PSYCHOMETRIKA—VOL. 37, NO. 1
MARCH, 1972

ESTIMATING ITEM PARAMETERS AND LATENT ABILITY WHEN RESPONSES ARE SCORED IN TWO OR MORE NOMINAL CATEGORIES*

R. DARRELL BOCK

UNIVERSITY OF CHICAGO

A multivariate logistic latent trait model for items scored in two or more nominal categories is proposed. Statistical methods based on the model provide 1) estimation of two item parameters for each response alternative of each multiple choice item and 2) recovery of information from "wrong" responses when estimating latent ability. An application to a large sample of data for twenty vocabulary items shows excellent fit of the model according to a chi-square criterion. Item and test information curves are compared for estimation of ability assuming multiple category and dichotomous scoring of these items. Multiple scoring proves substantially more precise for subjects of less than median ability, and about equally precise for subjects above the median.

1. *Introduction*

It is well known that subjects who answer a multiple-choice item incorrectly are unlikely to distribute their responses uniformly over the incorrect alternatives. This suggests that "wrong" responses contain information which might be applied to the estimation of latent ability. Such information could be readily recovered if there existed a measurement model which expresses the probability of response to each alternative of the item as a function of ability.

For the case in which the alternatives form a graded scale, a suitable model has been introduced by Samejima [1969]. Her model is an extension of one employed by Thurstone in the scaling of successive categories data [see Bock and Jones, 1968] and by Aitchison and Silvey [1957] in a generalization of probit analysis to graded data.

In the present paper, we propose a latent-trait model for alternatives which are purely *nominal.* This model exploits the multivariate generalization of the logistic response function [see Bock and Jones, 1968] which is implicit in the choice model of Luce [1959] and in the bivariate logistic distribution investigated by Gumbel [1961]. The proposed model defines operating characteristics for each response category such that the probability of response, conditional on ability, is restricted to sum to unity. As in the other latent-trait models, local independence is assumed and the probability of a

* Preparation of this paper was supported in part by N.S.F. Grant GS-2900.

given pattern of item response can be expressed as the product of the corresponding category characteristics conditional on ability.

The statistical procedures which we derive for this model include 1) unconditional and conditional estimation of item parameters, 2) estimation of latent ability and measurement error, 3) an item-by-item test of goodness-of-fit of the model, and 4) evaluation of information recovered from the correct and incorrect alternatives. An application of the model to data for a set of vocabulary items is presented.

Although the discussion in this paper is phrased in terms of multiple-choice items, it will be apparent that the results apply to nominal scoring of responses in any format, as, for example, categorical classification of free response in a projective test [see Beck, 1950, Vol. I].

2. The Data

Suppose that each of N subjects responds to n multiple-choice items, of which the j-th item has m_j alternatives. If some subjects omit or fail to complete some of the items, it may be well to include "no response" as one of the alternatives.

Let the alternatives for item j be indexed $k_j = 1, 2, \cdots, m_j$ in any arbitrary order. Assuming the subject is permitted at most one response to each item, we may then express his response pattern (vector) for the n items as $\mathbf{k} = [k_1, k_2, \cdots, k_n]$. There are $w = \prod_j^n m_j$ possible distinct patterns.

3. The Response Model

A response model suitable for categorically scoring items is contained in the general model for multinomial response relations proposed by Bock [1970]. For present purposes, this model specializes as follows:

Let θ be a value on the continuum of latent ability underlying the response to the test items. Then the probability that a subject of ability θ will respond to item j in category k_j is given by,

$$\Psi_{jk_j}(\theta) = \exp\,[z_{jk_j}(\theta)] \Big/ \sum_{h=1}^{m_j} \exp\,[z_{jh}(\theta)], \tag{1}$$

where

$$z_{jh}(\theta) = c_{jh} + a_{jh}\theta, \qquad h = 1, 2, \cdots, k_j, \cdots, m_j\,. \tag{2}$$

The quantities c_{jh} and a_{jh}, are item parameters associated with the h-th category of item j. The vector comprised of components $z_{j1}(\theta)$, $z_{j2}(\theta)$ and $z_{jm_j}(\theta)$ may be called a *multivariate logit.*

This model has a plausible psychological interpretation: Each alternative of item j is assumed to give rise to a quantitative "response tendency" in a given subject. In the population of subjects of ability θ, these tendencies are assumed to be normally and independently distributed. If each subject chooses the alternative for which his tendency is *maximal,* the proportion of subjects in the population who choose alternative k_j is closely approximated by (1). For the bivariate case, the accuracy of this approximation has been investigated by Gumbel [1961]. An empirical test of the approximation in a psychological application has been described by Bock and Jones [1968, p. 133].

An important property of (1) is its invariance with respect to translation of the logit. We may therefore subject the elements of this vector to an arbitrary linear restriction such as

$$\sum_h^{m_j} z_{jh}(\theta) = 0.$$

This implies that the item parameters in (2) are subject to the same restriction:

$$\sum_h^{m_j} c_{jh} = 0 \quad \text{and} \quad \sum_h^{m_j} a_{jh} = 0.$$

Our problem now is to estimate these restricted parameters for each item, and to estimate latent ability given a response pattern of the form described in Section 2.

4. Estimation

The estimation of item parameters may be approached from an *unconditional* or *conditional* point of view. The unconditional approach, exemplified by the work of Bock and Lieberman [1970] on the normal ogive model for dichotomous items, involves the following steps: 1) for purposes of calibrating items, the subjects are assumed to be randomly sampled from a population with a known distribution of latent ability, 2) the response function is integrated with respect to this distribution in order to obtain an expression for the likelihood of the item parameters free of θ, 3) values for the item parameters are estimated by maximum likelihood, and 4) these estimated values of the item parameters are then used to estimate the ability of given subjects in terms of scores which are automatically normed to the population from which the calibrating sample was drawn. Note that only with respect to estimation of item parameters is this approach referred to as *unconditional,* that is, the abilities of the subjects are not necessarily assumed to be estimated with respect to a population of items.

The alternative approach, of which the work of Lord [1968] on the logistic model for dichotomous items is an example, may be termed *conditional* estimation. Lord expresses the probability of the sample as a function of item parameters and the latent abilities of the given subjects from whom the data was obtained. He then estimates item parameters and latent ability simultaneously by maximum likelihood. The estimation of the item parameters is conditional in the sense that the abilities of the subjects are regarded as fixed unknowns rather than as sampled quantities. Abilities estimated in this way have arbitrary origin and unit of scale and are not normed to a specified population. In the present paper both unconditional and conditional estimation of the item parameters in (1) are discussed.

4.1 Unconditional Estimation

If the principle of local independence is assumed, *i.e.*, that responses of subjects of the same ability to different items are statistically independent, the probability of observing some response vector

$$\mathbf{k}_i = [k_{i1}, k_{i2}, \cdots, k_{ij}, \cdots, k_{in}]$$

conditional on θ is

$$P(\mathbf{k}_i \mid \theta) = \prod_{j=1}^{n} \Psi_{jk_{ij}}(\theta). \tag{4}$$

Now suppose that θ is distributed normally with mean zero and variance 1 in the population from which the calibration sample was drawn. Then the *unconditional* probability of observing a vector $\mathbf{k}_i$ is

$$P_i = P(\mathbf{k}_i) = [1/\sqrt{(2\pi)}] \int_{-\infty}^{\infty} P(\mathbf{k}_i \mid \theta) e^{-\theta^2/2} \, d\theta. \tag{5}$$

Using (5), we may compute the probability of the sample distribution of response vectors. Suppose all possible distinct vectors are indexed in any arbitrary order by $i = 1, 2, \cdots, w$. Then the probability of observing a sample in which r_i of N subjects respond with the i-th vector is given by

$$P = \frac{N!}{\prod_i^w r_i!} \prod_i^w P_i^{r_i} \tag{6}$$

From (6), we may express the likelihood of the item parameters given the observed r_i. Before doing so, however, it proves convenient to substitute, for restriction (3), a reparameterization in terms of $2(m_j - 1)$ linearly independent parameters. By reducing the number of parameters by $2n$, the approach leads to more economical computing procedures than alternative devices such as maximization subject to a restriction by means of Lagrange multipliers.

Using the parameterization presented in Bock [1970], we express the logits for item j as elements of the row vector

$$\underset{1 \times m_j}{\mathbf{z}_j'} = KB_jA_j ,$$

where

$$\underset{1\times 2}{K} = [1 \;\; \theta],$$

$$\underset{2\times m_j}{B_j} = \begin{bmatrix} c_{j1} & c_{j2} & \cdots & c_{jm_j} \\ a_{j1} & a_{j2} & \cdots & a_{jm_j} \end{bmatrix},$$

and

$$\underset{m_j\times m_j}{A_j} = \begin{bmatrix} 1-\dfrac{1}{m_j} & -\dfrac{1}{m_j} & \cdots & -\dfrac{1}{m_j} \\ -\dfrac{1}{m_j} & 1-\dfrac{1}{m_j} & \cdots & -\dfrac{1}{m_j} \\ \vdots & \vdots & & \vdots \\ -\dfrac{1}{m_j} & -\dfrac{1}{m_j} & \cdots & 1-\dfrac{1}{m_j} \end{bmatrix}.$$

The latter is the projection operator which implements restriction (3).

Now let $A_j = S_j T_j$ and reparameterize as follows:

$$\begin{aligned} \mathbf{z}_j &= K(B_j S_j)T_j\,, \\ &= K\Gamma_j T_j \end{aligned} \tag{7}$$

where

$$\underset{m_j\times(m_j-1)}{S_j} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & -1 \end{bmatrix}$$

$$\underset{2\times(m_j-1)}{\Gamma_j} = \begin{bmatrix} \gamma_{j1} & \gamma_{j2} & \cdots & \gamma_{j,m_j-1} \\ \alpha_{j1} & \alpha_{j2} & \cdots & \alpha_{j,m_j-1} \end{bmatrix}$$

and

$$\underset{(m_j-1)\times m_j}{T_j} = \begin{bmatrix} \dfrac{1}{m_j} & \dfrac{1}{m_j}-1 & \dfrac{1}{m_j} & \cdots & \dfrac{1}{m_j} \\ \dfrac{1}{m_j} & \dfrac{1}{m_j} & \dfrac{1}{m_j}-1 & \cdots & \dfrac{1}{m_j} \\ \vdots & \vdots & \vdots & & \vdots \\ \dfrac{1}{m_j} & \dfrac{1}{m_j} & \dfrac{1}{m_j} & \cdots & \dfrac{1}{m_j}-1 \end{bmatrix}$$

Proceeding with the maximum likelihood estimation, we write the log likelihood,

$$\log L = \sum_{i=1}^{w} \left[\log \frac{N!}{\prod_{i}^{w} r_i} + r_i \log P_i \right],$$

and obtain the likelihood equations,

$$\frac{\partial \log L}{\partial \gamma_{lg}} = \sum_{i=1}^{w} \frac{r_i}{P_i} \frac{\partial P_i}{\partial \gamma_{lg}} = 0, \qquad \frac{\partial \log L}{\partial \alpha_{lg}} = \sum_{i=1}^{w} \frac{r_i}{P_i} \frac{\partial P_i}{\partial \alpha_{lg}} = 0,$$

for

$$l = 1, 2, \cdots n \quad \text{and}$$

$$g = 1, 2, \cdots m_j - 1.$$

To express the derivatives of P_i, let k_{ij} and k_{il} be the j-th and l-th element of the i-th response vector, respectively. Then,

$$\frac{\partial P_i}{\partial u_{lg}} = [1/\sqrt{(2\pi)}] \int_{-\infty}^{\infty} \frac{\partial z_{l,k_{il}}(\theta)}{\partial u_{lg}} [1 - \Psi_{lk_{il}}(\theta)] \prod_{j \neq l} [\Psi_{jk_{ij}}(\theta)] e^{-\theta^2/2} \, d\theta. \tag{8}$$

Writing $t_{lgk_{il}}$ to represent the g, k_{il} element of the matrix T_l, we obtain from (7),

$$\frac{\partial z_{lk_{il}}(\theta)}{\partial \gamma_{lg}} = t_{lgk_{il}},$$

$$\frac{\partial z_{lk_{il}}(\theta)}{\partial \alpha_{lg}} = \theta t_{lgk_{il}},$$

according as γ_{lg} or α_{lg} replaces u_{lg}.

Let us assemble into a $2 \prod_j^n (m_j - 1) \times 1$ vector the first derivatives evaluated at some provisional values of the $\hat{\gamma}$ and $\hat{\alpha}$,

$$\begin{bmatrix} \varphi_{\hat{\gamma}} \\ \varphi_{\hat{\alpha}} \end{bmatrix}_p = \begin{bmatrix} \dfrac{\partial \log L}{\partial \gamma_{lg}} \\ \dfrac{\partial \log L}{\partial \alpha_{lg}} \end{bmatrix}_{\substack{\gamma = \hat{\gamma}_p \\ \alpha = \hat{\alpha}_p}}$$

with the second subscript varying first. An excellent approximation to the matrix of *second* derivatives may be computed from the provisional values of the first derivatives as follows [see Kendall and Stuart, Vol. II, 1961]:

$$\begin{bmatrix} \varphi_{\hat{\gamma}\hat{\gamma}} & \varphi_{\hat{\gamma}\hat{\alpha}} \\ \varphi_{\hat{\alpha}\hat{\gamma}} & \varphi_{\hat{\alpha}\hat{\alpha}} \end{bmatrix}_p = \sum_{i=1}^{w} \frac{1}{\hat{P}_{i(p)}} \begin{bmatrix} \dfrac{\partial P_i}{\partial \gamma_{lg}} \\ \dfrac{\partial P_i}{\partial \alpha_{lg}} \end{bmatrix}_{\substack{\gamma = \hat{\gamma}_p \\ \alpha = \hat{\alpha}_p}} \cdot \begin{bmatrix} \dfrac{\partial P_i}{\partial \gamma_{lg}} & \dfrac{\partial P_i}{\partial \alpha_{lg}} \end{bmatrix}_{\substack{\gamma = \hat{\gamma}_p \\ \alpha = \hat{\alpha}_p}}, \tag{9}$$

where $\hat{P}_{i(p)}$ is a provisional estimate of P_i .

Then improves estimates of the $\hat{\gamma}$ and $\hat{\alpha}$ may be obtained by a Newton–Raphson iteration,

$$(10) \qquad \begin{bmatrix} \hat{\gamma}_{p+1} \\ \hat{\alpha}_{p+1} \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_p \\ \hat{\alpha}_p \end{bmatrix} - \begin{bmatrix} \varphi_{\hat{\gamma}\hat{\gamma}} & \varphi_{\hat{\gamma}\hat{\alpha}} \\ \varphi_{\hat{\alpha}\hat{\gamma}} & \varphi_{\hat{\alpha}\hat{\alpha}} \end{bmatrix}^{-1} \begin{bmatrix} \varphi_{\hat{\gamma}} \\ \varphi_{\hat{\alpha}} \end{bmatrix}_p$$

A method of finding initial provisional values from which the Newton–Raphson iteration will converge is discussed in connection with Example 1 below.

4.1.1 Numerical Methods 1

a) The integrals in (5) and (8) are readily evaluated by Gauss–Hermite quadrature. In this method, the integral

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}\, dx$$

is approximated by a sum of the form

$$\sum_{t=1}^{n_q} G_t f(X_t)$$

Tables of the coefficients G_t corresponding to the n_q quadrature points X_t are given by Stroud and Secrest [1966].

b) Since the number of possible response patterns may be very large, it is not practical to evaluate all possible terms in the likelihood equations. In any practical sample, however, only a limited number of these patterns will occur. For patterns which do not occur, the frequency, r_i , will, of course, equal zero and the corresponding terms in the likelihood equations will not need to be calculated.

The approximation to the matrix of second derivatives, on the other hand, contains non-zero terms for all w patterns and cannot easily be computed when w is large. We might assume, however, that the terms which correspond to patterns occurring in the sample are representative of the remaining terms. If so, the sum in (9) over, say, v sample patterns, when multiplied by a factor w/v will approximate the matrix of second derivatives. This approximation has been used successfully in the example which follows.

*4.1.2. Example**

Responses of 557 students of the University of Chicago Laboratory School were obtained for 160 Vocabulary items from obsolete forms of the Cooperative Reading Tests. For purposes of the present example, four items

* Original computer programming by Ronald Skirmont, Center for Studies in Information Processing, Inc., Chicago, Illinois.

were selected for which the data show better than chance responding to at least one of the four incorrect alternatives in each item. In order to simplify the calculations, we assigned alternatives with little frequency to the same category and renumbered categories so that 1 represented the correct alternative, 2 the most used incorrect alternative, and 3 the three remaining alternatives and no response, combined.

Initial estimates of the item parameters were obtained by regarding the correct alternative vs. each category of incorrect alternatives as a binary item. It was then assumed that all $\alpha_{jh} = 1$, and the corresponding γ_{jh} were estimated by the log of the ratio of number of correct to the number of incorrect responses in the sample. From these estimates the Newton–Raphson iterations described by (10) converged in 9 iterations. IBM 7094 computing time was 4.18 minutes. The final estimates of item parameters, their standard errors as approximated by the square roots of the diagonal elements inverse matrix of second derivatives, are shown in columns 4 and 5 of Table 1.

In this solution $w = 3^4 = 81$ and v was set equal to 63; *i.e.*, frequencies of occurrence of the 63 frequent response patterns, plus a balance category, constitute the multinomial data to which the model was fitted. Goodness-of-fit was assessed by computing the Pearsonion chi-square for observed and expected frequency in these categories. The degrees of freedom in this case equals the number of independent categories (63) minus the number of parameters fitted (16).

There appears to be some indication of lack of fit of the model, although the chi-square value is by no means extreme in view of the large sample size. Furthermore, the chi-square should perhaps be discounted because of the extremely small expected values of many of the 64 multinomial frequencies. The fact that the conditional solution shows no sign of lack of fit encourages this view. The estimates show considerable variation in the parameters within items, suggesting that information about ability contained in the wrong responses will vary according to which alternative is chosen by the subject who gives a wrong response. This impression will be confirmed when we interpret the solution in terms of item category operating characteristics and information functions in Section 4.2.3.

The unconditional solution is obviously not practical computationally when more than a few items are involved. It is valuable, however, as an independent check on the conditional method, which is much less demanding both in its assumptions and in the extent of computation required.

4.2 Conditional Estimation

In this section, the data are re-assembled in a multinomial form which is different from that of 4.1. The conventional r and P notation for multinomial occurrences and probabilities is retained, but the reader is to understand that the meaning is not the same as that in 4.1. However, the notation for the item parameters, latent ability, and the reparameterization scheme remains the same throughout.

To obtain a practical procedure for conditional estimation of the item parameters, we find it expedient to relax the local independence principle by assuming that subjects whose latent ability is in the "neighborhood" of θ respond independently to different items. The purpose of this relaxation is to justify grouping subjects for whom provisional estimates of latent scores are similar. It is assumed that the actual latent scores of subjects in such groups are confined to a sufficiently small neighborhood to assure independent responses. The question of how small this neighborhood must be to justify the local independence assumption is left for later empirical study. For the moment, let us proceed by assigning to group i those subjects whose latent ability is in the neighborhood of θ_i. Then according to the relaxed local independence principle, the probability that $r_{ij1}, r_{ij2}, \cdots, r_{ijh}, \cdots, r_{ijm}$ out of the N_i subjects in this group will respond to item j in categories $1, 2, \cdots, h, \cdots, m_j$, respectively, is,

$$P(\mathbf{r}_{ij} \mid \theta_i) = \frac{N_i!}{\prod_h^m r_{ijh}!} \prod_h^{m_j} [\Psi_{jh}(\theta_i)]^{r_{ijh}},$$

where $\mathbf{r}_{ij}$ is the vector of response frequencies for group i and item j.

The log likelihood of γ_{jl}, α_{jl} and θ_i, given data for q groups and n items, is therefore,

$$\log L = \sum_{i=1}^{q} \sum_{j=1}^{n} \sum_{h=1}^{m_j} \left[\log \frac{N_i!}{\prod_h^{m_j} r_{ijh}!} + r_{ijh} \log \Psi_{jh}(\theta_i) \right]$$

To simplify notation, let $\Psi_{jh}(\theta_i) = \Psi_{ijh}$. Then the likelihood equations may be expressed as

$$\frac{\partial \log L}{\partial \gamma_{jl}} = \sum_i^q \sum_h^{m_j} \frac{r_{ijh}}{\Psi_{ijh}} \frac{\partial \Psi_{ijh}}{\partial \gamma_{jl}} = 0$$

$$\frac{\partial \log L}{\partial \alpha_{jl}} = \sum_i^q \sum_h^{m_j} \frac{r_{ijh}}{\Psi_{ijh}} \frac{\partial \Psi_{ijh}}{\partial \alpha_{jl}} = 0$$

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j^n \sum_h^{m_j} \frac{r_{ijh}}{\Psi_{ijh}} \frac{\partial \Psi_{ijh}}{\partial \theta_i} = 0$$

Now define the vectors $\mathbf{\Psi}'_{ij} = [\Psi_{ij1}\Psi_{ij2} \cdots \Psi_{ijm_j}]$, $\boldsymbol{\gamma}'_j = [\gamma_{j1}\gamma_{j2} \cdots \gamma_{j(m_j-1)}]$, $\boldsymbol{\alpha}'_j = [\alpha_{j1}\alpha_{j2} \cdots \alpha_{j(m_j-1)}]$ and $\mathbf{r}'_{ij} = [r_{ij1}r_{ij2} \cdots r_{ijm_j}]$. Then the likelihood equations for the item parameters may be expressed as the vector derivatives

$$\frac{\partial \log L}{\partial \boldsymbol{\gamma}_j} = \sum_i^q T_j(\mathbf{r}_{ij} - N_i \mathbf{\Psi}_{ij}) = 0, \tag{11}$$

$$(12) \qquad \frac{\partial \log L}{\partial \boldsymbol{\alpha}_j} = \sum_i^q \theta_i T_j(\mathbf{r}_{ij} - N_i \boldsymbol{\Psi}_{ij}) = 0,$$

and the likelihood equation for θ is

$$(13) \qquad \frac{\partial \log L}{\partial \theta_i} = \sum_j^n \boldsymbol{\alpha}_j' T_j(\mathbf{r}_{ij} - N_i \boldsymbol{\Psi}_{ij}) = 0,$$

where $\boldsymbol{\alpha}_j$ is the second row of Γ_j .

To determine under what conditions these equations have a solution, we must investigate the dimensionality and shape of the likelihood surface over the parameter space. In this problem each of the parameters can range over the entire real line, so that the parameter space is potentially the $d = 2\sum_j^n (m_j - 1) + q$ dimensional linear manifold. However, since the sum with respect to j of (11) pre-multiplied by $\boldsymbol{\alpha}_j'$ equals identically the sum with respect to i of (13), and the sum with respect to j of (12) pre-multiplied by $\boldsymbol{\alpha}_j'$ equals identically the sum with respect to i of (13) multiplied by θ_i , no more than $d - 2$ of the likelihood equations are independent. (We shall see below that exactly $d - 2$ of the likelihood equation are in general independent). Thus, the dimensionality of the parameter space must be reduced by two, for example, by arbitrarily assigning the origin and scale of θ.

Proceeding to investigate the shape of the surface, we obtain the second derivatives of the log likelihood in terms of the $m_j \times m_j$ matrices,

$$W_{ij} = N_i \cdot \begin{bmatrix} \Psi_{ij1}(1 - \Psi_{ij1}) & \Psi_{ij1}\Psi_{ij2} & \cdots & -\Psi_{ij1}\Psi_{ijm_j} \\ -\Psi_{ij2}\Psi_{ij1} & \Psi_{ij2}(1 - \Psi_{ij2}) & \cdots & -\Psi_{ij2}\Psi_{ijm_j} \\ \vdots & \vdots & & \vdots \\ -\Psi_{ijm_j}\Psi_{ij1} & -\Psi_{ijm_j}\Psi_{ij2} & \cdots & \Psi_{ijm_j}(1 - \Psi_{ijm_j}) \end{bmatrix}$$

The submatrices which comprise the $d \times d$ matrix of second derivatives are as follows:

$$(14) \qquad \frac{\partial^2 \log L}{\partial \boldsymbol{\gamma}_j \, \partial \boldsymbol{\gamma}_k} = \begin{cases} -\sum_i^q T_j W_{ij} T_j' & j = k \\ 0 & j \neq k \end{cases}$$

$$(15) \qquad \frac{\partial^2 \log L}{\partial \boldsymbol{\gamma}_j \, \partial \boldsymbol{\alpha}_k} = \begin{cases} -\sum_i^q \theta_i T_j W_{ij} T_j' & j = k \\ 0 & j \neq k \end{cases}$$

$$(16) \qquad \frac{\partial^2 \log L}{\partial \boldsymbol{\alpha}_j \, \partial \boldsymbol{\alpha}_k} = \begin{cases} -\sum_i^q \theta_i T_j W_{ij} T_j' & j = k \\ 0 & j \neq k \end{cases}$$

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\gamma}_j \, \partial \theta_i} = -T_j W_{ij} T_j' \boldsymbol{\alpha}_j \tag{17}$$

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\alpha}_j \, \partial \theta_i} = -\theta_i T_j W_{ij} T_{j-i}' + T_j(\mathbf{r}_{ij} - N_i \boldsymbol{\Psi}_{ij}) \tag{18}$$

$$\frac{\partial^2 \log L}{\partial \theta_h \, \partial \theta_i} = \begin{cases} -\sum_j^n \boldsymbol{\alpha}_j' T_j W_{ij} T_j' \boldsymbol{\alpha}_j & h = i \\ 0 & h \neq i \end{cases} \tag{19}$$

The only terms in this matrix which contain sample quantities occur in the cross-derivatives of $\boldsymbol{\alpha}$ and θ. For a fixed number of items, the expected values of these quantities in the population of subjects is zero.

$$T_j E(\mathbf{r}_{ij} - N_i \mathbf{P}_{ij}) = 0$$

Thus, upon reducing the matrix of second derivatives to echelon (triangular) form, we find that the expected value of the $d \times d$ matrix of second derivatives is negative semi-definite of rank $d - 2$, provided the derivatives are evaluated at any interior point of the parameter space (*i.e.*, for any finite parameter values). By omitting rows and columns corresponding to any two θ_i, values of which may be assigned arbitrarily, the matrix can be made a negative-definite of rank $d - 2$ (*i.e.*, the negative of a $(d - 2) \times (d - 2)$ positive-definite matrix).

The expected likelihood surface is therefore concave down for all interior points of the restricted parameter space and thus has a unique maximum when the maximum does not occur at a boundary point. This means that with increasing sample size and a fixed number of ability groups (*i.e.*, q fixed), the probability that the *sample* likelihood surface has a unique maximum approaches 1 if the maximum is at an interior point.

If the number of ability groups increases with the number of subjects, however, difficulties may arise. In order for the negative of the reduced matrix of second derivatives to be positive-definite, it is necessary that the negative of every cross-derivative satisfy the Cauchy inequality. In particular, we must have

$$\left[\frac{\partial^2 \log L}{\partial \alpha_{jh} \, \partial \theta_i}\right]^2 < \frac{\partial^2 \log L}{\partial \theta_i^2} \cdot \frac{\partial^2 \log L}{\partial \alpha_{ih}^2} ,$$

or

$$\frac{[\alpha_{jh}\theta_i \mathbf{t}_{jh}' W_{ij} \mathbf{t}_{jh} + \mathbf{t}_{jh}'(\mathbf{r}_{ijh} - N_i \boldsymbol{\Psi}_{ijh})]^2}{\left[N_i \sum_j^n \alpha_{jh}^2 \mathbf{t}_{jh}' W_{ij} \mathbf{t}_{jh}\right]\left[N_i \sum_i^q \theta_i^2 \mathbf{t}_{jh}' W_{ij} \mathbf{t}_{jh}\right]} < 1 \tag{20}$$

In the most extreme case, we would assign each subject to a distinct ability group. Then $N_i = 1$, and $r_{ijh} = 1$ or 0 according as the subject in group i responds to item j in category h or not. If subject happens to respond in this category when, by virtue of his ability and the difficulty and discriminating ability for this item and category contrast, the quantity Ψ_{ijh} is extreme and the elements of W_{ij} are small for all j and all i, it is quite possible for the numerator of (20) to be in the neighborhood of 1 while the denominator is fractional. Thus there is some risk that this necessary condition will be violated in given data. On the other hand, if extreme subjects or items are eliminated and the number of subjects and items are not too small, we might expect that the summing of terms in the denominator should guarantee that this condition (20) is met.

Although (20) is not, of course, sufficient for positive-definiteness, all other necessary conditions involve linear combinations of rows and columns in the matrix of second derivatives, and the fortuitous combinations which would violate the Cauchy inequality would appear to be less probable than violations by particular cross-derivatives. Thus, it seems likely that if (20) is met for all ijh, a unique maximum will exist and be attainable by suitable iterative numerical procedures. This conclusion appears to agree with the experience of Lord [1968], who after eliminating certain extreme subjects and items was able to obtain convergence of a repeated substitution procedure for maximum likelihood estimates of a logistic model for dichotomous items. Similarly, we present in 4.2.3 an example in which convergence was obtained after eliminating (*i.e.*, assigning a default value to) one of 557 subjects. No items were eliminated.

4.2.1. Numerical Methods 2

The preceeding analysis suggests that, by grouping subjects with respect to provisionally estimated ability, we may achieve a practical solution of the likelihood equations (11), (12) and (13) along the lines of Lord's [1968] method of repeated substitutions. Provisional values of item parameters would first be assumed in order to calculate provisional maximum likelihood estimates of ability. Then these provisional estimates would be used to group the subjects to obtain improved estimates of the item parameters. Using these item parameters, new estimates of ability would be calculated, etc., until the estimates ceased to change. This procedure is equivalent to a Newton–Raphson solution in which the cross-derivatives $\partial^2 \log L/\partial\gamma_j\, \partial\theta_i$ and $\partial^2 \log L/\partial\alpha_j\, \partial\theta_i$ are assumed nil. Specific computational steps in the procedure are as follows:

Step 1

In preparation for the assignment of subjects to ability groups, the maximum likelihood estimate of latent ability is calculated, using the provisional values for item parameters at that stage. At the first stage, satisfactory provisional values are

$$\alpha_{jl} = 0$$

and

$$(21) \qquad \gamma_{jl} = -\ln (p_{j1}/p_{jl}), \qquad l \neq 1,$$

where p_j is the proportion of subjects responding in the normally "correct" category indexed 1, say, and p_{jl} is the proportion responding in the l-th incorrect category.

The likelihood equation for the ability estimate is obtained from (13) by specializing N_i to 1, in which case the vector

$$\mathbf{r}_{ij} = [0, 0, \cdots, r_{ijk_j} = 1, \cdots, 0]$$

represents the observed response of subject i to item j. Equation (13) can then be solved by Newton–Raphson, with the value of the second derivative computed from (19). A satisfactory initial value of θ_i for these iterations is either $\theta_i = 0$, or the value from a previous cycle of the repeated substitutions. Except when the ability estimate is infinite, the likelihood equation for θ_i has a unique solution as discussed above. If it appears in the Newton iterations that a provisional estimate of the θ_i is tending to infinity, it may be set to an arbitrary large default value and the computations continued.

Step 2

Subjects are ranked according to their provisional estimated ability, and the ranking is divided into q fractiles. Each fractile contains N_i subjects, where all N_i are equal if N/q is integral, or, otherwise, one additional subject is assigned to each of remainder (N/q) central fractiles. The responses in each category of each item for each fractile are then counted to obtain vectors $\mathbf{r}_{ij}$. The value of θ_i assigned i-th fractile is the *median* value of the provisional ability estimates of subjects in that fractile. Provided less than half the subjects in the extreme fractiles have infinite ability, all of the median-θ_i are finite. The location and scale of the median-θ_i are assigned by setting the sample mean to zero and variance to unity, where these statistics are computed by the usual formulas for grouped data with the group medians as class marks.

Alternatively, a conditional solution based on the same assumptions as the unconditional solution may be obtained by regarding the subjects to be a random sample from a specified population. In this case, percentage points of the cumulative distribution function for the population may be assigned as the class marks of the fractiles. Thus, only the ordinal information in the provisional estimates of ability is used. If a normal distribution of ability is assumed, the centroid of the intervals under the normal curve corresponding to differences between fractiles is a suitable choice for the class mark [see Bock and Jones, 1968, p. 251].

Step 3

Given the N_i, $\mathbf{r}_{ij}$, and median-θ_i obtained in Step 2, maximum likelihood estimates of the item parameter contrasts may be obtained by solving (11) and (12) simultaneously by Newton–Raphson. Elements of the matrix of second derivatives for this solution are given by (14), (15) and (16). The solution is the same as the generalized logit analysis presented, with computational examples, in Bock [1970]. Since the cross-derivatives between items are null, the estimates for each item may be obtained separately. The provisional values for the item parameters which were used in estimating ability for subjects may be used to start the Newton–Raphson iterations. Again, these likelihood equations have a unique solution, provided the maximum likelihood estimates are finite. Usually infinite values of the estimates can be avoided by pooling response categories to avoid zero values in the vectors of response frequencies for the item involved.

Step 4

With the estimates obtained in Step 3 as the new provisional values of the item parameters, the computations return to Step 1. These repeated substitutions are continued until estimates in successive cycles agree to some predetermined number of decimal places. A successful application of this method is described in Section 4.2.3.

4.2.2. Standard Errors and Information

In maximum likelihood estimation, the limiting variance-covariance matrix of the estimates is given by the negative inverse of the matrix of second derivatives; large-sample standard errors are thus the square roots of the diagonal elements of this matrix. In the case of the item parameters, the formulas for the standard errors, both in unconditional and conditional estimation, are difficult to express explicitly, but numerical values may be obtained from the inverse matrix of second derivatives computed for the Newton iteration at the final stage of the solution. Examples of such values are shown in Table 1. These estimated standard errors become exact as the N_i increases. In contrast, the large-sample standard error of the estimated latent ability is readily expressed as

$$\text{(22)}\qquad \begin{aligned} \text{S.E. }(\hat{\theta}_i) &= \left[\sum_{j}^{n} \boldsymbol{\alpha}_j' T W_{ij} T \boldsymbol{\alpha}_j\right]^{-1/2} \\ &= \left[\sum_{j}^{n} \mathbf{a}_j' W_{ij} \mathbf{a}_j\right]^{-1/2}. \end{aligned}$$

In this case, the S.E. becomes exact as n, the number of items, increases. If the number of subjects is also large, the estimates of $\boldsymbol{\alpha}_j$ or $\mathbf{a}_j$, obtained either from the unconditional or conditional solution, may be substituted

TABLE 1

UNCONDITIONAL AND CONDITIONAL ESTIMATION OF ITEM PARAMETER CONTRASTS OF FOUR VOCABULARY ITEMS FROM THE COOPERATIVE ACHIEVEMENT TESTS (N=557)

Items	Category Code	Percent Response	Category Contrast	Unconditional Estimates and (S.E.)	Conditional Estimates and (S.E.)
No.8 mirth				γ_j	
Form merriment	1	61	1-2	1.34 (.15)	1.34 (.13)
C2T folly	2	14	1-3	1.63 (.31)	1.27 (.14)
weight	3	3		α_j	
perfume	3	5			
affection	3	2	1-2	1.63 (.41)	1.08 (.17)
NR	3	15	1-3	2.95 (.60)	2.02 (.18)
		100			
No.16 exaltation				γ_j	
Form					
C2T rejoicing	1	57	1-2	1.11 (.11)	1.07 (.12)
praise	2	19	1-3	1.08 (.13)	1.18 (.13)
disappointment	3	1		α_j	
forgiveness	3	3			
worship	3	10	1-2	.53 (.16)	.45 (.13)
NR	3	10	1-3	1.20 (.18)	1.43 (.15)
		100			
No.20 harassed				γ_j	
Form worried	1	37	1-2	-.60 (.10)	-.52 (.10)
C2T angered	2	51	1-3	1.83 (.23)	1.83 (.23)
fostered	3	5		α_j	
comforted	3	1			
rescued	3	1	1-2	1.46 (.23)	1.01 (.12)
NR	3	5	1-3	2.27 (.35)	1.96 (.22)
		100			
No.24 raiment				γ_j	
Form clothing	1	43	1-2	1.60 (.18)	1.61 (.17)
C2R humor	2	7	1-3	-.23 (.10)	-.23 (.10)
shadow	3	5			
light	3	6		α_j	
jewelry	3	5	1-2	.92 (.30)	.62 (.20)
NR	3	34	1-3	1.28 (.18)	1.21 (.12)
		100			

when evaluating (22) numerically. Note that (22) depends upon θ_i (becaues W_{ij} depends upon θ_i) but not upon $\mathbf{r}_{ij}$. Thus, the measurement error in estimating latent ability is the same for subjects of the same estimated ability independent of their individual patterns of response to the test items.

For certain purposes, it may be useful to calculate the average measurement error with respect to a specified distribution of ability. If θ is distributed normally with mean 0 and variance 1 in the population and the estimate of θ is scaled accordingly, the average error variance is

$$\bar{\sigma}_\theta^2 = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \left(\sum_j^n \mathbf{a}_j' W_j \mathbf{a}_j \right)^{-1} e^{-\theta^2/2} \, d\theta, \tag{23}$$

which may be readily evaluated by Gauss–Hermite quadrature. The average reliability of the test is the complement of (23).

The concepts of error variance and reliability are, however, more relevant to classical test theory than to the latent trait theory employed here. Birnbaum [1968] has shown in the dichotomous case that the precision of the maximum likelihood estimator may be investigated in more detail by means of the Fisherian information analysis. Samejima [1969] has extended this form of analysis to the polychotomous case. Her definition of the amount of information contained in a response to item j is, in the present context,

$$\begin{aligned} I_j(\theta) &= \mathbf{a}_j' W_j \mathbf{a}_j \sum_h^{m_j} \Psi_{jh}(\theta) \\ &= \mathbf{a}_j' W_j \mathbf{a}_j \end{aligned} \tag{24}$$

The item information may be partitioned among the response categories by computing individual terms in (24). Thus, the information due to response in category h of item j is

$$I_{jh}(\theta) = \mathbf{a}_j W_j \mathbf{a}_j \Psi_{jh}(\theta). \tag{25}$$

By virtue of the additive property of the information measure, the test information function for the maximum likelihood estimator of θ is the sum of the item information function,

$$I(\theta) = \sum_{j=1}^{n} I_j(\theta). \tag{26}$$

The additive property of item information facilitates its interpertation; for example, a two-fold increase in test information is equivalent to the gain in precision expected when test length is doubled.

*4.2.3 Example** (*continued*)

The conditional solution was carried out on 20 vocabulary items selected from those available in the Laboratory School data. The four items from the unconditional solution were used, and the remainder were chosen from among items which were sufficiently difficult to have some frequency of wrong responses. Ten fractiles ($q = 10$) were used in assigning the 557 subjects to

* Original programming by Donald Kolakowski, University of Connecticut. Program distributed by the Education Statistical Laboratory, University of Chicago.

provisional ability groups during the calculations. For purposes of comparison with the unconditional solution, group centroids corresponding to each decile of the normal curve were assigned as class marks.

After six cycles of repeated substitutions, the estimates of the γ_i and α_i were stable to four and three significant figures, respectively. Goodness-of-fit for each item was assessed by means of a Pearsonian chi-square computed after the sixth cycle. The resulting chi-square values and degrees of freedom for each of the twenty vocabulary items are shown in Table 2. The solution shows remarkedly good fit: none of the chi-square values are significant at the .05 level, and the total chi-square is almost exactly at expectation. A slightly fairer test would compare the total chi-square with $440 - 8 = 432$ degrees of freedom to allow for the 8 independent values assigned to the ten ability groups. Even by this criterion the chi-square is not significant at the .05 level.

Taken in account with the quite different approaches used in the un-

TABLE 2

CONDITIONAL ANALYSIS OF 20 VOCABULARY ITEMS FROM THE COOPERATIVE ACHIEVEMENT TESTS (N=557)

Item Form and Number	Number of Categories	χ^2	d.f.
1. C2R- 2	2	7.07	8
2. -11	2	6.33	8
3. -24	3	21.53	16
4. -31	4	11.24	24
5. -44	4	32.62	24
6. -50	5	35.92	32
7. C2T- 8	3	17.16	16
8. -13	4	20.68	24
9. -16	3	21.68	16
10. -19	4	23.97	24
11. -20	3	17.63	16
12. -27	3	17.67	16
13. -32	4	22.40	24
14. -34	5	38.61	32
15. -38	3	18.10	16
16. -39	4	22.50	24
17. -41	5	44.18	32
18. -44	5	15.84	32
19. -53	5	36.48	32
20. -60	4	26.15	24
		437.56	440

conditional solution, the agreement of parameter contrast estimates shown in columns 5 and 6 of Table 1 is reasonably good and serves to verify the correctness of both solutions.

Category characteristic curves and information functions for two typical items in the conditional solution are shown in Figures 1 and 2. Below each figure are the estimated item parameters, expressed in the restricted form, from which the curves were constructed. To facilitate computation, little-used alternatives have been combined and the alternatives assigned to categories

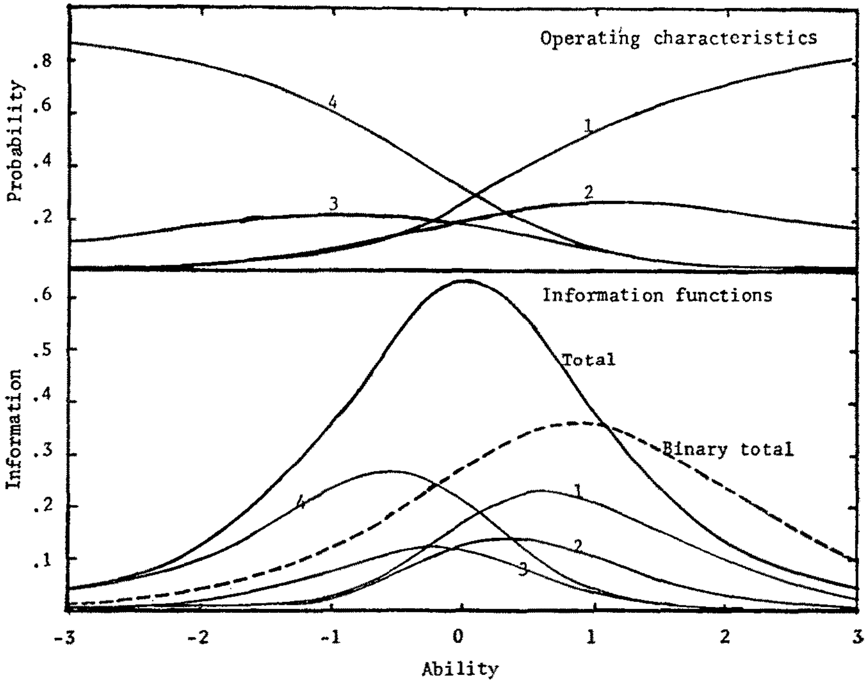


Figure 1. Operating characteristics and information functions for item C2T-19: "DOMICILE"

Estimated Item Parameters

Category	c	a
1. "Residence"	.126	.905
2. "Servant"	-.206	.522
3. "Legal document" "Hiding place" "Family"	-.257	-.469
4. No response	.336	-.959

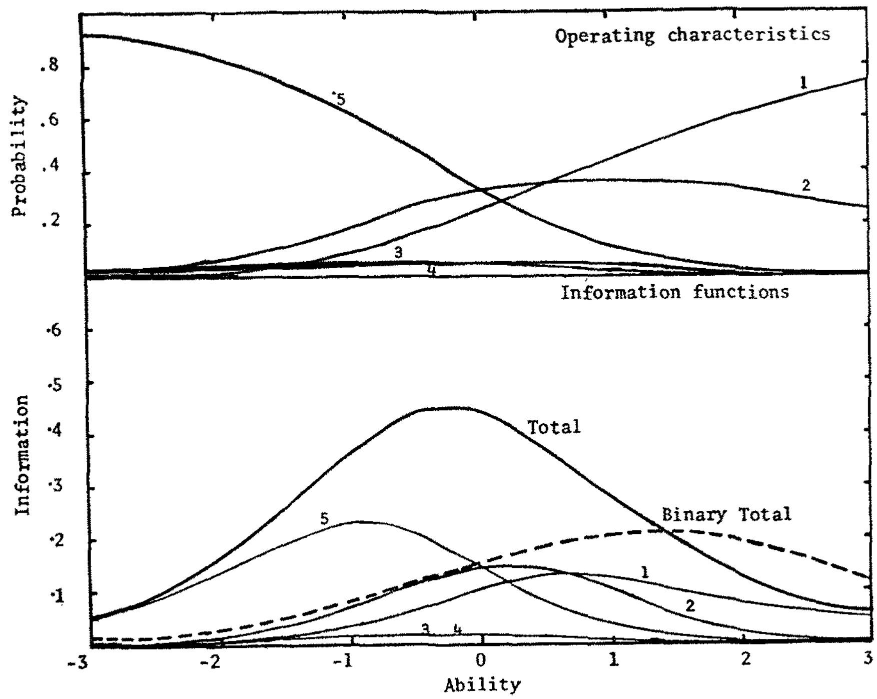


FIGURE 2. Operating characteristics and information functions for Item C2R-50, "ARTIFICE"

Estimated Item Parameters

Category	c	a
1. "Crafty device"	.565	.828
2. "Imitation"	.865	.375
3. "Careless error"	-1.186	-.357
4. "Flavor" "Accident"	-1.199	-.079
5. No response	.993	-.817

in such a way that category 1 corresponds to the correct response as assumed in (21).

To some extent, the form of the category characteristic curves may be deduced from the values of the corresponding restricted item parameters. As inspection of the derivative of the category characteristic with respect to θ shows, the largest algebraic value of $\hat{a}_i$ corresponds to the category with a monotonic increasing curve; whereas the smallest algebraic value corresponds to the category which is always monotonic decreasing. Curves of the remaining categories are non-monotonic and have a maximum at some finite value of θ. The shape of these curves and the locations of the maxima are functions of all the parameters for the item and appear difficult to specify without solving the stationary equations or computing the curves numerically.

For each of the items illustrated, the monotonic increasing curve corresponds to the correct answer. This indicates in each case that the nominal identification of the correct alternative is valid in the sense that it is the alternative chosen by subjects of indefinitely high ability. Conversely, the monotonic decreasing curve corresponds in each case to the "omit" category. This implies that subjects of low ability were omitting these items rather than marking them randomly. There is nothing in Figures 1 and 2 to support the idea, which is sometimes advanced, that subjects who omit have enough knowledge to know they are wrong and are actually of higher ability than those who mark incorrect alternatives.

This observation has a bearing on the question of whether or not subjects responding to multiple-choice items should be encouraged to guess when in doubt. It is clear that if the subjects who are omitting items distribute their responses randomly over the alternatives, it will be necessary to modify (1) so that the probability assigned to the omit category is distributed uniformly over the remaining categories. The model would then be the multi-category analogue of Birnbaum's three-parameter model for binary items. An extension of Birnbaum's result for the effect of guessing on the item information function would then apply—the effect of the guessing would be to diminish information, especially in the region of low ability. Thus, positive loss of information must be expected if subjects, who would otherwise omit, choose alternatives randomly when constrained to respond to all items.

It is, of course, entirely possible that the omitters have partial information which could be recovered if the guess-when-in-doubt instructions were used. The multiple category scoring would then be expected to yield a net gain when the subjects are required to respond to all items. In the present data, the characteristic curves for the omit category do not suggest that this is the case. However, the particular curves, depending as they do on specific item content, subject population, and test instructions, may not be typical.* Empirical study of the relative information yield of instructions which encourage or discourage guessing in a variety of settings is needed.

Continuing the inspection of the characteristic curves for the word "domicile" (Figure 1), we see that the response "servant," although incorrect, is largely indicative of positive ability. This contrasts with the curve for category 3, which shows that response in any remaining alternative of this item is similar to omit in its implication for ability.

The curves for the poorly discriminating item "artifice" (Figure 2) also

* The instructions in the present study were "Do not guess. If you do not know the correct answer, omit the item."

show that one of the alternatives, "imitation," corresponds to positive ability. Actually, this alternative has sufficiently high probability at high levels of ability to be considered partly correct. The remaining alternatives, on the other hand, have so little probability as to be ineffective as distractors.

The category information functions defined by (24) are depicted for these items in the bottom sections of Figures 1 and 2. The sum of category information functions gives the total item information (25) represented by the heavy line. For purposes of comparison, the total information for binary scoring of these items has been calculated and is shown in the same figures as a heavy broken line.

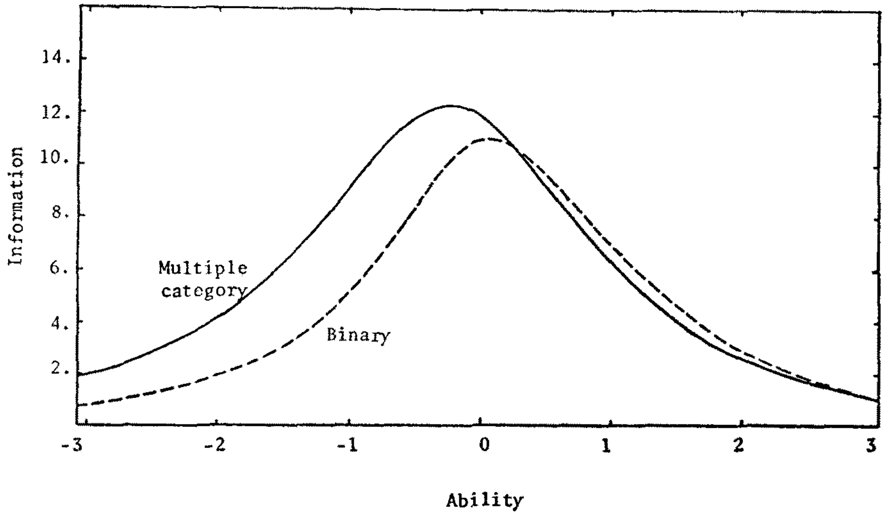
It is important to understand that the binary logistic model used to construct the later is in no sense a sub-set of the multiple categories model. Thus, the information about θ which is made available by the binary model is not necessarily contained in that provided by the multiple category model. This is apparent in Figures 1 and 2, where the information supplied by the former exceeds the latter at certain levels of ability. As we might expect with this type of item, the information gained in scoring the incorrect alternatives comes largely at middle and lower levels of ability where wrong responses occur with some frequency. Most of the gain in information due to multiple scoring appears to be the result of distinguishing category 2 responses, which are plausible or partly correct, from the omits. The practical implication of the information functions in Figures 1 and 2 appears to be that multiple category scoring with this type of item permits the test to be made more discriminating at higher levels of ability without sacrificing discriminating power at lower levels.

This interpretation is confirmed by the test information function (26) under multiple and binary scoring as shown in Figure 3. For the range below median ability, the multiple scoring has increased information (*i.e.*, reduced measurement error) one and a half to two times over binary scoring. Above median ability, the two scoring methods yield approximately equal information. In terms of test length, this means that the multiple categories scoring has, for about half the subject population, produced an increase in precision equivalent to a test half again to double the length of the binary scored test. Since in this particular example, the items are not constructed with multiple category scoring in mind, even better gains may be possible if varying degrees of plausibility are built into the alternatives.

The over-all information gain due to multiple category scoring is, of course, much more modest. The average measurement error variance, computed from (23) with 20 points of quadrature, is 0.139 for multiple scoring and 0.175 for binary scoring. Thus, multiple scoring has increased test reliability from .825 to .861.

Summary

A latent-trait model which retains the identity of individual response



Test information functions for multiple category and binary scoring of 20 multiple choice vocabulary items.

FIGURE 3

categories is proposed for tests consisting of multiple-choice items. The model makes use of a multivariate logistic function to describe the probability of response to each category in terms of two parameters characteristic of the category and an ability parameter characteristic of the subject. Procedures based on the method of maximum likelihood are proposed for estimating item parameters and ability, and for testing goodness-of-fit of the model. Probabilities of the categories for each item, conditional on ability, are constrained to sum to unity.

Two methods of estimating parameters of this model are presented. In the first, the likelihood function is integrated with respect to the distribution of ability in order to obtain unconditional maximum likelihood estimates of the item parameters. In the second method, estimates are obtained conditionally by maximizing the likelihood function with respect to the item parameters and subject abilities simultaneously. Because of the heavy computation required by the unconditional method, only the conditional method is recommended for practical use.

An application of these procedures to data consisting of the responses of 557 subjects to a twenty-item vocabulary test is reported. A comparison of the unconditional and conditional estimates of parameters of four of these items showed the results of the two methods to be closely comparable. An over-all and an item by item test of goodness-of-fit of the conditional solution shows no instance in which the model is rejected.

An information analysis is carried out in order to compare the precision of estimating ability using multiple category scoring of items under the proposed model, and using binary scoring under the conventional logistic latent-trait model. For subjects below median ability, multiple category scoring results in an increase in precision comparable to doubling the test length. For subjects above median ability, the two methods of scoring are essentially equal in precision. Estimates of test reliability computed from the average measurement error shows a modest gain in reliability due to multiple category scoring.

REFERENCES

Aitchison, J. and Silvey, S. D. The generalization of probit analysis to the case of multiple responses. *Biometrika*, 1957, 44, 131–140.

Beck, S. J. *Rorschach's test*, Vol. I. New York: Grune & Stratton, 1950.

Birnbaum, A. Estimation of ability. In F. M. Lord and M. Novick, *Statistical theories of test scores*. Reading, Mass.: Addison-Wesley, 1968.

Bock, R. D. Estimating multinomial response relations. In R. C. Bose, et al. (Eds.), *Essays in probability and statistics*. University of North Carolina Press, Chapel Hill, 1970, p. 453–479.

Bock, R. D. and Jones, L. V. *The measurement and prediction of judgment and choice*, San Francisco: Holden-Day, 1968.

Bock, R. D. and Lieberman, M. L. Fitting a response model for n dichotomously scored items. *Psychometrika*, 1970, 35, 179–197.

Gumbel, E. J. Bivariate logistic distributions. *Journal of the American Statistical Association*, 1961, 56, 335–349.

Kendall, M. G. and Stuart, A. *The advanced theory of statistics*, Vol. II, London: Griffin 1961.

Lord, F. M. An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 1968, 28, 989–1200.

Luce, R. D. *Individual choice behavior*, New York: Wiley, 1959.

Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika, Monograph Supplement*, No. 17, 1969.

Stroud, A. H. and Secrest, Don. *Gaussian quadrature formulas*, Englewood Cliffs, N. J.: Prentice-Hall, 1966.

Manuscript received 7/7/70
Revised manuscript received 3/5/71