RR-92-6

RESEARCH REPORT

A GENERALIZED PARTIAL CREDIT MODEL: APPLICATION OF AN EM ALGORITHM

Eiji Muraki

Educational Testing Service
Princeton, New Jersey
January 1992

January 6, 1992

A Generalized Partial Credit Model: Application of an EM Algorithm

by

Eiji Muraki

Educational Testing Service

Eiji Muraki

Educational Testing Service
Princeton, New Jersey 08541
(609) 734-1644

ABSTRACT

The Partial Credit model with a varying slope parameter has been developed, and it is called the Generalized Partial Credit model. The item step parameter of this model is decomposed to a location and a threshold parameter, following Andrich's Rating Scale formulation. The EM algorithm for estimating the model parameters was derived. The performance of this generalized model is compared with a Rasch family of polytomous item response models based on both simulated and real data.

Simulated data were generated and then analyzed by the various polytomous item response models. The results obtained demonstrate that the rating formulation of the Generalized Partial Credit model is quite adaptable to the analysis of polytomous item responses. The real data used in this study consisted of NAEP Mathematics data which was made up of both dichotomous and polytomous item types. The Partial Credit model was applied to this data using both constant and varying slope parameters. The Generalized Partial Credit model, which provides for varying slope parameters, yielded better fit to data than the Partial Credit model without such a provision.

Index terms: item response model
polytomous item response model
the Partial Credit model
the Rating Scale model
the Nominal Response model
NAEP

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, the Graded Response model (Samejima, 1969) and its rating scale version (Muraki, 1990) or the Partial Credit model (Masters, 1982) and its rating scale version (Andrich, 1978) are applicable. For a test item in which the response options are not necessarily ordered, Bock (1972) proposed the Nominal Response model. 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 happens to be two.

Partial Credit Model

A Rasch Family of Polytomous Item Response Models

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

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

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 is usually called an item difficulty. If the discrimination power a is assumed to vary among test items ($a=a_j$), then the model in Equation 1 is called Birnbaum's two-parameter logistic model. If the discrimination power a is assumed to be common for all items, the parameter a can be dropped from the model by arbitrarily setting a=1, and the 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 of the model invite a specialized parameter estimation - the conditional maximum likelihood estimation. If the model is viewed as the 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 the latent trait model with latent trait variable b, then the conditional likelihood of (θ_i), given the scores of item j (s_j), is independent of b. Therefore, the parameters (θ_j) 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 that all items have equal discriminating power and vary only in terms of difficulty is met, then the Rasch model has been shown to provide an elegant and simple solution for several technical applications for test analysis and construction (Wright and Stone, 1979).

The notable distinction between the Rasch polytomous item response models and a Thurstone family of polytomous item response models, such as the Graded Response model, is not the number of parameters but the difference in terms of the operating characteristic function (Samejima, 1972). The operating characteristic function 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. As Masters (1982) formulated his Partial Credit model by utilizing the Rasch dichotomous model, it is quite legitimate to construct the Partial Credit model based on the two-parameter logistic response model, following the same operating characteristic function Masters (1982) employs. Since the essential mechanism for constructing a general model is shared with Masters' Partial Credit model, the model constructed in this paper can be simply called the Generalized Partial Credit model.

The Generalized Partial Credit model is formulated based on the assumption that the probability of choosing the k-th category over the k-1st category is governed by the dichotomous response model. To develop the Partial Credit model, let us denote $P_{jk}(\theta)$ as the specific probability of choosing the k-th 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:

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

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

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

Note that $C_{jk}/(1-C_{jk})$ is the ratio of the two conditional probabilities, which may also be expressed as $\exp[a_j(\theta-b_{jk})]$. Equation 3 may be called the operating characteristic function for the Partial Credit model. If we start by determining

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

where G is called a normalizing factor which is defined below, we obtain the following probabilities by applying the operating characteristic function in Equation 3:

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

$$\cdots\cdots$$

$$P_{jg}(\theta) = \frac{\exp[\sum_{v=2}^{g} a_j(\theta-b_{jv})]}{G} \tag{5}$$

$$\cdots\cdots$$

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

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

$$\sum_{k=1}^{m_j} P_{jk}(\theta) = 1 \tag{6}$$

$$G = 1+\sum_{c=2}^{m_j} \exp[\sum_{v=2}^{c} a_j(\theta-b_{jv})] \tag{7}$$

Therefore, the Partial Credit model is formulated by

$$P_{jk}(\theta) = \frac{\exp[\sum_{v=1}^{k} a_j(\theta - b_{jv})]}{\sum_{c=1}^{m_j} \exp[\sum_{v=1}^{c} a_j(\theta - b_{jv})]} \tag{8}$$

where $b_{j1} \equiv 0$.

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

$$\begin{aligned} P_{jk}(\theta) &= \frac{\exp[Z_{j1}(\theta)] \cdot \exp[\sum_{v=2}^{k} Z_{jv}(\theta)]}{\exp[Z_{j1}(\theta)] + \sum_{c=2}^{m_j} \exp[Z_{j1}(\theta) + \sum_{v=2}^{c} Z_{jv}(\theta)]} \\ &= \frac{\exp[\sum_{v=2}^{k} Z_{jv}(\theta)]}{1 + \sum_{c=2}^{m_j} \exp[\sum_{v=2}^{c} Z_{jv}(\theta)]} \end{aligned} \tag{9}$$

where $Z_{jk}(\theta) = a_j(\theta - b_{jk})$. The Partial Credit model in Equation 8 reduces to the dichotomous item response model when $m_j=2$ and $k=1, 2$.

Masters (1982) calls the parameters b_{jk} in Equation 8 item step parameters. The b_{jk} are the points in 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 Characteristic Curves (ICCCs), intersect only once, and the intersection can occur anywhere along the θ scale. Thus,

$$\begin{aligned} &\textit{if} \quad \theta = b_{jk}, \quad P_{jk}(\theta) = P_{j,k-1}(\theta) \\ &\quad\quad\; \theta > b_{jk}, \quad P_{jk}(\theta) > P_{j,k-1}(\theta) \\ &\textit{and} \quad \theta < b_{jk}, \quad P_{jk}(\theta) < P_{j,k-1}(\theta) \end{aligned} \tag{10}$$

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 ICCCs of the Partial Credit model 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

$$\frac{\partial}{\partial\theta}P_{jk}(\theta) = a_j P_{jk}(\theta)\,[k-\sum_{c=1}^{m_j} cP_{jc}(\theta)] \tag{11}$$

Setting the first derivative in this equation to zero, we can see that

$$\sum_{c=1}^{m_j} c\; P_{jc}(\theta) = k \tag{12}$$

Equation 12 shows that the peak of ICCC, $P_{jk}(\theta)$, is affected by all of the other probabilities, $P_{jc}(\theta)$, c=1, 2, ..., k-1, k+1, ..., 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 ∞. For the dichotomous item response model, the discriminating power is a function of only the item discriminating power. However, for the polytomous item response model, the discriminating power of each ICCC depends on the combination of the slope and threshold parameters. Andrich (1978) distinguished these two kinds of a discriminating power and retained only the threshold discrimination in his Rasch family of the rating scale models. In our Partial Credit model, only the item discriminating power is included. This slope parameter indicates the degree to which categorical response varies among items as ability level changes. The concept of the 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, we can keep the continuity of Birnbaum's two parameter model from the dichotomous to the polytomous response case as well as the connection with the classical test concept. The model with this slope parameter can be also extended to the multidimensional form (Muraki, 1985).

Figures 1a, 1b, and 1c show the ICCCs for the Partial Credit model with four categorical responses. Figure 1a shows the ICCCs 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 -0.5, then the probability of responding to the second category decreases, as illustrated in Figure 1b. In other words, the range of the theta values of persons who are more likely to respond to the second category than the other categories decreases from (-2., 0.) to (-

0.5, 0.). If the slope parameter is changed from 1.0 to 0.7, as shown in Figure 1c, the intersection points of all trace lines are left unchanged and the curves become flatter. The item discriminating power of these ICCCs decreases for all categorical responses.

The ICCC in Figure 2a are for the Partial Credit model with three categorical responses. When the second item step parameter is made larger than the third item step parameter ($b_{j2}>b_{j3}$), the ICCC of P_{j2} drops, as shown in Figure 2a. For the whole range of theta values, the probability of the first or third categorical response is higher than the probability of the second categorical response. If you trace each of the ICCCs, you find that every person who is more likely to respond to the second category than 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. The interpretation of parameters in the polytomous item response models can be greatly facilitated by tracing the ICCCs along the θ axis.

If all item step parameters have the same value, as shown in Figure 2b, all ICCCs intersect at the same value of θ. Even though the values of item step parameters are not sequentially ordered, the Partial Credit model expresses the probabilities of ordered responses.

****

Insert Figures 1a, 1b, 1c, 2a, and 2b here

****

The Rating Scale model is derived from the Partial Credit model by assuming b_{jk} can be additively decomposed as $b_{jk}=b_j-d_k$,

$$P_{jk}(\theta) = \frac{\exp[\sum_{v=1}^{k} a_j(\theta-b_j+d_v)]}{\sum_{c=1}^{m}\exp[\sum_{v=1}^{c} a_j(\theta-b_j+d_v)]} \tag{13}$$

where $d_1=0$ and the parameter b_{jk} is resolved into two parameters b_j and d_k ($b_{jk}=b_j-d_k$). Historically, Masters (1982) modified Andrich's model (Andrich, 1978) in Equation 13 and called it the Partial Credit model. In their models, the slope parameter is assumed as a constant. For a Thurstone family of polytomous item response models, Samejima's model (Samejima, 1972) is extended to the Rating scale model by Muraki (1990). Both Andrich and Muraki separated the item category threshold parameter into an item parameter and a category parameter in the same manner. Since the Rating Scale model in Equation 13 is essentially identical with the Partial Credit model when single items are considered, this model is simply called a rating formulation of the Partial Credit model. The rating version of the Partial Credit model can be applied to any situations where the Partial Credit model 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, we also call the model in Equation 13 the Generalized Partial Credit model unless we want to specifically emphasize its rating aspect.

Andrich (1982) calls the parameters b_j and d_k in Equation 13 a item location parameter and a threshold parameter, respectively. Since 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,...,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 (U_{jk}), is the joint probability:

$$P_i((U_{jk})|\theta) = \prod_{j=1}^{n}\prod_{k=1}^{m}[P_{jk}(\theta)]^{U_{jki}} \qquad \textbf{(14)}$$

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

$$P_i(U_{jk}) = \int_{-\infty}^{\infty} P_i[(U_{jk})|\theta]\,\phi(\theta)\,d\theta \qquad \textbf{(15)}$$

If an examinee responds to n items with m categories, his/her 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 $P_i(U_{jk})$. This probability can be interpreted as the likelihood function of the parameters, a_j, b_j, and d_k:

$$L = \frac{N!}{\prod_{i=1}^{m^n} r!} \prod_{i=1}^{m^n} [P_i(U_{jk})]^{r_i} \tag{16}$$

Taking the natural logarithm of Equation 16 yields

$$lnL = lnN! - ln\sum_{i=1}^{m^n} r_i! + \sum_{i=1}^{m^n} r_i lnP_i(U_{jk}) \tag{17}$$

The likelihood equation for $\hat{a}_j$, $\hat{b}_j$, and $\hat{d}_k$ can be derived from the first partial derivative of Equation 17 with respect to each parameter, and setting them to 0.

Item Parameter Estimation

Let the two element vector υ_j represent the parameters a_j and b_j. With respect to υ_h, which is the parameter υ_j for the specific item j=h, the likelihood in Equation 17 can be differentiated as

$$\frac{\partial lnL}{\partial \upsilon_h} = \sum_{i=1}^{m^n} \frac{r_i}{P_i(U_{jk})} \int_{-\infty}^{\infty} P_i[(U_{jk})|\theta] \sum_{k=1}^{m} \frac{\partial [P_{hk}(\theta)]^{U_{hkl}}}{\partial \upsilon_h} \frac{\phi(\theta)\,d\theta}{[P_{hk}(\theta)]^{U_{hkl}}} \tag{18}$$

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

$$\sum_{l=1}^{S} r_l = N \tag{19}$$

The first derivative of the likelihood function in Equation 18 can be approximated by using the Gauss-Hermite quadrature, such that

$$\frac{\partial lnL}{\partial \upsilon_h} \approx \sum_{l=1}^{S} \sum_{f=1}^{F} \frac{r_l L_l(X_f) A(X_f)}{\tilde{P}_l} \sum_{k=1}^{m} \frac{\partial [P_{hk}(X_f)]^{U_{hkl}}}{\partial \upsilon_h} \frac{1}{[P_{hk}(X_f)]^{U_{hkl}}} \tag{20}$$

where

$$\tilde{P}_l = \sum_{f=1}^{F} L_l(X_f)A(X_f) \tag{21}$$

and

$$L_l(X_f) = \prod_{j=1}^{n}\prod_{k=1}^{m}[P_{jk}(X_f)]^{U_{jkl}} \tag{22}$$

In Equation 20, $A(X_f)$ 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

$$\sum_{f=1}^{F} A(X_f) = 1 \tag{23}$$

where F is the total number of quadrature points. Because U_{hkl} can take only two possible values, 1 or 0, Equation 20 can be rewritten as

$$\sum_{f=1}^{F}\sum_{k=1}^{m}\frac{\bar{r}_{hkf}}{P_{hk}(X_f)}\frac{\partial P_{hk}(X_f)}{\partial \upsilon_h} \tag{24}$$

where

$$\bar{r}_{hkf} = \sum_{l=1}^{S}\frac{r_l L_l(X_f)A(X_f)U_{hkl}}{\tilde{P}_l} \tag{25}$$

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, and Rubin, 1977) to estimate the parameters for each item individually, and then repeated 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

$$\boldsymbol{\upsilon}_q = \boldsymbol{\upsilon}_{q-1} + \boldsymbol{V}^{-1}\boldsymbol{t} \tag{26}$$

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

$$t_{\upsilon_h} = \sum_{f=1}^{F} \sum_{k=1}^{m} \frac{\overline{r}_{hkf}}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial \upsilon_h} \tag{27}$$

$$V_{\upsilon_h \omega_h} = \sum_{f=1}^{F} \overline{N}_f \sum_{k=1}^{m} \frac{1}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial \upsilon_h} \frac{\partial P_{hk}(X_f)}{\partial \omega_h} \tag{28}$$

where $\upsilon_h = a_h$ or b_h and $\omega_h = a_h$ or b_h. In Equation 28, $\bar{N}_f$ is called the provisional expected sample size at quadrature point, f, and is computed by

$$\overline{N}_f = \sum_{l=1}^{S} \frac{r_l L_l(X_f) A(X_f)}{\tilde{P}_l} \tag{29}$$

A rigorous proof of the approximation of the second derivatives in Equation 28 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

$$t_{a_h} = a_h^{-1} \sum_{f=1}^{F} \sum_{k=1}^{m} \overline{r}_{hkf} [Z_{hk}^{+}(X_f) - \sum_{c=1}^{m} Z_{hc}^{+}(X_f) P_{hc}(X_f)] \tag{30}$$

$$t_{b_h} = a_h \sum_{f=1}^{F} \sum_{k=1}^{m} \overline{r}_{hkf} [-k + \sum_{c=1}^{m} c P_{hc}(X_f)] \tag{31}$$

$$V_{a_h a_h} = a_h^{-2} \sum_{f=1}^{F} \overline{N}_f \sum_{k=1}^{m} P_{hk}(X_f) \, [Z_{hk}^{+}(X_f) - \sum_{c=1}^{m} Z_{hc}^{+}(X_f) P_{hc}(X_f)]^2 \tag{32}$$

$$V_{b_h b_h} = a_h^2 \sum_{f=1}^{F} \overline{N}_f \sum_{k=1}^{m} P_{hk}(X_f)\,[-k+\sum_{c=1}^{m} cP_{hc}(X_f)]^2 \tag{33}$$

and

$$V_{a_h b_h} = \sum_{f=1}^{F} \overline{N}_f \sum_{k=1}^{m} P_{hk}(X_f)\,[Z_{hk}^{+}(X_f) - \sum_{c=1}^{m} Z_{hc}^{+}(X_f)\,P_{hc}(X_f)]\,[-k+\sum_{c=1}^{m} cP_{hc}(X_f)] \tag{34}$$

where

$$\begin{aligned} Z_{hk}^{+}(X_f) &= \sum_{v=1}^{k} Z_{hv}(X_f) \\ &= \sum_{v=1}^{k} a_h(X_f - b_h + d_v) \end{aligned} \tag{35}$$

Threshold Parameter Estimation

Since 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, ..., m) as shown in Equation 13, the first derivative of the likelihood function in Equation 22 with respect to d_g is given by

$$\frac{\partial L_l(X_f)}{\partial d_g} = L_l(X_f) \sum_{j=1}^{n} \sum_{k=1}^{m} \frac{U_{jkl}}{P_{jk}(X_f)} \frac{\partial P_{jk}(X_f)}{\partial d_g} \tag{36}$$

According to the EM algorithm and Equation 36, the maximum likelihood function in Equation 17 with respect to d_g is written as

$$\begin{aligned} t_{d_g} &= \sum_{f=1}^{F} \sum_{j=1}^{n} \sum_{k=1}^{m} \frac{\overline{r}_{jkf}}{P_{jk}(X_f)} \frac{\partial P_{jk}(X_f)}{\partial d_g} \\ &= \sum_{f=1}^{F} \sum_{j=1}^{n} a_j \sum_{k=g}^{m} [\overline{r}_{jkf} - P_{jk}(X_f) \sum_{c=1}^{m} \overline{r}_{jcf}] \end{aligned} \tag{37}$$

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

$$V_{d_g d_{g'}} = \sum_{f=1}^{F} \overline{N}_f \sum_{j=1}^{n} \sum_{k=1}^{m} \frac{1}{P_{jk}(X_f)} \frac{\partial P_{jk}(X_f)}{\partial d_g} \frac{\partial P_{jk}(X_f)}{\partial d_{g'}}$$

$$= \sum_{f=1}^{F} \overline{N}_f \sum_{j=1}^{n} a_j^2 [\sum_{k=g}^{m} P_{jk}(X_f)] [1 - \sum_{k=g'}^{m} P_{jk}(X_f)] \tag{38}$$

Since 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)x(m-1) respectively.

Comparison with the Nominal Response Model

Let us rewrite the Partial Credit model by using $Z_{jk}^{+}(\theta)$ as defined in Equation 35, that is,

$$P_{jk}(\theta) = \frac{\exp[Z_{jk}^{+}(\theta)]}{\sum_{c=1}^{m} \exp[Z_{jc}^{+}(\theta)]} \tag{39}$$

Equation 39 is exactly the form of the Nominal Response model proposed by Bock (1972). His original formulation of the Nominal Response model is

$$P_{jk}^{*}(\theta) = \frac{\exp[Z_{jk}^{*}(\theta)]}{\sum_{c=1}^{m} \exp[Z_{jc}^{*}(\theta)]} \tag{40}$$

where

$$Z_{jc}^{*}(\theta)$$

$$= a_{jc}^{*}\theta + d_{jc}^{*} \tag{41}$$

$$= a_{jc}^{*}(\theta - b_{jc}^{*})$$

Therefore, the Nominal Response model becomes equivalent to the Partial Credit model if the following conditions are satisfied:

$$a_{jk}^{*} = k a_j \tag{42}$$

and

$$b_{jk}^{*} = \frac{\sum_{v=1}^{k} b_{jv}}{k} \tag{43}$$
$$= b_j - \frac{\sum_{v=1}^{k} d_v}{k}$$

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 one with lower θ. Then, Wainer (1991) constructed the following inequality:

$$\frac{P_{j,k+1}^{*}(\theta+\delta_\theta)}{P_{jk}^{*}(\theta+\delta_\theta)} > \frac{P_{j,k+1}^{*}(\theta)}{P_{jk}^{*}(\theta)} \tag{44}$$

where $\delta_\theta>0$. Inequality 44 can be rewritten as

$$\frac{P_{j,k+1}^{*}(\theta+\delta_\theta)\,P_{jk}^{*}(\theta)}{P_{jk}^{*}(\theta+\delta_\theta)\,P_{j,k+1}^{*}(\theta)} > 1 \tag{45}$$

Taking the natural logarithm of Inequality 45 yields

$$\ln P_{j,k+1}^{*}(\theta+\delta_\theta) + \ln P_{jk}^{*}(\theta) - \ln P_{jk}^{*}(\theta+\delta_\theta) - \ln P_{j,k+1}^{*}(\theta) > 0 \tag{46}$$

Substituting each $P_{jk}^{*}(\theta)$ with the Nominal Response model in Equation 40, we obtain

$$Z_{j,k+1}^{*}(\theta+\delta_\theta) + Z_{jk}^{*}(\theta) - Z_{jk}^{*}(\theta+\delta_\theta) - Z_{j,k+1}^{*}(\theta) \tag{47}$$
$$= a_{j,k+1}^{*}\delta_\theta - a_{jk}^{*}\delta_\theta > 0$$

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

$$a_{j,k+1}^{*} > a_{jk}^{*} \tag{48}$$

The increment of scaling factors along the consecutive categories, shown in Equation 42 as well as Equation 48 is a hidden feature of the Partial Credit model. Because of this nature, the Partial Credit model becomes the model for ordered response categories. The Partial Credit model is a special case of the Nominal response model. In the Partial Credit model, the degree of the expansion of scaling factors is expressed by Equation 42, that is, a_j, $2a_j$, $3a_j$ and so on. Andrich (1978) called this feature the linear scoring function.

The nominator of the Partial Credit model can be rewritten as

$$Z_{jk}^{+}(\theta) = a_j [k(\theta - b_j) + \sum_{v=1}^{k} d_v] \tag{49}$$

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

$$Z_{jk}^{+}(\theta) = a_j [T_k(\theta - b_j) + K_k] \tag{50}$$

where T_k is the scoring function and K_k is the category coefficient. Notice that Andrich's Rating scale model becomes the Nominal response model if we treat the scoring function, T_k, as an estimatable quantity from response data. In the Partial Credit model, T_k is set a priori as a series of sequential integers, which is shown in Equation 42. Andrich (1988) further extended his Rating model by reparameterizing the category coefficient so that the model incorporated binomial and Poisson response processes as well as linear and quadratic coefficients.

We have observed that the Partial Credit model is a special case of Bock's Nominal Response model. Andrich's scoring function is a key concept to understanding these model formulations. Andrich (1978, 1982, 1988) also demonstrated that the Partial Credit model can be further extended by reparameterizing the scoring function and the category coefficient. By utilizing these features, we may be able to construct the model expressing the partial order of the categorical responses or any specific response processes. 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 completely non-ordered and partially ordered. Their approach to the polytomous item response model utilizing the contrast is quite useful for the further developments in this particular field.

Constraints on the Threshold Parameters

Integrating out a nuisance variable, θ, from the likelihood with a fixed prior, as shown in Equation 15, 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 zero.

A block of items is defined here as a set of items which 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, we must impose a constraint on the estimation of threshold parameters. This constraint is called a location constraint. A location constraint is imposed so that the mean of threshold parameters within a categorical scale is constant over blocks. A natural choice is zero, that is,

$$\sum_{k=2}^{m_j} d_k = 0 \tag{51}$$

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 so 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 categorical trace line may have 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, we control the scaling factor due to the threshold parameters and extract the item discriminating power for each item. 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 to the analyses of polytomous item responses. The two-parameter dichotomous item response model becomes a special case of the Generalized Partial Credit model. More importantly, the model can be tested in a step-wise manner. If more than a single item is involved with a given block, we may estimate a common set of threshold parameters and slope and location parameters for each item. Then, we fit the model to each item again and obtain a separate set of threshold parameters for each item and test the assumption about the common threshold parameters for all items included in a block. If we can establish that a common set of threshold parameters for items in a block is reasonably fitted to polytomously scored response data, all methodologies based on the dichotomous item response models can be applicable without great difficulty. If we need the item step parameters, we fit the model in Equation 13 by imposing the location constraint on the threshold parameters. Then, the item step parameters, b_{jk}, can be simply computed by

$$b_{jk} = b_j - d_{jk} \tag{52}$$

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

The MML-EM Algorithm

The EM algorithm presented here is available in the PARSCALE computer program (Muraki and Bock, 1991). The PARSCALE program can estimate the parameters of the Graded Response model as well. Data analyses by this model were presented by Muraki (1990).

The MML-EM algorithm implemented in the PARSCALE program consists of two steps. The first is the expectation step (the E-step) where the provisional expected frequency and the provisional expected sample size are computed by using Equations 25 and 29, respectively. Then, in the maximization step (the M-step), the MML estimates are obtained by Equation 26. 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.

The PARSCALE program provides a likelihood ratio chi-square 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 a model fit is computed by summing the item fit statistics over items.

Data and Results

Simulated Data

Five thousand Likert-type response vectors for thirty items and three categories were generated with a standard normal distribution of θ. The original parameter values were obtained from the analysis of the knowledge-of-physics data, estimated by Masters (1982). The slope parameters of all items are 1.0. Masters' item step parameters, b_{jk}, vary, and, for the sake of brevity, the parameter values of every third item are presented in Table 1. The RESGEN computer program (Muraki, 1990) was used to generate the simulated data set. These data were then analyzed four times under various constraints. Ten quadrature points were used and the precision level 0.0001 was set for all estimations.

****

Insert Table 1 here

****

In Analysis 1, the simulated data were analyzed based on Masters' Partial Credit model. Slope parameters were kept constant during the estimation process. The location constraint was not applied to the estimation. Analysis 2 is identical with Analysis 1 except for the application of the location constraint. 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. Again, only the estimated values of every third item are presented in Table 1. As shown in Table 1, the EM algorithm successfully recovered original parameter values. Although the location estimates, b_j, are different between the results of Analyses 1 and 2, as shown in Table 2, the estimated values of b_{jk}, which were computed from b_j and d_{jk}, of Analysis 1 are indistinguishable from the estimated values of Analysis 2. Their -2 log likelihood statistics are almost identical. In other words, these two models are essentially the same. However, Analysis 2 needed fewer iterations to reach the convergence criterion because we eliminated the indeterminacy. In addition, the location estimates of Analysis 2 can be compared with each other because of its location constraint.

****

Insert Table 2 here

****

In Analysis 3, the thirty items were made into a block and only one set of threshold parameters was estimated. The slope parameters were again kept as constant. No location constraint was applied to this estimation. The -2 log likelihood was considerably higher than the solution of Analysis 1 or 2. In other words, the assumption about a common set of threshold parameters for all thirty items is not appropriate for these data, as would be expected based on the model used to specify the original set of parameter values. It should be also pointed out that the mean of b_j is zero in Analysis 3. The mean of the set of location estimates is completely absorbed into the threshold parameters. The threshold estimates are -0.257 and 0.249. When the data were analyzed by imposing the location constraint, the results showed that the threshold parameter estimates were shifted left by 0.004 and the mean of b_j was shifted in the opposite direction by the same amount. The location constraint again shortened the number of iterations.

In Analysis 4, we estimated the slope parameters. Since we set a block of thirty items 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, we generated another simulated data set. In this simulated data set, we used various slope parameter values. The original slope parameters were incremented from 0.3, 0.6, 0.9, 1.2, 1.5, and 1.8 for the set of six items. This set of slope parameters was, then, repeatedly applied to the remaining items. The Partial Credit model was fitted with a constant slope. The -2 log likelihood was 264366 as shown in Analysis 5 of Table 2. We also fitted the Generalized Partial Credit model and estimated the slope parameters (Analysis 6). The -2 log likelihood decreased to 256466. The difference is 7900, and the degree of freedom of this difference is 30. The difference of the model fit statistics is 7480, and the degree of freedom is 5. Thus, the model fit was significantly improved by applying the Generalized Partial Credit model.

The first twelve parameter estimates are presented in Table 3. It seemed that all slope parameters were underestimated, and consequently all threshold parameters were overestimated. Then, we repeatedly estimated the parameters of the first six items with various numbers of quadrature points. Our initial investigation suggests that parameters are not necessarily underestimated and the estimation bias seems to decrease if the number of quadrature points increases. A reasonable number of quadrature points may be determined by the number of items in a test and their number of response categories. This estimation problem should be studied further.

****
Insert Table 3 here
****

NAEP Mathematics Data

The National Assessment of Educational Progress (NAEP) is an ongoing survey designed to measure what students know and can do. In this section, NAEP Mathematics data of year 1989-90 was analyzed. The particular section for the assessment for Grade 8 contains 16 items. Twelve items were dichotomously scored, and the other four items were polytomous items of which the number of categories varies from three to six. The number of categorical responses of each item are the number of threshold parameters, which are shown in Table 4, plus 1. Item 10 originally had six categories, but no student responded to the fifth category. Therefore, we treated the item as a five-category item. The total number of students was 3699. Twenty students were excluded from the analysis since they omitted all items. Thus, the item responses of 3679 students were used for the analysis. Thirty three quadrature points and a convergence criterion 0.001 were used for the estimation.

****
Insert Table 4 here
****

The Partial Credit model with a constant slope was fitted to the data. The -2 log likelihood was 69688. The fit was significantly improved if we fit the Partial Credit model with varied slope parameters. The -2 log likelihood of this Generalized Partial Credit model was 68748. The difference is 940 and the degree of freedom is 16. The model fit statistics for these models were 1577 and 957, and their degrees of freedom were 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.

We applied the location constraint for each set of threshold parameters. Therefore, the location estimates are comparable. The higher location estimates indicate more difficult items. Item 6 is the easiest item and item 11 is the most difficult. More than 80 percent of the students correctly responded to item 6 while only 49 percent of the students could answer item 11 correctly. The slope parameters of polytomous items tend to be lower than the dichotomous items since 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 aide of the graphical presentation of the ICCCs. Figure 3a shows the ICCCs of item 13, and Figure 3b shows those of item 14. For item 13, the portions of categorical responses from 1 to 3 are 0.45, 0.10, and 0.46, respectively. For item 14, their proportions are 0.64, 0.01, and 0.35. Item 14 (b_j=0.603) is more difficult than item 13 (b_j=0.058). Therefore, the ICCCs of item 14 are shifted to the right compared to the ones of item 13. Since item 14 discriminates more highly between the first categorical response and the third categorical response, compared to item 13, its item discriminating power is higher (a_j=0.946) than that of item 13 (a_j=0.778). Both figures show that the ICCCs of the first and third categorical responses dominate over the middle category. The ICCC 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.

****
Insert Figures 3a and 3b here
****

Conclusion

In this study, we have demonstrated that the rating formulation of the Partial Credit model is quite flexible for analyzing polytomous item responses. We also found that the Rasch family of polytomous item response models is not appropriate 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 Generalized Partial Credit model can improve the model fit.

We assumed that the NAEP data were unidimensional. Since we can estimate the slope parameters without any constraints, we can extend the Generalized Partial Credit model to the multidimensional model in the same way as 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 items responses has increased. The polytomous item response model can facilitate this type of analysis and create further applications. Investigation has only recently begun on polytomous item response models. Some of the knowledge we have acquired through research about the dichotomous item response models can be directly applied to the polytomous item response models, but we also need to study the basic properties of the model parameters.

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. 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., and 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, Series B, 39, 1-38.

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). BILOG 3: Item analysis and test scoring with binary logistic models (2nd edition). Chicago, Ill.: Scientific Software, Inc.

Muraki, E. (1985). Full-information factor analysis for polytomous item response. A paper presented at the annual meeting of the American Educational Research Association, Chicago, Ill.

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

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

Muraki, E., & Bock, R. D. (1991). PARSCALE: Parameter scaling of rating data. Chicago, Ill.: Scientific Software, Inc.

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 & Steinberg (1986). A taxonomy of item response models, Psychometrika, 51, 567-577.

Wainer (1991). Personal Communication.

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

Acknowledgements

The author thanks Minhwei Wang at ETS and Raymond Baillargeon, a graduate student in the ETS summer research program, for their time in data analyses. Valuable comments from ETS research colleague as well as journal reviewers are also appreciated.

Eiji Muraki

Educational Testing Service
Princeton, New Jersey 08541
(609) 734-1644

Table 1

Original and Estimated Item Step Parameters
of the Knowledge-of-Physics Simulated Data

Analysis 1

Item	Item Step Parameters			
	Original		Estimated	
3	0.62	-0.78	0.669	-0.818
6	-0.22	-0.61	-0.177	-0.603
9	1.06	-0.28	1.137	-0.240
12	0.64	-0.37	0.590	-0.373
15	1.24	1.00	1.200	1.046
18	-0.24	-0.81	-0.264	-0.816
21	-0.02	-0.26	0.026	-0.277
24	0.14	0.31	0.151	0.317
27	-0.08	-0.53	-0.066	-0.554
30	0.87	0.78	0.834	0.795

Table 2

Means and Standard Deviations
of Estimated Slope and Location Parameters
and -2 Log Likelihoods of Analyses 1, 2, 3, 4, 5 and 6

Knowledge-of-Physics Simulated Data

Analysis	Block	Location Constraint	Estimated Parameters	Slope	Location	Log Likelihood
1	30	NO	Mean	1.000	-0.004	255826.337
			(S.D.)	(0.000)	(0.042)	
2	30	YES	Mean	1.000	0.002	255826.325
			(S.D.)	(0.000)	(0.582)	
3	1	NO	Mean	1.000	0.000	259762.896
			(S.D.)	(0.000)	(0.569)	
4	1	NO	Mean	0.883	-0.012	258341.066
			(S.D)	(0.155)	(0.658)	
5	30	YES	Mean	1.000	0.083	264365.585
			(S.D.)	(0.000)	(0.489)	
6	30	YES	Mean	0.913	0.015	256466.351
			(S.D.)	(0.434)	(0.644)	

Table 3

Estimated Values of Slope, Location, and Threshold Parameters and Standard Error of Estimates

Analysis 6

Item	Estimated Parameters			
	Slope (S.E.)	Location (S.E.)	Threshold (S.E.)	
1	0.254	-1.348	-0.729	0.729
	(0.015)	(0.103)	(0.155)	(0.134)
2	0.530	0.995	-0.295	0.295
	(0.018)	(0.045)	(0.066)	(0.079)
3	0.746	-0.063	-0.936	0.936
	(0.017)	(0.025)	(0.058)	(0.057)
4	1.065	0.380	-0.401	0.401
	(0.024)	(0.020)	(0.039)	(0.042)
5	1.332	0.288	-0.626	0.626
	(0.027)	(0.016)	(0.037)	(0.039)
6	1.580	-0.455	-0.237	0.237
	(0.035)	(0.016)	(0.032)	(0.029)
7	0.277	0.648	-1.356	1.356
	(0.013)	(0.068)	(0.130)	(0.138)
8	0.569	0.000	-0.308	0.308
	(0.018)	(0.032)	(0.065)	(0.065)
9	0.790	0.520	-0.804	0.804
	(0.019)	(0.025)	(0.053)	(0.057)
10	1.078	0.704	-0.749	0.749
	(0.023)	(0.021)	(0.044)	(0.048)
11	1.294	-0.379	-1.677	1.677
	(0.026)	(0.017)	(0.065)	(0.064)
12	1.475	0.156	-0.626	0.626
	(0.030)	(0.015)	(0.036)	(0.036)

Table 4

Estimated Slope, Location, and Threshold Parameters of the NAEP MATH Data

Item	Slope (S.E.)	Location (S.E.)	Threshold 1	2	3	4
1	0.777	-1.707	0.000			
	(0.046)	(0.098)	(0.000)			
2	1.483	-0.609	0.000			
	(0.058)	(0.032)	(0.000)			
3	1.032	-1.565	0.000			
	(0.053)	(0.073)	(0.000)			
4	0.948	-0.267	0.000			
	(0.044)	(0.041)	(0.000)			
5	1.985	-0.297	0.000			
	(0.068)	(0.022)	(0.000)			
6	0.839	-2.853	0.000			
	(0.067)	(0.203)	(0.000)			
7	0.625	-0.748	-0.278	-2.372	2.651	
	(0.015)	(0.029)	(0.088)	(0.128)	(0.117)	
8	1.211	0.113	0.000			
	(0.049)	(0.032)	(0.000)			
9	1.253	-0.284	0.000			
	(0.051)	(0.032)	(0.000)			
10	0.580	0.601	2.293	-4.470	3.285	-1.109
	(0.015)	(0.030)	(0.084)	(0.198)	(0.207)	(0.123)
11	0.704	1.384	0.000			
	(0.047)	(0.095)	(0.000)			
12	1.485	0.433	0.000			
	(0.056)	(0.029)	(0.000)			
13	0.778	0.058	-1.701	1.701		
	(0.020)	(0.028)	(0.081)	(0.081)		
14	0.946	0.603	-3.367	3.367		
	(0.026)	(0.025)	(0.157)	(0.158)		
15	1.825	0.242	0.000			
	(0.065)	(0.024)	(0.000)			
16	1.685	0.066	0.000			
	(0.062)	(0.025)	(0.000)			

Figure 1a
ICCC's for a Four–Category Item
Partial Credit Model with a = 1.0 and b = [–2.0, 0.0, 2.0]

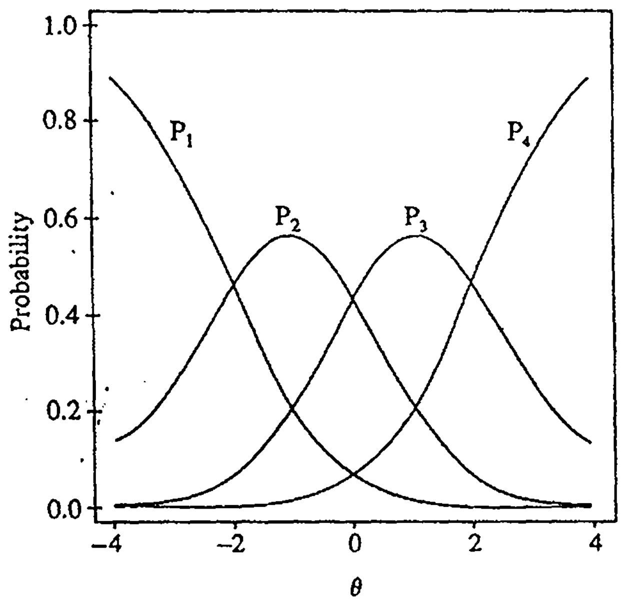

Figure 1b
ICCC's for a Four–Category Item
Partial Credit Model with a = 1.0 and b = [–0.5, 0.0, 2.0]

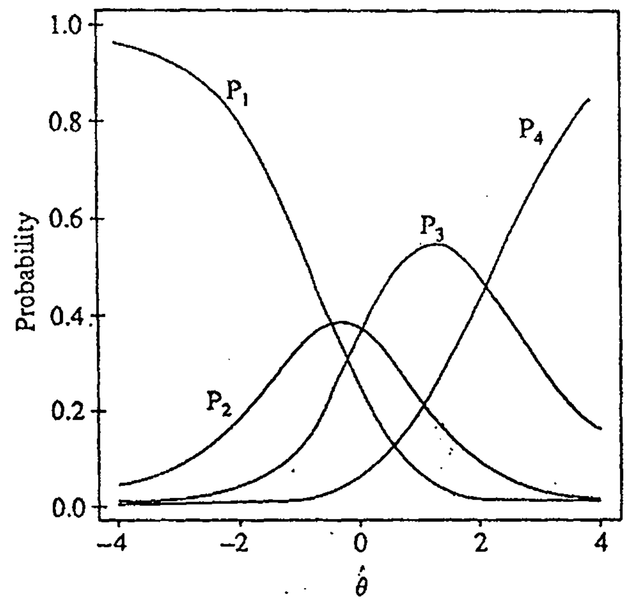

Figure 1c
ICCC's for a Four–Category Item
Partial Credit Model with a = 0.7 and b = [–0.5, 0.0, 2.0]

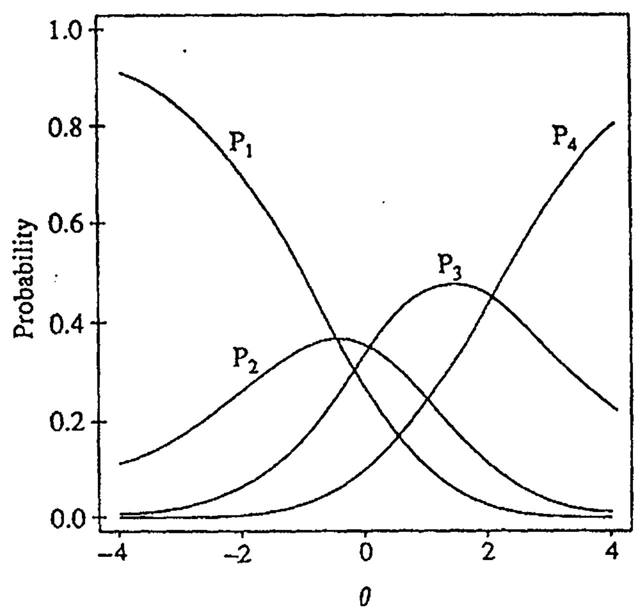

Figure 2a
ICCC's for a Three–Category Item
Partial Credit Model with a = 0.7 and b = [0.5, 0.0]

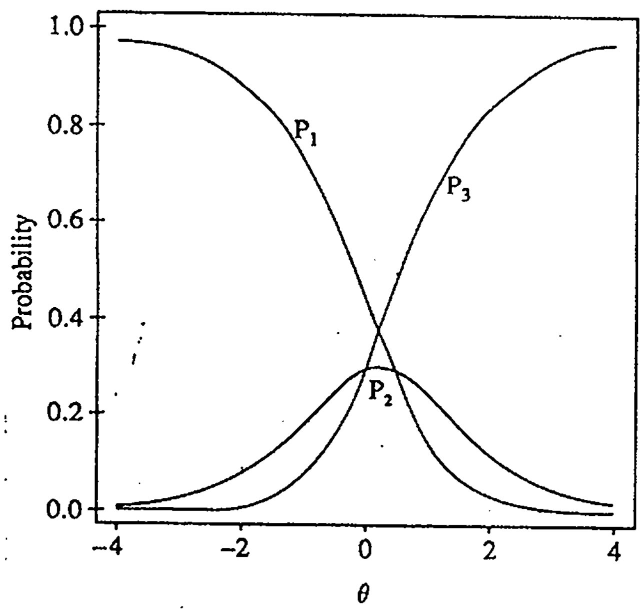

Figure 2b
ICCC's for a Four–Category Item
Partial Credit Model with a = 0.7 and b = [0.0, 0.0, 0.0]

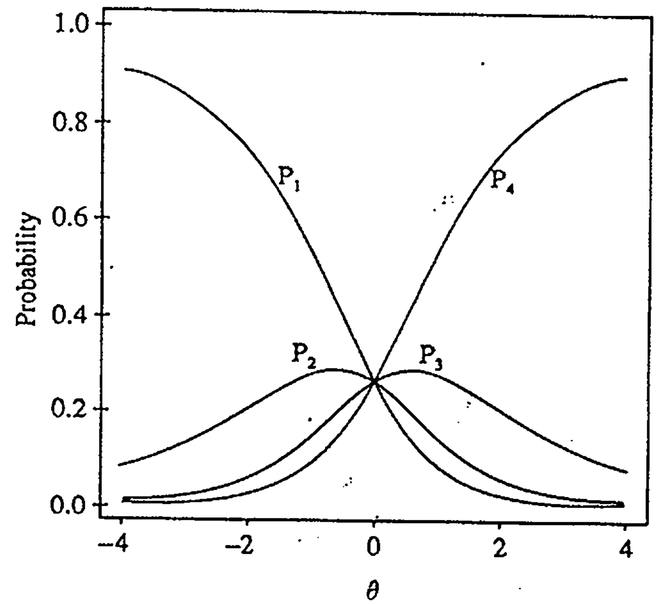

Figure 3a
ICCC's for a Three–Category Item
NAEP Mathematics Item 13
Partial Credit Model with a = 0.778 and b = [1.759, –1.643]

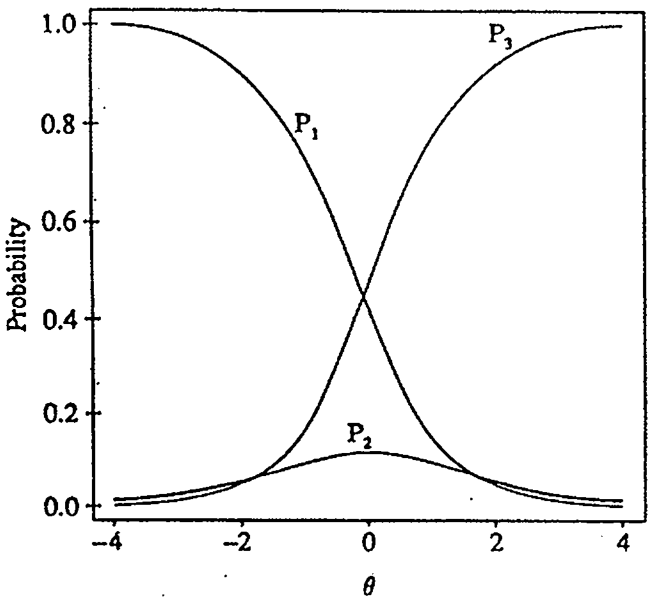

Figure 3b
ICCC's for a Three–Category Item
NAEP Mathematics Item 14
Partial Credit Model with a = 0.946 and b = [3.970, –2.764]