Best Procedures For Sample-Free Item Analysis

Benjamin D. Wright
University of Chicago

Graham A. Douglas
University of Western Australia

Wright's (1969) widely used "unconditional" procedure for Rasch sample-free item calibration is biased. A correction factor which makes the bias negligible is identified and demonstrated. Since this procedure, in spite of its superiority over "conditional" procedures, is nevertheless slow at calibrating 60 or more items, a simple approximation which produces comparable estimates in a few seconds is developed. Since no procedure works on data containing persons or items with infinite parameter estimates, an editing algorithm for preparing item response data for calibration is appended.

Wright and Panchapakesan (1969) described a procedure for sample-free item analysis based on Rasch's simple logistic response model (Rasch, 1960, 1966a, 1966b) and listed its essential FORTRAN segments. Since then that procedure has been incorporated in a number of computer programs for item analysis and widely used in this country and abroad.

The procedure described is incomplete in two minor but important ways. It mentions neither the editing of incoming item response data necessary to remove persons or items whose parameters will have infinite estimates nor the bias in estimation (Andersen, 1973) which the procedure entails, because the estimation equations are not conditioned for person ability. In addition, although the Wright-Panchapakesan procedure has proven far more efficient and practical than any other method reported in the literature (Rasch, 1960; Andersen, 1972), it can nevertheless take a quarter of an hour on a small computer to handle a large problem. An approximation which provided accurate calibrations, but executed more rapidly, would be useful.

We will refer to the Wright-Panchapakesan procedure as the "unconditional" solution, UCON. We will review UCON, discuss its inadequacies, report on an investigation of its bias, and describe a new "approximate" calibration procedure, PROX, which we have found to be as accurate as UCON under ordinary circumstances and much more rapid.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 1, No. 2 Spring 1977 pp. 281–295

The Unconditional Procedure

The Rasch model for binary observations defines the probability of a response x_{vi} to item i by person v as

$$P\{x_{vi}|\beta_v,\delta_i\} = e^{x_{vi}(\beta_v-\delta_i)}/(1 + e^{\beta_v-\delta_i}) \qquad [1]$$

where $x_{vi} = \begin{cases} 1 & \text{if correct} \\ 0 & \text{otherwise,} \end{cases}$

β_v = ability parameter of person v,

δ_i = difficulty parameter of item i.

The likelihood of the data matrix (x_{vi}) is the continued product of Equation 1 over all values of v and i, where L is the number of items and N is the number of persons with test scores between 0 and L, since scores of 0 and L lead to infinite ability estimates.

$$\Lambda = \prod_v^N \prod_i^L p_{vi} = e^{\sum_v^N \sum_i^L x_{vi}(\beta_v-\delta_i)} / \prod_v^N \prod_i^L (1 + e^{\beta_v-\delta_i}). \qquad [2]$$

Upon taking logarithms and letting

$\sum_i^L x_{vi} = r_v$ be the score of person v

and $\sum_v^N x_{vi} = s_i$ be the score of item i,

the log likelihood becomes

$$\lambda = \log\Lambda = \sum_v^N r_v\beta_v - \sum_i^L s_i\delta_i - \sum_i^L \sum_v^N \log(1 + e^{\beta_v-\delta_i}). \qquad [3]$$

The reduction of the data matrix (x_{vi}) to its margins (r_v) and (s_i) and the separation of $r_v\beta_v$ and $s_i\delta_i$ in Equation 3 establish the sufficiency of r_v for estimating β_v and of s_i for estimating δ_i.

It is important to recognize, of course, that although r_v and s_i lead to sufficient estimates of β_v and δ_i, they themselves are not satisfactory as estimates. Person score r_v is not free from the particular item difficulties encountered in the test. Nor is item score s_i free from the ability distribution of the persons who happen to be taking the item. To achieve independence from these local factors requires adjusting the observed r_v and s_i for the related item difficulty and person-ability distributions to produce the test-free person measures and sample-free item calibrations desired.

With a side condition such as $\sum_i^L \delta_i = 0$ to restrain the indeterminacy of origin in the response parameters, the first and second derivatives of λ with respect to β_v and δ_i become

$$\frac{\partial\lambda}{\partial\beta_v} = r_v - \sum_i^L \pi_{vi} \qquad v=1,N \qquad [4]$$

$$\frac{\partial^2\lambda}{\partial\beta_v^2} = -\sum_i^L \pi_{vi}(1 - \pi_{vi}) \qquad [5]$$

and

$$\frac{\partial L}{\partial\delta} = - s_i + \sum_v^N \pi_{vi} \qquad i=1,L \qquad [6]$$

$$\frac{\partial^2 L}{\partial\delta_i^2} = -\sum_v^N \pi_{vi}(1 - \pi_{vi}) \qquad [7]$$

where

$$\pi_{vi} = e^{\beta_v - \delta_i}/(1 + e^{\beta_v - \delta_i}).$$

These are the equations necessary for unconditional maximum likelihood estimation. The solutions for item difficulty estimates in Equations 6 and 7 depend on the presence of values for the person-ability estimates. Because unweighted test scores are the sufficient statistics for estimating people's abilities, all persons with identical scores obtain identical ability estimates. Hence, we may group persons by their score, letting

b_r be the ability estimate for any person with score r,
d_i be the difficulty estimate of item i,
n_r be the number of persons with score r,
and write the estimated probability that a person with a score r will succeed on item i as

$$p_{ri} = e^{b_r - d_i}/(1 + e^{b_r - d_i}). \qquad [8]$$

Then

$$\sum_v^N p_{vi} = \sum_r^{L-1} n_r p_{ri}, \text{ as far as estimates are concerned.}$$

A convenient algorithm for computing estimates (d_i) is:
(i) Define an initial set of (b_r) as

$$b_r^{(0)} = \ell og \left(\frac{r}{L-r}\right) \qquad r=1,L-1 \qquad [9]$$

(ii) Define an initial set of (d_i), centered at $d_{.} = 0$, as

$$d_i^{(0)} = \ell og\left(\frac{N-s_i}{s_i}\right) - \left[\sum_i^L \ell og\left(\frac{N-s_i}{s_i}\right)\right]/L \qquad i=1,L \qquad [10]$$

where $b_r^{(0)}$ is the maximum likelihood estimate of β_r for a test of L equivalent items centered at zero and $d_i^{(0)}$ is the similarly centered maximum likelihood estimate of δ_i for a sample of N equal-ability persons.

(iii) Improve each estimate d_i by applying the Newton-Raphson method to Equation 6; i.e.,

$$d_i^{(j+1)} = d_i^{(j)} - \left(\frac{-s_i + \sum_r^{L-1} n_r p_{ri}^{(j)}}{-\sum_r^{L-1} n_r p_{ri}^{(j)}(1-p_{ri}^{(j)})}\right) \qquad i=1,L \qquad [11]$$

until convergence at $|d_i^{(j+1)} - d_i^{(j)}| < .01$

$$\text{where} \quad p_{ri}^{(j)} = e^{b_r - d_i^{(j)}}/(1 + e^{b_r - d_i^{(j)}}) \qquad [12]$$

and convergence to .01 is usually reached in three or four iterations.

(iv) Recenter the set of (d_i) at $d_{.} = 0$.

(v) Using this improved set of (d_i), apply Newton-Raphson to Equation 4 to improve each b_r

$$b_r^{(m+1)} = b_r^{(m)} - \left(\frac{r - \sum_i^L p_{ri}^{(m)}}{-\sum_i^L p_{ri}^{(m)}(1 + p_{ri}^{(m)})}\right) \qquad r=1,L-1 \qquad [13]$$

until convergence at $|b_r^{(m+1)} - b_r^{(m)}| < .01$

$$\text{where} \quad p_{ri}^{(m)} = e^{b_r^{(m)} - d_i}/(1 + e^{b_r^{(m)} - d_i}). \qquad [14]$$

(vi) Repeat steps (iii) through (v) until successive estimates of the whole set of (d_i) become stable, i.e.,

$$\sum_{i}^{L}(d_i^{(k+1)} - d_i^{(k)})^2/L < .0001, \quad [15]$$

which usually takes three or four cycles.

(vii) Use the reciprocal of the square root of the negative of the second derivatives defined in Equation 7 as asymptotic estimates of the standard errors of difficulty estimates,

$$SE(d_i) = \left(\sum_{r}^{L-1} n_r p_{ri}(1 - p_{ri})\right)^{-\frac{1}{2}}. \quad [16]$$

Andersen (1973) has shown that the presence of the ability parameters (β_v) in the likelihood equation of this unconditional approach leads to biased estimates of item difficulties (δ_i). Because simulations undertaken to test UCON in 1966 indicated that the factor (L-1)/L would compensate for the bias, most versions in use today contain that unbiasing coefficient. However, this correction factor was not mentioned by Wright and Panchapakesan (1969) and no documentation was published.

The Bias in the Unconditional Procedure

Our evaluation of the bias in UCON is based on a comparison of the unconditional estimation, Equation 6, and its conditional counterpart.

From Equation 6 we have

$$s_i = \sum_{r}^{L-1} n_r e^{b_r - d_i^*}/(1 + e^{b_r - d_i^*}) \quad [17]$$

in which d_i^* is the biased UCON estimate of δ_i.

From the conditional probability of a response vector (x_{vi}) given a score r, namely,

$$P\{(x_{vi})|r,(\delta_i)\} = P\{(x_{vi})|\beta_v,(\delta_i)\}/ P\{r|\beta_v,(\delta_i)\}$$

$$= e^{-\sum_{i}^{L} x_{vi}\delta_i}/\gamma_r \quad [18]$$

in which r has replaced β_v so that Equation 18 is *not* a function of β_v,

$$\gamma_r = \sum_{(x_{vi})}^{r} e^{-\sum_{i}^{L} x_{vi}\delta_i} \quad [19]$$

and $\sum_{(x_{vi})}^{r}$ is the sum over all response vectors (x_{vi}) with $\sum_{i}^{L} x_{vi} = r$, we derive

$$s_i = \sum_{r}^{L-1} n_r e^{d_i} g_{r-1,i}/g_r \quad [20]$$

in which

$$g_r = \sum_{(x_{vi})}^{r} e^{-\sum_i^L x_{vi} d_i} \qquad [21]$$

and

$$g_{r-1,i} = \sum_{(x_{vj})}^{r-1} e^{-\sum_j^L x_{vj} d_j} \qquad j \neq i \quad . \qquad [22]$$

Since $g_r = e^{-d_i} g_{r-1,i} + g_{ri}$, we can write

$$(g_{r-1,i}/g_r) = (g_{r-1,i}/g_{ri})/(1+e^{-d_i}(g_{r-1,i}/g_{ri})) \quad . \qquad [23]$$

The conditional maximum likelihood estimate of the ability b_r that goes with a score r then becomes $b_r = \ell og(g_{r-i}/g_r)$. If we define the ability estimate that goes with a score of r, when item i is removed from the test, as b_{ri}, then

$$g_{r-1,i}/g_{ri} = e^{b_{ri}}$$

and the condition estimation equation comparable to Equation 17 becomes

$$s_i = \sum_r^{L-1} n_r e^{b_{ri}-d_i}/1+e^{b_{ri}-d_i}) \qquad [24]$$

Since Equations 17 and 24 estimate similar parameters for identical data, we can set .

$$b_{ri}-d_i = b_r-d_i^* \quad . \qquad [25]$$

Thus, the bias in the UCON estimate of item difficulty d_i^* at person score r, when compared with the unbiased d_i, becomes

$$d_i^*-d_i = b_r-b_{ri} = a_{ri} \quad . \qquad [26]$$

The simplest way to correct d_i^* would be with a correction factor $k= d_i/d_i^*$. If we define $c_{ri}=a_{ri}/d_i$ as the relative bias in the difficulty estimate of item i as score r, then

$$k = d_i/d_i^* = d_i/(d_i+a_{ri}) = 1/(1+c_{ri}) \quad , \qquad [27]$$

the utility of which depends on how much c_{ri} varies.

We explored the possibilities for k by studying how values of a_{ri} and c_{ri} vary over r and d_i in a variety of typical test structures.

For each test we

i) Specified test length L and item dispersion standard deviation Z for a normal distribution of item difficulties;

ii) Generated the consequent set of item difficulties (d_i);

iii) Calculated ability estimates from (d_i), namely,

$$b_r = \ell og(g_{r-1}/g_r) \qquad [28]$$

$$b_{ri} = \ell og(g_{r-1,i}/g_{ri}) \; ; \qquad [29]$$

iv) Arrived at the biases

$$a_{ri} = b_r - b_{ri} \qquad [30]$$

$$c_{ri} = a_{ri}/d_i, \quad \text{except } |d_i| < .5 \; ; \qquad [31]$$

v) Calculated the mean relative bias

$$c_{..} = \sum_r^{L-1} \sum_i^{L} c_{ri}/(L-1)(L), \quad \text{except } |d_i| < .5 \qquad [32]$$

$$\text{and approximated } k \text{ as } 1/(1+c_{..}) \qquad [33]$$

The results are summarized in Table 1 for normally shaped tests of lengths L = 20, 30, 40, 50, and 80 and item difficulty standard deviations of Z = 1.0, 1.5, 2.0, 2.5.

The values of $c_{..}$ in Table 1 are well approximated by 1/(L-1) regardless of Z.

If $$c_{..} \simeq 1/(L-1) \quad , \qquad [34]$$

then $$k \simeq 1/(1+1/(L-1)) = (L-1)/L \qquad [35]$$

TABLE 1

Relative Bias in the Unconditional Procedure
(Averaged over Items and Scores)

Test Length	Test Dispersion in Item Standard Deviations 1.0	1.5	2.0	2.5	$\frac{1}{L-1}$
20	.0525	.0525	.0560	.0522	.0526
30	.0345	.0320	.0349	.0349	.0345
40	.0238	.0258	.0261	.0252	.0256
50	.0203	.0192	.0195	.0199	.0204
80	.0126	.0124	.0121	.0127	.0126

This is the correction factor used in most versions of UCON and coincides with the correction deducible in the simple case where L=2 and the unconditional estimates can be shown to be exactly twice the conditional ones.

The correction factor (L-1)/L is based on averaging the relative bias c_{ri} over scores and items. This assumes that the distribution of item difficulties and of persons over scores is more or less uniform. In order to master the extent to which skewed item or person distributions might interfere with the accuracy of the corrected UCON procedure, we also examined how bias varies with item difficulty d_i and, at each difficulty, how it varies over scores. Table 2 shows how the means and standard deviations of a_{ri} over r vary with d_i for a test of L = 20 and Z = 1.5 and also indicates the residual bias in d_i^* after applying the correction factor (L-1)/L.

TABLE 2

Means and Standard Deviations of Bias in the Unconditional Procedure (For Items Averaged Over Scores)

Item Difficulty	Bias in UCON Mean	Bias in UCON Standard Deviation	Residual Bias After Correction
-2.46	-.12	.03	.010
-2.30	-.11	.04	.011
-1.93	-.09	.05	.010
-1.86	-.09	.05	.008
-1.28	-.06	.07	.007
-1.03	-.05	.07	.004
-0.69	-.03	.08	.006
-0.50	-.02	.08	.006
-0.39	-.01	.08	.010
-0.35	-.01	.08	.008
-0.29	-.01	.08	.005
-0.21	-.00	.08	.012
-0.08	+.00	.08	.004
0.09	+.01	.08	.005
0.26	+.02	.08	.006
0.50	+.03	.08	.004
0.75	+.05	.08	.010
1.28	+.07	.06	.002
2.11	+.12	.04	.009
2.93	+.16	.07	.006

$$\text{Residual Bias} = \left(\frac{L-1}{L}\right)\left(a_{\cdot i} - \frac{d_i}{L-1}\right) = \left(\frac{L-1}{L}\right) d_i^* - d_i$$

where $a_{\cdot i}$ = Mean Bias

d_i = Item Difficulty

The results in Table 2 are typical of the cases we studied. The bias represented by $a_{\cdot i}$ increases as d_i departs from zero and can become substantial in short, wide tests. But, as the residual bias column shows, the correction factor removes this bias for all practical purposes.

This leaves for investigation the fluctuation of bias over scores at a particular d_i which is most extreme when $d_i = 0$. In order to evaluate the threat of this fluctuation, we examined the maximum bias standard deviations over a wide variety of tests. Table 3 gives these maxima for the test covered in Table 1.

Two trends are apparent: as test length increases the fluctuation of UCON bias over item and score decreases, but as item dispersion increases the fluctuation increases. However, as Table 2 illustrates, the maximum fluctuation of bias is among items at the center of the test where precision in item calibration has least effect on precision of person measurement. Only for short, wide tests ($L<30$, $Z>1.5$) did the fluctuation in bias become detectable in the corrected estimates of item difficulty, and then only when the distribution of items or persons was distinctly skewed. For a typical test of $L=30$, $Z=1$ bias had standard deviations over an item or score of .02 or less, virtually nil, and was successfully corrected for. Thus, we found the corrected UCON procedure to be indistinguishable in results from the conditional procedure.

From a practical point of view, however, even the calibration accuracy available in UCON may be more than is needed for adequate measurement. When we studied (Wright and Douglas, 1975) the deterioration of measurement precision which arises when errors are introduced into the estimation of item parameters by generating disturbed item difficulties, with $d_i = \delta_i + e_i$ and e_i normally distributed around zero, we found that, for test designs encountered in practice, random disturbances in item calibration e_i can be as large as 1 logit before distortions in the estimates of β_r reached 0.1 logit. In view of this robustness with respect to measurement precision, we concluded that even the corrected UCON may be unnecessarily precise in many applications and that a further approximation would be useful.

TABLE 3

Maximum Standard Deviations of Bias in the Unconditional Procedure (For Items Over Scores)

Test Length	Test Dispersion in Item Standard Deviations			
	1.0	1.5	2.0	2.5
20	.06	.08	.12	.24
30	.02	.06	.17	.12
40	.02	.04	.07	.07
50	.02	.04	.06	.09
80	.01	.02	.04	.06

$$\text{Standard Deviations} = \left[\sum_{r}^{L-1} (a_{ri} - a_{\cdot i})^2/(L-2)\right]^{\frac{1}{2}}$$

where $a_{ri} = d^*_i - d_i$, the bias.

The Approximate Procedure

In order to discover a satisfactory shortcut to UCON we return to the crude initial estimates for UCON given by Equation 10

$$d_i^{(0)} = \ell og\left[\frac{N-s_i}{s_i}\right] - \sum_i^L \ell og\left[\frac{N-s_i}{s_i}\right]/L \qquad i=1,L \qquad [36]$$

and note that they are exact when the variation in ability is zero. The first row of Table 4 contains estimates of this kind for a typical test. Comparison with the UCON estimates in the second row shows that these estimates are too crude. The extreme items are not far enough out on the latent variable. What is needed is an expansion factor to take into account the dispersion in abilities. A simple approach to this dispersion (for which we are grateful to Leslie Cohen of the University of Southampton) is to assume that abilities are normally distributed, and so can be adequately described by their mean and variance. It is not necessary for abilities actually to be distributed in this way in order to use the assumption as the basis for developing an approximate calibration procedure, the utility of which can be investigated.

In our study of best test design (Wright and Douglas, 1975), we developed an approximation for the ability estimate b_r when the item difficulties (δ_i) are normally distributed.

Extensive simulations showed that this formula gave results extremely close to the maximum likelihood ability estimates. By interchanging items and abilities we arrived at a similar formula for approximating an item difficulty estimate d_i when people's abilities are normally distributed:

$$d_i = \sqrt{1+V_b/1.7^2}\;\ell og\left[\frac{N-s_i}{s_i}\right] \qquad i=1,L \qquad [37]$$

where V_b is the observed variance of the estimated abilities and the constant, 1.7, arises because the relation between the normal and logistic cumulative distributions, $|\phi(X)-\Psi(1.7x)|<.01$ for all x, is used as a basis for equating and exchanging them in the derivation of the approximation (see Wright and Douglas, 1975, for details).

TABLE 4

Item Difficulty Estimates Based on Initial, UCON and PROX Procedures, for a Moderately Skewed Set of 60 Items

Procedure	Item Sequence Number					
	5	10	20	40	50	60
INITIAL	-1.21	-1.17	.21	.38	.66	1.44
UCON	-1.36	-1.31	.24	.43	.75	1.63
PROX	-1.35	-1.30	.24	.43	.74	1.60

The way in which this formula can be used for approximate item calibration outlined in the following sequence:

i) Let

$$d_i^{(0)} = \ell og\left[\frac{N-s_i}{s_i}\right] - \sum_i^L \ell og\left[\frac{N-s_i}{s_i}\right]/L \quad . \quad i=1,L \qquad [38]$$

ii) Define initial ability estimates

$$b_r^{(0)} = \ell og\left[\frac{r}{L-r}\right] \qquad r=1,L-1 \qquad [39]$$

and find their variance

$$V_b^{(0)} = \sum_r^{L-1} n_r(b_r^{(0)} - b_\bullet^{(0)})^2/(N-1) \qquad [40]$$

where

$$b_\bullet^{(0)} = \sum_r^{L-1} n_r b_r^{(0)}/N \quad . \qquad [41]$$

iii) Set

$$d_i^{(j)} = \sqrt{1+V_b^{(j)}/1.7^2} \quad d_i^{(0)} \qquad i=1,L \qquad [42]$$

$$= a_b^{(j)} d_i^{(0)}$$

iv) Obtain the variance of these item estimates, $V_d^{(j)}$, an item expansion factor

$$\text{namely} \quad a_d^{(j)} = \sqrt{1+V_d^{(j)}/1.7^2}$$

arrive at a revised set of ability estimates

$$b_r^{(j+1)} = a_d^{(j)} b_r^{(0)} \qquad r=1,L-1 \qquad [43]$$

and repeat (iii) and (iv) until convergence to

$$|a_b^{(j+1)} - a_b^{(j)}| < .01 \quad . \qquad [44]$$

v) At convergence, estimate the standard errors of item difficulty estimates by using the analogous approximations:

$$SE(d_i) = (2\pi(1.7^2+V_b))^{1/4}/N^{1/2} e^{(d_i-b_\bullet)^2/4(1.7^2+V_b)} \quad . \qquad [45]$$

The final row of Table 4 shows the estimates produced by PROX for simulated data in which neither items nor persons were normally distributed. For all practical purposes, the PROX estimates are equivalent to the estimates obtained from UCON.

Comparisons of UCON and PROX for Accuracy

The simulations used to evaluate the performance of these procedures were set up to expose their biases. We compared their estimates with known generating parameters for calibrating samples which did not coincide with the tests on the latent variable. Samples for which the test was off-target, i.e., either too difficult or too easy, were used because that is the situation which produces the poorest item difficulty estimation.

i) For each "test" to be studied, a set of L normally distributed item difficulties was generated with a mean zero and a standard deviation Z. Although a great many combinations of L and Z were reviewed, tests of length 20 and 40 and Z's of 1 and 2 are sufficient to summarize the results.

ii) For each of these tests, 500 "persons" normally distributed in ability with a mean M, a standard deviation SD and an upper truncation of TR, were simulated for exposure to the test. Truncation was set to induce a pronounced skew in the ability distribution.

iii) Each ability was combined with each difficulty to yield a probability of success, and this probability was compared with a uniform random number to produce a stochastic response. These responses were accumulated into the data vectors (s_i) and (n_r) necessary for item calibration.

iv) Item parameters and their standard errors were estimated by each procedure.

v) Steps (ii)-(iv) were repeated 15 times for each test so that there were 15 replications of (s_i) and (n_r) with which to investigate the discrepancies among the methods.

Table 5 shows how well PROX performs in comparison with UCON, even for moderately skewed samples. Only in the extreme case, where items range from −4 to +4 and the sample has a mean of 2, standard deviation of 2, and is truncated at 4.5, does PROX become potentially unreliable. In this instance even UCON is not very satisfactory. When a reasonable editing of the data is undertaken before estimation so that extreme items such as those which interfered with successful calibration in case 12 in Table 5 are eliminated, then PROX is equivalent to UCON as far as accuracy of estimation is concerned.

As for computing efficiency, both UCON and PROX were applied to the 60-item test used for Table 4. The computer was a basic IBM 1130. The core storage used by the estimation portions of UCON and PROX were 740 and 672 bytes. The resulting item calibrations were within .03 of one another, as Table 4 suggests. But the times taken were dramatically different. UCON took 314 seconds to converge to its estimates. PROX took 8 seconds.

We found this ratio of about 40 to 1 to be typical for problems in the 50 to 60 item range. PROX is more that 40 times faster than UCON for larger problems. However, for small problems of 20 to 30 items, the time advantage of PROX is not so dramatic, and when the calibrating sample (or the distribution of item difficulties) is markedly skewed, PROX can err by several tenths of a logit in some of its item difficulty estimates, as case 12 in Table 5 documents.

Editing Data Before Calibration

The calibration procedures described in the previous pages depend upon the marginals of a data matrix (x_{vi}) for input, where v=1,N for persons and i=1,L for items. When we sum the entries (0 for incorrect and 1 for correct) in the v^{th} row of such a matrix, we obtain a score for the v^{th} person, r_v. When we sum the i^{th} column, we obtain a score s_i for the i^{th} item, reporting the number of persons who responded with a correct answer to item i.

Since the ability estimates are "item-free," we may group subjects according to their scores. Thus, the initial N pieces of information in the person marginals may be reduced to L+1 pieces by letting n_r be the number of persons at each raw score r. The vector (n_r) is the frequency distribution of scores.

TABLE 5

The Bias of UCON and PROX Procedures for Various Normal Shaped Tests and Truncated Samples, Based on 15 Replications of 500 Persons Each

Case	DATA					PROCEDURES					
	Test		Sample			UCON			PROX		
	L	Z	M	SD	TR	MAX DIFF	RMS	MEAN ABS	MAX DIFF	RMS	MEAN ABS
1	20	1	0	0.5	2.0	-.05	.02	.08	.07	.04	.09
2	20	1	0	1.0	2.0	-.06	.03	.08	.07	.04	.08
3	20	1	1	1.0	2.5	-.06	.03	.09	-.11	.05	.10
4	20	1	1	1.5	2.0	-.07	.03	.09	.12	.03	.09
5	20	1	1	1.5	2.5	-.07	.03	.10	.08	.04	.10
6	40	1	0	0.5	2.0	-.08	.03	.08	.11	.05	.09
7	40	1	0	1.0	2.0	-.07	.03	.08	-.08	.03	.08
8	40	1	1	1.0	2.5	-.11	.03	.09	-.14	.04	.09
9	40	1	1	1.5	2.0	.06	.03	.09	.08	.03	.09
10	40	1	1	1.5	2.5	.07	.03	.09	.07	.03	.09
11	40	2	0	2.0	4.5	-.12	.05	.10	.13	.07	.12
12	40	2	2	2.0	4.5	-.30	.06	.12	-.73	.13	.14

where
MAX DIFF: the maximum difference between a generating item parameter and the mean of 15 replications of its estimates,
RMS: the root mean square over items of these differences,
MEAN ABS: the mean over items of the absolute value of these differences.

Finite estimates of ability for persons who obtain zero and perfect scores, however, do not exist. When a person gets all items correct, we have no evidence that he is not perfect and can only estimate his ability as infinitely high. When a person gets all items incorrect, we have no evidence that he can succeed at all and can only estimate his ability as infinitely low. Similarly, it will be impossible to obtain finite estimates of item parameters for those items which nobody or everybody has answered correctly. These restrictions mean that we require a prior editing of the data matrix before a calibration procedure can be applied. This editing can be accomplished in the following manner:

i) Drop persons with initial $r = 0$ and $r = L$.

ii) Scan the item score vector (s_i) for $i=1,L$:

a) If $s_i = 0$, then

1. drop item *i* from the test and decrease test length by 1

$$L' = L-1 \qquad [46]$$

2. drop the n_{L-1} persons who now have perfect scores and decrease sample size by n_{L-1}

$$N' = N-n_{L-1} \;; \qquad [47]$$

3. remove their responses from the item scores in the item score vector (s_i)

$$s_j' = s_j - n_{L-1} \;; \qquad j=1,L \qquad [48]$$

4. return to (a) with $i = 1$.

b) If $s_i = N$, then

1. drop item *i* from the test and decrease test length by 1

$$L' = L-1 \;; \qquad [49]$$

2. drop the n_1 persons who now have zero scores and decrease sample size by n_i

$$N' = N-n_1 \;; \qquad [50]$$

3. shift each score group down one

$$n_{r-1} = n_r \;; \qquad r=2,L \qquad [51]$$

4. return to (b) with $i = 1$.

References

Andersen, E. B. The numerical solution of a set of conditional estimation equations. *The Journal of the Royal Statistical Society.* Series B, 1972, *34* (1), 42–54.

Andersen, E. B. *Conditional inference and models for measuring.* Copenhagen, Denmark: Mentalhygiejnisk Forlag, 1973.

Rasch, G. *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Danmarks Paedogogiske Institut, 1960.

Rasch, G. An individualistic approach to item analysis. In P. F. Lazarsfeld and N.W. Henry (Eds.), *Readings in mathematical social sciences.* Chicago: Science Research Associates, 1966. (a)

Rasch, G. An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 1966, *19*, 49–57. (b)

Wright B. D. & Douglas, G. A. *Best test design and self-tailored testing*. (Research Memorandum No. 19). University of Chicago, Department of Education, Statistical Laboratory, 1975.

Wright, B. D. & Panchapakesan, N. A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 1969, *29*, 23–48.

Author's Address

Benjamin D. Wright, Department of Education, 5835 Kimbark Ave., University of Chicago, Chicago, IL 60037.