PSYCHOMETRIKA—VOL. 54, NO. 4, 635–659
DECEMBER 1989

EXTENSIONS OF THE PARTIAL CREDIT MODEL

C. A. W. GLAS AND N. D. VERHELST

NATIONAL INSTITUTE FOR EDUCATIONAL MEASUREMENT (CITO)
ARNHEM, THE NETHERLANDS

The partial credit model, developed by Masters (1982), is a unidimensional latent trait model for responses scored in two or more ordered categories. In the present paper some extensions of the model are presented. First, a marginal maximum likelihood estimation procedure is developed which allows for incomplete data and linear restrictions on both the item and the population parameters. Secondly, two statistical tests for evaluating model fit are presented: the former test has power against violation of the assumption about the ability distribution, the latter test offers the possibility of identifying specific items that do not fit the model.

Key words: item response theory, Rasch model, partial credit, maximum likelihood, EM-algorithm, model test.

1. Introduction

The partial credit model is developed for a test situation in which the respondent has the opportunity to complete an item at different performance levels. The respondent receives a credit equal to the level of performance at which the item was completed. If only two categories are present, for instance "right" and "wrong", the model is equivalent with the Rasch model for dichotomous items. Desirable properties of the dichotomous model, such as separation of person and item parameters and the existence of sufficient statistics for both sets of parameters, are preserved in the partial credit model. For a detailed derivation of the model, one is referred to Masters (1982); only essentials will be given here.

Consider the response of a person, indexed n, to an item, indexed i, which has $m_i + 1$ response categories which will be indexed $j = 0, 1, \ldots, m_i$. Person n produces a m_i-dimensional response vector $\mathbf{x}_{ni}$ with elements

$$x_{nij} = \begin{cases} 1 \text{ if person } n \text{ scores in category } j \text{ on item } i, \\ 0 \text{ if this is not the case,} \end{cases}$$

for $j = 1, \ldots, m_i$. So if the respondent scores in category $j = 0$, $\mathbf{x}_{ni} = \mathbf{0}$. In the partial credit model it is assumed that the probability of a person scoring in category j rather than scoring in category $j - 1$ is a logistic function of a person parameter ϑ_n and a parameter δ_{ij} associated with category j of the item i. Thus if $j > 0$

$$\Pr(x_{nij} = 1 | x_{nij} = 1 \text{ or } x_{nij-1} = 1, \vartheta_n, \delta_{ij}) = \frac{\exp(\vartheta_n - \delta_{ij})}{1 + \exp(\vartheta_n - \delta_{ij})}. \tag{1}$$

The probability of a person with parameter ϑ_n scoring in category j, $j = 1, \ldots, m_i$, on an item with parameter $\boldsymbol{\delta}_i$, $\boldsymbol{\delta}_i' = (\delta_{i1}, \ldots, \delta_{ij}, \ldots, \delta_{im_i})$, is given by

The authors are indepted to professor Wim van der Linden and Huub Verstralen for their helpful comments.

Requests for reprints should be sent to C. A. W. Glas, Cito, PO Box 1034, 6801 MG Arnhem, THE NETHERLANDS.

0033-3123/89/1200-9090$00.75/0

$$\psi_{ij}(\vartheta_n) \overset{d}{=} \Pr(x_{nij} = 1 | \vartheta_n, \boldsymbol{\delta}_i) = \frac{\exp\left(\sum_{h=1}^{j} (\vartheta_n - \delta_{ih})\right)}{1 + \sum_{k=1}^{m_i} \exp\left(\sum_{h=1}^{k} (\vartheta_n - \delta_{ih})\right)}, \tag{2}$$

and it can be shown that

$$\Pr(\mathbf{x}_{ni} = \mathbf{0} | \vartheta_n, \boldsymbol{\delta}_i) = \frac{1}{1 + \sum_{k=1}^{m_i} \exp\left(\sum_{h=1}^{k} (\vartheta_n - \delta_{ih})\right)}. \tag{3}$$

Although the properties of the model are desirable. Molenaar (1983) has pointed out that interpretation of the parameters should be made carefully. From (1) it follows that the model is linear in the log of the odds of scoring in category j and scoring in category $j - 1$:

$$\ln\left\{\frac{\Pr(x_{nij} = 1 | \vartheta_n, \boldsymbol{\delta}_i)}{\Pr(x_{nij-1} = 1 | \vartheta_n, \boldsymbol{\delta}_i)}\right\} = \vartheta_n - \delta_{ij}. \tag{4}$$

So δ_{ij} cannot be interpreted as the difficulty parameter of category j alone, since the probability of completing the item in category $j - 1$ must also be taken into account. However, the advantages of having a model with parameter separation will, in most instances, prevail over the complicated interpretation of the parameters.

In the sequel it will prove convenient to introduce a reparametrization of the model. Consider the reparametrization $\eta_{ij} = \Sigma_{h=1}^{j}\delta_{ih}$. Then (2) can be rewritten as

$$\psi_{ij}(\vartheta_n) = \Pr(x_{nij} = 1 | \vartheta_n, \boldsymbol{\eta}_i) = \frac{\exp(j\vartheta_n - \eta_{ij})}{1 + \sum_{k=1}^{m_i} \exp(k\vartheta_n - \eta_{ik})}, \tag{5}$$

with $\boldsymbol{\eta}_i' = (\eta_{i1}, \ldots, \eta_{ij}, \ldots, \eta_{im_i})$ and

$$\Pr(\mathbf{x}_{ni} | \vartheta_n, \boldsymbol{\eta}_i) = \frac{\exp(\varphi(\mathbf{x}_{ni})\vartheta_n - \mathbf{x}_{ni}'\boldsymbol{\eta}_i)}{1 + \sum_{k=1}^{m_i} \exp(k\vartheta_n - \eta_{ik})}, \tag{6}$$

with $\varphi(\mathbf{x}_{ni}) = j$ if $x_{nij} = 1$ for some $j, j = 1, \ldots, m_i$, and $\varphi(\mathbf{x}_{ni}) = 0$ if $\mathbf{x}_{ni} = \mathbf{0}$.

Andersen (1977) has derived a general formulation for the class of latent trait models for polytomous items that allows for separate sets of minimal sufficient statistics for item and person parameters. From the inspection of (6) the reader may verify that the partial credit model is a special case of the general model presented by Andersen. The results in the present paper are more easily presented in the parametrization defined by (5) and (6). The translation of the results into the original parametrization proposed by Masters (1982) will be treated in the section on estimating linear functions of the parameters.

2. Estimation

The objective of the present paper is to develop an estimation and testing procedure for the partial credit model in a situation where the data matrix is incomplete. This incompleteness may be either accidental or intended by the test administrator. In the latter case it may be more precise to speak of an incomplete design. One may think of the calibration of a test battery or an item bank where it is impractical to confront all testees with all items.

For the estimation of the parameters in a complete design, i.e. a design where all persons take all items, two methods have been proposed: a so-called conditional (CML) and a so-called unconditional maximum likelihood (UML) estimation method. The UML estimation method (Masters, 1982; Wright & Masters, 1982) maximizes the likelihood of the data over item and person parameters simultaneously. This estimation procedure has the drawback that the person parameters act as so-called incidental parameters, which cause the estimates to be inconsistent (Andersen, 1973), and, as a consequence, asymptotic confidence intervals and the asymptotic distribution of likelihood ratio tests are hard to derive. In the case of a complete design, the bias in the estimation of the item parameters seems to be removed if they are multiplied by $(I - 1)/I$, where I is the number of items in the test (Wright & Masters). In an incomplete design however, an equivalent correction is not readily defined, because different groups of persons respond to different numbers of items. Because of the untractable bias of the estimators the author refrains from generalizing the UML method to the situation of incomplete data.

Masters (1982) has shown that conditioning on sufficient statistics for the person parameters results in a likelihood function that only depends on the item parameters. This makes it possible to compute CML estimates of the item parameters (see, for instance, Masters, 1985), which are known to be consistent (Andersen, 1973). A CML estimation procedure for the dichotomous Rasch model with incomplete data has been described by Fischer (1981) and this procedure could well be extended to the partial credit model with incomplete data. The present paper, however, does not pursue this approach, but deals with the problem of obtaining consistent estimates by using marginal maximum likelihood estimation methods (MML), as already developed for other latent trait models (Bock & Aitkin, 1981; Rigdon & Tsutakawa, 1983; Thissen, 1982).

To describe the estimation procedure, the following definitions are needed. Consider T tests indexed $t = 1, \ldots, T$ and I items indexed $i = 1, \ldots, I$. The composition of a test t is defined by a design vector $\mathbf{d}_t' = (d_{t1}, \ldots, d_{ti}, \ldots, d_{tI})$, where $d_{ti} = 1$ if item i figures in test t and $d_{ti} = 0$ if this is not the case. It will be assumed that every item figures at least in one test, so for $i = 1, \ldots, I$, $d_{ti} = 1$ for some $t = 1, \ldots, T$.

For the time being it will also be assumed that the tests are linked via common items. This means that for every two tests indexed t and t', there exists a sequence of indices $z_1, z_2, \ldots, z_h$ such that $\mathbf{d}_t'\mathbf{d}_{z_1} > 0$, $\mathbf{d}_{z_1}'\mathbf{d}_{z_2} > 0, \ldots, \mathbf{d}_{z_h}'\mathbf{d}_{t'} > 0$. In the sequel it will be shown that in some instances this restriction may be dropped. Let $\{n\}_t$ be the index set of the persons taking test t and let $\boldsymbol{\vartheta}_t$ be a N_t-dimensional vector of the parameters ϑ_n with $n \in \{n\}_t$. So test t is given to N_t persons. Every person produces a response vector $\mathbf{x}_n' = (\mathbf{x}_{n1}', \ldots, \mathbf{x}_{ni}', \ldots, \mathbf{x}_{nI}')$. If $d_{ti} = 1$, $\mathbf{x}_{ni}$ is the response vector for item i as defined in section 1, and if $d_{ti} = 0$, $\mathbf{x}_{ni}' = (c, c, \ldots, c)$, with c an arbitrary constant.

The data matrix X_t produced by the respondents given test t has rows $\mathbf{x}_n'$, $n \in \{n\}_t$. Finally the full data matrix is defined by $X' = [X_1', \ldots, X_t', \ldots, X_T']$ and the test administration design is represented by the matrix $D = [\mathbf{d}_1, \ldots, \mathbf{d}_t, \ldots, \mathbf{d}_T]$. Let r_n be the sum score of a person n taking test t, so $r_n = \Sigma_i \, d_{ti}\phi(\mathbf{x}_{ni})$.

The probability of X_t as a function of the item parameters $\boldsymbol{\eta}' = (\boldsymbol{\eta}_1', \ldots, \boldsymbol{\eta}_i', \ldots, \boldsymbol{\eta}_I')$ and the person parameters $\boldsymbol{\vartheta}_t$ is given by

$$\Pr(X_t|\mathbf{d}_t, \boldsymbol{\eta}, \boldsymbol{\vartheta}_t) = \prod_{n \in \{n\}_t} \Pr(\mathbf{x}_n|\mathbf{d}_t, \boldsymbol{\eta}, \vartheta_n)$$

$$= \prod_{n \in \{n\}_t} \exp(r_n\vartheta_n) \exp\left(-\sum_i d_{ti}x_{nij}\eta_{ij}\right)P_{t0}(\vartheta_n) \tag{7}$$

with

$$P_{t0}(\vartheta_n) = \prod_i \left(1 + \sum_{k=1}^{m_i} \exp(k\vartheta_n - \eta_{ik})\right)^{-d_{ti}}. \tag{8}$$

In the marginal approach it is assumed that the items are a fixed factor and that the parameters of the persons taking test t are randomly sampled from a normal distribution with expectation μ_t and variance σ_t^2. Let $g(\vartheta_n|\mu_t, \sigma_t)$ be the density function of the ability parameter of a person taking test t. Then the log of the likelihood of $(X, \boldsymbol{\vartheta})$, with $\boldsymbol{\vartheta}' = (\boldsymbol{\vartheta}_1', \ldots, \boldsymbol{\vartheta}_t', \ldots, \boldsymbol{\vartheta}_T')$ can be written as

$$\ln L(\boldsymbol{\lambda}|X, D, \boldsymbol{\vartheta}) = \sum_t \ln L_t(\boldsymbol{\lambda}|X_t, \mathbf{d}_t, \boldsymbol{\vartheta}_t)$$

$$= \sum_t \sum_{n \in \{n\}_t} \{\ln \Pr(\mathbf{x}_n|\mathbf{d}_t, \boldsymbol{\eta}, \vartheta_n) + \ln g(\vartheta_n|\mu_t, \sigma_t)\} \tag{9}$$

with $\boldsymbol{\lambda}' = (\boldsymbol{\eta}', \mu_1, \ldots, \mu_T, \sigma_1, \ldots, \sigma_T)$. Notice that $L_t(\boldsymbol{\lambda}|X_t, \mathbf{d}_t, \boldsymbol{\vartheta}_t)$ does not depend on all parameters in $\boldsymbol{\lambda}$ and $\Pr(\mathbf{x}_t|\mathbf{d}_t, \boldsymbol{\eta}, \vartheta_n)$ does not depend on all parameters in $\boldsymbol{\eta}$, but this notation will prove convenient in the sequel. Because $\boldsymbol{\vartheta}$ cannot be observed, the joint likelihood of X and $\boldsymbol{\vartheta}$ will be integrated over the range of $\boldsymbol{\vartheta}$ to obtain the so-called marginal likelihood

$$L^{(m)}(\boldsymbol{\lambda}|X, D) = \int \ldots \int L(\boldsymbol{\lambda}|X, D, \boldsymbol{\vartheta})\partial\boldsymbol{\vartheta}. \tag{10}$$

It is possible to maximize (10) directly by a Newton-Raphson procedure, but this leads to rather laborious computations. Therefore an "incomplete data" approach (Mislevy, 1984), where $\boldsymbol{\vartheta}$ is considered the incomplete data, is chosen and a version of the EM algorithm (Dempster, Laird & Rubin, 1977) can be used. Let $\boldsymbol{\lambda}^*$ be some estimate of $\boldsymbol{\lambda}$. This estimate can be improved by maximizing

$$\sum_t E(\ln L_t(\boldsymbol{\lambda}|X_t, \mathbf{d}_t, \boldsymbol{\vartheta}_t)|X_t, \mathbf{d}_t, \boldsymbol{\lambda}^*) \tag{11}$$

with respect to $\boldsymbol{\lambda}$. Setting $\boldsymbol{\lambda}^*$ equal to this estimate, this can be repeated until some convergence criterium is met. Applying the result of Dempster et al. to this particular problem, it can be shown that this procedure converges to the maximum of (10). Rigdon and Tsutakawa (1983) have worked out in detail the method for the dichotomous Rasch model. Since the procedure described here can be viewed as an extension of the one for the dichotomous model, only the major features of the method are given here. The quantity (11) can be written as

$$\sum_t \sum_{n \in \{n\}_t} \left\{ \int \ln \Pr(\mathbf{x}_n | \mathbf{d}_t, \boldsymbol{\eta}, \vartheta) f(\vartheta_n | \mathbf{x}_n, \mathbf{d}_t, \boldsymbol{\lambda}^*) \partial \vartheta_n + \int \ln g(\vartheta_n | \mu_t, \sigma_t) f(\vartheta_n | \mathbf{x}_n, \mathbf{d}_t, \boldsymbol{\lambda}^*) \partial \vartheta_n \right\} \tag{12}$$

with

$$f(\vartheta_n | \mathbf{x}_n, \mathbf{d}_t, \boldsymbol{\lambda}^*) = \frac{\Pr(\mathbf{x}_n | \mathbf{d}_t, \boldsymbol{\eta}^*, \vartheta_n) g(\vartheta_n | \mu_t^*, \sigma_t^*)}{\int \Pr(\mathbf{x}_n | \mathbf{d}_t, \boldsymbol{\eta}^*, \vartheta_n) g(\vartheta_n | \mu_t^*, \sigma_t^*) \partial \vartheta_n}. \tag{13}$$

Applying (7) and (8) to (13) results in

$$f(\vartheta_n | \mathbf{x}_n, \mathbf{d}_t, \boldsymbol{\lambda}^*) = \frac{\exp(r_n \vartheta_n) P_{t0}(\vartheta_n) g(\vartheta_n | \mu_t, \sigma_t)}{\int \exp(r_n \vartheta_n) P_{t0}(\vartheta_n) g(\vartheta_n | \mu_t, \sigma_t) \partial \vartheta_n}. \tag{14}$$

Since the right hand side of (14) only depends upon response pattern $\mathbf{x}_n$ via the total number of correct responses r_n, the posterior density $f(\vartheta_n | \mathbf{x}_n, \mathbf{d}_t, \boldsymbol{\lambda}^*)$ is equivalent with the density of ϑ given a sum score r, which will be denoted by $f(\vartheta | r, \mathbf{d}_t, \boldsymbol{\lambda}^*)$. Let n_{tr} be the number of respondents achieving a sum score r, $r = 0, \ldots, J(t)$ with $J(t)$ the maximum score that can be obtained on test t. Further s_{ij} is defined as the number of persons responding to item i in category j, for $j = 1, \ldots, m_i$, that is, $s_{ij} = \Sigma_t d_{ti} \Sigma_{n \in \{n\}_t} x_{nij}$.

Taking first order derivatives of (12) with respect to $\boldsymbol{\lambda}$ and setting these equal to zero results in the estimation equations

$$s_{ij} = \sum_t d_{ti} \sum_r n_{tr} \int \psi_{ij}(\vartheta) f(\vartheta | r, \mathbf{d}_t, \boldsymbol{\lambda}^*) \partial \vartheta, \qquad i = 1, \ldots, I \text{ and } j = 1, \ldots, m_i, \tag{15}$$

$$\mu_t = \left(\frac{1}{N_t}\right) \sum_r n_{tr} \int \vartheta f(\vartheta | r, \mathbf{d}_t, \boldsymbol{\lambda}^*) \partial \vartheta, \qquad t = 1, \ldots, T, \tag{16}$$

$$\sigma_t^2 = \left(\frac{1}{N_t}\right) \sum_r n_{tr} \int \vartheta^2 f(\vartheta | r, \mathbf{d}_t, \boldsymbol{\lambda}^*) \partial \vartheta - \mu_t^2, \qquad t = 1, \ldots, T. \tag{17}$$

It must be noted that the model with a parametrization where $\boldsymbol{\lambda}$ is completely free, is not identified. Some restriction must be imposed like $\mu_t = 0$ for some t or $\Sigma_{i,j} \eta_{ij} = 0$. For the existence of a finite solution to the estimation equations, it is also necessary that $s_{ij} \neq 0$ and $s_{ij} \neq \underline{\Sigma_t\, d_{ti} N_t}$. This can be verified by noticing that $0 < \int \psi_{ij}(\vartheta) f(\vartheta | r, \mathbf{d}_t, \boldsymbol{\lambda}^*) \partial \vartheta < 1$.

Evaluation of (14), (15), (16) and (17) can be done by using Gauss-Hermite quadrature. One can take advantage of the fact that the weight functions in the integrals, that is, the part of the integrand absorbed in the weights of the linear functional used for the approximation, are identical. This is in fact the essence of the computational short cut proposed by Thissen (1982) for the dichotomous model.

3. Asymptotic Confidence Intervals

In order to obtain an estimate of standard errors, the observed information matrix (see Efron & Hinkley, 1978) which is given by

$$I(\boldsymbol{\lambda}, \boldsymbol{\lambda}) = -\frac{\partial^2 \ln L^{(m)}(\boldsymbol{\lambda}|X, D)}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'} \tag{18}$$

must be computed. Louis (1982) has derived a procedure for extracting the observed information matrix when the EM algorithm is used.

This procedure will be applied to the present problem. Let $\mathbf{b}^{(m)}(\boldsymbol{\lambda}) = \partial \ln L^{(m)}(\boldsymbol{\lambda}|X, D)/\partial\boldsymbol{\lambda}$, $\mathbf{b}(\boldsymbol{\lambda}) = \partial \ln L(\boldsymbol{\lambda}|X, D, \boldsymbol{\vartheta})/\partial\boldsymbol{\lambda}$ and

$$B(\boldsymbol{\lambda}, \boldsymbol{\lambda}) = \frac{\partial^2 \ln L(\boldsymbol{\lambda}|X, D, \boldsymbol{\vartheta})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}.$$

Louis (1982) has shown that

$$I(\boldsymbol{\lambda}, \boldsymbol{\lambda}) = -E(B(\boldsymbol{\lambda}, \boldsymbol{\lambda})|X, D, \boldsymbol{\lambda}) - E(\mathbf{b}(\boldsymbol{\lambda})\mathbf{b}(\boldsymbol{\lambda})'|X, D, \boldsymbol{\lambda}) + \mathbf{b}^{(m)}(\boldsymbol{\lambda})\mathbf{b}^{(m)}(\boldsymbol{\lambda})' \tag{19}$$

If $\hat{\boldsymbol{\lambda}}$ is the maximum likelihood estimate of $\boldsymbol{\lambda}$ then $\mathbf{b}^{(m)}(\hat{\boldsymbol{\lambda}}) = \mathbf{0}$. A detailed derivation of the information matrix is given in appendix A. Since the model is not identified without some restriction on $\boldsymbol{\lambda}$ in the sequel it will be assumed that the model is identified by putting one of the item parameters or one mean of an ability distribution equal to zero. The relation between the resulting parametrization, which will be denoted by $\boldsymbol{\xi}$, and the original parametrization $\boldsymbol{\lambda}$ can be written as the linear transformation $\boldsymbol{\lambda} = F\boldsymbol{\xi}$ and so the observed information matrix for the model parameterized by $\boldsymbol{\xi}$ is given by

$$-\left[\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{\xi}'}\right]'\left[\frac{\partial^2 \ln L^{(m)}(\boldsymbol{\lambda}|X, D)}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\right]\left[\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{\xi}'}\right], \tag{20}$$

and the asymptotic covariance matrix of the maximum likelihood estimate of $\boldsymbol{\xi}$ is given by $(F'I(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\lambda}})F)^{-1}$.

4. Estimating Linear Functions of the Parameters

In many instances, the interest is not so much in the partial credit model as such, but in some more restricted case of the model. Masters and Wright (1984) have identified several models that can be derived from the partial credit model, by imposing linear restrictions on the item parameters.

It can also be useful to impose restrictions on the population parameters. In an incomplete design one may, for instance, formulate the hypothesis that all subgroups are a random sample from the same normal distribution by imposing the restrictions $\mu_t = \mu$ and $\sigma_t = \sigma$, for $t = 1, \ldots, T$. In this case it is no longer necessary that the test administration design is linked, for the common distribution now serves as a link between the different sets of item parameters.

In many instances it may also be useful to reparameterize the version of the model as defined by (6) to the original version as it was presented by Masters (1982). The reparameterization $\eta_{ij} = \Sigma_{h=1}^{j} \delta_{ih}$, $j = 1, \ldots, m_i$, introduced in the introduction of this paper can be written, for example, for $m_i = 4$ as

$$\boldsymbol{\eta}_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \boldsymbol{\delta}_i,$$

and so

$$\boldsymbol{\delta}_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \boldsymbol{\eta}_i.$$

The transformation to the original parametrization is particularly useful because special cases of the model are in most instances defined in relation to the original parametrization. A good example is the rating scale model proposed by Masters (1982), that can be formulated by imposing the linear restrictions $\delta_{ij} = \alpha_i + \tau_j$, for $i = 1, \ldots, I$ and $j = 1, \ldots, m_i$. All these examples can be viewed as special cases of the linear mapping

$$\boldsymbol{\xi} = G\boldsymbol{\beta}, \tag{21}$$

with dimension ($\boldsymbol{\beta}$) $\leq$ dimension ($\boldsymbol{\xi}$) and G of full column rank.

The estimation equations for these models can easily be derived by observing that

$$\frac{\partial \ln L^{(m)}(\boldsymbol{\lambda} | X, D)}{\partial \boldsymbol{\beta}'} = \frac{\partial \ln L^{(m)}(\boldsymbol{\lambda} | X, D)}{\partial \boldsymbol{\lambda}'} \left(\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\xi}'}\right)\left(\frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{\beta}'}\right) = G'F'\mathbf{b}^{(m)}(\boldsymbol{\lambda}),$$

and in the same manner $\partial \ln L(\boldsymbol{\lambda} | X, D, \boldsymbol{\vartheta})/\partial \boldsymbol{\beta}' = G'F'\mathbf{b}(\boldsymbol{\lambda})$. In matrix notation the generalized EM algorithm boils down to finding a maximum of

$$E(\ln L(\boldsymbol{\lambda} | X, D, \boldsymbol{\vartheta}) | X, D, \boldsymbol{\lambda}^*) \tag{22}$$

with respect to $\boldsymbol{\lambda}$, where $\boldsymbol{\lambda}^*$ is some initial estimate, setting $\boldsymbol{\lambda}^*$ equal to this maximum and repeating the process until convergence is obtained. So the procedure for estimating $\boldsymbol{\beta}$ is found by premultiplying (22) by $G'F'$ and the asymptotic covariance matrix of $\boldsymbol{\beta}$ is given by $(G'F'I(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\lambda}})FG)^{-1}$.

5. Testing the Model with a Focus on the Items

For the construction of a model test one can make use of the well-established framework of the multinomial model. Let $\{\mathbf{x}\}_t$ stand for the set of all possible response patterns $\mathbf{x}$ on test t and let $\mathbf{n}_t$ be the associated vector of frequency counts. So $\mathbf{n}_t$ has elements $n_{t\mathbf{x}}$, where $n_{t\mathbf{x}}$ is the number of persons taking test t and producing response pattern $\mathbf{x} \in \{\mathbf{x}\}_t$. It is easy to see that $\mathbf{n}_t$ has a multinomial distribution with parameters N_t and $\boldsymbol{\pi}_t$. The vector of theoretical probabilities $\boldsymbol{\pi}_t$ has elements $\pi_{t\mathbf{x}}$ defined by

$$\pi_{t\mathbf{x}} = \Pr(\mathbf{x} | \mathbf{d}_t, \boldsymbol{\lambda}) = \int \Pr(\mathbf{x} | \mathbf{d}_t, \boldsymbol{\eta}, \vartheta) g(\vartheta | \mu_t, \sigma_t) \partial\vartheta, \tag{23}$$

where $\Pr(\mathbf{x} | \mathbf{d}_t, \boldsymbol{\eta}, \vartheta)$ is the probability of response pattern $\mathbf{x}$ as a function of ϑ and $g(\vartheta | \mu_t, \sigma_t)$ is the normal probability density function. To keep things general, $\boldsymbol{\lambda}$ satisfies linear constraints of the form $\boldsymbol{\lambda} = F\boldsymbol{\xi}$ or $\boldsymbol{\lambda} = FG\boldsymbol{\beta}$. In general it will be assumed that s parameters have to be estimated.

Testing the assumed model against a general multinomial alternative can by done by applying Pearsons' X^2 test

$$X^2 = \sum_t \sum_{\{\mathbf{x}\}_t} \frac{(n_{t\mathbf{x}} - N_t \pi_{t\mathbf{x}})^2}{N_t \pi_{t\mathbf{x}}}. \tag{24}$$

It can be shown (see, for instance Bishop, Fienberg & Holland, 1975) that X^2 has an asymptotic χ^2 distribution with $v - s - 1$ degrees of freedom, where v is the number of possible response patterns and s is the number of parameters to be estimated.

This approach however, suffers from serious drawbacks. Even for a relatively small number of items, the number of possible response patterns is very large, so the vector of frequency counts $\mathbf{n}_t$ will have a large number of very small and zero elements, which will damage the asymptotic properties of the test. A second drawback of the approach sketched above is the fact that the identification of the sources of misfit is very difficult. Therefore statistical testing procedures will be presented, that are based on a higher aggregation level of the data, which give information with respect to specific model violations. In the present paper two sources of model violations will be considered: model violations caused by an improper assumption about the ability distribution and model violations caused by specific items that do not fit the model. In actual data analysis, checking the appropriateness of the assumption concerning the ability distribution must be carried out first, for there is little sense in searching for misfitting items, if the model is violated at such an essential level. For didactical reasons however, the subject of item fit will be treated first. So the objective of the present section is to derive a test of model fit, where the contribution of specific items to the outcome of the test can be identified. The test will be based on a comparison of predicted and observed item characteristic curves (ICC). The test is motivated by the following considerations.

Equations (2) and (6) give the probability of a person n scoring in category j on item i, $\psi_{ij}(\vartheta_n)$, with ϑ_n a fixed person parameter, for respectively the parametrizations $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$. If ϑ is varied across the range $\langle -\infty, \infty \rangle$, $\psi_{ij}(\vartheta)$ is called an ICC. Since the frequency distribution of sum scores is a sufficient statistic for the ability distribution, it can be assumed that persons who obtain the same sum score, form homogeneous subgroups on the latent continuum. Therefore the probability of scoring in category j of item i as a function of the sum score can be viewed as an approximation of the ICC. Thus, the difference between the observed and expected number of persons obtaining a certain sum score and responding to an item in a certain response category can be used to evaluate model fit.

To formalize these notions, let $I(t)$ be the number of items and let $\{i\}_t$ be the set of indices of the items in test t. Further $J(t)$ stands for the maximum score that can be attained and n_{trij} stands for the number of persons taking test t who obtain a sum score $r = 1, \ldots, J(t) - 1$ and score in category j, $(j = 1, \ldots, m_i)$ on an item $i \in \{i\}_t$.

The scores of $r = 0$ and $r = J(t)$ are excluded, because they are special in the sense that there exists only one response pattern to obtain them. The categories $j = 0$ are not considered, because the information they yield for every item i, is completely contained in the frequencies n_{trij}, for $j = 1, \ldots, m_i$. Notice that if $r < j$, $n_{trij} = 0$, because it is not possible to respond in category j and obtain a sum score $r < j$. In the same manner it is also not possible to respond in category j and obtain a sum score $r > J(t) - m_i + j$. To evaluate the expectation $E(N_{trij}|\mathbf{d}_t, \boldsymbol{\lambda})$ of n_{trij}, an elementary symmetric function $\Gamma_r(\boldsymbol{\varepsilon}, \mathbf{d}_t)$ of the parameters $\boldsymbol{\varepsilon}' = (\boldsymbol{\varepsilon}_1', \ldots, \boldsymbol{\varepsilon}_i', \ldots, \boldsymbol{\varepsilon}_I')$, where $\boldsymbol{\varepsilon}_i$ has elements $\varepsilon_{ij} = \exp(-\eta_{ij})$, is defined by

$$\Gamma_r(\boldsymbol{\varepsilon}, \mathbf{d}_t) = \sum_{\{\mathbf{x}\}_{tr}} \prod_{i,j} \varepsilon_{ij}^{d_{ti} x_{ij}}. \tag{25}$$

Here $\{\mathbf{x}\}_{tr}$ stands for the set of all possible response patterns $\mathbf{x}$ associated with score r. The following example may clarify (25). For a test of three items, $m_i = 2$, for $i = 1, \ldots, 3$ and $\mathbf{d}' = (1, 1, 1, 0, 0, 0, \ldots)$ the elementary symmetric functions are given by:

$$\Gamma_0(\boldsymbol{\varepsilon}, \mathbf{d}_t) = 1,$$

$$\Gamma_1(\boldsymbol{\varepsilon}, \mathbf{d}_t) = \varepsilon_{11} + \varepsilon_{21} + \varepsilon_{31},$$

$$\Gamma_2(\boldsymbol{\varepsilon}, \mathbf{d}_t) = \varepsilon_{11}\varepsilon_{21} + \varepsilon_{11}\varepsilon_{31} + \varepsilon_{21}\varepsilon_{31} + \varepsilon_{12} + \varepsilon_{22} + \varepsilon_{32},$$

$$\Gamma_3(\boldsymbol{\varepsilon}, \mathbf{d}_t) = \varepsilon_{11}\varepsilon_{21}\varepsilon_{31} + \varepsilon_{12}\varepsilon_{21} + \varepsilon_{11}\varepsilon_{22} + \varepsilon_{12}\varepsilon_{31} + \varepsilon_{11}\varepsilon_{32} + \varepsilon_{22}\varepsilon_{31} + \varepsilon_{21}\varepsilon_{32},$$

$$\Gamma_4(\boldsymbol{\varepsilon}, \mathbf{d}_t) = \varepsilon_{12}\varepsilon_{21}\varepsilon_{31} + \varepsilon_{12}\varepsilon_{22} + \varepsilon_{12}\varepsilon_{32} + \varepsilon_{11}\varepsilon_{22}\varepsilon_{31} + \varepsilon_{11}\varepsilon_{21}\varepsilon_{32} + \varepsilon_{22}\varepsilon_{32},$$

$$\Gamma_5(\boldsymbol{\varepsilon}, \mathbf{d}_t) = \varepsilon_{12}\varepsilon_{22}\varepsilon_{31} + \varepsilon_{12}\varepsilon_{21}\varepsilon_{32} + \varepsilon_{11}\varepsilon_{22}\varepsilon_{32},$$

$$\Gamma_6(\boldsymbol{\varepsilon}, \mathbf{d}_t) = \varepsilon_{12}\varepsilon_{22}\varepsilon_{32}.$$

The computation of elementary functions of the kind defined by (25) has been described by Andersen (1972) and Fischer (1974).

By summing (23) over the set of all possible response patterns leading to a sum score r and with $x_{tij} = 1$, it can be shown that for $j \geq 1$

$$E(N_{trij}|\mathbf{d}_t, \boldsymbol{\lambda}) = N_t \int \varepsilon_{ij}\Gamma^{(i)}_{r-j}(\boldsymbol{\varepsilon}, \mathbf{d}_t) \exp(r\vartheta)P_{t0}(\vartheta)g(\vartheta|\mu_t, \sigma_t)\partial\vartheta, \qquad (26)$$

where $\Gamma^{(i)}_{r-j}(\boldsymbol{\varepsilon}, \mathbf{d}_t)$ stands for an elementary symmetric function of order $r - j$ as defined by (25), of the parameters $(\boldsymbol{\varepsilon}'_1, \ldots, \boldsymbol{\varepsilon}'_{i-1}, \boldsymbol{\varepsilon}'_{i+1}, \ldots, \boldsymbol{\varepsilon}'_I)$.

The difference between the predicted and the observed ICC for item i and category j, as it appears in test t, can be evaluated by inspecting the sequence of deviates $n_{trij} - E(N_{trij}|\mathbf{d}_t, \hat{\boldsymbol{\lambda}})$, for $r = j, \ldots, J(t) - m_i + j$, or by inspecting the scaled deviate $z^0_{trij} = (n_{trij} - E(N_{trij}|\mathbf{d}_t, \hat{\boldsymbol{\lambda}}))/\text{var}\,(N_{trij}|\mathbf{d}_t, \hat{\boldsymbol{\lambda}})^{1/2}$. The sign of z^0_{trij} indicates whether the predicted ICC is higher or lower at a certain score level than should be expected. The interpretation of the magnitude of z^0_{trij} may be helped by the fact that if only one item, one category, and one score level are considered, and no parameters have to be estimated, z^0_{trij} is a standardized biniomial variable. Squaring and summing z^0_{trij} over the appropriate range of sum scores yields an index of item fit that is approximately χ^2 distributed if the assumptions given should hold. They do, of course, not hold, but for the identification of the items that least fit the model, the index of item fit serves its purpose.

What is also needed in conjunction with the item fit indices, however, is a more formal model test. The test that will be presented below is a generalization of the model test for the marginal Rasch model presented by Glas (1988).

For $t = 1, \ldots, T$ and $r = 1, \ldots, J(t) - 1$, let $\mathbf{z}_{tr}$ be a vector with elements z_{trij} defined by $z_{trij} = N_t^{-1/2}(n_{trij} - E(N_{trij}|\mathbf{d}_t, \hat{\boldsymbol{\lambda}}))$ for $i \in \{i\}_t$ and $j = \max(1, r + m_i - J(t)), \ldots, \min(m_i, r)$. Let e_{tr} stand for the dimension of $\mathbf{z}_{tr}$, so $e_{tr} = \min(m_i, r) - \max(1, r + m_i - J(t)) + 1$. The vectors $\mathbf{z}_{tr}$ will be combined into a quadratic form using the inverse of $e_{tr} \times e_{tr}$ matrices W_{tr}. Let z_{trij} be the k-th element of $\mathbf{z}_{tr}$, $z_{trij'}$ is the $\mathbf{k}'$-th element ($j' \neq j$) and $z_{tri'j''}$ is the k''-th element ($i' \neq i$) of $\mathbf{z}_{tr}$. The elements $W_{tr}(k, k)$, $W_{tr}(k, k')$ and $W_{tr}(k, k'')$ of W_{tr} are defined by:

$$W_{tr}(k, k) = \int \varepsilon_{ij}\Gamma^{(i)}_{r-j}(\boldsymbol{\varepsilon}, \mathbf{d}_t) \exp(r\vartheta)P_{t0}(\vartheta)g(\vartheta|\mu_t, \sigma_t)\partial\vartheta,$$

$$W_{tr}(k, k') = 0, \text{ and} \qquad (27)$$

$$W_{tr}(k, k'') = \int \varepsilon_{ij}\varepsilon_{i'j''}\Gamma^{(i,i')}_{r-j-j'}(\boldsymbol{\varepsilon}, \mathbf{d}_t) \exp(r\vartheta)P_{t0}(\vartheta)g(\vartheta|\mu_t, \sigma_t)\partial\vartheta,$$

with $\Gamma^{(i,i')}_{r-j-j'}(\boldsymbol{\varepsilon}, \mathbf{d}_t)$ an elementary symmetric function of order $r - j - j''$ as defined by (25), where the parameters $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\varepsilon}_{i'}$ have been removed from $\boldsymbol{\varepsilon}$.

To be able to derive the distribution of the model test, the number of persons obtaining a perfect score on test t, denoted by $n_{tJ(t)}$, and the number of persons who obtain a zero score, denoted by n_{t0}, must also be taken into consideration. Let $z_{t0} = N_t^{-1/2}(n_{t0} - E(N_{t0}|\mathbf{d}_t, \hat{\boldsymbol{\lambda}}))$, $z_{tJ(t)} = N_t^{-1/2}(n_{tJ(t)} - E(N_{tJ(t)}|\mathbf{d}_t, \hat{\boldsymbol{\lambda}}))$, π_{t0} is the probability of a zero response pattern and π_{t1} is the probability of a perfect response pattern. As before, $\boldsymbol{\lambda}$ may be subject to linear restrictions, so let s be the number of parameters that have to be estimated. In section 8 it will be shown that

$$R_1 = \sum_t^T \left\{ \frac{z_{t0}^2}{\hat{\pi}_{t0}} + \sum_{r=1}^{J(t)-1} \mathbf{z}'_{tr}\hat{W}_{tr}^{-1}\mathbf{z}_{tr} + \frac{z_{tJ(t)}^2}{\hat{\pi}_{t1}} \right\} \qquad (28)$$

has an asymptotic χ^2 distribution. The degrees of freedom are equal to the number of deviates on which the test is based, which is $\Sigma_t^T\Sigma_{r=1}^{J(t)-1}e_{tr} + 2T$, minus the number of parameters that have to be estimated, which is s, minus one degree of freedom for every one of the T multinomial distributions in the model. So R_1 has $(\Sigma_t^T\Sigma_{r=1}^{J(t)-1}e_{tr}) + T - s$ degrees of freedom.

In section 7 of this paper, some examples of application of the technique will be presented, but first the evaluation of the fit of the ability distribution will be discussed.

6. Testing the Model With a Focus on the Ability Distribution

In the present section a model test will be defined that has power against improper modeling of the ability distribution.

Since a respondent's sum score is a sufficient statistic for his/her ability parameter, the test will focus on clusters of response patterns leading to the same sum score. As in the previous sections, n_{tr} stands for the number of respondents who obtain a sum score r on test t. By summing (23) over the set of all possible response patterns leading to a sum score r, it can be verified that

$$E(N_{tr}|\mathbf{d}_t, \boldsymbol{\lambda}) = N_t \int \Gamma_r(\boldsymbol{\varepsilon}, \mathbf{d}_t) \exp(r\vartheta)P_{t0}(\vartheta)g(\vartheta|\mu_t, \sigma_t)\partial\vartheta. \qquad (29)$$

As in the previous section, the deviates $n_{tr} - E(N_{tr}|\mathbf{d}_t, \hat{\boldsymbol{\lambda}})$ can be transformed into scaled deviates with an approximate standard normal distribution, and inspecting the magnitude and the sign of these scaled deviates will yield information with respect to the appropriateness of the assumption about the ability distribution.

Combining the scaled deviates into a χ^2 distributed quadratic form is, however, somewhat more complicated than in the previous section. In the present section only the main result is presented, the details are given in section 8.

Let $\mathbf{z}$ be a vector defined $\mathbf{z}' = (\mathbf{z}'_1, \ldots, \mathbf{z}'_t, \ldots, \mathbf{z}'_T, \mathbf{z}'_{T+1})$, where, for $t = 1, \ldots, T$, $\mathbf{z}_t$ is an N_t-dimensional vector with elements $N_t^{-1/2}(n_{tr} - E(N_{tr}|\hat{\boldsymbol{\lambda}}))$ and $\mathbf{z}_{T+1}$ is a $(J - 1)$-dimensional vector with elements

$$\left(\sum_t d_{ti}N_t^{-1/2}\right)\left(s_{ij} - \sum_t N_t d_{ti} \int \hat{\psi}_{ij}(\vartheta)g(\vartheta|\hat{\mu}_t, \hat{\sigma}_t)\partial\vartheta\right),$$

for $i = 1, \ldots, I$ and $j = 1, \ldots, m_i$, where one element, say the element associated with $i = j = 1$ is omitted. Then it can be shown that $R_0 = \mathbf{z}'\hat{W}^{-1}\mathbf{z}$ has an asymptotic χ^2 distribution with $(2\Sigma_t J(t)) - s - T$ degrees of freedom, for some matrix W, which will be defined in section 8. In section 8 it is also argued that at $\hat{\boldsymbol{\lambda}}$, the elements of $\mathbf{z}_{T+1}$ are all equal to zero. Thus the influence of specific items on the outcome of the R_0 test cannot traced. In section 8 it will be shown that the vector $\mathbf{z}_{T+1}$ must be added to account for the loss of degrees of freedom caused by the estimation of the item parameters and the outcome of the R_0 test is influenced by the presence of $\mathbf{z}_{T+1}$ via the elements of W associated with $\mathbf{z}_{T+1}$.

As the Rasch model for dichotomous items is a special case of the partial credit model, the test presented here can, of course, also be used to evaluate the appropriateness of the assumption about the ability distribution for the marginal Rasch model for dichotomous items.

7. Some Examples

To illustrate the possibilities of the techniques described in the previous sections, a number of simulated examples will be presented. The first set of examples focusses on the effects of a non-normal ability distribution on the parameter estimates and the possibility to detect these model violations using the statistical testing procedure presented in the sections 5 and 6. The second set of examples will focus on selecting items that fit the partial credit model. For the clarity of the presentation, the simulation studies were carried out using one test of three items with five categories, so $m = m_i = 4$ for $i = 1, \ldots, 3$. The data were generated using the following algorithm:

Step 1. Choose a set of item parameters $\boldsymbol{\eta}$.

Step 2. Draw N person parameters ϑ_n from a distribution Θ.

Step 3. Compute $\psi_{ij}(\vartheta_n)$ for $i = 1, \ldots, 3, j = 1, \ldots, 4$ and $n = 1, \ldots, N$, using (5) and let $\psi_{i0}(\vartheta_n) = 1 - \Sigma_{j=1}^{m_i}\psi_{ij}(\vartheta_n)$.

Step 4. Draw $v_{ni} (i = 1, \ldots, 3$ and $n = 1, \ldots, N)$ from the uniform distribution on [0, 1].

Step 5. A data matrix X with entries x_{nij} is generated using

$$x_{nij} = \begin{cases} 1 & \text{if } v_{ni} > \sum_{h=0}^{j-1} \psi_{ih}(\vartheta_n) \text{ and } v_{ni} < \sum_{h=0}^{j} \psi_{ih}(\vartheta_n). \\ 0 & \text{in other instances.} \end{cases}$$

The artificial data were generated under three conditions: in the first condition Θ was normally distributed with a mean equal to zero and a variance equal to one, in the second condition Θ was distributed uniformly on (−2, 2) and in the third condition Θ had a χ^2 distribution with one degree of freedom, which was shifted such that its expectation equaled zero.

Table 1 gives the result of the parameter estimation for a typical simulation run

TABLE 1

Parameter Estimation For Data Confirming
The Partial Credit Model

item	category	η_{ij}	$\hat{\eta}_{ij}$	SE($\hat{\eta}_{ij}$)	S_{ij}
1	1	-1.000	-0.966	0.047	1208
	2	-0.500	-0.426	0.062	627
	3	0.000	-0.046	0.073	551
	4	0.500	0.525	0.078	763
2	1	-1.500	-1.447	0.049	1352
	2	-1.000	-0.905	0.063	667
	3	-0.500	-0.499	0.074	576
	4	0.000	0.037	0.078	776
3	1	0.000	-0.013	0.044	862
	2	0.500	0.423	0.052	556
	3	1.000	0.958	0.055	500
	4	1.500	1.473	0.055	709

σ=1.000, $\hat{\sigma}$=0.991, SE($\hat{\sigma}$)=0.0187, μ=0.000 (fixed)

with data generated in the first condition, that is, data conforming the partial credit model. The parameter estimation was carried out on a sample of $N = 4000$ response patterns, the third column of Table 1 gives the true values of the item parameters, the fourth column gives the estimated values. The last two columns give the standard errors computed using the method given in section 3 and the values of the sufficient statistics for the item parameters. Next, the statistical testing procedure presented in section 6 was carried out, Table 2 gives the results. The column marked "deviate" gives the difference between the observed and expected frequency of the score levels, the column marked "scaled deviate" gives the deviate divided by its standard deviation.

The value of the R_0 statistic defined in section 6 is given at the bottom of the Table. Since the 5% critical value of a χ^2 distributed variable with 10 degrees of freedom is 18.3 the hypothesis that the data fit the partial credit model can, as expected, not be rejected. To get some idea of the robustness of the estimation procedure, data were generated under the two other conditions: the uniform ability distribution condition and the shifted χ^2 ability distribution condition. In both instances a sample size $N = 4000$ was used.

Table 3 gives the results of the parameter estimation for both conditions, the results are averaged over ten replications. Inspection of Table 3 shows that for the uniform distribution condition the magnitude of the bias in the estimates is relatively small. The shifted χ^2 condition however yields a considerable bias. One of the main reasons for this phenomenon is the fact that the uniform distribution is symmetric with respect to its first moment, whereas the χ^2 distribution is not symmetric. For every replication in every condition, the model test R_0 was computed, and in all instances the model had to be rejected. So even though the bias in the estimates is small in the uniform distribution condition, it is serious enough to cause a rejection of the model.

The examples presented so far concern the appropriateness of the assumptions about the ability distribution. Lack of model fit, however, can also be caused by items that do not fit the model. Therefore in section 5 a statistical test was proposed that has

TABLE 2

Evaluation Of Model Fit For data Conforming The Partial Credit Model

score	observed frequency	expected frequency	deviate	scaled deviate
0	188	194.43	-6.44	-0.46
1	437	429.91	7.08	0.34
2	509	505.07	3.93	0.19
3	429	439.93	-10.93	-0.56
4	382	373.83	8.17	0.59
5	345	333.69	11.31	0.65
6	296	299.73	-3.73	0.22
7	266	271.19	5.19	-0.33
8	251	256.23	-5.23	-0.34
9	251	245.38	5.62	0.37
10	210	227.59	-17.59	-1.20
11	237	225.18	11.81	0.79
12	199	197.62	1.37	0.098

R_0 = 9.3826 df = 10

power against differences between the observed empirical ICC's and the ones predicted by the model. First it will be shown what kind of information is produced by the statistical testing procedure.

Using the same device and the same item parameters as above, several model conform data sets were generated. Table 4 gives the results for a typical simulation run with $N = 4000$. The results presented concern the second item, so the true value of the item parameter is $\eta_2' = (-1.50, -1.00, -0.50, 0.00)$. Estimation resulted in $\hat{\eta}_2' = (-1.48, -0.99. -0.55, -0.09)$. Notice that for the first category the range of scores only extends to 9, because larger scores cannot be obtained while scoring in the first category on some item. In the same manner it is for instance also not possible to score in the fourth category and obtain a score less than 4. Inspection of the column marked "expected frequency" reveals what the ICC's for the different categories look like: unimodal for the first three categories and monotone increasing for the last category. The magnitude of the column marked "scaled deviate" can be interpreted by assuming that every entry is a standardized normal variable.

The sensitivity of the "scaled deviates" to model violations caused by items can be illustrated by generating responses to items with ICC's which do not conform the partial credit model. This can be achieved by introducing a set of so-called discrimination parameters α_{ij}, for $i = 1, \ldots, 3$ and $j = 1, \ldots, 4$, into the equations for the ICC's, that is, $\psi_{ij}(\vartheta)$ is redefined as

$$\psi_{ij}(\vartheta) = \frac{\exp(\alpha_{ij}(j\vartheta - \eta_{ij}))}{1 + \sum_{k=1}^{m_i} \exp(\alpha_{ik}(k\vartheta - \eta_{ik}))}. \tag{30}$$

TABLE 3

Parameter Estimation For Data Generated Using A Non-Normal Ability Distribution

uniform ability distribution				
item	category	η_{ij}	$\hat{\eta}_{ij}$	$SE(\hat{\eta}_{ij})$
1	1	-1.000	-0.998	0.048
	2	-0.500	0.523	0.067
	3	0.000	-0.072	0.079
	4	0.500	0.488	0.082
2	1	-1.500	-1.448	0.050
	2	-1.000	-0.995	0.066
	3	-0.500	-0.721	0.053
	4	0.000	-0.050	0.082
3	1	0.000	0.044	0.048
	2	0.500	0.665	0.054
	3	1.000	0.956	0.066
	4	1.500	1.329	0.056
σ=1.000, $\hat{\sigma}$=0.993, SE($\hat{\sigma}$)=0.0182, μ=0.000 (fixed)				
shifted chi-square distribution				
item	category	η_{ij}	$\hat{\eta}_{ij}$	$SE(\hat{\eta}_{ij})$
1	1	-1.000	-0.884	0.044
	2	-0.500	-0.066	0.057
	3	0.000	0.652	0.077
	4	0.500	1.287	0.085
2	1	-1.500	-1.654	0.069
	2	-1.000	-0.682	0.061
	3	-0.500	0.255	0.089
	4	0.000	0.710	0.078
3	1	0.000	0.290	0.041
	2	0.500	1.182	0.052
	3	1.000	2.066	0.067
	4	1.500	2.590	0.068
σ=1.4142, $\hat{\sigma}$=1.0610, SE($\hat{\sigma}$)=0.0211, μ=0.000 (fixed)				

Using the algorithm described above, with the new definition of $\psi_{ij}(\vartheta)$, a number of data sets were generated, with $\alpha_{22} = 0.50$, $\alpha_{23} = 2.00$ and all other discrimination parameters equal to one. The item parameters were not altered.

Table 5 gives the results for item 2, for a randomly chosen replication. It can be seen that, compared with Table 5, the number of significant deviates increases considerably, especially for the categories two, three and four. The introduction of discrimination parameters resulted in a considerable bias in the estimation of the parameter of

TABLE 4

Evaluation Of Model Fit in A Data Set Conforming The Partial Credit Model

category	score	observed frequency	expected frequency	scaled deviate
1	1	226	235.90	-0.66
	2	346	337.94	0.46
	3	231	247.73	-1.09
	4	175	167.20	0.62
	5	126	125.84	0.01
	6	85	89.23	-0.45
	7	61	56.62	0.59
	8	45	37.81	1.17
	9	23	23.22	-0.05
2	2	58	57.87	0.02
	3	132	113.61	1.75
	4	95	108.87	-1.35
	5	93	93.11	-0.01
	6	97	86.81	1.10
	7	86	74.97	1.29
	8	52	57.72	-0.76
	9	39	47.27	-1.21
	10	25	36.24	-1.88
3	3	23	20.27	0.61
	4	41	52.03	-1.54
	5	52	63.18	-1.42
	6	54	66.93	-1.59
	7	80	76.02	0.46
	8	101	79.65	2.42
	9	70	75.19	-0.60
	10	87	76.90	1.16
	11	80	75.58	0.51
4	4	10	9.08	0.30
	5	35	29.55	1.00
	6	47	44.45	0.38
	7	51	57.36	-0.85
	8	74	79.03	-0.57
	9	108	101.55	0.65
	10	143	119.71	2.16
	11	139	156.93	-1.46
	12	199	207.13	-0.58

the second item, for instance $\hat{\eta}_2' = (-1.567, -0.679. -0.682, -0.237)$ for the replication under consideration. Computation of the test defined by (28) resulted in a rejection of the model for all ten replications: the average value over ten replications of the test was 275.05 with 22 degrees of freedom.

The artificial examples presented here, of course, far from exhaust all possible

TABLE 5

Evaluation Of Model Fit in A Data Set
Not Conforming The Partial Credit Model

category	score	observed frequency	expected frequency	scaled deviate
1	1	219	245.94	-1.77
	2	370	370.67	-0.04
	3	279	281.32	-0.14
	4	180	205.27	-1.81
	5	145	141.54	0.30
	6	102	93.54	0.89
	7	75	57.91	2.26
	8	36	28.73	1.36
	9	25	15.75	2.08
2	2	80	42.92	5.69
	3	126	87.20	4.20
	4	91	85.67	0.58
	5	57	78.67	-2.47
	6	32	66.90	-4.30
	7	46	53.82	-1.07
	8	38	40.48	-0.39
	9	15	24.57	-1.94
	10	10	16.69	-1.64
3	3	17	24.62	-0.60
	4	60	64.76	-0.60
	5	89	80.08	-1.01
	6	63	90.68	-2.94
	7	94	93.88	0.01
	8	97	91.77	0.55
	9	126	84.43	4.57
	10	64	63.47	0.07
	11	43	54.58	-1.58
4	4	7	11.68	-1.38
	5	35	38.66	-0.59
	6	35	58.95	-3.14
	7	72	81.27	-1.04
	8	91	102.23	-1.13
	9	149	122.22	2.46
	10	250	139.30	9.55
	11	99	132.56	-2.96
	12	106	149.79	-3.65

forms and combinations of model violations; they only serve to illustrate the techniques. A final remark must be made with respect to the application of the techniques presented here. In practical situations, the amount of information that is produced by the statistical testing procedures may become quite overwhelming, especially if the number of response categories becomes large. Therefore it seems most convenient to work bottom-up, starting with a relatively small subset of items that constitute a scale and adding items as long as the model holds. In many instances it is unrealistic to assume that one scale holds for all items, and one has to be content with a number of subscales. Also the computation of the statistical tests has its limitations, which are mainly caused by the computation of elementary symmetric functions and the inversion of matrices. In the computer program written by the author, a test length of 96 score points was chosen as an upper bound for the application of the model tests. Pushing this upper bound further will result in numerical problems that may be hard to overcome. For most applications, however, a maximum test length of 96 score points is not unreasonable. When analyzing several tests in an incomplete design simultaneously, the total number of parameters that can be estimated can be much larger than 96, since it is only the length of the separate tests that limits the applicability of the techniques presented. In the next section the mathematical framework for the development of the statistical testing procedures will be described.

8. The Construction of Asymptotically χ^2 Distributed Quadratic Forms

Consider a k-dimensional response vector $\mathbf{x}$, $\mathbf{x}' = (x_1, \ldots, x_i, \ldots, x_k)$. To keep the theory presented in this section as general as possible, it will be assumed that for $i = 1, \ldots, k$, x_i assumes values in the set $\{0, 1, \ldots, v_i\}$, with $v_i > 0$.

Let N observations $\mathbf{x}$ be sampled and let the vector of observed proportions $\mathbf{p}$, $\mathbf{p}' = (p_1, \ldots, p_v)$ have a multinomial distribution with parameters N and $\boldsymbol{\pi}$, $\boldsymbol{\pi}' = (\pi_1(\boldsymbol{\phi}), \ldots, \pi_v(\boldsymbol{\phi}))$ and $\boldsymbol{\phi}$ a s-dimensional vector of parameters, $v = \Pi_{i=1}^{k} v_i$ and $s < v - 1$.

Let $\mathbf{y}$ be defined by $\mathbf{y} = N^{1/2}(\mathbf{p} - \hat{\boldsymbol{\pi}})$, where $\hat{\boldsymbol{\pi}}$ stands for the maximum likelihood estimate of $\boldsymbol{\pi}$. It is a well known result of asymptotic theory (see for instance Rao, 1973, p. 392; or Bishop, Fienberg, & Holland, 1975, p. 517) that under a number of regularity conditions defined by Birch (1964), which are assumed to be fulfilled in the sequel, the distribution of $\mathbf{y}$ converges to a multivariate normal distribution with expectation $\mathbf{0}$ and a covariance matrix given by

$$\Sigma = D_\pi - \boldsymbol{\pi}\boldsymbol{\pi}' - D_\pi^{1/2}A(A'A)^{-1}A'D_\pi^{1/2},$$

with D_π a $v \times v$ diagonal matrix of the elements $\pi_i(\boldsymbol{\phi})$, $i = 1, \ldots, v$ and A a $v \times s$ matrix defined by $A = D_\pi^{-1/2}(\partial\boldsymbol{\pi}/\partial\boldsymbol{\phi}')$. The parametrization of the model must be such that A is of full column rank.

From this result, it can be derived that $\mathbf{y}'\hat{D}_\pi^{-1}\mathbf{y}$ has an asymptotic χ^2 distribution with $v - s - 1$ degrees of freedom (Rao, 1973; Bishop et al., 1975). The objective of the present paper is to alter the aggregation level of the model test, in such a way that the model violations under consideration may show. This is accomplished by defining the transformation $\mathbf{z} = X\mathbf{y}$, where X is a $u \times v$ matrix of rank u and $\mathbf{z}$ is a u-dimensional vector of so-called deviates. It easily follows that $\mathbf{z}$ coverges in distribution to a u-variate normal distribution with expectation $\mathbf{0}$ and covariance matrix $X\Sigma X'$. The model tests considered in the next sections will all have the basic form

$$R = \mathbf{z}'\hat{W}^{-1}\mathbf{z}, \tag{31}$$

with $W = XD_\pi X'$ and $\hat{D}_\pi$ equal to D_π evaluated using the m.l. estimates of the model parameters.

The objective of the present section is to derive conditions which are sufficient for the asymptotic χ^2 distribution of R.

Using Rao (1973, p. 186) R has an asymptotic χ^2 distribution if $B =$

$W^{-1/2}X\Sigma X'(W')^{-1/2}$ is idempotent and the degrees of freedom are given by trace (B). B can be written as $B = I_u - B_1 - B_2$, with I_u an identity matrix of order u,

$$B_1 = W^{-1/2}X\boldsymbol{\pi}\boldsymbol{\pi}'X'(W')^{-1/2} \quad \text{and} \quad B_2 = W^{-1/2}XD_{\pi}^{1/2}A(A'A)^{-1}A'D_{\pi}^{1/2}X'(W')^{-1/2}.$$

Since $B^2 = I_u - B_1 - B_2 - B_1 + B_1^2 + B_1B_2 - B_2 + B_2B_1 + B_2^2$, it is sufficient to show that B_1 and B_2 are idempotent and $B_1B_2 = 0$.

Consider the matrix $T = D_{\pi}^{1/2}X'W^{-1}XD_{\pi}^{1/2}$. Since T is idempotent, it is a projection and its manifold is given by $M(D_{\pi}^{1/2}X')$ (see for instance Rao, 1973, sec. 1C). Let 1_v be a v-dimensional vector with all elements equal to one and let $\mathbf{c}$ be a u-dimensional vector of constants. It will be proved that R has an asymptotic χ^2 distribution if the following two conditions are fulfilled:

Condition 1.

$$\text{The columns of } A \text{ belong to } M(D_{\pi}^{1/2}X'). \tag{32}$$

Condition 2.

$$\text{There exist a vector of constants } \mathbf{c} \text{ such that } X'\mathbf{c} = 1_v. \tag{33}$$

Using the two conditions, the following three lemma's can be proved.

Lemma 1. $B_1^2 = B_1$ and trace $(B_1) = 1$.

Proof. Since X can be multiplied with a vector of constants $\mathbf{c}$ such that $X'\mathbf{c} = 1_v$ it follows that $D_{\pi}^{1/2}X'\mathbf{c} = \boldsymbol{\pi}^{1/2}$ and so $\boldsymbol{\pi}^{1/2} \in M(D_{\pi}^{1/2}X')$. But $B_1^2 = W^{-1/2}X\boldsymbol{\pi}\boldsymbol{\pi}'^{1/2}T\boldsymbol{\pi}^{1/2}\boldsymbol{\pi}'X'(W')^{-1/2} = W^{-1/2}X\boldsymbol{\pi}\boldsymbol{\pi}'^{-1/2}\boldsymbol{\pi}^{1/2}\boldsymbol{\pi}'X'(W')^{-1/2} = B_1$ and trace $(B_1) =$ trace $(\boldsymbol{\pi}'X'W^{-1}X\boldsymbol{\pi}) =$ trace $(\boldsymbol{\pi}'^{1/2}T\boldsymbol{\pi}^{1/2}) =$ trace $(\boldsymbol{\pi}'^{1/2}\boldsymbol{\pi}^{1/2}) = 1$. □

Lemma 2. $B_2^2 = B_2$ and trace $(B_2) = s$.

Proof. $B_2^2 = W^{-1/2}XD_{\pi}^{1/2}A(A'A)^{-1}A'TA(A'A)^{-1}A'D_{\pi}^{1/2}X'(W')^{-1/2} = B_2$ by Condition 1 and trace $(B_2) =$ trace $(TA(A'A)^{-1}A') =$ trace $((A'A)^{-1}A'A) = s$. □

Lemma 3. $B_1B_2 = 0$.

Proof. $B_1B_2 = W^{-1/2}X\boldsymbol{\pi}\boldsymbol{\pi}'^{1/2}TA(A'A)^{-1}A'D_{\pi}^{1/2} = 0$, by using condition 1 and $\boldsymbol{\pi}'^{1/2}A = \mathbf{0}$. □

From the three lemma's given, it follows that $(X\mathrm{y})'(\mathrm{X}D_{\pi}\mathrm{X}')^{-1}(\mathrm{X}\mathrm{y})$ has an asymptotic χ^2 distribution since $B^2 = B$, and the degrees of freedom are equal to trace $(B) =$ trace $(I_u - B_1 - B_2) = u - 1 - s$. From the well known fact that $\mathbf{y}'\hat{D}_{\pi}^{-1}\mathbf{y}$ converges in distribution to $\mathbf{y}'D_{\pi}^{-1}\mathbf{y}$ (see for instance Bishop et al., 1975, p. 515) it follows that $R = (X\mathbf{y})'(X\hat{D}_{\pi}X')^{-1}(X\mathbf{y})$ converges in distribution to $(X\mathbf{y})'(XD_{\pi}X')^{-1}(X\mathbf{y})$. The result is summarized in the following theorem.

Theorem. Let $N\mathbf{p}$ have a multinomial distribution with parameters N and $\boldsymbol{\pi}$, let $\boldsymbol{\pi}$ be parameterized by $\boldsymbol{\phi}$ such that $\partial\boldsymbol{\pi}/\partial\boldsymbol{\phi}'$ is of full column rank and the conditions of Birch are fulfilled. Then $R = \mathbf{z}'W^{-1}\mathbf{z}$ has an asymptotic χ^2 distribution with $u - s - 1$ degrees of freedom, if the Conditions 1 and 2 are valid.

Notice that the number of degrees of freedom is equal to the number of deviates on which the test is based, minus the number of paramers that have to be estimated, minus one. The theorem can be applied to the R_1 test in the following way. To keep things relatively simple, it will be first assumed that the design consists of only one test, the generalization to more than one test will be treated later.

Let the theoretical probability of response pattern $\mathbf{x}$ is given by

$$\pi_{\mathbf{x}} = \int \exp\,(r(\mathbf{x})\vartheta)\,\exp\,(-\mathbf{x}'\boldsymbol{\eta})P_0(\vartheta)g(\vartheta|\sigma)\partial\vartheta, \tag{34}$$

with $r(\mathbf{x})$ the sum score associated with $\mathbf{x}$, $P_0(\vartheta) = \Pi_{i=1}^{I}\,(1 + \Sigma_{k=1}^{m_i}\exp\,(k\vartheta - \eta_{ij}))^{-1}$, $\boldsymbol{\eta}$ a vector of item parameters with elements η_{ij} ($i = 1, \ldots, I$ and $j = 1, \ldots, m_i$) and $g(\vartheta|\sigma)$ the normal density function with $\mu = 0$. Let X_r be a $J \times v(r)$ matrix with as columns all $v(r)$ possible response patterns leading to a sum score r, and let $\boldsymbol{\pi}_r$ and $\mathbf{p}_r$ be the vector of, respectively, the theoretical and the empirical probabilities of the response patterns leading to a sum score r, for $r = 1, \ldots, J - 1$, where J stands for the maximum score that can be obtained on the test. It is assumed that the orderings of the response patterns used to define X_r, $\boldsymbol{\pi}_r$ and $\mathbf{p}_r$ are identical. Consider the expression $\mathbf{z}_r = N^{1/2}X_r(\mathbf{p}_r - \boldsymbol{\pi}_r)$. It can easily be verified that $\mathbf{z}_r$ has elements z_{rij}($i = 1, \ldots, I$ and $j = 1, \ldots, m_i$) defined by $z_{rij} = N^{1/2}\Sigma_{(*)}p_{\mathbf{x}} - \pi_{\mathbf{x}}$, where the summation runs over the set $\{\mathbf{x}|r(\mathbf{x}) = r \text{ and } x_{ij} = 1\}$. Since $\Sigma_{(*)}\exp\,(-\mathbf{x}'\boldsymbol{\eta}) = \varepsilon_{ij}\Gamma_{r-j}^{(i)}(\boldsymbol{\varepsilon})$ an elementary symmetric function as defined in section 5, of the parameters $\varepsilon_{kh} = \exp\,(-\eta_{kh})$, for $k = 1, \ldots, I$ and $h = 1, \ldots, m_k$, z_{rij} can also be given by

$$z_{rij} = N^{-1/2}\left(n_{rij} - N\int \varepsilon_{ij}\Gamma_{r-j}^{(i)}(\boldsymbol{\varepsilon})\,\exp\,(r\vartheta)P_0(\vartheta)g(\vartheta|\sigma)\partial\vartheta\right), \tag{35}$$

where n_{rij} stands for the number of respondents who obtain a sum score r and complete item i in category j.

In section 6 it was argued that $z_{rij} = 0$ automatically, for certain combinations of i, j and r. Therefore these uninformative elements are removed from the statistical testing procedure by introducing a matrix X_r^*, which can be derived from X_r by removing all rows which which are associated with a combinations of indices (i, j) such that $r < j$ or $r > J(t) - m_i + j$. So if X_r had $\Sigma_i m_i$ rows, X_r will have $e_r^* = \Sigma_i \min\,(m_i, r) - \max\,(1, J - m_i + j) + 1$ rows. Let X be defined by

$$X = \begin{bmatrix} 1 & & & & & \\ & X_1^* & & & & \\ & & X_2^* & & 0 & \\ & 0 & & X_r^* & & \\ & & & & X_{J-1}^* & \\ & & & & & 1 \end{bmatrix}. \tag{36}$$

If v stands for the number of possible response patterns, X has $u = 2 + \Sigma_{r=1}^{J-1}e_r^*$ rows and v columns.

If $\mathbf{z} = N^{1/2}X(\mathbf{p} - \hat{\boldsymbol{\pi}})$ and $W = XD_\pi X'$, with D_π a diagonal matrix of the elements of $\boldsymbol{\pi}$, the model test defined by (28) can now be given as $R_1 = \mathbf{z}'\hat{W}^{-1}\mathbf{z}$. The theoretical probabilities are evaluated using m.l. estimates of the parameters $\boldsymbol{\xi}$ with $\boldsymbol{\xi}' = (\boldsymbol{\eta}', \sigma)$, or $\boldsymbol{\beta}$ with $\boldsymbol{\xi} = G\boldsymbol{\beta}$, whatever the case may be. The following lemma concerns the first condition.

Lemma 4. The columns of A belong to $M(D_\pi^{1/2}X')$.

Proof. T has a block diagonal form diag $(1, T_1, \ldots, T_r, \ldots, T_{J-1}, 1)$, where $T_r = D_{\pi(r)}^{1/2}X_r^{*\prime}(X_r^*D_{\pi(r)}X_r^{*\prime})^{-1}X_r^*D_{\pi(r)}^{1/2}$, with $D_{\pi(r)}$ a diagonal matrix of the elements of

$\boldsymbol{\pi}_r$. It can be verified that $\partial\boldsymbol{\pi}_r/\partial\boldsymbol{\eta}' = D_{\pi(r)}X'_r - \boldsymbol{\alpha}_r\boldsymbol{\xi}'_r$, with $\boldsymbol{\alpha}_r$ a vector with elements exp $(-\mathbf{x}'\boldsymbol{\eta})$ for all $\mathbf{x}$ with $r(\mathbf{x}) = r$, and $\boldsymbol{\xi}_r$ a vector with elements $\int\psi_{ij}(\vartheta)$ exp $(r\vartheta)P_0(\vartheta)P_0(\vartheta)g(\vartheta|\sigma)\partial\vartheta$ for $i = 1, \ldots, I$ and $j = 1, \ldots, m_i$. It can also be verified that $\partial\boldsymbol{\pi}_r/\partial\sigma = \boldsymbol{\alpha}_r\sigma^{-3}\int\vartheta^2$ exp $(r\vartheta)P_0(\vartheta)g(\vartheta|\sigma)\partial\vartheta - \sigma^{-1}\boldsymbol{\pi}_r$. Define $A_r = D_{\pi(r)}^{-1/2}(\partial\boldsymbol{\pi}_r/\partial\boldsymbol{\xi}')(\partial\boldsymbol{\xi}/\partial\boldsymbol{\beta}')$. Then $T_rA_r = A_r$, since $D_{\pi(r)}^{1/2}X_r^*$, $D_{\pi(r)}^{1/2}\boldsymbol{\alpha}_r$ and $D_{\pi(r)}^{1/2}\boldsymbol{\pi}_r$ are all elements of $M(T_r)$. □

The second condition is dealt with in the next lemma.

Lemma 5. There exist a vector of constants $\mathbf{c}$ such that $X'\mathbf{c} = \mathbf{1}_v$.

Proof. Consider the J-dimensional vector $\mathbf{a}$ which is defined by $(1, 2, \ldots, m_1, \ldots, 1, 2, \ldots, m_i, \ldots, 1, 2, \ldots, m_I)$. Since every response pattern in X_r leads to a sum score r, $X'_r\mathbf{a} = r\mathbf{1}_{v(r)}$, with $\mathbf{1}_{v(r)}$ *a* $v(r)$-dimensional vector with all elements equal to one. Let $\mathbf{c}_r$ be equivalent to $\mathbf{a}$ with the elements associated with the removed rows form X_r also removed, such that the product $X_r^{*\prime}\mathbf{c}_r$ is well defined. Then if $\mathbf{c}$ is a u-dimensional vector defined by $c' = (1, \ldots, (1/r)\mathbf{c}_r, \ldots, 1)$, it follows that $X'\mathbf{c} = \mathbf{1}_v$. □

So in the case of a test design with one test and one population only, the model test defined by (28) has $u - s - 1$ degrees of freedom, since both the conditions specified by (32) and (33) are valid.

The theory presented so far, concerns situations where the sampling model has a multinomial form. In the case of a design with more than one test however, the sampling model is a product of multinomial models, because within every test the theoretical probabilities of the response patterns sums to one. Birch (1963) and Haberman (1974) have shown that the asymptotic theory on m.l. estimation procedures and statistical testing procedures for parameterized product multinomial models can be translated into equivalent procedures for multinomial models. This can also be applied to the present problem.

For $t = 1, \ldots, T$, let N_t be the size of a sample and $\boldsymbol{\pi}_t$ be the vector of theoretical probabilities. Further N stands for the total sample size. Assume that the stochastic variables N_t, $t = 1, \ldots, T$, have a multinomial distribution with theoretical probabilities ω_t, for $t = 1, \ldots, T$. Notice that $\hat{\omega}_t = N_t/N$. Using this definition it can easily be seen that the elements of the amalgated vector of probabilities $(\omega_1\boldsymbol{\pi}'_1, \ldots, \omega_t\boldsymbol{\pi}'_t, \ldots, \omega_T\boldsymbol{\pi}'_T)$ sum to one. So by adding $T - 1$ "dummy" parameters a product-multinomial model is transformed into a multinomial one. With this re-definition of the model, the theory developed for the multinomial model directly applies and the quadratic form $R = \mathbf{z}'\hat{W}^{-1}\mathbf{z}$ defined above has an asymptotic χ^2 distribution with $u - s - T$ degrees of freedom. The derivation of the R_0 test proceeds as follows.

Assume that only one test is considered and that $\boldsymbol{\xi}$ is not subject to linear restrictions. Consider a matrix of contrasts X^* defined by

$$X^* = \begin{bmatrix} \mathbf{0} & X_1 & X_2 & \ldots & X_r & \ldots & X_{J-1} & \mathbf{1} \\ 1 & \mathbf{0}' & \mathbf{0}' & \ldots & \mathbf{0}' & \ldots & \mathbf{0}' & 0 \\ 0 & \mathbf{1}'_{v(1)} & \mathbf{0}' & \ldots & \mathbf{0}' & \ldots & \mathbf{0}' & 0 \\ 0 & \mathbf{0}' & \mathbf{1}'_{v(2)} & \ldots & \mathbf{0}' & \ldots & \mathbf{0}' & 0 \\ 0 & \mathbf{0}' & \mathbf{0}' & \ldots & \mathbf{1}'_{v(r)} & \ldots & \mathbf{0}' & 0 \\ 0 & \mathbf{0}' & \mathbf{0}' & \ldots & \mathbf{0}' & \ldots & \mathbf{1}'_{v(J-1)} & 0 \\ 0 & \mathbf{0}' & \mathbf{0}' & \ldots & \mathbf{0}' & \ldots & \mathbf{0}' & 1 \end{bmatrix} \tag{37}$$

where X_r is a $J \times v(r)$ matrix with as columns all response patterns leading to a sum score r and $\mathbf{1}_{v(r)}$ stands for a $v(r)$-dimensional vector with all elements equal to one. The definition of the other elements in X^* is now obvious. Notice that X^* has $2J + 1$ rows and v columns. The construction of X^* can be motivated in the following manner.

Consider the vector $X^*\boldsymbol{\pi}$. The elements $\mathbf{1}_r\boldsymbol{\pi}_r$ are given by $\mathbf{1}'_r\boldsymbol{\pi}_r = \int\Gamma_r(\boldsymbol{\varepsilon}) \exp(r\vartheta)P_0(\vartheta)g(\vartheta|\mu, \sigma)\partial\vartheta$, so these elements are proportional to the expectation of the number of persons obtaining a sum score r. The other elements of $X^*\boldsymbol{\pi}$ are given by $X_\eta\boldsymbol{\pi} = [\mathbf{0}, X_1, X_2, \ldots, X_r, \ldots, X_{J-1}, 1]\boldsymbol{\pi}$, so $X_\eta\boldsymbol{\pi}$ has elements

$$\sum_{r=1}^{J} \int \varepsilon_{ij}\Gamma_{r-j}^{(i)}(\boldsymbol{\varepsilon}) \exp(r\vartheta)P_0(\vartheta)g(\vartheta|\sigma)\partial\vartheta = \int \psi_{ij}(\vartheta)g(\vartheta|\sigma)\partial\vartheta. \tag{38}$$

In the sequel it will become clear, that these elements must be added, to account for the degrees of freedom associated with the estimation of item parameters. The matrix X_η contains the sufficient statistics for the item parameters, which also motivates its subscript. Let Y be a matrix such that $X' = [X'_\eta, Y']$. The theory presented above cannot be applied to X^*, because this matrix is not of full row rank: a v-dimensional zero vector can be constructed by summing all rows in X_η and subtracting J times the J-th row in Y. Therefore a matrix of contrasts X will be constructed by removing the first row of X^*.

Lemma 6. The columns of A belong to $M(D_\pi^{1/2}X')$.

Proof. Inspection of Lemma 4 reveals that $\partial\boldsymbol{\pi}/\partial\boldsymbol{\eta}'$ can be written as $\partial\boldsymbol{\pi}/\partial\boldsymbol{\eta}' = D_\pi X'_\eta - \int\boldsymbol{\pi}(\vartheta)\boldsymbol{\pi}(\vartheta)'g(\vartheta|\sigma)\partial\vartheta X'_\eta$, with $\boldsymbol{\pi}(\vartheta)$ a vector of the probabilities of all response patterns as a function of ϑ. So it easily follows that $D_\pi^{-1/2}\partial\boldsymbol{\pi}/\partial\boldsymbol{\eta}'$ is an element of $M(D_\pi^{1/2}X'_\eta)$. In the same manner it can be shown that $D_\pi^{-1/2}\partial\boldsymbol{\pi}/\partial\sigma \in M(D_\pi^{1/2}Y')$. Observing that $D_\pi^{1/2}X^{*\prime}$ and $D_\pi^{1/2}X'$ have the same linear manifold concludes the proof. □

Since the rows of Y sum to $\mathbf{1}_v$, both the Conditions (32) and (33) are fulfilled and if $\mathbf{z} = N^{1/2}X(\mathbf{p} - \hat{\boldsymbol{\pi}})$ and $W = XD_\pi X'$, $R_0 = \mathbf{z}'\hat{W}^{-1}\mathbf{z}$ has an asymptotic χ^2 distribution with $J - 2$ degrees of freedom. With respect to the computation of the model test, the following remarks are in order. Let X_η^* be equal to X_η with the first row removed. Let $\mathbf{z}$ be defined by $\mathbf{z}' = (\mathbf{z}'_1, \mathbf{z}'_2)$ with $\mathbf{z}_1 = NY(\mathbf{p} - \hat{\boldsymbol{\pi}})$ and $\mathbf{z}_2 = NX_\eta(\mathbf{p} - \hat{\boldsymbol{\pi}})$ and let W be partitioned as

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} = \begin{bmatrix} YD_\pi Y' & YD_\pi X_\eta^{*\prime} \\ X_\eta^* D_\pi Y' & X_\eta^* D_\pi X_\eta^{*\prime} \end{bmatrix}. \tag{39}$$

By carrying out the summations over the probabilities of response patterns defined by X_η^* and Y, it can be verified that W_{22} is a $(J - 1) \times (J - 1)$ matrix with diagonal elements

$$\sum_{r=0}^{k} \int \varepsilon_{ij}\Gamma_{r-j}^{(i)}(\boldsymbol{\varepsilon}) \exp(r\vartheta)P_0(\vartheta)g(\vartheta|\sigma)\partial\vartheta, \tag{40}$$

for $i = 1, \ldots, I$ and $j = 1, \ldots, m_i$, without the element $i = j = 1$. W_{22} has off-diagonal elements

$$\sum_{r=0}^{k} \int \varepsilon_{ij}\varepsilon_{i'j''}\Gamma_{r-j'-j}^{(i,i')}(\boldsymbol{\varepsilon}) \exp(r\vartheta)P_0(g(\vartheta|\sigma)\partial\vartheta, \tag{41}$$

if $i \neq i'$ and off-diagonal elements equal to zero if $i = i'$. W_{21} is a $(J - 1) \times J$ matrix with elements $\int \varepsilon_{ij}\Gamma_{r-j}^{(i)}(\boldsymbol{\varepsilon}) \exp(r\vartheta)P_0(\vartheta)g(\vartheta|\sigma)\partial\vartheta$, for all relevant combinations of i and j and $r = 0, \ldots, J$ and W_{11} is a diagonal matrix of the elements $\int \Gamma_r(\boldsymbol{\varepsilon}) \exp(r\vartheta)P_0(\vartheta)g(\vartheta|\sigma)\partial\vartheta$. So all matrices are functions of elementary symmetric functions and can be computed. In actual data analysis, both with real and artificial data, it turns out that at $\hat{\xi}s_{ij} = N\int \hat{\psi}_{ij}(\vartheta)g(\vartheta|\hat{\sigma})\partial\vartheta$, with s_{ij} the number of responses in category j on item i. But

$$\int \psi_{ij}(\vartheta)g(\vartheta|\sigma)\partial\vartheta = \sum_{r=0}^{J}\int \varepsilon_{ij}\Gamma_{r-j}^{(i)}(\boldsymbol{\varepsilon}) \exp(r\vartheta)P_0(\vartheta)g(\vartheta|\sigma)\partial\vartheta$$

and so it follows that $X_\eta \mathbf{p} = X_\eta \hat{\boldsymbol{\pi}}$ and the elements of $\mathbf{z}_2$ are equal to zero.

Although the conjecture that $\mathbf{z}_2 = \mathbf{0}$ at $\hat{\boldsymbol{\xi}}$ is never rejected in the data analysis carried out by the author, a proof of the conjecture can, as yet, not be given. Assuming that the conjecture is correct, R can also be given by $R_0 = \mathbf{z}_1'(W_{11} - W_{12}W_{22}^{-1}W_{21})^{-1}\mathbf{z}_1$.

With respect to the proof, the consequences of linear restrictions of the form $\boldsymbol{\xi} = G\boldsymbol{\beta}$ can be handled by redefining A as $A = D_\pi^{-1/2}(\partial\boldsymbol{\pi}/\partial\boldsymbol{\xi}')(\partial\boldsymbol{\xi}/\partial\boldsymbol{\beta}') = D_\pi^{-1/2}(\partial\boldsymbol{\pi}/\partial\boldsymbol{\xi}')G$ and lemma 6 still holds. However, at $\hat{\boldsymbol{\beta}}$ the identity $X_\eta \mathbf{p} = X_\eta \pi$ no longer holds, so the test statistic will also contain information with respect to the degrees in which the sufficient statistics of the original item parameters and their expected values have been matched under the restricted model.

9. Discussion

The partial credit model has several appealing properties which compare favourably with the properties of other latent trait models for polytomous data (Molenaar, 1983). Sufficient statistics for the item parameters exist and the scoring rule used to assess the ability of a subject corresponds with common practice in education and research. The existence of sufficient statistics for the item parameters and the person parameters makes it possible to obtain consistent estimates of one set of parameters, not depending on the parameters in the other set. One way to obtain these estimates is maximizing a likelihood function which is conditional upon sufficient statistics for one set of parameters. Although this approach has several theoretical and practical advantages, and is definitely worth pursuing, the present paper deals with the problem of obtaining consistent estimates of one set of the parameters, the item parameters, by assuming that the other set of parameters, the ability parameters, is a random sample from a normal distribution. The most important reason for developing this procedure as an alternative option to the conditional approach, is the fact that the so-called marginal approach can handle a more general class of test administration designs.

The present paper shows that adopting the marginal approach leads to a comprehensive framework for parameter estimation, the derivation of asymptotic confidence intervals and the derivation of statistical testing procedures, both for the partial credit model applied to data sampled in an incomplete design and for models that can be derived from the partial credit model by imposing linear restrictions on the parameters.

In the last section of the present paper it is shown that an incorrect assumption about the ability distribution causes serious bias in the parameter estimates. The incorrect assumption is however easily detected by the proposed statistical testing procedure. Besides a model test that has power against an incorrect specification of the ability distribution, a model test is presented that makes it possible to identify misfitting items. As a final remark, it must be noted that the present paper gives some conditions which are necessary for the existence of a unique solution of the likelihood equations, but it fails to identify sufficient conditions. In practical situations parameter estimation does not encounter any serious trouble, but from a theoretical point of view both necessary and sufficient conditions for the existence of a unique solution to the likelihood equations would be very welcome.

Appendix A: The Derivation of the Information Matrix for the Marginal Partial Credit Model

When using the EM algorithm, the observed information matrix evaluated at the maximum likelihood estimate of the parameters $\hat{\boldsymbol{\lambda}}$, is given by

$$I(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\lambda}}) = -E(B(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\lambda}})|X, D, \hat{\boldsymbol{\lambda}}) - E(\mathbf{b}(\hat{\boldsymbol{\lambda}}), \mathbf{b}(\hat{\boldsymbol{\lambda}})'|X, D, \hat{\boldsymbol{\lambda}}), \tag{A1}$$

where $\mathbf{b}(\boldsymbol{\lambda})$ stands for the vector of first order derivatives of $\ln L(\boldsymbol{\lambda}|X, D, \boldsymbol{\vartheta})$ with respect to $\boldsymbol{\lambda}$ and $B(\boldsymbol{\lambda}, \boldsymbol{\lambda})$ stands for the matrix of second order derivatives of $\ln L(\boldsymbol{\lambda}|X, D, \boldsymbol{\vartheta})$. First, the general result will be given, then some examples of the derivation of the elements will be given. The derivation of the remaining elements can be easily reconstructed by applying the same principles as used in the examples. The information matrix has the following no-zero elements:

$$I(\hat{\eta}_{ij}, \hat{\eta}_{ij}) = \sum_t d_{ti} \sum_r n_r \{E(\hat{\psi}_{ij}(\vartheta)(1 - \hat{\psi}_{ij}(\vartheta))|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}}) - \text{var}\,(\hat{\psi}_{ij}(\vartheta)|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}})\}, \tag{A2}$$

$$I(\hat{\eta}_{ij}, \hat{\eta}_{il}) = \sum_t d_{ti} \sum_r n_r \{-E(\hat{\psi}_{ij}(\vartheta)\hat{\psi}_{il}(\vartheta)|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}}) - \text{cov}\,(\hat{\psi}_{ij}(\vartheta), \hat{\psi}_{il}(\vartheta)|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}})\}, \ (j \neq 1), \tag{A3}$$

$$I(\hat{\eta}_{ij}, \hat{\eta}_{kl}) = -\sum_t d_{ti} d_{tk} \sum_r n_r \,\text{cov}\,(\hat{\psi}_{ij}(\vartheta), \hat{\psi}_{kl}(\vartheta)|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}}), \ (i \neq k), \tag{A4}$$

$$I(\hat{\eta}_{ij}, \hat{\sigma}_t) = -d_{ti}\hat{\sigma}_t^{-2} \sum_r n_r \,\text{cov}\,(\hat{\psi}_{ij}(\vartheta), \vartheta|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}}), \tag{A5}$$

$$I(\hat{\mu}_t, \hat{\mu}_t) = \sigma_t^{-4} \sum_r n_r (E(\vartheta|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}}) - \mu_t)^2, \tag{A6}$$

$$I(\hat{\sigma}_t, \hat{\sigma}_t) = 2N_t/\hat{\sigma}_t^2 - \hat{\sigma}_t^{-6} \sum_r n_r \,\text{var}\,(\vartheta^2|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}}), \tag{A7}$$

$$I(\hat{\mu}_t, \hat{\sigma}_t) = -\hat{\sigma}_t^{-5} \sum_r n_r \,\text{cov}\,(\vartheta, \vartheta^2|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}}), \tag{A8}$$

where $i = 1, \ldots, I$, $k = 1, \ldots, I$, $j = 1, \ldots, m_i$, $l = 1, \ldots, m_i$ and $t = 1, \ldots, T$. First, the diagonal elements for the item parameters will be derived.

Lemma A1. $I(\hat{\eta}_{ij}, \hat{\eta}_{ij}) = \Sigma_t d_{ti} \Sigma_r n_r \{E(\hat{\psi}_{ij}(\vartheta)(1 - \hat{\psi}_{ij}(\vartheta))|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}}) - \text{var}\,(\hat{\psi}_{ij}(\vartheta|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}})\}$.

Proof. Using $b(\eta_{ij}) = -s_{ij} + \Sigma_t \Sigma_{n \in \{n\}_t} \psi_{ij}(\vartheta_n)$ and

$B(\eta_{ij}, \eta_{ij}) = \partial b(\eta_{ij})/\partial \eta_{ij} = \sum_t \sum_{n \in \{n\}_t} \psi_{ij}(\vartheta_n)(1 - \psi_{ij}(\vartheta_n))$ results in

$$E(B(\hat{\eta}_{ij}, \hat{\eta}_{ij})|X, D, \boldsymbol{\lambda}) = -\sum_t d_{ti} \sum_r n_r E(\hat{\psi}_{ij}(\vartheta)(1 - \hat{\psi}_{ij}(\vartheta))|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}})$$

$$\text{for } i = 1, \ldots, I \text{ and } j = 1, \ldots, m_i. \quad (9)$$

Also $E(b(\hat{\eta}_{ij})^2|X, D, \hat{\lambda}) = E((-s_{ij} + \sum_t \sum_{n \in \{n\}_t} \hat{\psi}_{ij}(\vartheta_n))^2|X, D, \hat{\boldsymbol{\lambda}})$, but due to (15),

$$s_{ij} = \sum_t d_{ti} \sum_{n \in \{n\}} E(\hat{\psi}_{ij}(\vartheta_n)|X, D, \hat{\boldsymbol{\lambda}}),$$

$$\text{so } E(b(\hat{\eta}_{ij})^2|X, D, \hat{\boldsymbol{\lambda}}) = \sum_t d_{ti} \sum n_r \text{ var } (\hat{\psi}_{ij}(\vartheta)|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}}). \quad \text{(A10)}$$

Combination of (A9) and (A10) gives the desired diagonal element. □

The elements (A3) and (A4) can be derived in the same manner.

Lemma A2. $I(\hat{\mu}_t, \hat{\mu}_t) = \underline{\sigma_t^{-4}}\Sigma_r n_r(E(\vartheta|r, \mathbf{d}_t, \hat{\boldsymbol{\lambda}}) - \mu_t)^2$.

Proof. The summations below only concern $n \in \{n\}_t$ and $m \in \{n\}_t$. For μ_t the information matrix has an element

$$-E(B(\hat{\mu}_t, \hat{\mu}_t)|X, D, \hat{\boldsymbol{\lambda}}) - E(b(\hat{\mu}_t)^2|X, D, \hat{\boldsymbol{\lambda}})$$

$$= \left(\frac{N_t}{\hat{\sigma}_t^2}\right) - E\left(\left(\sum_n \frac{\vartheta_n - \hat{\mu}_t}{\hat{\sigma}_t^2}\right)^2 |X, \hat{D}, \boldsymbol{\lambda}\right)$$

$$= \left(\frac{N_t}{\hat{\sigma}_t^2}\right) - \sigma_t^{-4}\left\{E\left(\sum_n(\vartheta_n - \mu_t)^2 + \sum_n\sum_m(\vartheta_n - \mu_t)(\vartheta_m - \mu_t)|X, D, \hat{\boldsymbol{\lambda}}\right)\right\}$$

$$= \left(\frac{N_t}{\hat{\sigma}_t^2}\right) - \left(\frac{N_2}{\hat{\sigma}_t^2}\right) - \hat{\sigma}_t^{-4}\left\{\sum_n E((\vartheta_n - \mu_t)|X, D, \hat{\boldsymbol{\lambda}})\right\}^2 + \hat{\sigma}_t^{-4}\sum_n(E(\vartheta_n - \mu_t|X, D, \hat{\boldsymbol{\lambda}}))^2.$$

The fact that the first two terms cancel and the third term is equal to zero due to estimation equation (16) gives the result. □

The elements (A5), (A7) and (A8) can be derived in the same manner.

References

Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society, Series B*, *34*, 42–50.

Andersen, E. B. (1973). Conditional inference and models for measuring. Kopenhagen: Mentalhygienjnisk Forskningsinstitut.

Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika*, *42*, 69–81.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W., (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: M.I.T. Press.

Birch, M. W. (19863). Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society, Series B*, *25*, 220–233.

Birch, M. W. (1964). A new proof of the Pearson-Fisher theorem. *Annals of Mathematical Statistics*, *35*, 817–824.

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

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

Efron, B., & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, *65*, 457–487.

Fischer, G. H. (1974). *Einführung in die theorie psychologischer tests*. Bern: Verlag Hans Huber.

Fischer, G. H. (1981). On the existence and uniqueness of maximum likelihood estimates in the Rasch model. *Psychometrika*, *46*, 59–77.

Glas, C. A. W. (1988). The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika*, *53*, 525–546.

Haberman, S. J. (1974). The analysis of frequency data. Chicago: University of Chicago Press.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society*, *Series B*, *44*, 226–233.

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

Masters, G. N. (1985). Comparing latent trait and latent class analysis of Likert type data. *Psychometrika*, *50*, 69–82.

Masters, G. N., & Wright, B. D. (1984). The essential process in a family of measurement models. *Psychometrika*, *49*, 529–544.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika*, *49*, 359–381.

Molenaar, I. W. (1983). Item steps. *Heymans Bulletins* (Report No. HB-83-630-EX). Groningen: Rijks Universiteit Groningen.

Rao, C. R. (1973). Linear statistical inference and its applications (2nd ed.). New York: Wiley.

Rigdon, S. E., & Tsutakawa, R. K. (1983). Parameter estimation in latent trait models. *Psychometrika*, *48*, 567–574.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, *47*, 175–186.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, MESA Press.

Manuscript received 4/1/87
Final version received 8/13/88