Full-Information Factor Analysis for Polytomous Item Responses

Eiji Muraki and James E. Carlson

Educational Testing Service

A full-information item factor analysis model for multidimensional polytomously scored item response data is developed as an extension of previous work by several authors. The model is expressed both in factor-analytic and item response theory parameters. Reckase's multidimensional parameters for the model also are discussed as well as the related geometry. An EM algorithm for estimation of the model parameters is presented and results of the analysis of item response data by a computer program incorporating this algorithm are presented. *Index terms: EM algorithm, full-information item factor analysis, multidimensional item response theory, polytomous response data.*

The use of Likert items on questionnaires is very popular in psychological and sociological measurement. In the typical analysis of data from such a questionnaire, frequencies of endorsements to points on the scale are tabulated and percentages are computed separately for each item and point of the scale. These descriptive statistics are inconvenient when used to compare and interpret persons' responses as measures on the construct being assessed. Comparisons among items become more cumbersome as the number of items or the number of points on the scale increase. The points on the scale are frequently assumed to constitute an equal interval scale and persons' scores are treated as constituting a continuous variable, although they are in fact discrete. Item response theory (IRT) models have provided a solution to problems such as these in other contexts.

A variety of IRT models applicable to scales consisting of dichotomously scored items measuring a single trait have been developed and are now in widespread use. The model developed here is based on two extensions of the basic IRT model. Models that can incorporate polytomously scored items have been proposed and used by several researchers (Andrich, 1978, 1982, 1988; Bock, 1972; Masters, 1982; Muraki, 1990, 1992; Samejima, 1969, 1972). Bock & Aitkin (1981) extended IRT models for dichotomously scored items to the multidimensional case (several traits) and developed an EM algorithm (Dempster, Laird, & Rubin, 1977) to estimate the parameters of the model based on the normal ogive. McKinley & Reckase (1983) proposed a multidimensional model based on the logistic function. In this paper, a multidimensional IRT model for polytomously scored items, based on Samejima's graded response model (GRM) and using the normal ogive, is developed. An EM algorithm that may be used to estimate the parameters of the model also is discussed.

The Polytomous Full-Information Factor Analysis Model

Development of the Model

Bock & Aitkin (1981) and Bock, Gibbons, & Muraki (1988) assumed that the interaction of item i and person j results in a response process variable, y_{ij}, that is a linear combination of M latent traits. Using vector notation in which $\boldsymbol{\theta}$ is an M-dimensional vector of latent traits (common factors), and $\boldsymbol{\alpha}$ is a vector

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 19, No. 1, March 1995, pp. 73–90

0146-6216/95/010073-18\$2.15

of factor loadings:

$$\theta_j' = (\theta_{1j}, \theta_{2j}, \ldots, \theta_{nj}) \tag{1}$$

and

$$\alpha_i' = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iM}) . \tag{2}$$

This combination can be written as

$$y_{ij} = \alpha_{i1}\theta_{1j} + \alpha_{i2}\theta_{2j} + \ldots + \alpha_{iM}\theta_{Mj} + \epsilon_{ij} = \alpha_i'\theta_j + \epsilon_{ij} , \tag{3}$$

where ϵ_{ij} is an unobserved random variable (a unique factor in factor analytic terms) that is assumed to be distributed $N(0, \sigma_i^2)$. In conventional factor analysis, it is assumed that the distribution of θ is $N(\mathbf{0}, \mathbf{I})$ and that y_i is distributed with mean 0 and variance 1. Hence, the unique variance is

$$\sigma_i^2 = 1 - \alpha_i'\alpha_i . \tag{4}$$

Classical factor analysis for continuous variables is based on the assumption that the response process is directly observable. In contrast, the factor analytic model for categorical variables is based on the assumption that the response process variable, y_{ij}, is latent and realized into a vector of polytomous item responses for n items,

$$\mathbf{W}_j = (W_{1j}, W_{2j}, \ldots, W_{nj}), \tag{5}$$

according to the psychological mechanism

$$\begin{gathered} W_{ij} = k \text{ if } \gamma_{i,k-1} \le y_{ij} < \gamma_{ik} \quad (k = 1, 2, \ldots, K_i) \\ \gamma_{i0} = -\infty, \quad \gamma_{iK_i} = \infty , \end{gathered} \tag{6}$$

where γ_{ik} is a threshold parameter associated with category k of a K_i-category Likert-type item, i. The process generates a categorical response of k for person j to item i when y_{ij} equals or exceeds the threshold, $\gamma_{i,k-1}$, but does not reach the next threshold, γ_{ik}. Assuming a normal ogive model, the probability of categorical response k by person j to item i given his/her M-dimensional latent trait is expressed as

$$P(W_{ij} = k|\theta_j) = \frac{1}{(2\pi)^{1/2}\sigma_i} \int_{\gamma_{i,k-1}}^{\gamma_{ik}} \exp\left[-\frac{1}{2}\left(\frac{y_{ij} - \alpha_i'\theta_j}{\sigma_i}\right)^2\right] dy . \tag{7}$$

The item response model in Equation 7 can be rewritten by defining

$$t_{ij} = \frac{y_{ij} - \alpha_i'\theta_j}{\sigma_i} . \tag{8}$$

Then

$$dy_{ij} = \sigma_i dt_{ij} , \tag{9}$$

and at $y_{ij} = \gamma_{ik}$

$$t_{ij} = \frac{\gamma_{ik} - \alpha_i'\theta_j}{\sigma_i} , \tag{10}$$

and at $y_{ij} = \gamma_{i,k-1}$

$$t_{ij} = \frac{\gamma_{i,k-1} - \alpha_i'\theta_j}{\sigma_i} . \tag{11}$$

Following Bock et al. (1988), the slope and item-category parameters can be defined in terms of the factor analysis model parameters as

$$a_{im} = \frac{\alpha_{im}}{\sigma_i} \tag{12}$$

and

$$b_{ik} = -\frac{\gamma_{ik}}{\sigma_i} \quad (k = 1, 2, \ldots, K_i - 1). \tag{13}$$

The functions are defined as

$$Z_{ik}(\boldsymbol{\theta}) = \mathbf{a}_i'\boldsymbol{\theta} + b_{ik} \quad (k = 1, 2, \ldots, K_i - 1). \tag{14}$$

The model in Equation 7 then can be rewritten as

$$P_{ik}(\boldsymbol{\theta}) = \int_{-Z_{i,k-1}(\boldsymbol{\theta})}^{-Z_{ik}(\boldsymbol{\theta})} \phi(t)\mathrm{d}t = \int_{Z_{ik}(\boldsymbol{\theta})}^{Z_{i,k-1}(\boldsymbol{\theta})} \phi(t)\mathrm{d}t = \int_{-\infty}^{Z_{i,k-1}(\boldsymbol{\theta})} \phi(t)\mathrm{d}t - \int_{-\infty}^{Z_{ik}(\boldsymbol{\theta})} \phi(t)\mathrm{d}t, \tag{15}$$

where

$$\phi(t) = \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{t^2}{2}\right). \tag{16}$$

Because the probability of any categorical response is 0 or positive,

$$Z_{i,k-1}(\boldsymbol{\theta}) \geq Z_{ik}(\boldsymbol{\theta}), \tag{17}$$

and thus

$$b_{i,k-1} \geq b_{ik}. \tag{18}$$

The model in Equation 15 can be expressed as

$$P_{ik}(\boldsymbol{\theta}) = P_{i,k-1}^{+}(\boldsymbol{\theta}) - P_{ik}^{+}(\boldsymbol{\theta}) \quad (k = 1, 2, \ldots, K_i), \tag{19}$$

where $P_{i0}^{+}(\boldsymbol{\theta}) = 1$ and $P_{iK_i}^{+}(\boldsymbol{\theta}) = 0$.

In Equation 19 $P_{ik}^{+}(\boldsymbol{\theta})$ represents the probability, given $\boldsymbol{\theta}$, of obtaining a score on item i that is above the kth threshold. The values for the 0th and K_ith categories are defined as 1 and 0, respectively (Samejima, 1972), because γ_{i0} and γ_{iK_i} are at $-\infty$ and $+\infty$, respectively.

In the univariate case, Equation 19 is referred to as the operating characteristic of the GRM (Samejima, 1972). Because the model in Equation 19 is a multidimensional extension of the GRM, it is called a multidimensional GRM (MGRM).

From Equation 19,

$$P_{i1}(\boldsymbol{\theta}) = 1 - P_{i1}^{+}(\boldsymbol{\theta}), \tag{20}$$

$$P_{iK_i}(\boldsymbol{\theta}) = P_{i,K_i-1}^{+}(\boldsymbol{\theta}), \tag{21}$$

and

$$P_{ik}^{+}(\boldsymbol{\theta}) = \sum_{c=k+1}^{K_i} P_{ic}(\boldsymbol{\theta}) = 1 - \sum_{c=1}^{k} P_{ic}(\boldsymbol{\theta}). \tag{22}$$

In the case of a dichotomous response where $K_i = 2$, the probability of an incorrect response, $P_{i1}(\boldsymbol{\theta})$, is given by the complement of the probability of a correct response, $P_{i2}(\boldsymbol{\theta})$. For the unidimensional dichotomous item response model, the item-category parameter is generally called an intercept parameter. The item difficulty parameter is the negative of the intercept parameter divided by the slope parameter.

The direction of the θ axis is arbitrary. It is customary to adopt the convention that a larger θ value indicates a higher trait level. Using this convention, a lower item-category parameter means that, given θ, the item category is more difficult or less likely to be responded to or endorsed. Because the threshold, γ, is a negative function of the item-category parameter, as seen in Equation 13, a more difficult item category has a higher threshold value. Also, because $\gamma_{ik} \geq \gamma_{i,k-1}$, as defined in Equation 6, the kth item category becomes a higher or more difficult category than the $k-1$st category. Consequently, the item-category response surfaces, $P_{ik}(\boldsymbol{\theta})$, are located higher along the θ dimensions with increasing k. Although the operating characteristic function can be defined in the opposite direction (Muraki, 1992), such that $P_{ik}(\boldsymbol{\theta}) = P^{+}_{ik}(\boldsymbol{\theta}) - P^{+}_{i,k-1}(\boldsymbol{\theta})$, the present convention provides a parameter interpretation that is directly analogous to the case of dichotomous item responses. These two systems, however, become congruent by either reversing the signs of all model parameters including the θ parameters, or recoding the categorical responses as their complements ($k' = K_i - k + 1$).

As shown in Equation 19, the probability of responding in a particular middle category for a given $\boldsymbol{\theta}$ is a function of the difference between adjacent item-category parameters, $b_{i,k-1}$ and b_{ik}. Similar to the incorrect response in the dichotomous case, a person with a fixed $\boldsymbol{\theta}$ more likely responds to the lowest category with a lower value of b_{i1} (a higher value of γ_{i1}). Conversely, a person with a fixed $\boldsymbol{\theta}$ more likely responds to the highest category with a higher value of b_{i,k_i-1} (a lower value of γ_{i,k_i-1}).

Once the estimates of slope and item-category parameters are obtained, the corresponding estimates of factor loadings and threshold parameters can be computed from relationships analogous to

$$\alpha_{im} = \frac{a_{im}}{\xi_i} \tag{23}$$

and

$$\gamma_{ik} = -\frac{b_{ik}}{\xi_i}, \tag{24}$$

where

$$\xi_i^2 = \frac{1}{\sigma_i^2} = 1 + \mathbf{a}_i'\mathbf{a}_i. \tag{25}$$

These are derived from Equations 4, 12, and 13.

Interpretation of Parameters in Multidimensional IRT Models

Reckase (1985) and Reckase & McKinley (1991) developed multidimensional parameters that aid in understanding the nature of items generated according to the multidimensional logistic model, and Carlson (1987) further elaborated on their meaning. Here similar ideas are applied to items generated by multidimensional models for polytomously scored items and, in particular, the normal ogive MGRM given in Equation 15.

Following Reckase & McKinley's (1991) definition, the multidimensional discrimination of a polytomously scored item can be defined as η_i, where

$$\eta_i^2 = \mathbf{a}_i'\mathbf{a}_i. \tag{26}$$

The multidimensional item-category parameter can be defined as

$$\beta_{ik} = -\frac{b_{ik}}{\eta_i} . \tag{27}$$

Reckase & McKinley (1991) showed that the direction of steepest slope of the response surface can be expressed in terms of direction cosines of angles with the θ axes as

$$\cos\omega_{im} = \frac{a_{im}}{\eta_i} = \lambda_{im} . \tag{28}$$

The slope in the direction specified by angles ω_{im} is at its maximum when the item response surface (IRS) of $P^+_{ik}(\boldsymbol{\theta})$ crosses the .5 probability hyperplane.

The item-category parameter of the cumulative probability in Equation 15 is the θ value at the point of .5 probability of the kth or higher categorical response where

$$\sum_{m=1}^{M} a_{im}\theta_m = -b_{ik} . \tag{29}$$

If a particular dimension is denoted as m' $(1 \le m' \le M)$, then

$$\theta_m = \frac{\theta_{m'}}{a_{im'}} a_{im} . \tag{30}$$

Equations 29 and 30 can be solved for the points in the θ dimensions by locating the item-category parameter

$$\theta_{m'} = -a_{im'}\frac{b_{ik}}{\eta_i^2} = \frac{a_{im'}}{\eta_i}\beta_{ik} = \beta_{ik}\cos\omega_{im'} . \tag{31}$$

From Equations 28 and 31,

$$a_{im} = \eta_i \cos\omega_{im} \tag{32}$$

and

$$\tau_{ikm} = \beta_{ik}\cos\omega_{im} \tag{33}$$

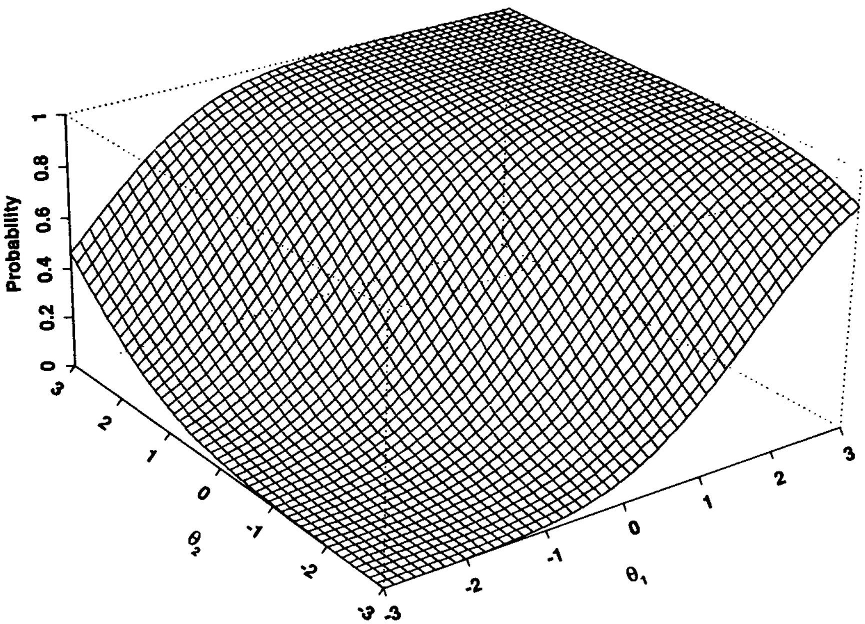
are obtained. Because the direction cosine is $\cos\omega_{im}$ for both η_i and β_{ik}, they must reside along the same axis. This axis is referred to as the direction of measurement of the item. The slope parameter a_{im} is the mth coordinate of the point where the slope of the IRS is the steepest. The value τ_{ikm} indicates the mth coordinate of the point on the line. In other words, it is the projection of β_{ik} onto the mth θ dimension. Thus, the parameter, τ_{ikm}, can be referred to as the mth component of β_{ik}. The IRSs have their steepest slopes in the direction of measurement of the item at the point having coordinates τ_{ikm}.

The concepts discussed above can be best explained in the context of two dimensions. Figure 1 shows a two-dimensional IRS for $P^+_{ik}(\boldsymbol{\theta})$ with $a_1 = .8$, $a_2 = .6$, and $b_{i1} = .5$. Note that the locus of all points on the response surface for which $\theta_1 = 0$ defines a line in the surface that is a unidimensional item response function (IRF) above the θ_2 axis. Similarly, for $\theta_2 = 0$ there is a two-parameter unidimensional IRF over the θ_1 axis. For item i in Figure 1, $\eta_i = 1.0$ and $\beta_{i1} = -.5$. The direction cosines of the line of maximum slope in reference to θ_1 and θ_2 are .8 and .6, respectively. Because these θ axes are orthogonal, $\cos^2\omega_{i1} + \cos^2\omega_{i2} = 1$.

Consider the geometry in Figure 2. θ_1 and θ_2 are two orthogonal axes representing the two dimensions underlying item i. The linear combination in Equation 14 indicates the combination of the two θs that the item is considered to be measuring:

$$\mathbf{a}_i'\boldsymbol{\theta} = a_{i1}\theta_1 + a_{i2}\theta_2 . \tag{34}$$

Figure 1
IRS for the MGRM ($a_1 = .8$, $a_2 = .6$, and $b_1 = .5$)



The equation of the line defining the direction of measurement, as is shown in Figure 2, can be written as

$$\theta_2 = \frac{a_{i2}}{a_{i1}}\theta_1 . \tag{35}$$

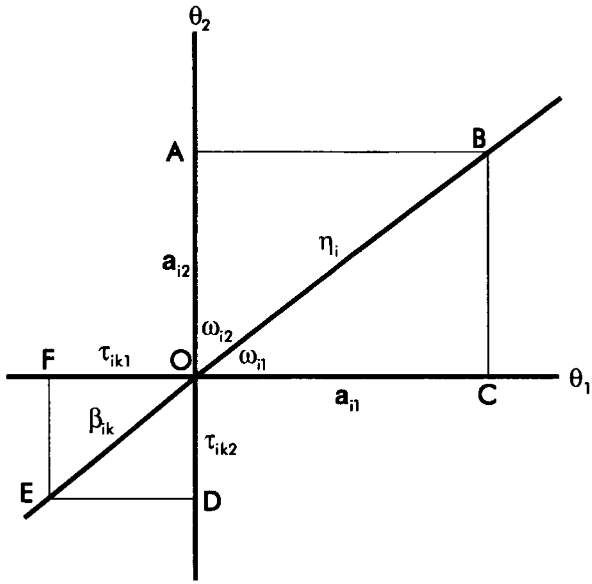
The two a parameters are the lengths of two sides ($a_{i1} = \text{OC}$ and $a_{i2} = \text{OA}$) of a right triangle. The length of the hypotenuse (OB) is equal to Reckase & McKinley's (1991) multidimensional discrimination parameter, η_i. The angle BOC is ω_{i1}, and the angle AOB is its orthogonal complement, ω_{i2}. The two τ parameters are the lengths of two sides ($\tau_{ik1} = \text{OF}$ and $\tau_{ik2} = \text{OD}$) of a right triangle with hypotenuse (OE) of length equal to the absolute value of the multidimensional difficulty parameter (β_{ik}). The multidimensional IRS representing the cumulative probability function is a surface above the $\theta_1 - \theta_2$ plane (Figure 1), and a slice through that surface along the line of measurement of the item is a unidimensional IRF for an item with discrimination parameter equal to η_i and item-category parameter equal to β_{ik} (Carlson, 1987).

By using the multidimensional parameters defined above, the function in Equation 14 can be rewritten as

$$\begin{aligned} Z_{ik}(\theta) &= a_{i1}\theta_1 + a_{i2}\theta_2 + \ldots + a_{iM}\theta_M + b_{ik} = \eta_i(\cos\omega_{i1}\theta_1 + \cos\omega_{i2}\theta_2 + \ldots + \cos\omega_{iM}\theta_M - \beta_{ik}) \\ &= \eta_i(\lambda_{i1}\theta_1 + \lambda_{i2}\theta_2 + \ldots + \lambda_{iM}\theta_M - \beta_{ik}) = \eta_i(\theta_i^* - \beta_{ik}) . \end{aligned} \tag{36}$$

The unidimensional latent trait θ_i^* is a composite of M-dimensional latent traits θ_m ($m = 1, 2, ..., M$). Reckase & McKinley (1991) pointed out that an item, although scaled in a multidimensional context, can be considered to be measuring along a single dimension. That dimension is a linear combination of the uncorrelated θ dimensions. The test containing the item may, however, be multidimensional if it consists of items that measure along different directions in the θ space. Because it is assumed that the

Figure 2
Multidimensional Discrimination for a Two-Dimensional Space



distribution of $\boldsymbol{\theta}$ is N$(\mathbf{0}, \mathbf{I})$ and the sum of squared direction cosines in the orthogonal space is 1, the variance of θ_i^* is also 1.

The function in Equation 36 also can be expressed as a linear combination of factor loadings and thresholds:

$$Z_{ik}(\boldsymbol{\theta}) = \xi_i(\alpha_{i1}\theta_1 + \alpha_{i2}\theta_2 + \ldots + \alpha_{iM}\theta_M - \gamma_{ik}) = \frac{\theta_i^{**} - \gamma_{ik}}{\sigma_i}. \tag{37}$$

Unlike the composite latent trait θ_i^* in Equation 36, the variance of a linear combination of the weighted individual θ_m, θ_i^{**}, in Equation 37 is not 1. The variance of this composite of latent traits is called the communality, h_i^2. Because the θs have variance 1 and are uncorrelated, this composite variance is

$$h_i^2 = \boldsymbol{\alpha}_i'\boldsymbol{\alpha}_i = 1 - \sigma_i^2 . \tag{38}$$

From Equation 38 and using Equations 23 and 26, the communality also may be expressed as

$$h_i^2 = 1 - \frac{1}{\xi_i^2} = \frac{\eta_i^2}{\xi_i^2} . \tag{39}$$

Using Equations 24 and 39,

$$Z_{ik}(\boldsymbol{\theta}) \sim \mathrm{N}(-\eta_i\beta_{ik}, \eta_i^2), \tag{40}$$

which also may be expressed as

$$Z_{ik}(\boldsymbol{\theta}) \sim \mathrm{N}\left(-\frac{\gamma_{ik}}{\sigma_i}, \frac{h_i^2}{\sigma_i^2}\right). \tag{41}$$

Using the above relationships, the multidimensional discrimination and direction cosines can be expressed in terms of factor-analytic parameters as

$$\eta_i = \frac{h_i}{\sigma_i} \tag{42}$$

and

$$\lambda_{im} = \frac{a_{im}}{\eta_i} = \frac{\alpha_{im}}{h_i} . \tag{43}$$

Comparing Equations 12 and 42 demonstrates that the multidimensional discrimination can be interpreted as a transformation of the square root of the communality by the same factor, $1/\sigma_i$, as that in the transformation of factor loadings to the slope parameters. Furthermore, the direction cosine also can be computed as the ratio of the factor loadings to the square root of the communality.

In summary, the multidimensional parameters defined for the Reckase-McKinley multidimensional two-parameter logistic model (McKinley & Reckase, 1983; Reckase, 1985; Reckase & McKinley, 1991) can be adopted for the MGRM because the cumulative probability, $P^+_{ik}(\boldsymbol{\theta})$, expresses the dichotomous categorical response of the kth or below versus above the kth category, and the logistic function is only an approximate form of the normal ogive function. The difference is that the multidimensional polytomous item response model yields a set of $K_i - 1$ IRSs of $P^+_{ik}(\boldsymbol{\theta})$ rather than one such surface (as it does for the dichotomous item response model). These IRSs are parallel along the line defined by a set of direction cosines, λ_{im}, $m = 1, 2, ..., M$. The probability of the specific middle category k, $P_{ik}(\boldsymbol{\theta})$, is defined by subtracting $P^+_{ik}(\boldsymbol{\theta})$ from $P^+_{i,k-1}(\boldsymbol{\theta})$ as shown in Equation 19. Because these IRSs of cumulative probabilities are parallel, the multidimensional parameters are still meaningful even for specific middle categories.

Bock et al. (1988) established the relationships between the parameters of the factor analysis model and the parameters of the item response model. In this paper, relationships between the factor loadings and multidimensional parameters similar to those in McKinley & Reckase's (1983) model (Reckase, 1985; Reckase & McKinley, 1991) have been established. The multidimensional parameters can provide useful interpretations of parameters of other multidimensional item response models (Luecht, 1993; Luecht & Miller, 1992; Miller & Hirsch, 1992; Reckase, 1985; Reckase & McKinley, 1991). These parameters can be computed directly from the factor loadings.

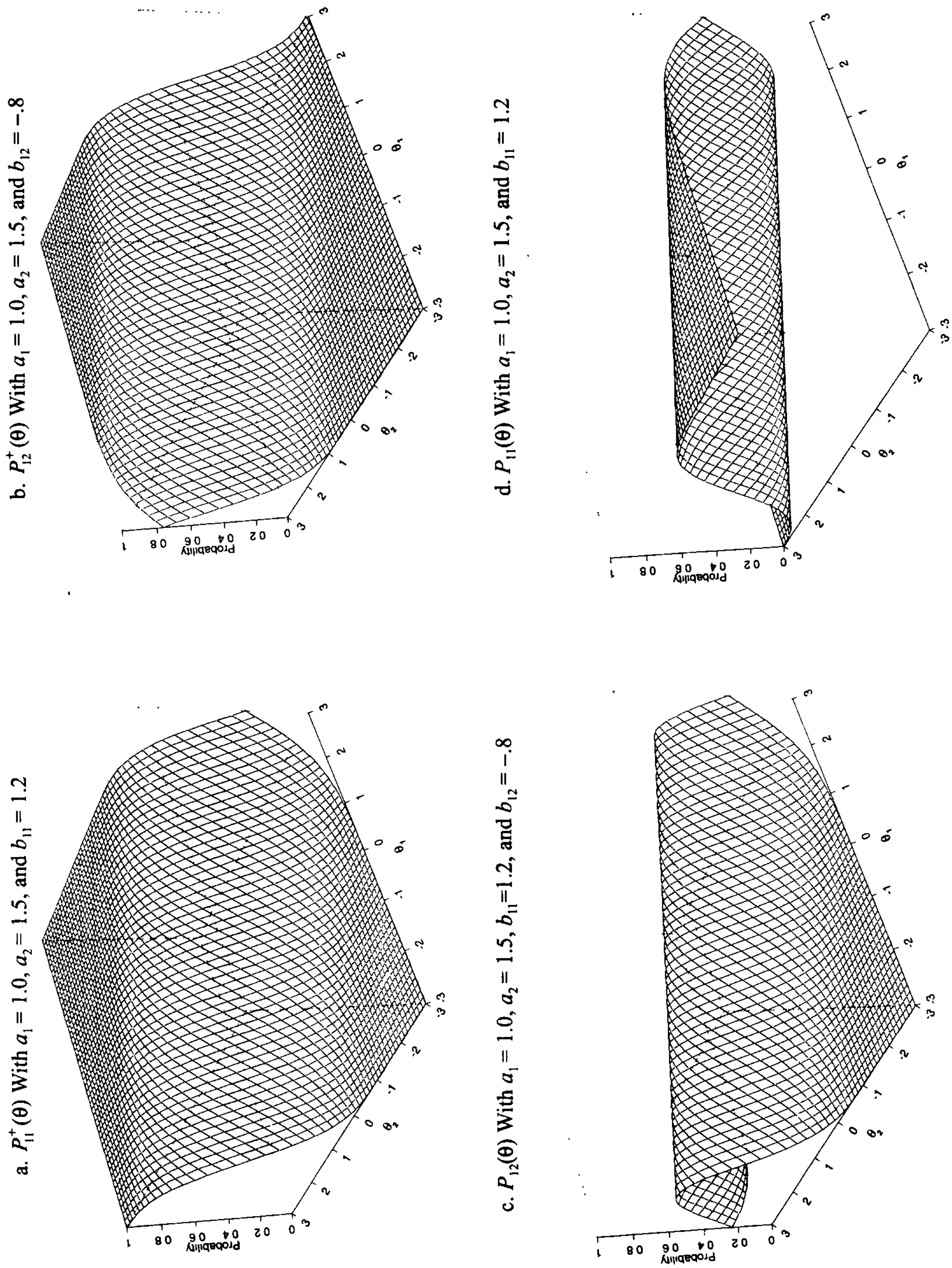
Figure 3 shows the IRSs of a two-dimensional polytomously scored three-category item with $a_{11} = 1.0$, $a_{12} = 1.5$, $b_{11} = 1.2$, and $b_{12} = -.8$. Figure 3a shows the IRS of $P^+_{11}(\boldsymbol{\theta})$, and Figure 3b shows the IRS of the second cumulative probability, $P^+_{12}(\boldsymbol{\theta})$. This is the same probability as $P_{13}(\boldsymbol{\theta})$. These two IRSs are parallel to each other. The model probability of the middle categorical response, $P_{12}(\boldsymbol{\theta}) = P^+_{11}(\boldsymbol{\theta}) - P^+_{12}(\boldsymbol{\theta})$, is shown in Figure 3c, and $P_{11}(\boldsymbol{\theta}) = 1 - P^+_{11}(\boldsymbol{\theta})$ is plotted in Figure 3d.

If the distance of the item-category parameters is shortened, the probability of the middle categorical response is uniformly decreased. Consequently, the IRS becomes flatter. Figure 4a shows the IRS of the middle categorical response of a second item with the same parameter values as Item 1, except that $b_{22} = .8$. Because the distance of the item-category parameters becomes shorter given the same slope parameters, the IRS of $P_{22}(\boldsymbol{\theta})$ is pushed downward. The observed response frequencies of the middle category are expected to be smaller for Item 2 than for Item 1. The shape of the IRS does not change if the original IRS is rotated around a point in the θ plane. This is demonstrated in Figure 4b, which shows the IRS of the middle categorical response, $P_{32}(\boldsymbol{\theta})$, with slope parameters changed to $a_{31} = 1.5$ and $a_{32} = 1.0$.

Parameter Estimation

Let U_{ijk} represent an element in the matrix of the observed response pattern j. $U_{ijk} = 1$ if item i is rated by the jth respondent in the kth category of a Likert scale, otherwise $U_{ijk} = 0$. By the principle of local independence (Birnbaum, 1968), the conditional probability of a response pattern j, given $\boldsymbol{\theta}$ and K_i re-

Figure 3
IRSs for the MGRM



a. $P^+_{11}(\theta)$ With $a_1 = 1.0$, $a_2 = 1.5$, and $b_{11} = 1.2$

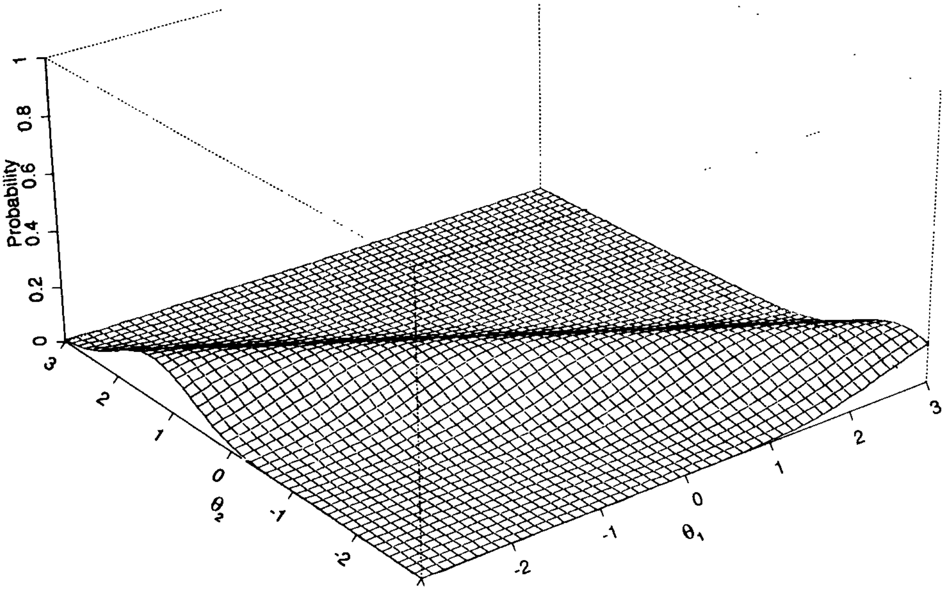
b. $P^+_{12}(\theta)$ With $a_1 = 1.0$, $a_2 = 1.5$, and $b_{12} = -.8$

c. $P_{12}(\theta)$ With $a_1 = 1.0$, $a_2 = 1.5$, $b_{11}=1.2$, and $b_{12} = -.8$

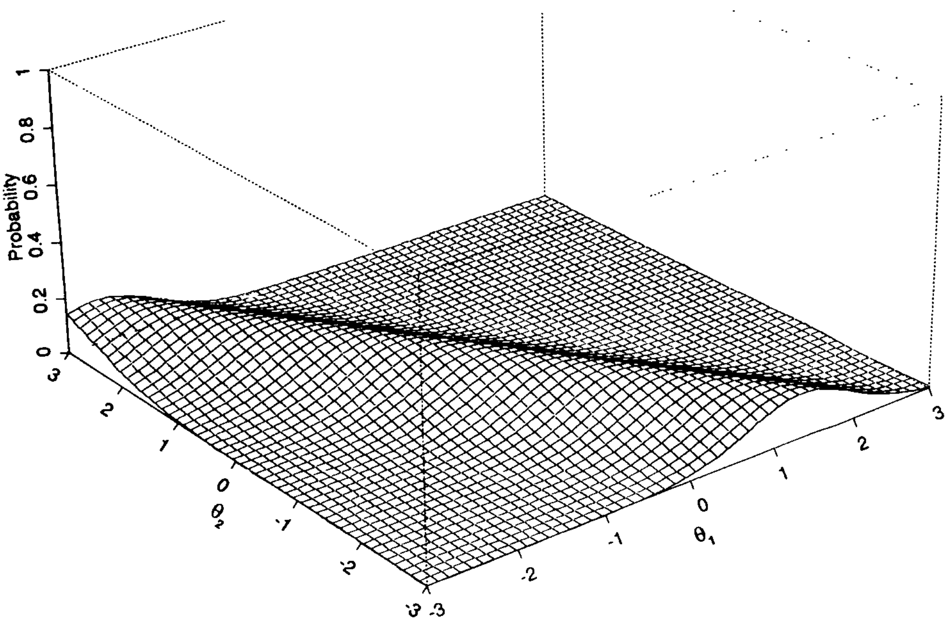
d. $P_{11}(\theta)$ With $a_1 = 1.0$, $a_2 = 1.5$, and $b_{11} = 1.2$

Figure 4
IRSs for the MGRM

a. $P_{22}(\theta)$ With $a_1 = 1.0$, $a_2=1.5$, $b_{21} = 1.2$, and $b_{22} = .8$



b. $P_{32}(\theta)$ With $a_1 = 1.0$, $a_2=1.5$, $b_{31} = 1.2$, and $b_{32} = .8$



sponse categories, for item i and n items, as denoted by response matrix $\mathbf{U}$, is the joint probability:

$$P_j(\mathbf{U}|\boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{k=1}^{K_i}\left[P_{ik}(\boldsymbol{\theta})\right]^{U_{ijk}} . \tag{44}$$

For examinees randomly sampled from a population with a multivariate normal distribution of the latent trait variable, $\psi(\boldsymbol{\theta})$, the marginal probability of the observed response pattern j is

$$P_j(\mathbf{U}) = \int_{\boldsymbol{\theta}} P_j(\mathbf{U}|\boldsymbol{\theta})\psi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} . \tag{45}$$

If an examinee responds to each of n items that has K_i categories, his/her response pattern, j, then can be assigned to one of

$$J = \prod_{i=1}^{n} K_i \tag{46}$$

mutually exclusive patterns. Let r_j represent the number of examinees observed in pattern j, and let N be the total number of examinees sampled from the population. Then r_j is multinomially distributed with parameters N and $P_j(\mathbf{U})$. The marginal likelihood of the parameters a_{im} and b_{ik} is therefore

$$L = \frac{N!}{\prod_{j=1}^{J} r_j!}\prod_{j=1}^{J}\left[P_j(\mathbf{U})\right]^{r_j} . \tag{47}$$

Taking the natural logarithm of Equation 47 yields

$$\ln L = \ln N! - \sum_{j=1}^{J}\ln r_j! + \sum_{j=1}^{J} r_j \ln P_j(\mathbf{U}) . \tag{48}$$

Bock & Aitkin (1981) applied the EM algorithm (Dempster et al., 1977) to estimate the parameters for each item individually, and then repeated the iterative process over n items until the estimates of all the items became stable to the required number of decimal places. This is in contrast to the Fisher scoring procedure of Bock & Lieberman (1970). The qth cycle of the iterative process can be expressed as

$$\mathbf{v}^{q} = \mathbf{v}^{q-1} + \mathbf{V}^{-1}\mathbf{t} \tag{49}$$

for each item, i. The vector of estimates $\mathbf{v}$ represents the model parameters. When the number of response categories is K_i, only $K_i - 1$ item-category parameters can be estimated. Thus, the orders of the parameter vector $\mathbf{v}$ and gradient vector $\mathbf{t}$ are both $M + K_i - 1$, and the order of the information matrix $\mathbf{V}$ is $(M + K_i - 1) \times (M + K_i - 1)$. The information matrix is the negative expectation of the matrix of second derivatives. Because the covariance between any categories that differ by more than two points is 0, the partitioned information matrix for the category parameter estimation becomes a tridiagonal symmetric matrix (Bock & Jones, 1968; Muraki, 1990).

The likelihood equations for $\hat{a}_{im}$ and $\hat{b}_{ik}$ can be derived from equations resulting from setting the first derivatives of Equation 48 with respect to each parameter to 0. With respect to a_{im}, the likelihood in Equation 48 can be differentiated as

$$\frac{\partial \ln L}{\partial a_{im}} = \sum_{j=1}^{J}\frac{r_j}{P_j(\mathbf{U})}\int_{\boldsymbol{\theta}} P_j(\mathbf{U}|\boldsymbol{\theta})\sum_{k=1}^{K_i}\frac{\partial\left[P_{ik}(\boldsymbol{\theta})\right]^{U_{ijk}}}{\partial a_{im}}\psi(\boldsymbol{\theta})\frac{\mathrm{d}\boldsymbol{\theta}}{\left[P_{ik}(\boldsymbol{\theta})\right]^{U_{ijk}}} . \tag{50}$$

Now let the observed score patterns be indexed by $l = 1, 2, \ldots, S$, where $S \le \min(N, J)$. If the number of examinees with response pattern l is denoted by r_l, then

$$\sum_{l=1}^{s} r_l = N . \tag{51}$$

The derivative in Equation 50 can be approximated by using Gauss-Hermite quadrature (Stroud & Secrest, 1966), as

$$\frac{\partial \ln L}{\partial a_{im}} \approx \sum_{l=1}^{s} \sum_{f_M=1}^{F_M} \sum_{f_2=1}^{F_2} \sum_{f1=1}^{F_1} \frac{r_l L_l(\mathbf{X}) A(X_{f_1}) A(X_{f_2}) \ldots A(X_{f_M})}{\tilde{P}_l} \sum_{k=1}^{K_i} \frac{\partial [P_{ik}(\mathbf{X})]^{U_{ilk}}}{\partial a_{im}} \frac{1}{[P_{ik}(\mathbf{X})]^{U_{ilk}}} , \tag{52}$$

where

$$\tilde{P}_l = \sum_{f_M=1}^{F_M} \cdots \sum_{f_2=1}^{F_2} \sum_{f_1=1}^{F_1} \prod_{i=1}^{n} \prod_{k=1}^{K_i} [P_{ik}(\mathbf{X})]^{U_{ilk}} A(X_{f_1}) A(X_{f_2}) \ldots A(X_{f_M}) \tag{53}$$

and

$$L_l(\mathbf{X}) = \prod_{i=1}^{n} \prod_{k=1}^{K_i} [P_{ik}(\mathbf{X})]^{U_{ilk}} , \tag{54}$$

where $\mathbf{X} = X_{f_1}, X_{f_2}, \ldots, X_{f_M}$.

In Equation 52, $A(X_f)$ is the weight of the Gauss-Hermite quadrature and X_f is the quadrature point (Stroud & Secrest, 1966). $A(X_f)$ is approximately the standard normal probability density at the point X_f for each dimension. Because U_{ilk} can take only two possible values, 1 or 0, the element of the gradient vector $\mathbf{t}_a$ can be written as

$$t_{a_{im}} = \sum_{f_M=1}^{F_M} \cdots \sum_{f_2=1}^{F_2} \sum_{f_1=1}^{F_1} \sum_{k=1}^{K_i} \frac{\bar{r}_{ikf_1f_2\ldots f_M}}{P_{ik}(\mathbf{X})} \frac{\partial P_{ik}(\mathbf{X})}{\partial a_{im}} , \tag{55}$$

where

$$\bar{r}_{ikf_1f_2\ldots f_M} = \sum_{l=1}^{s} \frac{r_l L_l(\mathbf{X}) A(X_{f_1}) A(X_{f_2}) \ldots A(X_{f_M}) \, U_{ilk}}{\tilde{P}_l} , \tag{56}$$

and $\bar{r}_{ikf_1f_2\ldots f_M}$ is called the provisional expected frequency of the kth categorical response of item i at the fth quadrature point on the M-dimensional axes.

The item-category parameter, b_{ik}, is contained in both $P_{ik}(\boldsymbol{\theta})$ and $P_{i,k+1}(\boldsymbol{\theta})$ as shown in Equations 14 and 15. The first derivative of Equation 44 with respect to b_{ik} is given by

$$\frac{P_j(\mathbf{U}|\boldsymbol{\theta})}{\partial b_{ik}} = P_j(\mathbf{U}|\boldsymbol{\theta}) \left[\frac{U_{ij,k+1}}{P_{i,k+1}(\boldsymbol{\theta})} - \frac{U_{ijk}}{P_{ik}(\boldsymbol{\theta})} \right] \frac{\partial P_{ik}^{+}(\boldsymbol{\theta})}{\partial b_{ik}} . \tag{57}$$

Therefore, the element of the gradient vector, $\mathbf{t}_b$, is numerically computed as

$$t_{b_{ik}} = \sum_{f_M=1}^{F_M} \cdots \sum_{f_2=1}^{F_2} \sum_{f_1=1}^{F_1} \left[\frac{\bar{r}_{i,k+1,f_1f_2\ldots f_M}}{P_{i,k+1}(\mathbf{X})} - \frac{\bar{r}_{ikf_1f_2\ldots f_M}}{P_{ik}(\mathbf{X})} \right] \frac{\partial P_{ik}^{+}(\mathbf{X})}{\partial b_{ik}} . \tag{58}$$

The elements of the information matrix are given by

$$V_{a_{im}a_{im'}} = \sum_{f_M=1}^{F_M} \cdots \sum_{f_2=1}^{F_2} \sum_{f_1=1}^{F_1} \bar{N}_{if_1f_2\ldots f_M} \sum_{k=1}^{K_i} \frac{1}{P_{ik}(\mathbf{X})} \frac{\partial P_{ik}(\mathbf{X})}{\partial a_{im}} \frac{\partial P_{ik}(\mathbf{X})}{\partial a_{im'}} , \tag{59}$$

$$V_{b_{ik}b_{ik}} = \sum_{f_M=1}^{F_M} \cdots \sum_{f_2=1}^{F_2} \sum_{f_1=1}^{F_1} \overline{N}_{if_1f_2\ldots f_M} \left[\frac{1}{P_{ik}(\mathbf{X})} + \frac{1}{P_{i,k+1}(\mathbf{X})} \right] \left[\frac{\partial P_{ik}^{+}(\mathbf{X})}{\partial b_{ik}} \right]^2, \tag{60}$$

and

$$V_{b_{i,k-1}b_{ik}} = -\sum_{f_M=1}^{F_M} \cdots \sum_{f_2=1}^{F_2} \sum_{f_1=1}^{F_1} \overline{N}_{if_1f_2\ldots f_M} \frac{1}{P_{ik}(\mathbf{X})} \frac{\partial P_{i,k-1}^{+}(\mathbf{X})}{\partial b_{i,k-1}} \frac{\partial P_{ik}^{+}(\mathbf{X})}{\partial b_{ik}}, \tag{61}$$

and, when $|k - k'| \geq 2$,

$$V_{b_{ik}b_{ik'}} = 0. \tag{62}$$

Finally,

$$V_{a_{im}b_{ik}} = \sum_{f_M=1}^{F_M} \cdots \sum_{f_2=1}^{F_2} \sum_{f_1=1}^{F_1} \overline{N}_{if_1f_2\ldots f_M} \left[\frac{1}{P_{i,k+1}(\mathbf{X})} \frac{\partial P_{i,k+1}(\mathbf{X})}{\partial a_{im}} - \frac{1}{P_{ik}(\mathbf{X})} \frac{\partial P_{ik}(\mathbf{X})}{\partial a_{im}} \right] \frac{\partial P_{ik}^{+}(\mathbf{X})}{\partial b_{ik}}, \tag{63}$$

where

$$\overline{N}_{if_1f_2\ldots f_M} = \sum_{l=1}^{S} \frac{r_l L_l(\mathbf{X}) A(X_{f_1}) A(X_{f_2}) \ldots A(X_{f_M})}{\tilde{P}_l}. \tag{64}$$

The algorithm presented above was implemented in the POLYFACT computer program (Muraki, 1993). The program computes the factor loadings using the principal factor method applied to the product-moment correlation matrix. Item responses are treated as a continuous variable. Because the factors in the principal factor solution are orthogonal, their loadings are suitable for the initial values of the full-information solution after conversion to slopes. Slope estimates based on the full-information method are then converted again into factor loadings. The resulting full-information factor loadings may be rotated orthogonally to the varimax criterion (Kaiser, 1958) and, with the varimax solution as the target, rotated obliquely by the promax method (Hendrickson & White, 1964).

Test of Goodness of Fit

Although the sparseness of the response pattern frequencies in these applications usually precludes a test of the fit of the factor model versus the general multinomial alternative, a likelihood-ratio test of the significance of an added factor, based on successive maxima of the marginal likelihoods, is available (Haberman, 1977). In addition, the goodness of fit of the MGRM can be tested approximately item by item. If the test is sufficiently long, the method used in BILOG 3 (Mislevy & Bock, 1990) can be used with its multidimensional extension. In this method, the respondents in a sample of size N are assigned to intervals of the θ continuum. The expected a posteriori (EAP) estimator is used to estimate each respondent's θ. The EAP estimate is the mean of the posterior distribution of θ, given the observed response pattern l (Bock & Mislevy, 1982). The EAP estimate of the response pattern l on the mth dimension, $\tilde{\theta}_{lm}$, is approximated using the quadrature points, X_f, and weights, $A(X_f)$; that is,

$$\tilde{\theta}_{lm} = \frac{\sum_{f_M=1}^{F_M} \cdots \sum_{f_2=1}^{F_2} \sum_{f_1=1}^{F_1} X_{f_m} L_l(\mathbf{X}) A(X_{f_1}) A(X_{f_2}) \ldots A(X_{f_M})}{\sum_{f_M=1}^{F_M} \cdots \sum_{f_2=1}^{F_2} \sum_{f_1=1}^{F_1} L_l(\mathbf{X}) A(X_{f_1}) A(X_{f_2}) \ldots A(X_{f_M})}, \tag{65}$$

where $L_l(\mathbf{X})$ is the probability of observing a particular response pattern given $\mathbf{X}$. After all respondents' M-dimensional vectors of EAP estimates are assigned to any one of the predetermined H^M cells on the multidimensional θ continuum, the observed frequency of the kth categorical responses to item i in each of H^M cells—$r_{h_1h_2h_3\ldots h_Mik}$—and the number of respondents assigned to item i in the h_1, h_2, ..., h_Mth interval—$N_{h_1h_2h_3\ldots h_Mik}$—are computed. The estimated θs are rescaled so that the variance of the sample distribution equals that of the latent distribution on which the marginal maximum likelihood estimation of the item parameters is based [N($\mathbf{0}$, $\mathbf{I}$) is usually the default]. Thus, the $H^M \times K_i$ contingency table for each item i is obtained. For each cell, the cell mean, $\bar{\boldsymbol{\theta}}_{h_1h_2h_3\ldots h_Mik}$, and the value of the fitted response function, $P_{ik}(\bar{\boldsymbol{\theta}})$, are computed. Finally, a likelihood-ratio χ^2 statistic for each item is computed by

$$G_i^2 = 2\sum_{h_M=1}^{H_M} \cdots \sum_{h_2=1}^{H_2} \sum_{h_1=1}^{H_1} \sum_{k=1}^{K_i} r_{h_1h_2\ldots h_Mik} \ln \frac{r_{h_1h_2\ldots h_Mik}}{N_{h_1h_2\ldots h_Mi} P_{ik}(\bar{\boldsymbol{\theta}})}, \tag{66}$$

where h_M is the number of intervals in the mth dimension remaining after neighboring intervals are merged, if necessary, to avoid expected values—$N_{h_1h_2h_3\ldots h_Mik} P_{ik}(\bar{\boldsymbol{\theta}})$—less than 5. The number of degrees of freedom (df) is equal to the number of cells (after combining) in the M-dimensional trait space multiplied by $K_i - 1$. The likelihood-ratio χ^2 statistic of the test is simply the summation of the separate χ^2 statistics. The number of df is also the sum of the df for each item. The fit statistics applied to IRT models are generally quite large because of the fragmentation of residuals and the complexity of the models. They are, however, useful for comparing the fit of different models to the same response data when the models are nested in their parameters (Bock et al., 1988).

Example

Data

For illustrative purposes, a subset of the data collected in the 1992 National Assessment of Educational Progress (NAEP) main writing assessment at the fourth-grade level was used. A large number of writing prompts (items) were used in the assessment and each examinee was administered items in a balanced incomplete block (BIB) design. The subset of data analyzed here was comprised of nine writing items. Each examinee responded to two items. Three types of writing were assessed in the 1992 NAEP–informative, narrative, and persuasive. Each item was created to assess students on one of the three types. The writing assessments were scored on a six-point scale using an enhanced primary trait approach. Further details about the NAEP writing assessment, scoring, and scale development may be found in the NAEP 1992 Writing Report Card (Applebee, Langer, Mullis, Latham, & Gentile, 1994) and the NAEP 1992 National Technical Report (Johnson, Carlson, & Kline, 1994).

The dataset contained three informative items (I1, I2, and I3), three persuasive items (P1, P2, and P3), and three narrative items (N1, N2, and N3). Of the 9,552 examinees selected to respond to these items, 416 did not do so. Hence, the responses of 9,136 examinees were analyzed. Because the scoring resulted in some very small frequencies of certain categories on some items, some of the categories were combined so that the resulting categories had frequencies of at least 10. Table 1 displays the frequencies of occurrence of each categorical score for each of the nine items, as well as the number of analyzed score categories for each item.

Analyses were performed using one, two, and three dimensions. For all solutions, 64 quadrature points were used. Hence, there were four points per dimension in the three-dimensional solutions and eight in the two-dimensional solutions. A normal ogive model was assumed using a normal prior for the θ distributions (an orthogonal fixed prior was used in all iterations). No priors were used for the item parameters.

A smoothed correlation matrix as described in Bock et al. (1988) was used to compute the initial factor loadings, which then were used as initial values in the iterative process. Because of the incomplete nature

Table 1
Observed and Expected Response Frequencies for the Two-Factor Solution for Categories 1–6

Item	Observed Frequencies 1	2	3	4	5	6	Expected Frequencies 1	2	3	4	5	6
I1	83	208	897	553	104	22	110.894	220.149	842.114	532.270	123.886	35.812
I2	98	237	892	446	160	28	110.740	245.147	874.095	436.350	163.289	31.379
I3	164	213	922	449	94		193.344	225.562	886.803	429.652	106.640	
P1	240	820	547	126	14		236.116	789.909	554.839	146.032	18.108	
P2	392	552	601	213	33		416.457	505.172	569.581	242.019	56.690	
P3	236	815	595	158			244.948	766.246	586.497	205.843		
N1	116	694	764	370	55		197.147	658.054	629.388	394.947	117.206	
N2	174	461	815	456	61	10	294.337	375.024	618.430	501.757	125.465	59.696
N3	195	691	712	307	35		293.317	590.650	609.611	348.919	96.297	

of the design, some product-moment correlation coefficients were missing. In the full-information procedure, missing data is not a problem for estimating slope parameters and factor loadings. Note that the correlation matrix is used only for computing initial values–the full information estimation procedure uses all the frequency data.

Results

The one-dimensional solution converged to a criterion of .0005 (value of the fitting function) in 128 iterations. The two- and three-factor solutions took 151 and 270 iterations, respectively. The χ^2 fit statistic for one dimension and two dimensions were 14,484 with 535 *df* and 10,078 with 347 *df*, respectively. The difference between these solutions (4,406 with 188 *df*) was highly significant ($p < .001$). The χ^2 fit statistic for three dimensions (12,674 with 340 *df*) did not represent a decrease from that for two dimensions. Therefore, the two-dimensional solution was considered most appropriate for these data.

Table 1 also provides the expected frequencies of occurrence given the two-dimensional factor-analytic solution. Table 2 shows the slope and item-category parameter estimates for the two-factor solution. Table 3 contains the multidimensional parameter estimates.

Table 2
Slope Parameter Estimates and Item-Category Parameter Estimates for the Two-Factor Solution

Item	Slope Parameters a_1	a_2	Item-Category Parameters b_1	b_2	b_3	b_4	b_5
I1	.873	.226	2.255	1.334	–.503	–2.051	–3.082
I2	.516	.380	1.917	1.074	–.497	–1.521	–2.589
I3	.613	.286	1.624	.994	–.656	–1.978	
P1	.724	.031	1.313	–.369	–1.755	–2.988	
P2	.888	.176	1.036	–.096	–1.475	–2.835	
P3	.787	.158	1.409	–.300	–1.767		
N1	.347	1.002	2.281	.353	–1.150	–2.759	
N2	.505	1.449	2.391	.776	–1.224	–3.379	–4.767
N3	.559	1.210	2.103	.183	–1.516	–3.429	

Table 4 shows the varimax-rotated factor loadings. Figure 5 shows the projection of the nine vectors representing the items onto the two-dimensional common factor space with axes defined by the varimax rotation. Figure 5 and the varimax factor loadings in Table 4 show that the two-factor solution did an excellent job of separating the items representing the three types of writing. Factors 1 and 2 separate the narrative items (N1–N3) from the informative (I1–I3) and persuasive (P1–P3) items very well. There is, however, some overlap between the informative and persuasive items.

Table 3
Multidimensional Discrimination (η), Direction Cosines (λ_1 and λ_2), and Multidimensional Difficulties (β_1, β_2, β_3, β_4, and β_5) for the Two-Factor Solution

Item	η	λ_1	λ_2	β_1	β_2	β_3	β_4	β_5
I1	.902	.968	.250	−2.500	−1.479	.557	2.274	3.418
I2	.640	.805	.593	−2.994	−1.677	.776	2.376	4.044
I3	.677	.906	.423	−2.401	−1.468	.970	2.923	
P1	.725	.999	.042	−1.811	.509	2.421	4.122	
P2	.906	.981	.194	−1.144	.106	1.629	3.130	
P3	.803	.980	.197	−1.755	.373	2.201		
N1	1.060	.327	.945	−2.152	−.333	1.085	2.602	
N2	1.534	.329	.944	−1.558	−.506	.798	2.202	3.107
N3	1.333	.420	.908	−1.578	−.137	1.137	2.572	

An oblique solution also was obtained using the promax criterion. For the two-dimensional solution the interfactor correlation was .584. This correlation represents an angle of 54.3° between the axes representing the two oblique factors. The promax factor loadings in Table 4 show that the axis for the first factor goes through the cluster containing both the information and the persuasive items, and the second factor goes through the cluster of narrative items.

Table 4
Varimax and Promax Factor Loadings for the Two-Factor Solution

	Varimax		Promax	
Item	α_1	α_2	α_1	α_2
I1	.630	.226	.662	.127
I2	.403	.358	.347	.257
I3	.484	.283	.469	.137
P1	.582	.078	.658	−.140
P2	.644	.190	.691	−.035
P3	.600	.179	.643	−.030
N1	.174	.706	−.047	.754
N2	.203	.813	−.051	.867
N3	.268	.754	.047	.772

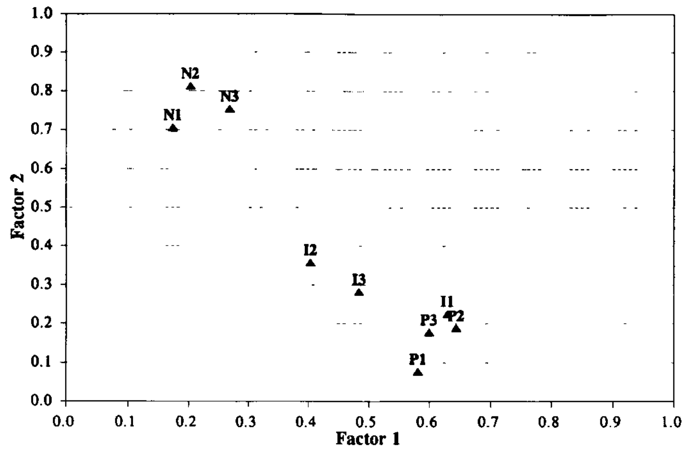
Discussion

The full-information item factor procedure described here has several advantages over alternative linear or nonlinear factor-analytic procedures. One advantage is that it uses all of the information contained in the response category patterns, rather than first reducing the data to correlation coefficients. Thus, the method is not strictly a linear factor analytic procedure. McDonald (1965) and McDonald & Ahlawat (1974) discussed reasons for using nonlinear procedures in item factor analysis. Another advantage is that the method can handle the case of matrix sampling of the examinee-item matrix, as is done in NAEP with a BIB design. That is, not all examinees need be administered all items in order to use the technique.

Several extensions of the model discussed here are possible, and the authors are currently developing software for some of these extensions. One extension is to the multigroup situation, which will be useful in the study of differential item functioning in different subgroups of a population. Another extension is to the multidimensional partial credit model. Some work is currently being planned to develop improved fit statistics methodology for the models.

It must be emphasized that full-information multidimensional item factor analytic solutions require considerable computing resources and are not practical for more than four dimensions at the present time.

Figure 5
Varimax Two-Factor Solution



For example, the number of provisional expected frequencies ($\bar{r}_{ikf_1f_2f_3\cdots f_M}$) computed at each E step is

$$\left(\sum_{i=1}^{n} k_i\right) F^M . \tag{67}$$

If these quantities are evaluated for five items with 4 categories each using 10 quadrature points in four dimensions, then 200,000 points must be computed. This number doubles because $\bar{N}_{if_1f_2f_3\cdots f_M}$ also must be computed. The number increases exponentially as the number of dimensions increases. Although four-dimensional solutions are theoretically possible they are computationally impractical, even when using a SUN workstation to analyze the data.

Another problem associated with the MGRM is that the fit statistics are quite sensitive to the choice of intervals, H_m. If H is too small, the EAP scores are concentrated in certain intervals and the χ^2 is underestimated. On the other hand, if H is too large, the frequencies of the EAP scores in each interval become scarce and the χ^2 becomes unstable. This problem also exists for unidimensional IRT models, but the problem is exacerbated as the number of dimensions increases. Further studies are necessary on this topic.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561–573.

Andrich, D. (1982). An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika, 47,* 105–113.

Andrich, D. (1988). A general form of Rasch's extended logistic model for partial credit scoring. *Applied Measurement in Education, 1,* 363–378.

Applebee, A. N., Langer, J. A., Mullis, I. V. S., Latham, A. S., & Gentile, C. A. (1994). *NAEP writing report card.* Washington DC: National Center for Education Statistics.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (pp. 397–479). Reading MA: Addison-Wesley.

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

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46,* 443–459.

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12,* 261–280.

Bock, R. D., & Jones, L. V. (1968). *Measurement and prediction of judgment and choice.* San Francisco: Holden-Day.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psych-*

ometrika, 35, 179–197.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6,* 431–444.

Carlson, J. E. (1987). *Multidimensional item response theory estimation: A computer program* (Research Report 87-19). Iowa City IA: The American College Testing Program.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, 39,* (Series B), 1–38.

Haberman, S. J. (1977). Log-linear models and frequency tables with small expected cell counts. *Annals of Statistics, 5,* 1148–1169.

Hendrickson, A. E., & White, P. O. (1964). PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Mathematical and Statistical Psychology, 17,* 65–70.

Johnson, E. G., Carlson, J. E., & Kline, D. L. (1994). *The NAEP 1992 technical report.* Washington DC: National Center for Education Statistics.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23,* 187–200.

Luecht, R. M. (1993, April). *A marginal maximum likelihood approach to deriving multidimensional composite abilities under the generalized partial credit model.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Luecht, R. M., & Miller, T. R. (1992). Unidimensional calibrations and interpretations of composite traits for multidimensional tests. *Applied Psychological Measurement, 16,* 279–293.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149–174.

McDonald, R. P. (1965). Difficulty factors and non-linear factor analysis. *British Journal of Mathematical and Statistical Psychology, 18,* 11–23.

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology, 27,* 82–99.

McKinley, R. L., & Reckase, M. D. (1983). *An extension of the two-parameter logistic model to the multidimensional latent space* (Research Report ONR 83-2). Iowa City IA: The American College Testing Program.

Miller, T. R., & Hirsch, T. M. (1992). Cluster analysis of angular data in applications of multidimensional item-response theory. *Applied Measurement in Education, 5,* 193–211.

Mislevy, R. J., & Bock, R. D. (1990). *BILOG 3: Item analysis and test scoring with binary logistic models* (2nd ed.). Mooresville IN: Scientific Software.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement, 14,* 59–71.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159–176.

Muraki, E. (1993). *POLYFACT* [Computer program]. Princeton NJ: Educational Testing Service.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9,* 401–412.

Reckase, M. D., & McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement, 15,* 361–373.

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

Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph Supplement,* No. 18.

Stroud, A. H., & Secrest, D. (1966). *Gaussian quadrature formulas.* Englewood Cliffs NJ: Prentice-Hall.

Acknowledgments

The work reported in this paper was partially supported under the National Assessment of Educational Progress as administered by the National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education. The authors thank R. Darrell Bock for his input through discussion of the topic with the first author, and Chung Park for data analysis and preparation of figures and tables.

Author's Address

Send requests for reprints or further information to Eiji Muraki, Educational Testing Service, Rosedale Road, Princeton NJ 08541, U.S.A. Internet: emuraki@ets.org.