Open access • Journal Article • DOI:10.1177/014662169201600206

A Generalized Partial Credit Model: Application of an EM Algorithm — Source link

Eiji Muraki

Published on: 01 Jun 1992 - Applied Psychological Measurement (Sage Publications)

Topics: Polytomous Rasch model, Item response theory, Item analysis and Rasch model

A Generalized Partial Credit Model: Application of an EM Algorithm

Eiji Muraki, Educational Testing Service

The partial credit model (PCM) with a varying slope parameter is developed and called the generalized partial credit model (GPCM). The item step parameter of this model is decomposed to a location and a threshold parameter, following Andrich's (1978) rating scale formulation. The EM algorithm for estimating the model parameters is derived. The performance of this generalized model is compared on both simulated and real data to a Rasch family of polytomous item response models. Simulated data were generated and then analyzed by the various polytomous item response models. The results demonstrate that the rating formulation of the GPCM is quite adaptable to the analysis of polytomous item responses. The real data used in this study consisted of the National Assessment of Educational Progress (Johnson & Allen, 1992) mathematics data that used both dichotomous and polytomous items. The PCM was applied to these data using both constant and varying slope parameters. The GPCM, which provides for varying slope parameters, yielded better fit to the data than did the PCM. *Index terms: item response model, National Assessment of Educational Progress, nominal response model, partial credit model, polytomous response model, rating scale model.*

If responses to a test item are classified into two categories, dichotomous item response models can be applied. When responses to an item have more than two categories, a polytomous item response model is appropriate for the analysis of the responses. If the options on a rating scale are successively ordered, applicable models include the graded response model (GRM) (Samejima, 1969) and its rating scale version (Muraki, 1990a), or the partial credit model (PCM) (Masters, 1982) and its rating scale version (Andrich, 1978). For a test item in which the response options are not necessarily ordered, Bock (1972) proposed the nominal response model (NRM). The dichotomous item response model can be thought of as a special case of the polytomous item response model in which the number of categories is two.

The Partial Credit Model

A Rasch Family of Polytomous Item Response Models

Although the Rasch (1960) dichotomous model was developed independently of the latent trait models of Birnbaum (1968) and Lord (1980), the basic difference between the Rasch and the other models is the introduction of the assumption about the discriminating power of test items. These models share the following common form:

$$P_i(U_j = 1|\theta) = \frac{\exp[a(\theta - b_j)]}{1 + \exp[a(\theta - b_j)]} \quad , \tag{1}$$

which expresses the probability of person i, whose ability is parameterized by latent trait θ, correctly responding to an item j ($U_j = 1$). The parameter b_j usually refers to item difficulty. If the

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 16, No. 2, June 1992, pp. 159-176

0146-6216/92/020159-18$2.15

discrimination a is assumed to vary among test items, then the model in Equation 1 is called Birnbaum's two-parameter logistic model. If a is assumed to be common for all items, then it can be eliminated from the model by arbitrarily setting $a = 1$. This model is known as Rasch's dichotomous response model.

The separability of the model parameters and the existence of the minimal sufficient statistics of the column-wise and row-wise analyses of the response data matrix (Wright & Stone, 1979) are distinct mathematical properties of the Rasch model. These features permit a specialized parameter estimation procedure—conditional maximum likelihood estimation. If the model is viewed as a latent trait model with latent trait variable θ, the conditional likelihood of b_j, given the scores of examinee i (r_i), is independent of θ. Therefore, the parameters b_j can be estimated from the conditional likelihood involving no person parameters. From this point of view, the inflection points of the model are b_j. On the other hand, if the model is viewed as a latent trait model with latent trait variable b, then the conditional likelihood of θ_i, given the scores of item j, is independent of b. Therefore, the parameters θ_i can be estimated from this conditional likelihood involving no item parameters. From this point of view, the inflection points of the models are θ_i. If the assumption is met that all items have equal discriminations and vary only in terms of difficulty, then the Rasch model provides an elegant and simple solution for several technical applications to test analysis and construction (Wright & Stone, 1979).

The notable distinction between the Rasch polytomous item response models (Andrich, 1978; Masters, 1982) and the GRM (Muraki, 1990a; Samejima, 1969) is not the number of parameters but the difference in terms of the operating characteristic function (OCF) (Samejima, 1972). The OCF is central to the polytomous item response models. This function expresses how the probability of a specific categorical response is formulated according to the law of probability, as well as psychological assumptions about item response behavior. Masters (1982) formulated his PCM by using the Rasch dichotomous model; therefore, it is legitimate to construct the PCM based on the two-parameter logistic response model following the same OCF Masters employed. Because the essential mechanism for constructing a general model is shared with Masters' PCM, the model constructed here can be called the generalized partial credit model (GPCM).

The Generalized Partial Credit Model

The GPCM is formulated based on the assumption that the probability of selecting the kth category over the k minus first (k - 1) category is governed by the dichotomous response model. To develop the PCM, denote $P_{jk}(\theta)$ as the specific probability of selecting the kth category from m_j possible categories of item j.

For each of the adjacent categories, the probability of the specific categorical response k over k - 1 is given by the conditional probability, which is the same as Equation 1:

$$C_{jk} = P_{jk|k-1,k}(\theta) = \frac{P_{jk}(\theta)}{P_{j,k-1}(\theta) + P_{jk}(\theta)} = \frac{\exp[a_j(\theta - b_{jk})]}{1 + \exp[a_j(\theta - b_{jk})]} \quad , \tag{2}$$

where $k = 2, 3, \ldots, m_j$. Equation 2 then becomes

$$P_{jk}(\theta) = \frac{C_{jk}}{1 - C_{jk}} P_{j,k-1}(\theta) \quad . \tag{3}$$

Note that $C_{jk}/(1 - C_{jk})$ is the ratio of the two conditional probabilities, which also may be expressed as $\exp[a_j(\theta - b_{jk})]$. Equation 3 may be called the OCF for the PCM.

If

$$P_{j1}(\theta) = \frac{1}{G} \quad , \tag{4}$$

where G is called a normalizing factor (defined below), the following probabilities are obtained by applying the OCF in Equation 3:

$$P_{j2}(\theta) = \frac{\exp[a_j(\theta - b_{j2})]}{G} \quad , \tag{5}$$

$$\vdots$$

$$P_{jg}(\theta) = \frac{\exp\left[\sum_{v=2}^{g} a_j(\theta - b_{jv})\right]}{G} \quad , \tag{6}$$

$$\vdots$$

and

$$P_{jm_j}(\theta) = \frac{\exp\left[\sum_{v=2}^{m_j} a_j(\theta - b_{jv})\right]}{G} \quad , \tag{7}$$

where g is a subscript for a specific categorical response $k = g$. Because

$$\sum_{k=1}^{m_j} P_{jk}(\theta) = 1 \quad , \tag{8}$$

and

$$G = 1 + \sum_{c=2}^{m_j} \exp\left[\sum_{v=2}^{c} a_j(\theta - b_{jv})\right] \quad , \tag{9}$$

the PCM is formulated by

$$P_{jk}(\theta) = \frac{\exp\left[\sum_{v=1}^{k} a_j(\theta - b_{jv})\right]}{\sum_{c=1}^{m_j} \exp\left[\sum_{v=1}^{c} a_j(\theta - b_{jv})\right]} \quad , \tag{10}$$

where $b_{j1} \equiv 0$.

Note that b_{j1} is arbitrarily defined as 0. This value is not a location factor. It could be any value, because the term including this parameter is canceled from the numerator and denominator of the model:

$$P_{jk}(\theta) = \frac{\exp[Z_{j1}(\theta)] \cdot \exp\left[\sum_{v=2}^{k} Z_{jv}(\theta)\right]}{\exp[Z_{j1}(\theta)] + \sum_{c=2}^{m_j} \exp\left[Z_{j1}(\theta) + \sum_{v=2}^{c} Z_{jv}(\theta)\right]} = \frac{\exp\left[\sum_{v=2}^{k} Z_{jv}(\theta)\right]}{1 + \sum_{c=2}^{m_j} \exp\left[\sum_{v=2}^{c} Z_{jv}(\theta)\right]} \quad , \tag{11}$$

where $Z_{jk}(\theta) = a_j(\theta - b_{jk})$. The PCM in Equation 10 reduces to the dichotomous item response model when $m_j = 2$ and $k = 1, 2$.

Masters (1982) calls the parameters b_{jk} in Equation 10 item step parameters. The b_{jk} are the points on the θ scale at which the plots of $P_{j,k-1}(\theta)$ and $P_{jk}(\theta)$ intersect. These two curves—which can be referred to as the item category response functions (ICRFs)—intersect only once, and the intersection can occur anywhere along the θ scale. Thus,

$$\begin{aligned} &\text{if } \theta = b_{jk}, \quad P_{jk}(\theta) = P_{j,k-1}(\theta); \\ &\text{if } \theta > b_{jk}, \quad P_{jk}(\theta) > P_{j,k-1}(\theta); \\ &\text{and if } \theta < b_{jk}, \quad P_{jk}(\theta) < P_{j,k-1}(\theta); \end{aligned} \tag{12}$$

under the assumption $a_j > 0$. It should be noted that b_{jk} is not sequentially ordered within item j because the parameter represents the relative magnitude of the adjacent probabilities $P_{j,k-1}(\theta)$ and $P_{jk}(\theta)$.

Although the intersection points of ICRFs of the PCM are easily interpretable, the peak points of these curves for the middle categories are not. The first derivative of $P_{jk}(\theta)$ is expressed by

$$\frac{\partial}{\partial\theta} P_{jk}(\theta) = a_j P_{jk}(\theta)\left[k - \sum_{c=1}^{m_j} c P_{jc}(\theta)\right] \quad . \tag{13}$$

By setting the first derivative in this equation to 0,

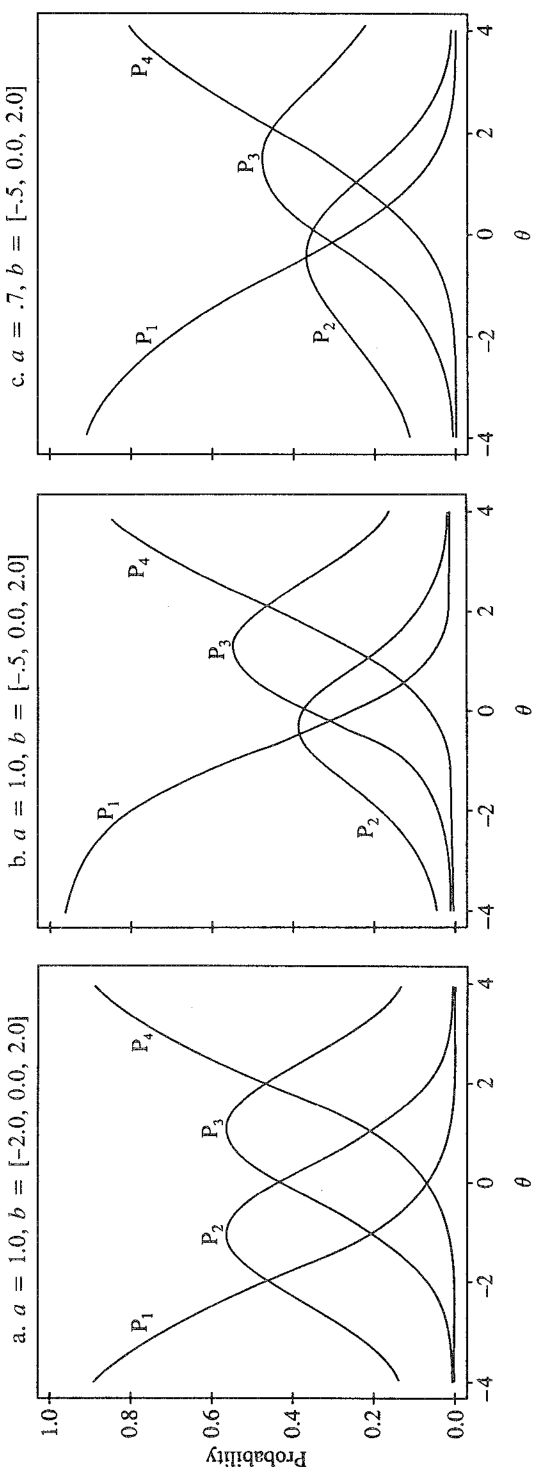
$$\sum_{c=1}^{m_j} c P_{jc}(\theta) = k \quad . \tag{14}$$

Equation 14 shows that the peak of the ICRF, $P_{jk}(\theta)$, is affected by all the other probabilities, $P_{jc}(\theta)$, $c = 1, 2, \ldots, k-1, k+1, \ldots, m_j$.

The parameter a_j is a slope parameter for item j. The range of a_j is generally assumed to be from 0 to ∞. In contrast to the dichotomous models, for the polytomous item response model the discriminating power of each ICRF depends on a combination of the slope and threshold parameters. Andrich (1978) distinguished between the discriminating powers of the two types of models and retained only the threshold discrimination in his Rasch family of rating scale models (RSMs).

In the PCM, only the item discriminating power is included. This slope parameter indicates the degree to which categorical responses vary among items as θ level changes. This concept of item discriminating power is closely related to the item reliability index in classical test theory. Thus, by retaining the item discriminating power in the model, the continuity of Birnbaum's (1968) two-parameter model from the dichotomous to the polytomous response case, as well as the connection with the classical test concept, is retained. The model with this slope parameter also can be extended to the multidimensional form (Muraki, 1985, 1990b).

Figure 1
PCM ICRFs for a Four-Category Item

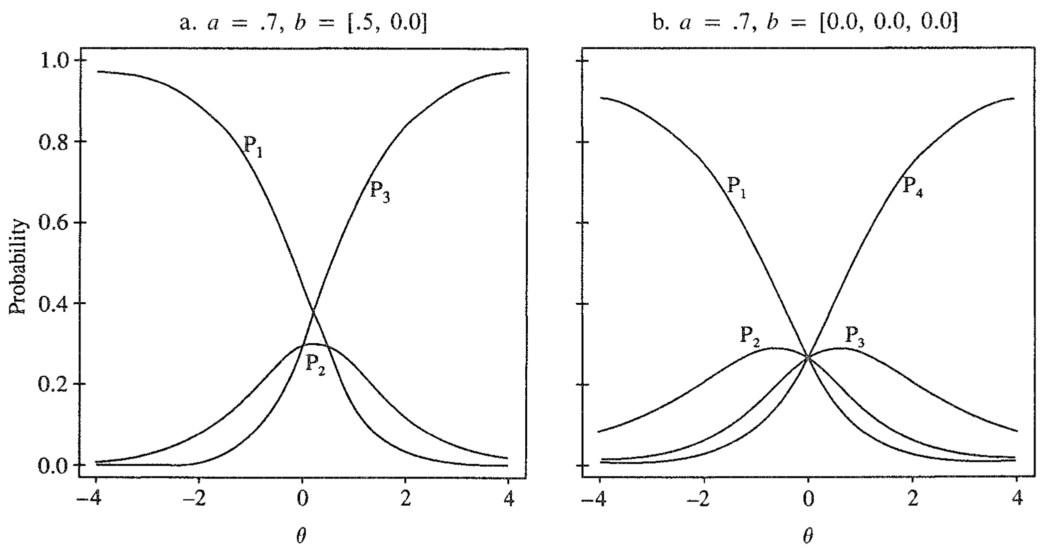


Figures 1a, 1b, and 1c show the ICRFs for the PCM with four categorical responses. Figure 1a shows the ICRFs for an item with $a_j = 1.0$, $b_{j2} = -2.0$, $b_{j3} = 0.0$, and $b_{j4} = 2.0$. If b_{j2} and b_{j3} are brought closer together by changing b_{j2} to $-.5$, then the probability of responding to the second category decreases, as illustrated in Figure 1b. In other words, the range of the θ values of persons who are more likely to respond to the second category than to the other categories decreases from (-2, 0) to (-.5, 0). If the slope parameter is changed from 1.0 to .7, as shown in Figure 1c, the intersection points of all ICRFs are left unchanged, and the curves become flatter. The discriminating power of these ICRFs decreases for all categorical responses.

Figure 2a shows the ICRFs for the PCM with three categorical responses. When the second item step parameter is made larger than the third item step parameter ($b_{j2} > b_{j3}$), the ICRF of P_{j2} drops, as shown in Figure 2a. For the entire range of θ values, the probability of the first or third categorical response is higher than the probability of the second categorical response. Figure 2a shows that every person who is more likely to respond to the second category than to the first (or third) category, is most likely to respond to the third (or first) category. Consequently, the marginal frequency of the second categorical response becomes quite small compared to the other response frequencies.

If all item step parameters have the same value, as shown in Figure 2b, all ICRFs intersect at the same value of θ. Although the values of item step parameters are not sequentially ordered, the PCM expresses the probabilities of ordered responses.

Figure 2
PCM ICRFs for a Three-Category Item



The RSM is derived from the PCM by assuming b_{jk} can be decomposed additively as $b_{jk} = b_j - d_k$,

$$P_{jk}(\theta) = \frac{\exp\left[\sum_{v=1}^{k} a_j(\theta - b_j + d_v)\right]}{\sum_{c=1}^{m} \exp\left[\sum_{v=1}^{c} a_j(\theta - b_j + d_v)\right]} , \tag{15}$$

where $d_1 \equiv 0$ and the parameter b_{jk} is resolved into two parameters b_j and d_k ($b_{jk} = b_j - d_k$). Masters

(1982) modified Andrich's (1978) model in Equation 15 and called it the PCM. In their models, the slope parameter is assumed to be a constant. Samejima's (1972) GRM also was extended to the RSM by Muraki (1990a). Both Andrich and Muraki separated the item category threshold parameter into an item parameter and a category parameter in the same manner. Because the RSM in Equation 15 is essentially identical to the PCM when single items are considered, this model is simply called a rating formulation of the PCM. The rating version of the PCM can be applied to any situations where the PCM is fitted. The parameter b_{jk} can be recomputed from the estimates of b_j and d_k (or d_{jk}) after the parameters are estimated. Therefore, the model in Equation 15 is called the GPCM, unless its rating aspect is specifically emphasized.

Andrich (1982) calls b_j and d_k in Equation 15 an item location parameter and a threshold parameter, respectively. Because the values of the item step parameters (b_{jk}) are not necessarily ordered within item j, the threshold parameters (d_k) are not sequentially ordered for $k = 1, 2, \ldots, m$. The parameter d_k is interpreted as the relative difficulty of step k in comparing other steps within an item.

Parameter Estimation

Let U_{jki} represent an element in the matrix of the observed response pattern i. $U_{jki} = 1$ if the response to item j is in the kth category, otherwise $U_{jki} = 0$. By the principle of local independence (Birnbaum, 1968), the conditional probability of a response pattern i, given θ, for m response categories and n items, as denoted by a response matrix $[\mathbf{U}_i = (U_{jk})_i]$, is the joint probability:

$$P(\mathbf{U}_i|\theta) = P_i[(U_{jk})|\theta] = \prod_{j=1}^{n}\prod_{k=1}^{m}[P_{jk}(\theta)]^{U_{jki}} \quad , \tag{16}$$

where $U_{jk} = 1$ if the kth category of item j is selected, otherwise $U_{jk} = 0$.

For examinees randomly sampled from a population with a normal distribution of the latent trait variable, $\phi(\theta)$, the marginal probability of the observed response pattern i is

$$P_i[(U_{jk})] = \int_{-\infty}^{\infty} P_i[(U_{jk})|\theta]\ \phi(\theta)\,d\theta \quad . \tag{17}$$

If an examinee responds to n items with m categories, the examinee's response pattern i can then be assigned to one of m^n mutually exclusive patterns. Let r_i represent the number of examinees observed in such a pattern i, and let N be the total number of examinees sampled from the population. Then r_i is multinomially distributed with parameters N and $P_i[(U_{jk})]$. Thus,

$$L = \frac{N!}{\prod_{i=1}^{m^n} r!}\prod_{i=1}^{m^n}\{P_i[(U_{jk})]\}^{r_i} \quad . \tag{18}$$

Taking the natural logarithm of Equation 18 yields

$$\ln L = \ln N! - \sum_{i=1}^{m^n}\ln r_i! + \sum_{i=1}^{m^n} r_i \ln P_i[(U_{jk})] \quad . \tag{19}$$

The likelihood equation for $\hat{a}_j$, $\hat{b}_j$, and $\hat{d}_k$ can be derived from the first partial derivative of Equation 19 with respect to each parameter and setting them to 0.

Item parameter estimation. Let υ_j represent the parameter a_j or b_j. With respect to υ_h, which is the parameter υ_j for the specific item $j = h$, the likelihood in Equation 19 can be differentiated as

$$\frac{\partial \ln L}{\partial \upsilon_h} = \sum_{i=1}^{m^n}\frac{r_i}{P_i(U_{jk})}\int_{-\infty}^{\infty} P_i[(U_{jk})|\theta]\sum_{k=1}^{m}\frac{\partial[P_{hk}(\theta)]^{U_{hki}}}{\partial \upsilon_h}\ \frac{\phi(\theta)\,d\theta}{[P_{hk}(\theta)]^{U_{hki}}} \quad . \tag{20}$$

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

$$\sum_{\ell=1}^{S} r_\ell = N \quad . \tag{21}$$

The first derivative of the likelihood function in Equation 20 can be approximated by using the Gauss-Hermite quadrature, such that

$$\frac{\partial \ln L}{\partial \upsilon_h} \approx \sum_{\ell=1}^{S} \sum_{f=1}^{F} \frac{r_\ell L_\ell(X_f) A(X_f)}{\tilde{P}_\ell} \sum_{k=1}^{m} \frac{\partial [P_{hk}(X_f)]^{U_{hk\ell}}}{\partial \upsilon_h} \frac{1}{[P_{hk}(X_f)]^{U_{hk\ell}}} \quad , \tag{22}$$

where

$$\tilde{P}_\ell = \sum_{f=1}^{F} L_\ell(X_f) A(X_f) \tag{23}$$

and

$$L_\ell(X_f) = \prod_{j=1}^{n} \prod_{k=1}^{m} [P_{jk}(X_f)]^{U_{jk\ell}} \quad . \tag{24}$$

In Equation 22, $A(X_f)$ is the weight of the Gauss-Hermite quadrature, and X_f is the quadrature point (Stroud & Secrest, 1966). The quadrature weight $A(X_f)$ is approximately the standard normal probability density at the point X_f, such that

$$\sum_{f=1}^{F} A(X_f) = 1 \quad , \tag{25}$$

where F is the total number of quadrature points. Because $U_{hk\ell}$ can take only two possible values, 1 or 0, Equation 22 can be rewritten as

$$\sum_{f=1}^{F} \sum_{k=1}^{m} \frac{\bar{r}_{hkf}}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial \upsilon_h} \quad , \tag{26}$$

where

$$\bar{r}_{hkf} = \sum_{\ell=1}^{S} \frac{r_\ell L_\ell(X_f) A(X_f) U_{hk\ell}}{\tilde{P}_\ell} \quad , \tag{27}$$

and $\bar{r}_{hkf}$ is the provisional expected frequency of the kth categorical response of item h at the fth quadrature point.

Bock and Aitkin (1981) applied the EM algorithm (Dempster, Laird, & Rubin, 1977) to estimate the parameters for each item individually, and then repeat the iteration process over n items until the estimates of all items become stable to the required number of decimal places. The qth cycle of the iterative process can be expressed as

$$\upsilon_q = \upsilon_{q-1} + \mathbf{V}^{-1}\mathbf{t} \quad , \tag{28}$$

where υ_q and υ_{q-1} are the parameter estimates of the qth and $q-1$st cycles respectively, $\mathbf{V}^{-1}$ is the inverse of the information matrix, and $\mathbf{t}$ is the gradient vector. For item parameter estimation, the elements of $\mathbf{t}$ and $\mathbf{V}$ are

$$t_{\upsilon_h} = \sum_{f=1}^{F} \sum_{k=1}^{m} \frac{\bar{r}_{hkf}}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial \upsilon_h} \tag{29}$$

and

$$V_{\upsilon_h \omega_h} = \sum_{f=1}^{F} \bar{N} \sum_{k=1}^{m} \frac{1}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial \upsilon_h} \frac{\partial P_{hk}(X_f)}{\partial \omega_h} \quad , \tag{30}$$

where $\upsilon_h = a_h$ or b_h and $\omega_h = a_h$ or b_h. In Equation 30, $\bar{N}_f$ is called the provisional expected sample size at quadrature point f, and is computed by

$$\bar{N}_f = \sum_{\ell=1}^{S} \frac{r_\ell L_\ell(X_f) A(X_f)}{\tilde{P}_\ell} \quad . \tag{31}$$

A rigorous proof of the approximation of the second derivatives in Equation 30 by the product of the first derivatives is given by Kendall and Stuart (1973).

The model, $P_{jk}(X_f)$, is a logistic function so that the evaluation of several functions stated above becomes relatively simple in comparison with the normal ogive model. The elements of the gradient vector and the information matrix are given by

$$t_{a_h} = a_h^{-1} \sum_{f=1}^{F} \sum_{k=1}^{m} \bar{r}_{hkf} \left[Z_{hk}^{+}(X_f) - \sum_{c=1}^{m} Z_{hc}^{+}(X_f) P_{hc}(X_f) \right] \quad , \tag{32}$$

$$t_{b_h} = a_h \sum_{f=1}^{F} \sum_{k=1}^{m} \bar{r}_{hkf} \left[-k + \sum_{c=1}^{m} c P_{hc}(X_f) \right] \quad , \tag{33}$$

$$V_{a_h a_h} = a_h^{-2} \sum_{f=1}^{F} \bar{N}_f \sum_{k=1}^{m} P_{hk}(X_f) \left[Z_{hk}^{+}(X_f) - \sum_{c=1}^{m} Z_{hc}^{+}(X_f) P_{hc}(X_f) \right]^2 \quad , \tag{34}$$

$$V_{b_h b_h} = a_h^2 \sum_{f=1}^{F} \bar{N}_f \sum_{k=1}^{m} P_{hk}(X_f) \left[-k + \sum_{c=1}^{m} c P_{hc}(X_f) \right]^2 \quad , \tag{35}$$

and

$$V_{a_h b_h} = \sum_{f=1}^{F} \bar{N}_f \sum_{k=1}^{m} P_{hk}(X_f) \left[Z_{hk}^{+}(X_f) - \sum_{c=1}^{m} Z_{hc}^{+}(X_f) P_{hc}(X_f) \right] \left[-k + \sum_{c=1}^{m} c P_{hc}(X_f) \right] \quad , \tag{36}$$

where

$$Z_{hk}^{+}(X_f) = \sum_{v=1}^{k} Z_{hv}(X_f) = \sum_{v=1}^{k} a_h (X_f - b_h + d_v) \quad . \tag{37}$$

Threshold parameter estimation. Because the threshold parameter d_g, which is the parameter d_k for the specific category $k = g$, is contained in all $P_{jk}(\theta)$ ($k = 1, 2, \ldots, m$) as shown in Equation 15, the first derivative of the likelihood function in Equation 24 with respect to d_g is given by

$$\frac{\partial L_\ell(X_f)}{\partial d_g} = L_\ell(X_f) \sum_{j=1}^{n} \sum_{k=1}^{m} \frac{U_{jk\ell}}{P_{jk}(X_f)} \frac{\partial P_{jk}(X_f)}{\partial d_g} \quad . \tag{38}$$

According to the EM algorithm and Equation 38, the maximum likelihood function in Equation 19 with respect to d_g is written as

$$t_{d_g} = \sum_{f=1}^{F} \sum_{j=1}^{n} \sum_{k=1}^{m} \frac{\bar{r}_{jkf}}{P_{jk}(X_f)} \frac{\partial P_{jk}(X_f)}{\partial d_g} = \sum_{f=1}^{F} \sum_{j=1}^{n} a_j \sum_{k=g}^{m} [\bar{r}_{jkf} - P_{jk}(X_f) \sum_{c=1}^{m} \bar{r}_{jcf}] \quad . \tag{39}$$

The entry of the information matrix for $g' \le g$, then becomes

$$V_{d_g d_{g'}} = \sum_{f=1}^{F} \bar{N}_f \sum_{j=1}^{n} \sum_{k=1}^{m} \frac{1}{P_{jk}(X_f)} \frac{\partial P_{jk}(X_f)}{\partial d_g} \frac{\partial P_{jk}(X_f)}{\partial d_{g'}} = \sum_{f=1}^{F} \bar{N}_f \sum_{j=1}^{n} a_j^2 \left[\sum_{k=g}^{m} P_{jk}(X_f) \right] \left[1 - \sum_{k=g'}^{m} P_{jk}(X_f) \right] \quad . \tag{40}$$

Because d_1 is defined to be 0, the orders of the gradient vector **t** and the information matrix **V** are $m - 1$ and $(m - 1) \times (m - 1)$, respectively.

Comparison with the Nominal Response Model

The PCM can be rewritten by using $Z_{jk}^{+}(\theta)$ as defined in Equation 37, that is,

$$P_{jk}(\theta) = \frac{\exp[Z_{jk}^{+}(\theta)]}{\sum_{c=1}^{m} \exp[Z_{jc}^{+}(\theta)]} \quad . \tag{41}$$

Equation 41 is exactly the form of the NRM proposed by Bock (1972). His original formulation of the NRM is

$$P_{jk}^{*}(\theta) = \frac{\exp[Z_{jk}^{*}(\theta)]}{\sum_{c=1}^{m} \exp[Z_{jc}^{*}(\theta)]} \quad , \tag{42}$$

where

$$Z_{jc}^{*}(\theta) = a_{jc}^{*}\theta + d_{jc}^{*} = a_{jc}^{*}(\theta - b_{jc}^{*}) \quad . \tag{43}$$

Therefore, the NRM becomes equivalent to the PCM if the following conditions are satisfied:

$$a_{jk}^{*} = k a_j \tag{44}$$

and

$$b_{jk}^{*} = \frac{\sum_{v=1}^{k} b_{jv}}{k} = b_j - \frac{\sum_{v=1}^{k} d_v}{k} \quad . \tag{45}$$

If the item response model is for ordered categories, the odds of being in a higher score category should be greater for an examinee with higher θ than for an examinee with lower θ. Wainer (personal communication, May, 1991) constructed the following inequality:

$$\frac{P_{j,k+1}^{*}(\theta + \delta_\theta)}{P_{jk}^{*}(\theta + \delta_\theta)} > \frac{P_{j,k+1}^{*}(\theta)}{P_{jk}^{*}(\theta)} \quad , \tag{46}$$

where $\delta_\theta > 0$. Inequality 46 can be rewritten as

$$\frac{P_{j,k+1}^{*}(\theta + \delta_\theta) P_{jk}^{*}(\theta)}{P_{jk}^{*}(\theta + \delta_\theta) P_{j,k+1}^{*}(\theta)} > 1 \quad . \tag{47}$$

Taking the natural logarithm of Inequality 47 yields

$$\ln P_{j,k+1}^{*}(\theta + \delta_\theta) + \ln P_{jk}^{*}(\theta) - \ln P_{jk}^{*}(\theta + \delta_\theta) - \ln P_{j,k+1}^{*}(\theta) > 0 \quad . \tag{48}$$

Substituting each $P_{jk}^{*}(\theta)$ with the NRM in Equation 42, yields

$$Z^*_{j,k+1}(\theta + \delta_\theta) + Z^*_{jk}(\theta) - Z^*_{jk}(\theta + \delta_\theta) - Z^*_{j,k+1}(\theta) = a^*_{j,k+1}\delta_\theta - a^*_{jk}\delta_\theta > 0 \quad . \tag{49}$$

Therefore, the more general condition for the NRM to be the model for ordered response categories is

$$a^*_{j,k+1} > a^*_{jk} \quad . \tag{50}$$

The increment of scaling factors along the consecutive categories, shown in Equations 44 and 50, is a hidden feature of the PCM. Because of this characteristic, the PCM becomes the model for ordered response categories. The PCM is a special case of the NRM. In the PCM, the degree of the expansion of scaling factors is expressed by Equation 44, that is, a_j, $2a_j$, $3a_j$, etc. Andrich (1978) called this feature the linear scoring function.

The numerator of the PCM can be rewritten as

$$Z^+_{jk}(\theta) = a_j\left[k(\theta - b_j) + \sum_{v=1}^{k} d_v\right] \quad . \tag{51}$$

Andrich's RSM (Andrich, 1978), with a varying slope parameter, a_j, is written as

$$Z^+_{jk}(\theta) = a_j[T_k(\theta - b_j) + K_k] \quad , \tag{52}$$

where T_k is the scoring function, and K_k is the category coefficient. Note that Andrich's RSM becomes the NRM if the scoring function, T_k, is treated as an estimable quantity from response data. In the PCM, T_k is set a priori as a series of sequential integers, which is shown in Equation 44. Andrich (1988) further extended his RSM by reparameterizing the category coefficient so that the model incorporated binomial and Poisson response processes as well as linear and quadratic coefficients.

As observed above, the PCM is a special case of Bock's (1972) NRM. Andrich's (1978) scoring function is a key concept to understanding these model formulations. Andrich (1978, 1982, 1988) also demonstrated that the PCM can be extended further by reparameterizing the scoring function and the category coefficient. By using these features, the model expressing the partial order of the categorical responses or any specific response processes may be constructed. Thissen and Steinberg (1986) demonstrated that the nominal item response model is a basic model which can be extended further to the ordered response model or other models, including the completely nonordered and the partially ordered models. Their approach to the polytomous item response model using the contrast is quite useful for further developments with respect to these models.

Constraints on the Threshold Parameters

Integrating out a nuisance variable, θ, from the likelihood with a fixed prior, as shown in Equation 17, eliminates the indeterminacy of the item parameters, a_j and b_{jk}. Thus, for dichotomous item responses, both slope and location parameters can be estimated without constraints. In this case, d_1 is already defined to be 0.

A block of items is defined here as a set of items that share the same set of threshold parameters. For the polytomous item response models, there is an indeterminacy between a set of threshold parameters and location parameters of items within a block. To obtain a unique set of parameters, a constraint, called a location constraint, must be imposed on the estimation of threshold parameters. A location constraint is imposed so that the mean of threshold parameters within a categorical scale is constant over blocks. A natural choice is 0, that is,

$$\sum_{k=2}^{m_j} d_k = 0 \quad . \tag{53}$$

If there are more than two blocks of items in a given questionnaire or cognitive test, estimated location parameters can be compared within each block, but not among blocks. The location constraint makes the comparison of location parameters over the blocks possible.

For the dichotomous item response models, a slope parameter represents the discriminating power of the item. However, for the polytomous item response models, the discriminating power of each item is a combination of a slope parameter and a set of threshold parameters. In other words, each ICRF may have a different discriminating power within an item. The slope parameters are directly comparable only when the items share the same set of threshold parameters. In this case, the scaling factor due to the threshold parameters is controlled, and the item discriminating power for each item is extracted. Separability of the effects due to the slope and threshold parameters on the discriminating power needs to be investigated further.

Reparameterization of a Rasch family of polytomous item response models provides more flexibility for the analyses of polytomous item responses. The two-parameter dichotomous item response model becomes a special case of the GPCM. More importantly, the model can be tested in a step-wise manner. If more than a single item is involved with a given block, a common set of threshold parameters and slope and location parameters for each item can be estimated. The model then can be fitted again to each item, a separate set of threshold parameters for each item can be obtained, and the assumption about the common threshold parameters for all items included in a block can be tested. If a common set of threshold parameters for items in a block is reasonably fitted to polytomously scored response data, then all methodologies based on the dichotomous item response models can be applicable without great difficulty. If item step parameters are necessary, the model in Equation 15 can be fit by imposing the location constraint on the threshold parameters. Then, the item step parameters, b_{jk}, can be computed by

$$b_{jk} = b_j - d_{jk} \quad . \tag{54}$$

Thus, the model can be equally applicable to situations in which each block contains only one item and to situations in which a test contains a mixture of dichotomous and polytomous item responses.

The Marginal Maximum Likelihood (MML) EM Algorithm

The EM algorithm presented here is available in the PARSCALE computer program (Muraki & Bock, 1991). PARSCALE also can estimate the parameters of the GRM. Data analyses by this model were presented by Muraki (1990a).

The MML-EM algorithm implemented in the PARSCALE program consists of two steps. The first is the expectation step (the E step) in which the provisional expected frequency and the provisional expected sample size are computed by Equations 27 and 31, respectively. Then, in the maximization step (the M step), the MML estimates are obtained by Equation 28. Both the E step and the M step are repeated (the EM cycle) until all estimates become stable.

Each EM cycle consists of two estimation processes. First, the threshold parameters, d_k, are estimated one block at a time with or without constraints, and then the item parameters within the block are estimated one item at a time. Each estimation process is repeated until the values become stable at a specified level of precision.

PARSCALE provides a likelihood ratio χ^2 for each item, which is computed based on the method Mislevy and Bock (1990) described in the manual of PC-BILOG 3. An index of model fit is computed by summing the item fit statistics over items.

Example Applications

Simulated Data

5,000 Likert-type response vectors for 30 items and three categories were generated with a standard normal distribution of θ. The original parameter values were obtained from the analysis of knowledge of physics data (Masters, 1982). The slope parameters of all items were 1.0. Masters' item step parameters, b_{jk}, varied (see Table 1). The RESGEN computer program (Muraki, 1990b) was used

Table 1
Original and Estimated Item Step Parameters for Analysis 1

	Item Step Parameters			
	Original		Estimated	
Item	b_{j2}	b_{j3}	$\hat{b}_{j2}$	$\hat{b}_{j3}$
1	-.64	-1.75	-.687	-1.751
2	1.10	.53	1.096	.615
3	.62	-.78	.669	-.818
4	.68	-.01	.695	-.055
5	.84	-.30	.782	-.340
6	-.22	-.61	-.177	-.603
7	1.76	-.66	1.697	-.550
8	.41	-.24	.351	-.340
9	1.06	-.28	1.137	-.240
10	1.29	-.04	1.277	-.033
11	1.02	-1.75	1.033	-1.706
12	.64	-.37	.590	-.373
13	.66	.43	.671	.406
14	.40	-.54	.494	-.612
15	1.24	1.00	1.200	1.046
16	-.30	-.46	-.237	-.456
17	-.02	.26	-.065	.315
18	-.24	-.81	-.264	-.816
19	-1.97	-.05	-2.001	-.081
20	.34	-1.13	.330	-1.135
21	-.02	-.26	.026	-.277
22	-.54	-.57	-.515	-.599
23	-.43	-1.29	-.401	-1.270
24	.14	.31	.151	.317
25	.41	.66	.415	.630
26	-.45	-.45	-.453	-.485
27	-.08	-.53	-.066	-.554
28	-.37	-.58	-.373	-.572
29	.57	.79	.555	.842
30	.87	.78	.834	.795

to generate the simulated dataset. These data were then analyzed four times under various constraints. 10 quadrature points were used, and the precision level .0001 was set for all estimations.

In Analysis 1, the simulated data were analyzed based on Masters' (1982) PCM. Slope parameters were kept constant during the estimation process. The location constraint was not applied to the estimation. Analysis 2 was identical to Analysis 1, except that the location constraint was applied. The item step parameters, b_{jk}, were computed from b_j and d_{jk}. Original and estimated item step parameter values are presented in Table 1. As shown in Table 1, the EM algorithm successfully recovered original parameter values.

Although the location estimates, $\hat{b}_j$, were different between the results of Analyses 1 and 2, as shown in Table 2, the values of $\hat{b}_{jk}$ were computed from b_j and d_{jk} of Analysis 1, and they were found to be indistinguishable from the estimated values of Analysis 2. Their −2 log likelihood statistics were almost identical. In other words, these two models were essentially the same. However, Analysis 2 needed fewer iterations to reach the convergence criterion because the indeterminacy was eliminated. In addition, the location estimates of Analysis 2 can be compared with each other because of the location constraint.

Table 2
Means and Standard Deviations (SD) of Estimated Slope and Location Parameters and −2 Log Likelihoods for Analyses 1 Through 6

			Estimated Parameters				
		Location	Slope		Location		−2 Log
Analysis	Block	Constraint	Mean	SD	Mean	SD	Likelihood
1	30	No	1.00	0.00	−.004	.042	255,826.34
2	30	Yes	1.00	0.00	.002	.582	255,826.33
3	1	No	1.00	0.00	0.000	.569	259,762.90
4	1	No	.88	.16	−.012	.658	258,341.07
5	30	Yes	1.00	0.00	.083	.489	264,365.59
6	30	Yes	.91	.43	.015	.644	256,466.35

In Analysis 3, the 30 items were made into a block and only one set of threshold parameters was estimated. The slope parameters were again kept constant. No location constraint was applied to this estimation. The −2 log likelihood was considerably higher than the solution found using Analyses 1 or 2. In other words, the assumption about a common set of threshold parameters for all 30 items was not appropriate for these data, as would be expected based on the model used to specify the original set of parameter values. It also should be pointed out that the mean of $\hat{b}_j$ was 0 in Analysis 3. The mean of the set of location estimates was completely absorbed into the threshold parameters. The threshold estimates were −.257 and .249. When the data were analyzed by imposing the location constraint, the results showed that the threshold parameter estimates were shifted left by .004, and the mean of $\hat{b}_j$ was shifted in the opposite direction by the same amount. The location constraint again shortened the number of iterations.

Slope parameters were estimated in Analysis 4. Because a block of 30 items was set and only one set of threshold parameters was estimated, the −2 log likelihood was higher than that obtained in Analysis 1 or 2, but the model fit was improved compared to Analysis 3. In other words, some portion of the categorical discriminating power was absorbed by the slope parameters.

Using the same item step parameters, another simulated dataset was generated in Analysis 5. In this simulated dataset, various slope parameter values were used. The original slope parameters were .3, .6, .9, 1.2, 1.5, and 1.8 for the set of six items. This set of slope parameters was applied repeatedly to the remaining items. The PCM was fitted with a constant slope. The −2 log likelihood was 264,366, as shown in Table 2. When the GPCM was fit and the slope parameters were estimated (Analysis 6), the −2 log likelihood decreased to 256,466. The difference was 7,900 with 30 degrees of freedom (*df*). The difference of the model fit statistics was 7,480 (df = 5). Thus, the model fit was significantly improved by applying the GPCM.

Parameter estimates for the first 12 items are presented in Table 3. All slope parameters were

Table 3
Estimated Values of Slope, Location, and Threshold Parameters for Two Thresholds, and Standard Errors (SE) of the Estimates for Analysis 6

	Slope		Location		1		2	
Item	Est.	SE	Est.	SE	Threshold	SE	Threshold	SE
1	.254	.015	−1.348	.103	−.729	.155	.729	.134
2	.530	.018	.995	.045	−.295	.066	.295	.079
3	.746	.017	−.063	.025	−.936	.058	.936	.057
4	1.065	.024	.380	.020	−.401	.039	.401	.042
5	1.332	.027	.288	.016	−.626	.037	.626	.039
6	1.580	.035	−.455	.016	−.237	.032	.237	.029
7	.277	.013	.648	.068	−1.356	.130	1.356	.138
8	.569	.018	0.000	.032	−.308	.065	.308	.065
9	.790	.019	.520	.025	−.804	.053	.804	.057
10	1.078	.023	.704	.021	−.749	.044	.749	.048
11	1.294	.026	−.379	.017	−1.677	.065	1.677	.064
12	1.475	.030	.156	.015	−.626	.036	.626	.036

underestimated; consequently, all threshold parameters were overestimated. The parameters of the first six items with various numbers of quadrature points then were repeatedly estimated. This initial investigation (Muraki, 1992) suggested that parameters are not necessarily underestimated, and estimation bias seems to decrease as the number of quadrature points increases. A reasonable number of quadrature points may be determined by the number of items in a test and the number of response categories. This estimation problem should be studied further.

National Assessment of Educational Progress (NAEP) Mathematics Data

NAEP mathematics data for 1989–90 were analyzed, based on 16 items from the Grade 8 assessment. 12 items were dichotomously scored, and the other four items were polytomous items for which the number of categories varied from three to six. The number of categorical responses for each item is the number of threshold parameters, which is shown in Table 4, plus 1. Item 10 originally had six categories, but no student responded to the fifth category. Therefore, the item was treated as a five-category item. The item responses of 3,679 students were used for the analysis. (The total number of students was 3,699, but 20 students were excluded from the analysis because they omitted all items.) 33 quadrature points and a convergence criterion .001 were used for the estimation.

The PCM with a constant slope was fit to the data. The −2 log likelihood was 69,688. The fit was significantly improved when the PCM was fit with varied slope parameters. The −2 log likelihood of this GPCM was 68,748. The difference was 940 (df = 16). The model fit statistics for these models were 1,577 and 957 (df = 188 and 189, respectively). Thus, the difference of the fit statistics, 620, also indicates a significant improvement. The estimated parameters are presented in Table 4.

The location constraint for each set of threshold parameters was applied. Therefore, the location estimates are comparable. The higher location estimates indicate more difficult items. Item 6 was the easiest item, and Item 11 was the most difficult. More than 80% of the students correctly responded to Item 6, and only 49% of the students could answer Item 11 correctly. The slope parameters of the polytomous items tended to be lower than the dichotomous items, because the overall discriminating power of these polytomous items was shared by the dispersion of threshold parameters as well as the slope parameters.

For the polytomous item response model, the parameter values must be interpreted with the aid of the graphical presentation of the ICRFs. Figure 3a shows the ICRFs of Item 13, and Figure 3b

Table 4
Estimated Slope, Location, and Threshold (T) Parameters and Their SEs for the NAEP Math Data

	Slope		Location		Threshold							
					1		2		3		4	
Item	Est.	SE	Est.	SE	T	SE	T	SE	T	SE	T	SE
1	.78	.05	−1.71	.098	0.0	0.0						
2	1.48	.06	−.61	.032	0.0	0.0						
3	1.03	.05	−1.57	.073	0.0	0.0						
4	.95	.04	−.27	.041	0.0	0.0						
5	1.99	.07	−.30	.022	0.0	0.0						
6	.84	.07	−2.85	.203	0.0	0.0						
7	.63	.02	−.75	.029	−.28	.09	−2.37	.13	2.65	.12		
8	1.21	.05	.11	.032	0.0	0.0						
9	1.25	.05	−.28	.032	0.0	0.0						
10	.58	.02	.60	.030	2.29	.08	−4.47	.20	3.29	.21	−1.11	.12
11	.70	.05	1.38	.095	0.0	0.0						
12	1.49	.06	.43	.029	0.0	0.0						
13	.78	.02	.06	.028	−1.70	.08	1.70	.08				
14	.95	.03	.60	.025	−3.37	.16	3.37	.16				
15	1.83	.07	.24	.024	0.0	0.0						
16	1.69	.06	.07	.025	0.0	0.0						

shows ICRFs for Item 14. For Item 13, the proportions of categorical responses from 1 to 3 were .45, .10, and .46, respectively. For Item 14, the proportions were .64, .01, and .35, respectively. Item 14 ($\hat{b}_{14}$ = .603) was more difficult than Item 13 ($\hat{b}_{13}$ = .058). Therefore, the ICRFs of Item 14 are shifted to the right compared to those of Item 13. Because Item 14 discriminated more highly between the first categorical response and the third categorical response, compared to Item 13, its item discriminating power was higher ($\hat{a}_{14}$ = .946) than that of Item 13 ($\hat{a}_{13}$ = .778). Both figures show that the ICRFs of the first and third categorical responses dominate over the middle category. The ICRF of the middle category in Item 14 is flatter than that of Item 13 because fewer students responded to the middle category of Item 14 than Item 13.

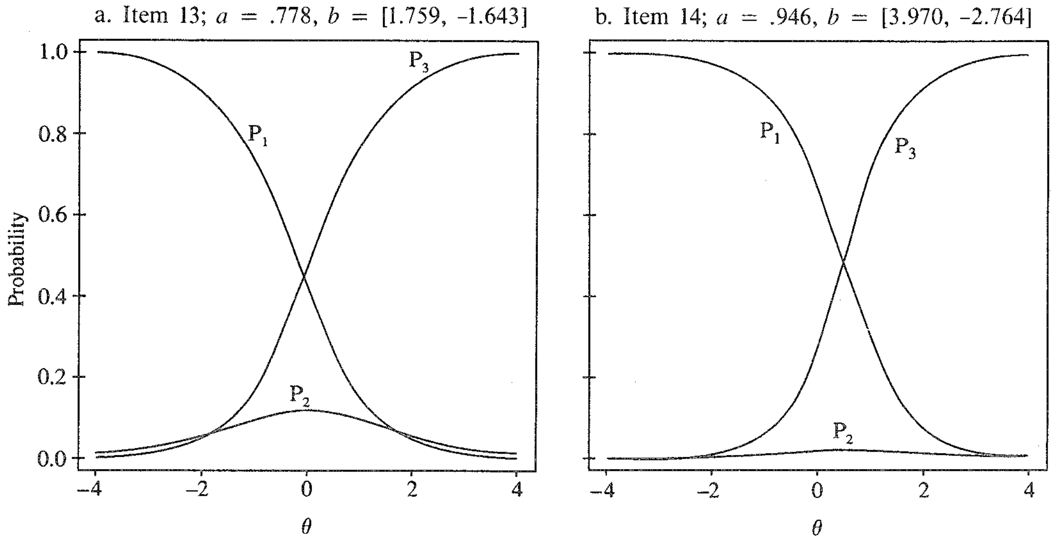
Conclusions

This study has demonstrated that the rating formulation of the PCM is quite flexible for analyzing polytomous item responses. The Rasch family of polytomous item response models was found to be inappropriate if the response data contain varying slope parameters. For this type of data, the marginal maximum likelihood estimation method with the EM algorithm can recover the slope parameters, and fitting the GPCM can improve the model fit.

It was assumed that the NAEP data were unidimensional. Because the slope parameters can be estimated without any constraints, the GPCM can be extended to the multidimensional model in the same way that Bock, Gibbons, and Muraki (1988) developed the multidimensional item response model based on the dichotomous model. The EM algorithm for the full-information factor analysis model for polytomous item responses was derived by Muraki (1985).

Polytomous item response data are often analyzed by assigning numeric scores to the response categories, based on the assumption that the observed categorical responses are quantitative and continuous. However, the actual intervals between adjacent categories are generally unknown in advance. Recently, the demand for the analysis of polytomous item responses has increased. The polytomous item response model can facilitate this type of analysis and create further applications. Investigation

Figure 3
PCM ICRFs for Three-Category NAEP Mathematics Items



has begun only recently on polytomous item response models. Some of the knowledge acquired through research about the dichotomous item response models can be applied directly to the polytomous item response models, but the basic properties of the model parameters also need to be studied.

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.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and 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 factor analysis. *Applied Psychological Measurement, 12,* 261–280.

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.

Johnson, E. G., & Allen, N. L. (1992). *The NAEP 1990 technical report* (No. 21-TR-20). Princeton NJ: Educational Testing Service, National Assessment of Educational Progress.

Kendall, M. G., & Stuart, A. (1973). *The advanced theory of statistics* (Vol. 2). New York: Hafner.

Lord, F. M. (1980). *Application of item response theory to practical testing problems.* Hillsdale NJ: Erlbaum.

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

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

Muraki, E. (1985, March). *Full-information factor analysis for polytomous item response.* Paper presented at the annual meeting of the American Educational Research Association, Chicago IL.

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

Muraki, E. (1990b). *RESGEN: Item response generator.* Princeton NJ: Educational Testing Service.

Muraki, E., & Bock, R. D. (1991). *PARSCALE:*

Parameter scaling of rating data. Chicago IL: Scientific Software, Inc.

Muraki, E., & Wang, M. (1992, April). *Issues relating to the marginal maximum likelihood estimation of the partial credit model.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco CA.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Denmarks Paedagogiske Institut (Reprinted by University of Chicago Press, 1980).

Samejima, F. (1969). Estimation of latent trait 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,* No. 18.

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

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 51,* 567–577.

Wright, B. D., & Stone, M. H. (1979). *Best test design.* Chicago: MESA.

Acknowledgments

The author thanks Minhwei Wang at ETS and Raymond Baillargeon for assistance with data analyses. Valuable comments from other ETS colleagues and journal reviewers are also appreciated.

Author's Address

Send requests for reprints or further information to Eiji Muraki, Educational Testing Service, Princeton NJ 08541, U.S.A.