CONDITIONAL VERSUS UNCONDITIONAL PROCEDURES FOR SAMPLE-FREE ITEM ANALYSIS

BENJAMIN D. WRIGHT
University of Chicago
GRAHAM A. DOUGLAS
University of Western Australia

Published in *Educational and Psychological Measurement,* 1977, 37:3, 573-586

Two procedures for Rasch, sample-free, item calibration are reviewed and compared for accuracy. Andersen's (1972) theoretically ideal "conditional" procedure is impractical for calibrating more than 10 or 15 items. A simplified alternative procedure for conditional estimation practical for 20 or 30 items which produces equivalent estimates is developed. When more than 30 items are analyzed recourse to Wright's (1969) widely used "unconditional" procedure is inevitable but that procedure is biased. A correction factor which makes the bias negligible is identified and demonstrated.

IN 1969 Wright and Panchepakesan described a procedure for sample-free item analysis based on the simple logistic response model (Rasch, 1960, 1961, 1966a, 1966b) and listed the FORTRAN segments essential to apply this procedure. Since then the procedure has been incorporated in a number of computer programs for item analysis and widely used in this country and abroad.

The procedure described, however, is incomplete in two minor but important ways. 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 which the procedure is known to entail (because the estimation equations are not conditioned for person ability) are mentioned.

In this article we will refer to the Wright-Panchepakesan procedure as the "unconditional" solution, UCON. We will review UCON;

discuss its inadequacies and the theoretically correct alternative, the "conditional" procedure; report on an investigation of the bias in UCON and make a comparison of the accuracy of the two procedures.

The Unconditional Procedure

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

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

$$x_{vi} = \text{binary response} = \begin{cases}1 \text{ if correct}\\ 0 \text{ otherwise}\end{cases}$$

$$\beta_v = \text{ability parameter of person } v,$$

$$\delta_i = \text{difficulty parameter of item } i.$$

The likelihood of the data matrix $((x_{vi}))$ is the continued product of Q_{vi} over all values of v and i:

$$\Lambda = \prod_v^N \prod_i^L Q_{vi} = e^{\sum_v^N \sum_i^L x_{vi}(\beta_v - \delta_i)} \Big/ \prod_v^N \prod_i^L (1 + e^{\beta_v-\delta_i}). \tag{2}$$

Upon taking logarithms and letting

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

and

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

the log likelihood becomes

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

With a side condition such as $\sum_i^L \delta_i = 0$ to restrain the indeterminacy of origin in the response parameters $(\beta_v - \delta_i)$ and with the results:

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

$$\frac{\partial \log (1 + e^{\beta_v-\delta_i})}{\partial\delta_i} = - e^{\beta_v-\delta_i}/(1 + e^{\beta_v-\delta_i}) = - \pi_{vi},$$

the first and second derivatives of λ with respect to β_v and δ_i become

$$\frac{\partial\lambda}{\partial\beta_v} = r_v - \sum_i^L \pi_{vi} \quad \text{v} = 1, \text{N} \tag{4}$$

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

and

$$\frac{\partial \lambda}{\partial \delta_i} = - s_i + \sum_v^N \pi_{vi} \quad \mathrm{i} = 1, \mathrm{L} \tag{6}$$

$$\frac{\partial^2 \lambda}{\partial \delta_i^2} = - \sum_v^N \pi_{vi} (1 - \pi_{vi}) \tag{7}$$

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 the data are binary, abilities can only be estimated for integer scores between O and L. Hence we may group persons by their score and let

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

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

Then

$$\sum_v^N p_{vi} = \sum_r^{L-1} n_r p_{ri}$$

as far as estimates are concerned.

A convenient algorithm for computing estimates (d_i) of (δ_i) is as follows:

(i) Define an initial set of (b_r) as

$$b_r^0 = \log\left(\frac{r}{L - r}\right) \qquad r = 1, L - 1$$

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

$$d_i^0 = \log\left(\frac{N - s_i}{s_i}\right) - \sum_i^L \log\left(\frac{N - s_i}{s_i}\right) /L \qquad i = 1, L$$

(iii) Improve each estimate d_i by applying Newton-Raphson to equation (6), i.e.,

$$d_i^{j+1} = d_i^j - \frac{-s_i + \sum_r^{L-1} n_r p_{ri}^j}{-\sum_r^{L-1} n_r p_{ri}^j (1 - p_{ri}^j)} \quad i = 1, L \tag{9}$$

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

where $p_{ri}^j = e^{b_r} - d_i^j/(1 + e^{b_r} - d_i^j)$.

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

$$b_r^{m+1} = b_r^m - \frac{r - \sum_i^L p_{ri}^m}{-\sum_i^L p_{ri}^m(1 - p_{ri}^m)} \quad r = 1, L - 1 \tag{10}$$

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

where $p_{ri}^m = e^{b^m_r} - d_i/(1 + e^{b^m_r} - d_i)$.

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

$$\sum_i^L (d_i^{k+1} - d_i^k)^2/L < .0001.$$

which usually takes three or four cycles.

(*vii*) Use the negative reciprocal square root of the second derivatives defined in equation (7) and found in the denominator of equation (9) as asymptotic estimates of the standard errors of each difficulty estimate, i.e.,

$$SE(d_i) = \left(\sum_r^{L-1} n_r p_{ri}\,(1 - p_{ri})\right)^{-1/2} \tag{11}$$

The algorithm described above is similar to the calibration techniques proposed by Birnbaum (1968) and Bock (1972) in that the person abilities are estimated simultaneously with the item difficulties so that the estimation procedure is *not* conditioned for the incidential ability parameters. However, Andersen (1970, 1971, 1972, 1973) has shown that this unconditional approach results in inconsistent estimates of the item parameters. The presence of the ability parameters (β_v) in the likelihood equation leads to biased estimates of item difficulties (δ_i). For a procedure to produce consistent and unbiased estimates requires a conditional approach in which the solution equations are conditioned for the ability parameters before maximization. We will describe the conditional solution and report on a study which supports the use of an unbiasing coefficient.

The Conditional Procedure

A conditional maximum likelihood procedure produces consistent estimates of the set of item parameters (δ_i) (Andersen, 1973). To develop this procedure we obtain from equation (1) the probability of the response vector (x_{vi}) for a person whose ability parameter is β_v as

$$
\begin{aligned}
P\{(x_{vi})|\beta_v,(\delta_i)\} &= \prod_i^L P\{x_{vi}|\beta_v, \delta_i\} \qquad (12)\\
&= \prod_i^L e^{x_{vi}(\beta_v-\delta_i)}/(1+e^{\beta_v-\delta_i})\\
&= e^{\sum_i^L x_{vi}\beta_v - \sum_i^L x_{vi}\delta_i} \Big/ \prod_i^L (1+e^{\beta_v-\delta_i})\\
&= e^{r_v\beta_v - \sum_i^L x_{vi}\delta_i} \Big/ \prod_i^L (1+e^{\beta_v-\delta_i}).
\end{aligned}
$$

Next we derive the probability of obtaining any score r.

A person v may obtain a score r in $\binom{L}{r}$ different ways. Hence we must sum all the different probabilities like (12) based on values of the response vector (x_{vi}) which sum to r.

If we let $\sum_{(x_{vi})}^{r}$ be the sum over all response vectors (x_{vi}) with $\sum_i^L x_{vi} = r$ and represent the elementary symmetric functions of the set of $(e^{-\delta_i})$ as

$$\gamma_r = \sum_{(x_{vi})}^{r} e^{-\sum_i^L x_{vi}\delta_i}, \qquad (13)$$

then the probability of a score r becomes

$$
\begin{aligned}
P\{r|\beta_v,(\delta_i)\} &= \sum_{(x_{vi})}^{r} P\{(x_{vi})|\beta_v,(\delta_i)\} \qquad (14)\\
&= e_r^{\beta_v} \sum_{(x_{vi})}^{r} e^{-\sum_i^L x_{vi}\delta_i} \Big/ \prod_i^L (1+e^{\beta_v-\delta_i})\\
&= e^{r\beta_v}\gamma_r \Big/ \prod_i^L (1+e^{\beta_v-\delta_i}).
\end{aligned}
$$

Finally, the "conditional" probability of the response vector (x_{vi}), given the score r is obtained by dividing (12) by (14) to yield

$$P\{(x_{\nu\iota})|r,(\delta_\iota)\} = P\{(x_{\nu\iota})|\beta_\nu,(\delta_\iota)\}/P\{r|\beta_\nu,(\delta_\iota)\} \tag{15}$$

$$= e^{-\sum_\iota^L x_{\nu\iota}\delta_\iota}/\gamma_r$$

in which r has replaced β_ν so that (15) is *not* a function of β_ν.

If we define $\gamma_{r-1,\iota}$ as the $r-1^{th}$ symmetric function with the element $e^{-\delta_\iota}$ omitted, i.e., the sum of all the ways responses to the other $L-1$ items can add up to a score of $r-1$, then the probability of a person with a score r getting item i correct and incorrect become

$$P\{x_{\nu\iota} = 1|r,(\delta_\iota)\} = e^{-\delta}\gamma_{r-1,\iota}/\gamma_r \tag{16}$$

and

$$P\{x_{\nu\iota} = 0|r,(\delta_\iota)\} = \gamma_{r\iota}/\gamma_r.$$

What we have done is to condition the response model on the minimal sufficient statistic r for the incidental parameter β_ν. The result is a conditional probability dependent only on the structural parameters (δ_ι).

For a group of $N = \sum_r^{L-1} n_r$ persons the conditional likelihood of their data (s_ι) and (n_r) is from (15)

$$\Lambda = e^{-\sum_\iota^L s_\iota\delta_\iota} \Big/ \prod_r^{L-1} \gamma_r^{n_r} \tag{17}$$

Taking logarithms we have

$$\lambda = \log\Lambda = -\sum_\iota^L s_\iota\delta_\iota - \sum_r^{L-1} n_r \log\gamma_r \tag{18}$$

The derivatives of $\log\gamma_r$ are obtained by factoring $e^{-\delta_\iota}$ and $e^{-\delta_j}$ out of γ_r, i.e.,

$$\gamma_r = e^{-\delta_\iota}\gamma_{r-1,\iota} + \gamma_{r,\iota} \tag{19}$$

Then

$$\frac{\partial \log\gamma_r}{\partial\delta_\iota} = \frac{1}{\gamma_r}\frac{\partial\gamma_r}{\partial\delta_\iota} = -e^{-\delta_\iota}\gamma_{r-1,\iota}/\gamma_r$$

$$\frac{\partial^2 \log\gamma_r}{\partial\delta_\iota^2} = \frac{1}{\gamma_r}\frac{\partial^2\gamma_r}{\partial\delta_\iota^2} - \frac{1}{\gamma_r^2}\left(\frac{\partial\gamma}{\partial\delta_\iota}\right)^2$$

$$= e^{-\delta_\iota}\gamma_{r-1,\iota}/\gamma_r - (e^{-\delta_\iota}\gamma_{r-1,\iota}/\gamma_r)^2$$

$$\frac{\partial^2 \log\gamma_r}{\partial\delta_\iota\partial\delta_j} = \frac{1}{\gamma_r}\frac{\partial^2\gamma_r}{\partial\delta_\iota\partial\delta_j} - \frac{1}{\gamma_r^2}\left(\frac{\partial\gamma_r}{\partial\delta_\iota}\right)\left(\frac{\partial\gamma_r}{\partial\delta_j}\right)$$

$$= e^{-\delta_\iota-\delta_j}\gamma_{r-2,\iota j}/\gamma_r - (e^{-\delta_\iota}\gamma_{r-1,\iota}/\gamma_r)(e^{-\delta_j}\gamma_{r-1,j}/\gamma_r)$$

If we now represent estimates of the symmetric functions γ by g with similar subscripts and define the estimated probability for a person with score r of getting item i correct as

$$q_{ri} = e^{-d_i} g_{r-1,i}/g_r,$$

and the estimated probability for a person with score r of getting both items i and j correct as

$$q_{rij} = e^{-d_i-d_j} g_{r-2,ij}/g_r$$

then the three derivatives of equation (18) can be rewritten for maximum likelihood estimation as

$$f_i^* = -s_i + \sum_r^{L-1} n_r q_{ri} \qquad i = 1, L \tag{20}$$

$$c_{ii}^* = \sum_r^{L-1} n_r q_{ri}(1 - q_{ri}) \qquad i = 1, L \tag{21}$$

$$c_{ij}^* = -\sum_r^{L-1} n_r (q_{rij} - q_{ri} q_{rj}) \qquad \begin{array}{l} i = 1, L \\ j = 1, L \\ i \neq j \end{array} \tag{22}$$

A multi-parameter algorithm for obtaining the estimates and their standard errors for the polychotomous case is described by Andersen (1972). For the binary case the essential steps are:

(*i*) Initialize item difficulty at

$$d_i^0 = \log\left[\frac{N - s_i}{s_i}\right] - \sum_i^L \log\left[\frac{N - s_i}{s_i}\right] \Big/ L$$

(*ii*) Use the current set of (d_i) to calculate the symmetric function ratios $(g_{r-1,i}/g_r)$ and $((g_{r-2,ij}/g_r))$ and hence (f_i^*) and $((c_{ij}^*))$ over $i = 1, L$ and $j = 1, L$.

(*iii*) Reduce (f_i^*) and $((c_{ij}^*))$ by one item to restrain the estimation equations to a unique solution, e.g.,

$$f_i = f_i^* - f_1^* \quad i = 2, L$$

$$c_{ij} = c_{ij}^* - c_{i1}^* - c_{i1}^* + c_{11}^* \qquad i = 2, L \qquad j = 2, L$$

(*iv*) Improve (d_i) by the multi-parameter Newton-Raphson procedure

$$D^{m-1} = D^m - [C^m]^{-1} F^m$$

in which

$$D = \begin{bmatrix} d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_L \end{bmatrix}, C = \begin{bmatrix} C_{22} & \cdots & C_{2L} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ C_{L2} & \cdots & C_{LL} \end{bmatrix}, F = \begin{bmatrix} f_2 \\ \cdot \\ \cdot \\ \cdot \\ f_L \end{bmatrix}$$

(*v*) Repeat steps (*ii*) through (*iv*) until the successive estimates in D converge, e.g.,

$$\sum_{2}^{L} (d_i^{m+1} - d_i^{m})^2 < .001$$

The estimates of item difficulty are contained in the final D vector and the associated asymptotic estimates of their standard errors are the square roots of the diagonal elements of the negative inverse of the final C matrix.

Our description of this full conditional procedure, FCON, has been brief since more complete details are available in Andersen (1973) and because there is little demand for FCON in practical situations. FCON has a fatal fault, namely the accumulation of round-off error in the calculation of the symmetric functions. As test length exceeds 10 or 15 items, round-off error accumulation rapidly reaches proportions which prohibit the achievement of accurate results even in extended precision.

Because of this limitation we revived an approximation to FCON developed by Wright in 1966 which simplifies the calculation of the symmetric functions. The resulting approximation, ICON (for "incomplete conditional"), does not use the covariance terms in the C matrix of FCON. This avoids the repeated calculation of the $L(L - 1)/2$ off-diagonal symmetric functions and their matrix inversion necessary in FCON.

The ICON procedure is based on the following argument:

From (20) we have the one parameter maximum likelihood solution

$$s_i = \sum_{r}^{L-1} n_r q_{ri} = e^{-d_i} \sum_{r}^{L-1} n_r g_{r-1,i}/g_r \qquad i = 1, L \tag{23}$$

from which

$$d_i = -\log\left(s_i \Big/ \left(\sum_{r}^{L-1} n_r g_{r-1,i}/g_r\right)\right) \qquad i = 1, L \tag{24}$$

The symmetric function ratios $g_{r-1,i}/g_r$ are not easy to compute as they stand. But the related ratios $g_{ri}/g_{r-1,i}$ can be calculated with a recursive application of (19). Then

$$\begin{aligned} g_{r-1,i}/g_r &= g_{r-1,i}/(e^{-d_i} g_{r-1,i} + g_{ri}) \\ &= 1/(e^{-d_i} + g_{ri}/g_{r-1,i}) \end{aligned} \tag{25}$$

so that

$$d_i = -\log\left(s_i \Big/ \left(\sum_{r}^{L-1} n_r/(e^{-d_i} + g_{ri}/g_{r-1,i})\right)\right) \tag{26}$$

The steps are:

(*i*) Initialize (d_i) at d_i°.

(*ii*) Use the current set of (d_i) in the right hand side of (26) and obtain a revised estimate for each d_i on the left side.

(*iii*) Recenter the (d_i).

(*iv*) Use the revised set (d_i) to recalculate the $g_{ri}/g_{r-1,i}$ and repeat steps (*ii*) and (*iv*) until convergence is obtained.

(*v*) At convergence, estimate the standard error from the second derivative of the likelihood equation as in (21).

ICON takes less computer space and less time for each iteration than FCON, but the total number of iterations required is slightly greater because we have ignored the covariances in C. In trials under a variety of conditions we have found no difference between the estimates obtained by FCON and ICON. For example, in three replications of the simulated administration of a 40-item test (in which we generated uniformly distributed item parameters with center at $\delta = 0$ and range ± 2 and then disturbed them by random fluctuations with standard deviation 0.1) to 100 subjects (with mean ability $\beta = 2.0$, standard deviation .5 and skewed by truncating $\beta < 3$) the maximum discrepancy between an FCON estimate and an ICON estimate was .03 logits.

As a result we dropped FCON from further consideration as an efficient and reasonable estimation algorithm and concentrated on ICON. Although convergence was always obtained from ICON we were limited in our study of its performance by the fact that fatal round-off errors accumulated when there were more than 20 or 30 items in the test, especially in the presence of extreme item parameters. This produced estimates which were biased when compared with the parameters used to generate the simulated data. Thus even ICON was not practical enough.

We had hoped that a useful algorithm could be derived from our study of the theoretically ideal conditional procedure. It is conceivable that fuller use of extended precision and a different organization of the iteration sequence might lead to the practical conditional procedure we were seeking. But we were unable to achieve this. So we turned our attention to an investigation of the extent and direction of the bias in the unconditional procedure UCON.

The Bias in the Unconditional Procedure

Our approach to evaluating the bias in the unconditional estimation procedure commences with the likelihood equations (6) and (20) which connect parameters and data for FCON and UCON. From (6) we have

$$s_i = \sum_r^{L-1} n_r e^{b_r - d_i^*}/(1 + e^{b_r - d_i^*}) \tag{27}$$

in which d_i^* is the biased UCON estimate of δ_i.
From (20) we have

$$s_i = \sum_r^{L-1} n_r e^{-d_i} g_{r-1,i}/g_r.$$

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

$$\begin{aligned}(g_{r-1,i}/g_r) &= (g_{r-1,i}/g_{ri})/(g_r/g_{ri}) \\ &= (g_{r-1,i}/g_{ri})/(1 + e^{-d_i}(g_{r-1,i}/g_{ri})).\end{aligned}$$

Were we to extend our analysis of FCON to derive functions of (d_i) which yield optimal ability estimates b_r we would find $b_r = \log(g_{r-1}/g_r)$ to be a maximum likelihood solution. 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 we may let

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

and the FCON estimation equation becomes

$$s_i = \sum_r^{L-1} n_r e^{b_{ri} - d_i}/(1 + e^{b_{ri} - d_i}) \tag{28}$$

Since (27) and (28) estimate similar parameters for identical data,

$$b_{ri} - d_i = b_r - d_i^*.$$

Thus the bias in the UCON estimate of item i difficulty at score r, when compared with the unbiased d_i from FCON is

$$b_r - b_{ri} = a_{ri}. \tag{29}$$

Our aim is to find a way to correct d_i^* for this bias. The simplest scheme would be to find 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 at score r then

$$k = d_i/d_i^* = d_i/(d_i + a_{ri}) = 1/(1 + c_{ri}). \tag{30}$$

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

(*i*) Specified test length L and standard deviation Z for a normal distribution of item difficulties.

(*ii*) Generated the consequent set of item difficulties (d_i).
(*iii*) Calculated from (d_i) the log symmetric function ratios

$$b_r = \log (g_{r-1}/g_r)$$

$$b_{ri} = \log (g_{r-1,i}/g_{ri}),$$

(*iv*) and hence the biases

$$a_{ri} = b_r - b_{ri}$$

$$c_{ri} = a_{ri}/d_i \quad \text{except } |d_i| < .5.$$

(*v*) Calculated the average relative bias

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

as a basis for estimating k,

The results were systematic. They are summarized in Table 1 for a series of 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 $$c_{..} \simeq 1/(L-1),$$

then $$k \simeq 1/(1 + 1/(L-1)) = (L-1)/L. \tag{31}$$

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 UCON estimates can be shown to be exactly twice the FCON ones. Thus the corrected UCON procedure is indistinguishable from ICON or FCON in the practice of item calibration.

TABLE 1
Relative Bias in the Unconditional Procedure
(Averaged over Items and Scores)

Test Length	Test Dispersion in Item Std. Dev.s 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

Mean Relative Bias $c_{..} = \sum_r^{L-1} \sum_i^{L} c_{ri}/(L-1)(L)$

where $c_{ri} = a_{ri}/d_i$

$a_{ri} = b_r - b_{ri} = d_i^* - d_i$, the bias.

Comparisons of ICON and UCON for Accuracy

Our simulation study of these procedures was set up to expose their biases. We did this by comparing their estimates with known generating parameters for calibrating samples which did not coincide with the tests on the latent variable. We focused on samples for which the test was either too difficult or too easy because that is the situation which produces the poorest item difficulty estimation. To achieve the required comparisons among procedures we set up our study according to the following steps:

(*i*) For each case 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 person abilities, normally distributed with a mean M, a standard deviation SD and an upper truncation point TR, were simulated and exposed to the test. Truncation was set at values which induced 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 generator to produce a stochastic response. These responses were accumulated into the data vectors (s_i) and (n_r) necessary for item calibration. An editing routine ensured that these vectors satisfied the algebraic requirements necessary for finite estimates.

(*iv*) Item parameters and their standard errors were estimated by each procedure: ICON and UCON.

(*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.

A variety of summary statistics were computed for each method. The ones reported in Table 2 are:

(a) MAX DIFF: the maximum difference between a generating item parameter and the mean, over the 15 replications, of its estimates,

(b) RMS: the root mean square of these differences over items,

(c) MEAN ABS: the mean of the absolute value of these differences over items.

These are sufficient to evaluate the relative accuracy of the two procedures.

TABLE 2

Comparison of ICON and UCON for Normal Tests and Truncated Samples (Based on 15 Replications of 500 Persons Each)

	DATA					PROCEDURES					
	Test		Sample			ICON			UCON		
Case	*L*	*Z*	*M*	*SD*	*TR*	MAX DIFF	RMS	MEAN ABS	MAX DIFF	RMS	MEAN ABS
1	20	1	0	0.5	2.0	.05	.03	.08	−.05	.02	.08
2	20	1	0	1.0	2.0	−.07	.03	.08	−.06	.03	.08
3	20	1	1	1.0	2.5	.08	.03	.09	−.06	.03	.09
4	20	1	1	1.5	2.0	−.05	.03	.09	−.07	.03	.09
5	20	1	1	1.5	2.5	.24	.06	.10	−.07	.03	.10
6	40	1	0	0.5	2.0	−.08	.03	.08	−.08	.03	.08
7	40	1	0	1.0	2.0	.13	.03	.08	−.07	.03	.08
8	40	1	1	1.0	2.5	.11	.03	.08	−.11	.03	.09
9	40	1	1	1.5	2.0	.15	.04	.09	.06	.03	.09
10	40	1	1	1.5	2.5	.27	.06	.09	.07	.03	.09
11	40	2	0	2.0	4.5	.31	.07	.10	−.12	.05	.10
12	40	2	2	2.0	4.5	.42	.11	.13	−.30	.06	.12

Where MAX DIFF: the maximum difference between a generating item parameter and the mean, over 15 replications, of its estimates,

RMS: the root mean square of these differences over items,

MEAN ABS: the mean of the absolute value of these differences over items.

In terms of the RMS and MEAN ABS there is little to choose between the procedures, no matter what the test and sample characteristics, so our discussion is confined to the MAX DIFF's. When the items are most appropriate for the calibrating sample (cases 1, 2, 6 and 7) there are no discernible differences between the two procedures. As the mean of the sample shifts away from zero, however (M = 1 or 2), or the severity of truncation and hence skew increases, then the MAX DIFF's tend to increase for both algorithms and particularly for ICON. This latter phenomenon was found to be due to the increasing discrepancy between item and sample characteristics which produced extreme item parameters that are never reasonably estimated by ICON because of accumulated round-off error in the calculation of the symmetric functions.

REFERENCES

Andersen, E. B. Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society*. 1970, 32, 283–301.

Andersen, E. B. The asymptotic distribution of conditional likelihood ratio tests. *Journal of the American Statistical Association,* 1971, 66 (335), 630–33.

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.* Cophenhagen, Denmark: Mentalhygiejnisk Forlag, 1973.

Birnbaum, A. Estimation of an ability. In F. Lord and M. Novick (Eds.), *Statistical Theories of Mental Test Scores.* Reading, Mass.: Addison-Wesley, 1968.

Bock, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika,* 1972, 37, 29–51.

Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen, Denmark: Denmarks Paedagogiske Institut, 1960.

Rasch, G. On general laws and the meaning of measurement in psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics,* 1961, 4, 321–33.

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. and Panchapakesan, N. A procedure for sample-free item analysis, EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 1969, 29, 23–48.