means in the trial-and-error creation of the four skew distributions.

**The qualifier "host" is used here to indicate that these distributions housed the experimental Ss who were common to all of them.

of ACT scores change in skew in a certain direction was noted. And then, either the low or the high scores, as was noted, were removed or added randomly to the score-group of the host distribution under manipulation.

a,b) ii) Procedure to test person-free and item-free ability estimation by the Rasch model. Of the total 51 blanks or items, 11 were discarded to satisfy the constant-discrimination-power assumption demanded by the Rasch model. The proportion-right easiness values were ranked and divided into three thirds to give three item-sets of hard, medium and easy items. The Ss in the 4 skew groups (host distributions) were "administered" all the three sets of items. Due to the particular manner of the construction of the 4 skew groups, there happened to be 26 Ss who belonged to all the 4 skew groups and had nonzero and non perfect scores in all the three item-sets. These formed the "experimental Ss" who yielded repeated measures on the pseudo-factors of skew and easiness (below). The Rasch ability estimates were the dependent variable(s). The correlation between these RAE's and the ECT raw score totals was controlled by making the latter a predictor in a linear prediction model. The correlation <u>among</u> the dependent variable(s) for the 12 conditions of three easiness levels and four skew levels was taken into account by making these 12 sets as dependent variables, with the assumption that they obey a 12-variate normal distribution with a common variance-covariance matrix (tables 6 a,b).

The design is, thus, a multivariate linear prediction model as represented by the scalar equation:

$$Y_{ij} = \alpha_i + \beta_i X_j + \varepsilon_{ij}$$

or the vector equation:

$$\vec{Y}_i = \alpha_i + \beta_i \vec{X} + \vec{\varepsilon}_i$$

Here,

Y_{ij} = the dependent variable (Rasch ability estimate in the logarithmic scale)

$i = 1, 2...12$; index for the 12 conditions of 4 skews and 3 easiness levels.

$j = 1, 2...35$ (maximum reached on the 50 blank ECT); index for raw total scores on the ECT

$X_j =$ the predictor; the ECT, assumed to be error-free

$\alpha_i =$ the population value of intercept-like parameter; determines the Y_{ij} zero value of X_j

$\beta_i =$ the population value of slope-like parameter (direction number); determines the rate of variation of Y for unit variation in X_j

$\varepsilon_{ij} =$ population value of residuals; deviation of the predicted Y_{ij} from the observed Y_{ij}; assumed to be normally and independently distributed.

This linear prediction model would permit the testing of equality of RAE's for the individual rather than the over-all mean by dint of the use of the concept of "fit" of the regression lines. The null and the research hypotheses are that the intercepts and the slopes should be the same for the experimental Ss immaterial of what skew host they belong to and irrespective of how hard or easy items that are administered to them. In essense, the hypotheses are meant to test whether the prediction lines are collinear. If there is noncollinearity owing to errors of measurement and estimation, the effect size of the departure in the population should be minute. Obviously, a large sample size should be used for this kind of goodness-of-fit like hypothesis. In other words, consideration of power or beta (type II) error is very relevant here. High power or low beta error should be aimed at. As far as type I error goes, we can set the alpha probability high (say, 0.4) to give the data at hand every chance of rejecting the proposed null, even though we are really interested in retaining it or failing to reject it. The multiple hypotheses involved (in testing separately for the equality of intercepts and the slopes and then these, separately within each skew level for all the easiness levels and also within each easiness level for all the skew levels) should automatically increase the alpha error for the whole experiment, even if that for the individual hypothesis is set low. The multivariate design is illustrated in figure 1.

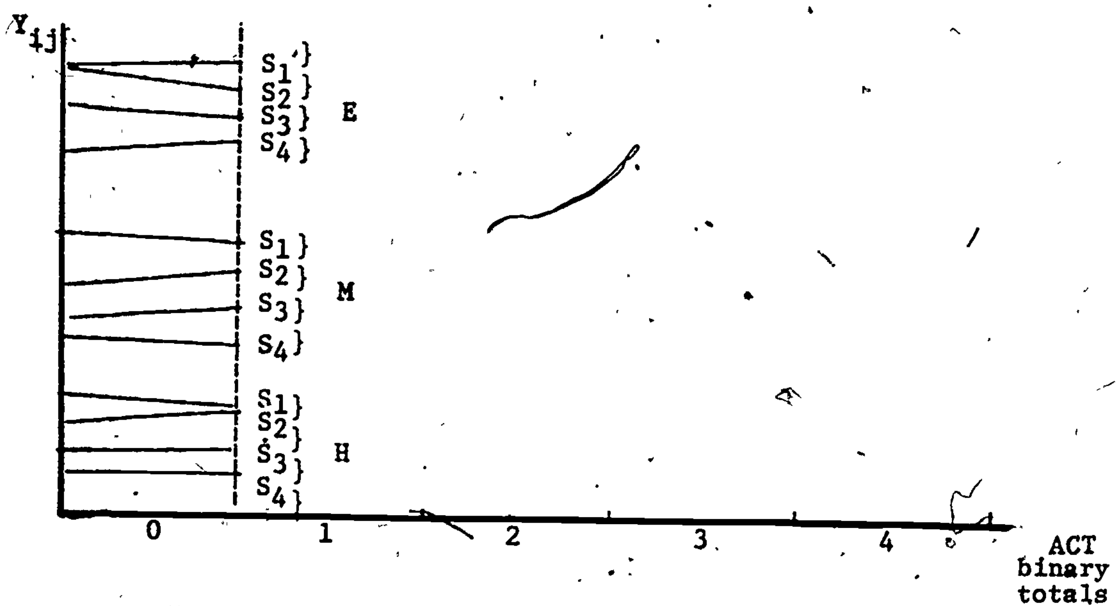

Fig. 1: Schematic diagram showing the linear prediction model for 12 dependent variables of 3 easiness levels and 4 skew levels; illustration for one score group of the auxiliary cloze tests' dichotomized score totals.

Legend: S_1, S_2,...S_4 indicate the four skew levels; E, M, & H denote the three easiness levels.

The null or the research hypotheses are symbolically represented in the following equations:* for any one score group of the ACT (see fig. 1),**

---

*In principle, a third hypothesis should also be tested on the equality of correlated variances (residual or predicted) or equivalently the equality of tightness of fit; all the three hypotheses together will go to test for the collinearity of the prediction lines. The pertinent tests for the last of these (to test the equality of correlated variances) is given by Anderson (1955) in the form of a hypothesis equating two variance-covariance matrices.

**If there is any reason to believe that the prediction lines are of equal intercept and slope across all the five score groups of the auxiliary cloze tests, then the appropriate tests are the likelihood ratio tests given by Gulliksen & Wilks (1950). This kind of different test is dictated for the reason that the assumption of random sampling does not hold good across these score groups which are merely different sections of the same bivariate distribution of the predictor and the predictand in the prediction system.

First, we test whether the intercepts are equal:

$$\alpha_{s1e1} = \alpha_{s1e2} = \ldots = \alpha_{s4e3}$$

Second, we test whether the corresponding slopes are equal:

$$\beta_{s1e1} = \beta_{s1e2} = \ldots = \beta_{s4e3} \quad \text{in the population.}$$

Here, "s" subscripts denote the skew levels and the "e" ones the easiness levels.

To get the 12 sets of the Rasch ability estimates, 12 runs of the Rasch item analysis program (Wright & Panchapekesan, 1968) were made, administering Ss in the four skew distributions items in the three item-sets. As it was thought that the degrees of freedom for the denomenator in the F-tests would be larger if only 8 dependent variables were used, the number of easiness levels was decreased from 3 to 2 for the important tests of person- and item-free ability estimation.

The Fortran program called "Mulgen" was used for the tests of contrast and the over-all-hypotheses. Only the experimental Ss in the zero group of ACT were used for the analysis. The latter was done on these 26 Ss common to the 4 skew groups and with nonzero and nonperfect scores on the 3 item-sets.

I c,d) Procedure to test person-free and item-free easiness estimation by the Rasch model: The procedure here was somewhat similar to that in testing Rasch ability estimation. Of the 30 items in the parent distribution, four host distributions were created, each containing 15 items. By trial-and-error, using the proportion correct easiness values as the measures for consideration, these 4 distributions "skew-sets" were constructed (table 1c). Five items common to these skew-sets formed the "experimental items". Three ability groups were created by ranking the ECT total scores and taking the three thirds (one of the groups had 76 Ss while the other two had 75 Ss each, making the total of 226 Ss). Rasch easiness estimates* were the dependent variables for this part of the study. These estimates under 4 different skew conditions enabled the testing of item-free (skew-

---

*The correlations among these Rasch easiness estimates are in table 7.

free) easiness estimation by the Rasch model. The same estimates measured under three ability conditions made possible the testing of person-free Rasch easiness estimation. The twelve sets of Rasch easiness were obtained by twelve runs of the Rasch item analysis program as prepared by Wright & Panchapekesan (1968). Again, the fortran program, meant for testing of multivariate general linear hypotheses, was used in the testing of equality of intercepts and slopes of the prediction lines.

6. RESULTS

Irrespective of the total number of individual contrast hypotheses tested, the type I error rate for these individual tests was set at .05 level (Conclusions will not change even if it is set at .01 level, excepting that there is no tabled F, for the degrees of freedom, 3 and 1, at .01 level for the tests of Rasch easiness estimation). Assuming our criterion of collinear lines does indeed go to prove Rasch's claims, we would expect failure to reject the nulls in all the individual contrast-tests, if his claims were true.

Now, going to table 2, where the case of person-free ability estimation is reported, we find that rejection of the null is in order. Within each level of the item-sets, the Rasch ability estimates are not statistically the same across the skew-groups. In other words, skew does have an influence on the Rasch ability estimation; or, Rasch model is not skew-free in the aspect of ability estimation.

Looking at table 3, where item-free ability estimation by Rasch model is tested, we see that the slope and the intercept hypotheses are not rejected while the over-all hypothesis is. Since these two types (individual and the over-all) of tests have different degrees of freedom, they cannot be compared. With reservations, it might be said that Rasch abilities are free of what items are employed to arrive at them.

The fact that at least one set of hypotheses for the tests of ability estimation was rejected (as in table 2), is reflected in the over-all test of both person- and item-free ability estimation, as in table 2'-3'.

Table 4 gives the tests for the case of person-free easiness estimation. Noting that the degrees of freedom are rather small especially for the denomenator of the F-statistic, we find that the data used do not support the null that the Rasch easiness estimates are unaffected by the diversity of person scores employed.

In table 5 are reported the tests of item-free easiness estimation by the Rasch model. The intercepts are statistically different while the slopes are not (at the level of significance chosen). Hence these parallel lines may be interpreted to mean that there is the effect of an additive constant in this case. By our criterion of collinear lines, Rasch easiness estimates are <u>not</u> item-free or skew-free.

To summarize, we conclude that item-free ability estimation by the Rasch model alone may have been supported by the data used in this study.

While the estimates of intercepts and the slopes appear to be practically about the same, they are not statistically the same in most cases. They need to be closer in magnitudes in the light of high correlations among the dependent variables (namely, the Rasch estimates for the conditions of person-score-skew, item-easiness-skew, the ability, and the easiness levels). But taking into account these correlations puts probably too great a constraint on the verifications.

7. LIMITATIONS OF THE STUDY

a) The sample size is small for the Rasch model verifications and this can be considered serious only in those tests where there was failure to reject the null. This is to ensure or ascertain that the true null is retained, in these cases, even though both the null and the research hypotheses were the same (like in the goodness-of-fit tests) in all verifications.

b) There is, of course, the logical axiom that one can only disprove any proposition and not prove it. On this account, the very basis of this study is shaky. More proper would have been a mathematical proof of the invariance of the Rasch estimates with respect to the considered manipulated variables using one or more of the invariance theorems.

c) The creation of the various distributions of the different skews from one parent distribution can produce the observed skew differences merely by the effect of random sampling; the different skews that were produced from one parent distribution were solely an artifact of the sampling fluctuations and for this reason, the skew magnitudes were close to one another. This problem could have been circumvented by generating artificial data to create several parent distributions which themselves could be used to host the experimental Ss.

d) The manner of creation of the host distributions of several skew magnitudes might be questionable. Trial-and-error method of manipulating the score distributions might be considered as quasi-random and not completely random. This might have affected the assumption of the independent chi-squares in the formation of the F-statistic.

e) In retrospective reflection, it is thought that it was not necessary to have attempted to keep the first two moments (namely, the mean and the standard deviation) of the parameter distributions constant. Any combination of these might be varied as would be true in real life norm distributions. Studying the effect of the variation of the skew of the distributions alone or their mean alone would be only of theoretical or academic interest.

f) A plot of the Rasch abilities against the raw scores will be a logistic curve. But in this study, straight lines were fitted and may not ~~not~~ have been appropriate.

g) In the ability verifications, the predictor variable for the fit of the straight lines was the totals in the 30 item experimental cloze test, even though the Rasch abilities had been derived from the respective subtest scores (easy, medium and hard subtests). These subtest totals would have been better predictors. In fact, going to table 6b and looking at the middle two columns, we see that the Pearson correlation between the medium subtest scores and the corresponding Rasch ability estimates is more or less uniformly .99 and that between the hard subtest scores and their corresponding Rasch abilities is very uniformly .98. But the correlations between the whole test scores and the medium-subtest Rasch abilities or the hard-subtest Rasch abilities are lower (nearly .79 and .59 respectively). In Appendix is reported an approximate repeat of all the previous inferential analyses using these subtest scores as predictors. It will be seen that our conclusion on the possibility of item-free Rasch ability estimation was nullified.

In view of the above limitations, the results of this study can, at best, be regarded as tentative and of restricted value. It is regrettable that even this elaborate effort could not pronounce the last word on the issue of the veracity of the Rasch model claims.

APPENDIX

In the Rasch ability verifications, we had used the 30 item scores as the predictors instead of the 10 item subtest scores. In view of the higher correlations of the Rasch abilities with the latter, the predictive power (predictable variance) will be increased if these subtest scores were used as predictors. Moreover, subtest scores are the more legitimate predictors for the reason that the Rasch abilities were in reality derived from them.

Therefore, a reanalysis was made on the inferences of the equality of "intercepts" alone at various specific values of the predictor. As expected the error variances dropped dramatically (to the range .02-.56) from what they were (range 2-12) when the whole-test scores were the predictor. The "intercepts", slopes and the F-statistic for the test of the equality of these intercepts (actually, the predicted means of the conditional distributions of the Rasch ability at specified predictor values) are given in tables 2', 3' and 2"-3". It is seen that the null hypotheses of the equal predicted-mean-Rasch abilities are contradicted by the data in all cases of the 4 skew levels within the 2 easiness levels and of the 2 easiness levels within each of the 4 skew levels. Going back to table 3, the rejection of the null in the over-all tests seems to be supported by this reanalysis. Thus, Rasch's claims are contradicted in all the four respects of person- and item-free ability and easiness estimations. But a word of caution. The high and sometimes inordinately high F values are due to the high and sometimes perfect correlations among the Rasch estimates (as in the bottom line of table 6c) and between the latter and the predictors.

Table 1a. Distribution statistics for the testing of person-free (skew-free) ability estimation: (auxiliary cloze test - ACT)

statistic	skew group-1	skew group-2	skew group-3	skew group-4
mean	1.08	1.03	1.07	1.25
standard deviation	1.14	1.09	1.21	1.28
SKEW	.71	.80	.95	.53
kurtosis	-.39	-.22	-.12	-.92

Table 1b. Frequency distribution in the four skew-groups (26 Ss common to all the 4 skew-groups)

total binary score in ACT	skew-group-1	skew group-2	skew group-3	skew group-4
0	82	79	82	82
1	36	50	50	22
2	52	40	29	52
3	12	15	17	22
4	7	5	11	11
total	189	189	189	189

Table 1c. Distribution statistics for the testing of item-free (skew-free) easiness estimation: (raw easiness in units of proportion correct) (5 items common to all the 4 skew-sets)

statistic	skew-set - 1	skew-set - 2	skew-set - 3	skew-set - 4
mean	.39	.40	.37	.37
standard deviation	.17	.25	.20	.17
SKEW	.46	.32	.44	.17
kurtosis	-.49	-1.11	-.99	-.95

Table 1b': Rasch ability estimates for the 4 skew levels and three item-sets: easy, medium and hard , each of ten items: (zero and perfect scores are omitted in the Rasch estimation procedure):

RAW SCORE	easy item-set		medium item-set		hard item-set	
	log	antilog	log	antilog	log	antilog
skew-group-1						
1	-2.356	.095	-2.337	.097	-2.288	.102
2	-1.495	.224	-1.483	.227	-1.453	.234
3	-0.911	.402	-0.905	.405	-0.891	.410
4	-0.429	.651	-0.426	.653	-0.426	.653
5	0.014	1.014	0.012	1.013	0.002	1.002
6	0.453	1.573	0.448	1.565	0.430	1.537
7	0.926	2.524	0.917	2.502	0.893	2.442
8	1.493	4.452	1.481	4.396	1.453	4.275
9	2.330	10.279	2.314	10.117	2.284	9.815
skew-group-2						
1	-2.336	.097	-2.355	.095	-2.323	.098
2	-1.483	.227	-1.496	.224	-1.482	.227
3	-0.905	.405	-0.912	.402	-0.912	.402
4	-0.427	.653	-0.427	.651	-0.439	.645
5	0.011	1.011	0.014	1.015	0.001	.999
6	0.447	1.563	0.454	1.575	0.437	1.547
7	0.916	2.500	0.927	2.526	0.911	2.487
8	1.481	4.397	1.493	4.451	1.482	4.403
9	2.315	10.129	2.329	10.268	2.325	10.231
skew-group-3						
1	-2.336	.097	-2.367	.938	-2.281	.102
2	-1.485	.226	-1.504	.222	-1.449	.235
3	-0.909	.413	-0.916	.400	-0.889	.411
4	-0.431	.650	-0.430	.651	-0.426	.653
5	0.008	1.008	0.016	1.016	0.001	1.001
6	0.445	1.561	0.458	1.581	0.427	1.532
7	0.917	2.502	0.932	2.540	0.889	2.433
8	1.484	4.410	1.500	4.483	1.449	4.258
9	2.321	10.186	2.337	10.352	2.280	9.776
skew-group-4						
1	-2.332	.097	-2.348	.096	-2.318	.099
2	-1.482	.227	-1.492	.225	-1.478	.228
3	-0.905	.404	-0.910	.402	-0.909	.403
4	-0.429	.651	-0.429	.651	-0.436	.646
5	0.009	1.000	0.012	1.012	0.000	1.000
6	0.445	1.561	0.450	1.569	0.436	1.546
7	0.915	2.497	0.922	2.516	0.908	2.480
8	1.480	4.393	1.489	4.434	1.478	4.383
9	2.315	10.125	2.326	10.236	2.318	10.160

Table 2. Slopes, intercepts and observed & critical F values: case of person-free ability estimation by the Rasch model across 4 skew levels in ability distribution: 4 dependent variables of Rasch ability estimates (RAE's): (the predictors were the totals in the 30-item ECT - sum of easy, medium and hard subtest scores):

	estimates for	skew group-1	skew group-2	skew group-3	skew group-4	observed	critical
						F(3,22)	
edium set	intercept	-3.1	-3.0	-3.2	-3.2	22.4 (13.8* (6,44))	
	slope	.18	.17	.19	.19	18.0	1.5 (75%) 2.4 (90%) 3.1 (95%) ← 4.2 (99%)
ard set	intercept	-3.1	-3.2	-3.1	-3.2	5.7 (33.6* (6,44))	
	slope	.10	.11	.10	.10	5.7	

*significant at .05 level

Table 3. Slopes, intercepts and observed & critical F values: case of item-free abil estimation by the Rasch model across 2 easiness levels: 2 dependent variables of Rasch ability estimates (RAE's): (the predictors or X's were the totals in the 30-item ECT; that is, the sum of easy, medium and hard subtests):

	estimates of	medium-easiness item set	low-easiness (hard) item set	observed	critical
				F(1,24)	
kew roup-1	intercept	-3.1	-3.1	.00 (10.5 (2,24))*	
	slope	.18	.10	2.4	
kew roup-2	intercept	-3.0	-3.2	.04 (9.4 ")*	1.4 (75%)
	slope	.17	.11	1.6	2.9 (90%)
skew group-3	intercept	-3.2	-3.1	0.02 (10.1(2,24))*	4.3 (95%) 7.8 (99%)
	slope	.19	.10	2.6	
skew group-4	intercept	-3.2	-3.2	.00 (10.7(2,24))*	
	slope	.19	.10	2.4	

*significant at .05 level

Table 2'-3' Slopes, intercepts and observed & critical F values: case of person- & item-free ability estimation by the Rasch model across 4 skew levels and 2 easiness levels 8 dependent variables of Rasch ability estimates (R'E's):

estimates of	skew-1 med.	hard	skew-2 med.	hard	skew-3 med.	hard	skew-4 med.	hard	observed F(7,18)	critical
intercept	-3.2	-3.1	-3.1	-3.1	-3.2	-3.1	-3.2	-3.1	27.6 (40.7* (14,36))	1.4 (75%) 2.1 (90%) 2.6 (95%) ← 3.8 (99%)
slope	.18	.11	.17	.11	.18	.11	.18	.11	12.8	

*significant at .05 level

Table 4. Slopes, intercepts and observed & critical F values: case of person-free easiness estimation by the Rasch model; 3 ability levels (in 4 separate sets of items): 3 dependent variables

	estimate for	low ability	medium ability	high ability	observed F(3,1)	critical
skew set-1	intercept	-2.3	-1.9	-2.1	4051.2 (49.6 (6,2))*	
	slope	.07	.05	.05	3526.4	
skew set-2	intercept	-2.7	-2.2	-2.0	5016.4 (59.9 ")*	8.2 (75%)
	slope	.07	.05	.05	3461.1	53.6 (90%) 216.0 (95%)
skew set-3 b	intercept	-2.2	-1.8	-1.9	3908.5 (42.7 ")*	
	slope	.07	.05	.05	3836.4	
skew set-4 b	intercept	-2.4	-1.8	-1.9	3986.9 (47.6 ")*	
	slope	.07	.05	.05	3825.9	

*significant at .05 level

Table 5. Slopes, intercepts and observed & critical F values: case of item-free easiness estimation by the Rasch model across 4 skew values in easiness distribution: 4 dependent variables

	estimate for	skew set-1	skew set-2	skew set-3	skew set-4	observed F(3,1)	critical
l low group	intercept	-2.3	-2.7	-2.2	-2.4	> 2000 (> 2000 (6,2))*	8.2 (75%) 53.6 (90%) 216.0 (95%)
	slope	.07	.07	.07	.07	3.5	
m medium group	intercept	-1.9	-2.2	-1.8	-1.8	> 2000 (> 2000 ")	
	slope	.05	.05	.05	.05	7.2	
h high group	intercept	-2.1	-2.0	-1.9	-1.9	> 2000 (> 2000)*	
	slope	.05	.05	.05	.05	.73	

*significant at .05 level

Table 2' Slopes, intercepts and observed & critical F values: case of person-free ability estimation by the Rasch model across 4 skew levels in ability distribution: 4 dependent variables of Rasch ability estimates (RAE's): ($X=X_1$ or X_2 according to whether medium or hard subtest is used; X_1 = medium subtest scores; X_2 = hard subtest scores) (inferential test was at the predicted means of RAE's at the given "X" values)

	estimates for	skew group-1	skew group-2	skew group-3	skew group-4	observed F(3,22)	critical F(3,22)
medium set	intercept	-2.70	-2.67	-2.74	-2.71	2456.1 (X=2) 36.2 (X=5) 429.6 (X=8)	1.5 (75%) 2.4 (90%) 3.1 (95%) ← 4.2 (99%)
	slope	.55	.53	.56	.55		
hard set	intercept	-2.80	-2.84	-2.79	-2.83	> 10^7 (X=2) § > 10^7 (X=5) § > 10^7 (X=8) §	
	slope	.57	.58	.57	.58		

§These inordinately high F's are due to the perfect correlations among the Rasch estimates of ability (see bottom line of table 6c).

Table 3' Slopes, intercepts and observed & critical F values: case of item-free abili estimation by the Rasch model across 2 easiness levels: 2 dependent variables of Rasch ability estimates (RAE's): ($X = X_1$ and X_2 are the predictors; X_1 = medium subtest scores; X_2 = hard subtest scores) (inferential test was at the predicted means of RAE's at the specified "X" values)

	estimates of	medium-easiness item set	low-easiness (hard) item set	observed F(1,23)	critical F(1,24)
skew group-1	intercept	-2.66	-2.72	439.5 (X=2)	
	coefficient-1	.56	- .03	1367.1 (X=5)	
	coefficient-2	- .05	.58	1228.1 (X=8)	
skew group-2	intercept	-2.57	-2.76	680.8 (X=2)	
	coefficient-1	.56	- .03	1884.9 (X=5)	1.4 (75%)
	coefficient-2	- .11	.59	1641.0 (X=8)	2.9 (90%)
skew group-3	intercept	2.69	-2.71	417 (X=2)	4.3 (95%) ←
	coefficient-1	.57	- .03	1355.2 (X=5)	
	coefficient-2	- .05	.58	1230.9 (X=8)	
skew group-4	intercept	-2.67	-2.76	463.6 (X=2)	
	coefficient-1	.56	- .03	1409.9 (X=5)	
	coefficient-2	- .05	.59	1259.2 (X=8)	

Table 2"-3". Intercepts, slopes, observed and critical F values: case of person- and item free ability estimation by the Rasch model across 4 skew levels and 2 easiness levels: 8 dependent variables of Rasch ability estimates (RAE's): (X = X_1 and X_2 are the predictors; X_1 = medium subtest scores; X_2 = hard subtest scores) (inferential tests were at the predicted means of RAE's at the specified values of "X"):

estimates of	skew-1		skew-2		skew-3		skew-4		observed	critical
	med.	hard	med.	hard	med.	hard	med.	hard	F(7,17)	
intercept	-2.66	-2.72	-2.57	-2.76	-2.69	-2.71	-2.67	-2.76	$>10^7$§ (X=2)	1.5 (75%)
coefficient-1	.56	-.03	.56	-.03	.57	-.03	.56	-.03	$>10^7$§ (X=5)	2.1 (90%) 2.6 (95%) ←
coefficient-2	-.05	.58	-.11	.59	-.05	.58	-.05	.59	$>10^7$§ (X=8)	3.9 (99%)

---

§ These tests are meaningless since the inordinately high F's are forced by the the perfect correlations as in the bottom line of table 6c among Rasch ability estimates.

6a) Correlations among the independent variables: X_1=medium subtest scores; X_2= hard subtest scores; X=whole test scores:

	X_1 and X_2	X_1 and X	X_2 and X
Pearson r	.31	.80	.62

6b) Correlations between the dependent and independent variables: Raschy1=Rasch abilities for medium subtest; Raschy2=Rasch abilities for hard subtest; s1.. ...s4=skew-group-1 ... skew-group-4:

Pearson r for	X_1 and Raschy2	X_1 and Raschy1	X_2 and Raschy2	X_2 and Raschy
s1	.24	.99 (.79)*	.98 (.58)*	.25
s2	.24	.98 (.76)*	.98 (.59)*	.17
s3	.24	.99 (.79)*	.98 (.58)*	.25
s4	.24	.99 (.79)*	.98 (.59)*	.25

*parenthetical r's are those between X (whole test scores) and the Rasch abilities

Typed

6c) Correlations among the dependent variables, the Rasch ability estimates, for 4 skew levels and 2 easiness levels:

	skew-group-1		skew-group-2		skew-group-3		skew-group-4
Raschy1		.996		.996		1.0	
	.18		.11		.18		.18
Raschy2		1.0		1.0		1.0	

7) Typical Correlations among the dependent variables of Rasch easiness estimates

ability group	skew-set-1		skew-set-2		skew-set-3		skew-set-4
low		1.0		1.0		1.0	
	.59		.60		.60		.60
medium	(.08)*	1.0	(.07)*	1.0	(.08)*	1.0	(.08)*
	.49		.49		.49		.49
high		1.0		1.0		1.0	

*parenthetical r's are those between the low and the high ability levels but within a particular skew-set

REFERENCES:

1. Anderson, T.W., An introduction to multivariate statistical analysis. New Delhi: Wiley Eastern Private limited, 1972.

2. Anderson, J., Kearney, G.E., & Everret, A.V., An evaluation of Rasch's structural model for test items. The British Journal of Mathematical and Statistical Psychology, 1968, 21, 231-238.

3. Brink, N.E., Characteristics of the Rasch's logistic model. ERIC microfiche, ED 042 804, 1970.

4. Brooks, R.D., An empirical investigation of the Rasch ratio-scale model for item-difficulty indices. Unpublished doctoral dissertation, University of Iowa, 1965.

5. Cypress, B.K., The effects of diverse test score distribution characteristics on the estimation of the ability parameter of the Rasch measurement model. Unpublished doctoral dissertation, Florida State University, 1972.

6. Gulliksen, H., & Wilks, S.S., Regression tests for several samples. Psychometrika, 1950, 15, 91-114.

7. Rasch, G., An item analysis which takes individual differences into account. The British Journal of Mathematical and Statistical Psychology, 1966, 19, 49-57.

8. Tinsley, H.E.A., & Dawis, R.V., Test-free person measurement with the Rasch simple logistic model. ERIC microfiche, ED 069 785, 1973.

9. Wright, B., Sample-free test calibration and person measurement. Unpublished manuscript, University of Chicago, 1968.

9. Wright, B., & Panchapekesan, N., A procedure for sample-free item analysis. Unpublished manuscript, University of Chicago, 1968.