Polytomous IRT models and monotone likelihood ratio of the total score

Hemker, B.T.; Sijtsma, K.; Molenaar, I.W.; Junker, B.W.

Published in:
Psychometrika

Publication date:
1996

Citation for published version (APA):
Hemker, B. T., Sijtsma, K., Molenaar, I. W., & Junker, B. W. (1996). Polytomous IRT models and monotone likelihood ratio of the total score. *Psychometrika*, *61*(4), 679-693.

PSYCHOMETRIKA—VOL. 61, NO. 4, 679–693
DECEMBER 1996

POLYTOMOUS IRT MODELS AND MONOTONE LIKELIHOOD RATIO OF THE TOTAL SCORE

BAS T. HEMKER AND KLAAS SIJTSMA
UTRECHT UNIVERSITY

IVO W. MOLENAAR
UNIVERSITY OF GRONINGEN

BRIAN W. JUNKER
CARNEGIE MELLON UNIVERSITY

In a broad class of item response theory (IRT) models for dichotomous items the unweighted total score has monotone likelihood ratio (MLR) in the latent trait θ. In this study, it is shown that for polytomous items MLR holds for the partial credit model and a trivial generalization of this model. MLR does not necessarily hold if the slopes of the item step response functions vary over items, item steps, or both. MLR holds neither for Samejima's graded response model, nor for nonparametric versions of these three polytomous models. These results are surprising in the context of Grayson's and Huynh's results on MLR for nonparametric dichotomous IRT models, and suggest that establishing stochastic ordering properties for nonparametric polytomous IRT models will be much harder.

Key words: IRT models, monotone likelihood ratio, nonparametric IRT models, parametric IRT models, polytomous items.

Introduction

In the behavioral and social sciences, tests and questionnaires are often used to measure the position of respondents on a latent trait θ. Let a test consist of L dichotomously scored items. The score on item i is denoted X_i: $X_i = 1$ for correct responses and $X_i = 0$ otherwise. Both in classical test theory and in item response theory (IRT), the number-correct score or count of the positive answers X_+ is used to order subjects with respect to θ. It is thus desirable that a higher total score corresponds to a higher expected latent trait value.

Under the very mild conditions of unidimensionality, local independence and item response functions (IRFs) $P(X_i = 1|\theta)$ that are nondecreasing in θ, Grayson (1988) and Huynh (1994) have shown that X_+ has monotone likelihood ratio (MLR) in θ. This means that for $0 \leq C < K \leq L$

$$g(K, C; \theta) = \frac{P(X_+ = K|\theta)}{P(X_+ = C|\theta)} \tag{1}$$

is a nondecreasing function of θ. Grayson also uses the requirements that $0 < P(X_i = 1|\theta) < 1$ and that $dP(X_i = 1|\theta)/d\theta$ exists to prove MLR of X_+. The first requirement is not very

Hemker's research was supported by the Netherlands Research Council, Grant 575-67-034. Junker's research was supported in part by the National Institutes of Health, Grant CA54852, and by the National Science Foundation, Grant DMS-94.04438.

Requests for reprints should be sent to Bas T. Hemker, National Institute for Educational Measurement (Cito), PO Box 1034, 6801 MG Arnhem, THE NETHERLANDS.

0033-3123/96/1200-95272$00.75/0

strong in practice because every IRF that does not meet this requirement can be replaced by an IRF that closely resembles it and that does meet the requirement. The second requirement is not needed in the proof of MLR of X_+ by Huynh (1994).

Because MLR implies stochastic ordering (SO) (Lehmann, 1959, p. 74; see also, Junker, 1993, Proposition 4.1), the three mild conditions imply for dichotomous items that θ is stochastically ordered by X_+, that is, for any constant value w of θ,

$$P(\theta > w|X_+ = C) \leq P(\theta > w|X_+ = K).$$

Because the MLR property is symmetric in its arguments, one may also infer that X_+ is stochastically ordered by θ, that is, for any two respondents a and b with $\theta_a < \theta_b$,

$$P(X_+ \geq K|\theta_a) \leq P(X_+ \geq K|\theta_b). \tag{2}$$

Note that because (2) takes the ordering on θ as a starting point, it is probably of less interest to the practical use of tests where only the ordering on X_+ can be observed and inferences with respect to θ are drawn on the basis of X_+. It should be noted that the SO property does not imply MLR (Lehmann, 1959, sec. 3.3; see also, Junker, 1993, Example 4.1; Rosenbaum, 1985).

It may be noted that the MLR/SO results hold for the Rasch (1960) model in which X_+ is a sufficient statistic for the maximum likelihood estimation of θ. In addition, the MLR/SO results also hold for the two- and three-parameter logistic models (e.g., Lord, 1980) and the One-Parameter Logistic Model (OPLM; Verhelst & Glas, 1995) in which X_+ is not sufficient for the maximum likelihood estimation of θ. This implies that MLR and SO are not equivalent with sufficiency of X_+ and may hold for models in which the full data patterns are used to produce an optimal estimate of θ. MLR and SO even hold for nonparametric models such as the Mokken models (Meijer, Sijtsma, & Smid, 1990; Mokken & Lewis, 1982) since proofs of MLR (e.g., Grayson, 1988; Huynh, 1994) use the nondecreasingness of the IRF, but do not require a particular parametric choice of this function.

The generalization of Grayson's results to IRT models for polytomous items with ordered answer categories will be studied in this paper. Like Grayson, we will investigate MLR for a simple total score X_+, which is the sum of the individual item scores. Again this total score is a sufficient statistic in some models such as the Partial Credit Model (PCM; Masters, 1982), whereas other models such as the Graded Response Model (GRM; Samejima, 1969) use the full data patterns to estimate θ. The use of X_+ has a long history that goes back to Likert (1932) and is very popular with researchers who construct tests and questionnaires, and with test users with whom results are more easily communicated in terms of the simple unweighted total scores. The interest in X_+ is further motivated by the result that X_+ correlates highly with many weighted sums of the item scores that may—strictly speaking—be more appropriate to estimate θ under particular IRT models. Additional inducement to the use of X_+ can be found in the ordinal consistency results of Stout (1990) and Junker (1991).

In this paper we first consider several parametric models, in particular, the PCM (Masters, 1982), a generalization of this model to which a discrimination parameter is added (the two-parameter PCM: 2p-PCM), and the GRM (Samejima, 1969). Second, if the MLR property holds for one or more parametric models, nonparametric versions based on weaker assumptions will be defined and it will be studied whether these weaker models also possess the MLR property.

Monotone likelihood ratio in polytomous IRT models

For notational convenience we will assume throughout that all items from a test have equal numbers of answer categories. It can easily be checked that our results remain valid for unequal numbers. Let each of the L items have $m + 1$ ordered answer categories which are scored $X_i = 0, \ldots, m$, respectively. We will consider the unweighted sum score on these L items, $X_+ = \Sigma_{i=1}^{L} X_i$, which is also known as the Likert score. The definition of MLR of X_+ in θ is almost the same as in the dichotomous case, see (1), with the only difference being that in the polytomous case $X_+ = 0, \ldots, mL$ rather than $X_+ = 0, \ldots, L$, and thus $0 \leq C < K \leq mL$.

Let the first derivative of a function with respect to θ be denoted by means of a prime. In the sequel, all derivatives are with respect to θ. Differentiating the likelihood ratio $g(K, C; \theta)$, we see that MLR holds if and only if

$$P'(X_+ = K|\theta) * P(X_+ = C|\theta) - P'(X_+ = C|\theta) * P(X_+ = K|\theta) \geq 0.$$

The numbers of score vectors that yield $X_+ = K$ or $X_+ = C$ are denoted by R_K and R_C, respectively. Vectors containing scores on the L items of the test are denoted $\mathbf{X}$; if the elements of a vector sum to K, this vector is denoted $\mathbf{X}_{(u)}$ where the index u ($u = 1, \ldots, R_K$) identifies the particular vector. Similarly, $\mathbf{X}_{(v)}$ denotes the vectors summing to C, where index v ($v = 1, \ldots, R_C$) identifies the particular vector. Using this notation, MLR holds if and only if

$$\sum_{u=1}^{R_K} P'(\mathbf{X}_{(u)}|\theta) * \sum_{v=1}^{R_C} P(\mathbf{X}_{(v)}|\theta) - \sum_{v=1}^{R_C} P'(\mathbf{X}_{(v)}|\theta) * \sum_{u=1}^{R_K} P(\mathbf{X}_{(u)}|\theta) \geq 0. \tag{3}$$

For a more compact notation, let the conditional probability $P(X_i = x_i|\theta)$ be denoted by π_{ix}, with $x_i = 0, 1, \ldots, m$. Due to local independence, $P(\mathbf{X}|\theta) = \Pi_{i=1}^{L} \pi_{ix}$. Consequently, (3) can be written as

$$\sum_{u=1}^{R_K} \left[\prod_{i=1}^{L} \pi_{ix(u)}\right]' * \sum_{v=1}^{R_C} \prod_{i=1}^{L} \pi_{ix(v)} - \sum_{v=1}^{R_C} \left[\prod_{i=1}^{L} \pi_{ix(v)}\right]' * \sum_{u=1}^{R_K} \prod_{i=1}^{L} \pi_{ix(u)} \geq 0,$$

which is equivalent to

$$\sum_{u=1}^{R_K} \sum_{v=1}^{R_C} \left(\left[\prod_{i=1}^{L} \pi_{ix(u)}\right]' * \prod_{i=1}^{L} \pi_{ix(v)} - \left[\prod_{i=1}^{L} \pi_{ix(v)}\right]' * \prod_{i=1}^{L} \pi_{ix(u)} \right) \geq 0. \tag{4}$$

Since

$$\left[\prod_{i=1}^{L} \pi_{ix}\right]' = \sum_{i=1}^{L} \frac{\pi'_{ix}}{\pi_{ix}} * \prod_{i=1}^{L} \pi_{ix},$$

(4) can be written as

$$\sum_{u=1}^{R_K} \sum_{v=1}^{R_C} \left(\sum_{i=1}^{L} \frac{\pi'_{ix(u)}}{\pi_{ix(u)}} * \prod_{i=1}^{L} \pi_{ix(u)} * \prod_{i=1}^{L} \pi_{ix(v)} - \sum_{i=1}^{L} \frac{\pi'_{ix(v)}}{\pi_{ix(v)}} * \prod_{i=1}^{L} \pi_{ix(u)} * \prod_{i=1}^{L} \pi_{ix(v)} \right) \geq 0.$$

Thus, MLR of X_+ in θ holds if and only if

$$\sum_{u=1}^{R_K} \sum_{v=1}^{R_C} \left(\sum_{i=1}^{L} \left[\frac{\pi'_{ix(u)}}{\pi_{ix(u)}} - \frac{\pi'_{ix(v)}}{\pi_{ix(v)}} \right] * \prod_{i=1}^{L} \pi_{ix(u)} * \pi_{ix(v)} \right) \geq 0. \tag{5}$$

This equation is used in the sequel to show whether a model implies MLR.

Monotone likelihood ratio in parametric IRT models

The Partial Credit Model

In the PCM (Masters, 1982), π_{ix} is defined as

$$\pi_{ix} = \frac{\exp\left[\sum_{j=1}^{x_i} (\theta - \delta_{ij})\right]}{\sum_{k=0}^{m} \exp\left[\sum_{j=1}^{k} (\theta - \delta_{ij})\right]}, \tag{6}$$

where δ_{ij} is the difficulty of step j on item i, and $\Sigma_{j=1}^{0} (\theta - \delta_{ij}) \equiv 0$ for notational convenience. Because the PCM is a member of the exponential family (Andersen, 1980, pp. 272–274) and X_+ is the sufficient statistic for θ (Masters, 1982), the MLR property holds by implication. In this section, we will provide our own line of proof using (5) which does not use the fact that the PCM is an exponential family model. This line of proof will then be used to investigate MLR in other, nonexponential family models.

Straightforward algebra shows that the derivative of log π_{ix} is

$$\frac{\pi'_{ix}}{\pi_{ix}} = x_i - \frac{\sum_{k=0}^{m} k \exp\left[\sum_{j=1}^{k} (\theta - \delta_{ij})\right]}{\sum_{k=0}^{m} \exp\left[\sum_{j=1}^{k} (\theta - \delta_{ij})\right]},$$

which is a ratio that can be substituted in (5). The summation of this ratio across all L items yields

$$\sum_{i=1}^{L} \frac{\pi'_{ix}}{\pi_{ix}} = X_+ - \sum_{i=1}^{L} \frac{\sum_{k=0}^{m} k \exp\left[\sum_{j=1}^{k} (\theta - \delta_{ij})\right]}{\sum_{k=0}^{m} \exp\left[\sum_{j=1}^{k} (\theta - \delta_{ij})\right]},$$

with $X_+ = K$ for $\Sigma_i (\pi'_{ix(u)}/\pi_{ix(u)})$ and $X_+ = C$ for $\Sigma_i (\pi'_{ix(v)}/\pi_{ix(v)})$. This means that in the case of the PCM, the left-hand side of (5) becomes

$$\sum_{u=1}^{R_K} \sum_{v=1}^{R_C} \left(\sum_{i=1}^{L} \left[\frac{\pi'_{ix(u)}}{\pi_{ix(u)}} - \frac{\pi'_{ix(v)}}{\pi_{ix(v)}} \right] * \prod_{i=1}^{L} \pi_{ix(u)} * \pi_{ix(v)} \right) = \sum_{u=1}^{R_K} \sum_{v=1}^{R_C} \left((K - C) * \prod_{i=1}^{L} \pi_{ix(u)} * \pi_{ix(v)} \right).$$

Since $K > C$, this means that the left-hand side of (5) is positive; therefore, X_+ has MLR in θ in the PCM.

The Two-Parameter Partial Credit Model

A more flexible model than (6) can be defined by adding a positive discrimination parameter α_{ij}. This model can be denoted the two-parameter partial credit model (2p-PCM), in which

$$\pi_{ix} = \frac{\exp\left[\sum_{j=1}^{x_i} \alpha_{ij}(\theta - \delta_{ij})\right]}{\sum_{k=0}^{m} \exp\left[\sum_{j=1}^{k} \alpha_{ij}(\theta - \delta_{ij})\right]}. \tag{7}$$

Note that this definition of π_{ix} is identical to the nominal response model (NRM; Bock, 1972) if nominal response categories are assumed. It will be shown that the validity of MLR for the 2p-PCM depends on α_{ij}.

If α_{ij} has the same value for all items and for all item steps, say $\alpha_{ij} = \alpha$ (Andrich, 1978), then the MLR property still holds. This can easily be seen by noting that if $\alpha_{ij} = \alpha$, the left-hand side of (5) becomes

$$\sum_{u=1}^{R_K} \sum_{v=1}^{R_C} \left(\sum_{i=1}^{L} \left[\frac{\pi'_{ix(u)}}{\pi_{ix(u)}} - \frac{\pi'_{ix(v)}}{\pi_{ix(v)}} \right] * \prod_{i=1}^{L} \pi_{ix(u)} * \pi_{ix(v)} \right)$$

$$= \sum_{u=1}^{R_K} \sum_{v=1}^{R_C} \left(\alpha(K - C) * \prod_{i=1}^{L} \pi_{ix(u)} * \pi_{ix(v)} \right)$$

for all $R_K * R_C$ possible combinations of $\mathbf{X}_{(u)}$ and $\mathbf{X}_{(v)}$. Because $\alpha > 0$, the derivative of the likelihood ratio $g(K, C; \theta)$ is nonnegative; therefore, MLR holds. Note that the case with $\alpha_{ij} = \alpha$ is a trivial generalization of the original PCM (in which $\alpha_{ij} = 1$) and that this model is also a member of the exponential family. If α_{ij} varies over items, or item steps, or both, however, the 2p-PCM does not in general imply the MLR property.

First, we will show that MLR does hold for the likelihood ratio $g(K, C; \theta)$ with $K = mL$ and C free ($K > C$). Along the same lines it can be shown that MLR holds for $g(K, C; \theta)$ with $C = 0$ and K free ($K > C$). We will omit the proof for $C = 0$ because it is similar to the proof for $K = mL$. Next, we will show by means of two numerical counterexamples that MLR does not hold in general if K and C are not maximal or minimal, respectively.

MLR for maximum K or minimum C. In the 2p-PCM the ratio π'_{ix}/π_{ix} from (5) can be written as

$$[\log \pi_{ix}]' = \sum_{j=1}^{x_i} \alpha_{ij} - \psi_i; \text{ with } \psi_i = \left\{ \log \sum_{k=0}^{m} \exp\left[\sum_{j=1}^{k} \alpha_{ij}(\theta - \delta_{ij})\right]\right\}'.$$

Note that if $K = mL$ then the item scores are all equal to m. Note also that there is only one score vector $\mathbf{X}$ with total sum mL. This means that the left-hand side of (5) can be simplified as

$$\sum_{v=1}^{R_C} \left(\sum_{i=1}^{L} \left[\frac{\pi'_{im}}{\pi_{im}} - \frac{\pi'_{ix(v)}}{\pi_{ix(v)}} \right] * \prod_{i=1}^{L} \pi_{im} * \pi_{ix(v)} \right).$$

It has to be shown that the difference term between brackets is nonnegative for all items i and all score vectors v:

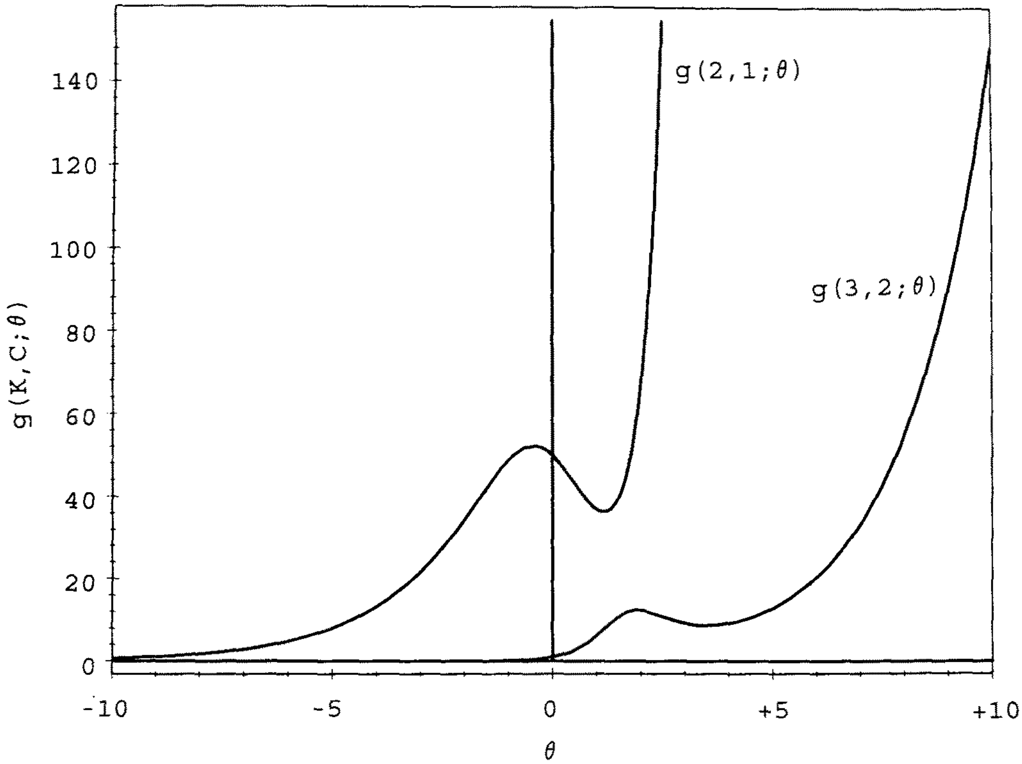

FIGURE 1.
MLR functions based on two items from the 2p-PCM with equal discrimination parameters across item steps per item ($\alpha_{ij} = \alpha_i$; i.e., the g-PCM). Precise parameter values in text.

$$\frac{\pi'_{im}}{\pi_{im}} - \frac{\pi'_{ix(v)}}{\pi_{ix(v)}} = \sum_{j=1}^{m} \alpha_{ij} - \psi_i - \sum_{j=1}^{x_{i(v)}} \alpha_{ij} + \psi_i$$

$$= \sum_{j=x_{i(v)}+1}^{m} \alpha_{ij}.$$

This sum equals 0 if $x_{i(v)} = m$ and is positive otherwise because $\alpha_{ij} > 0$. This completes the proof. A similar argument works for $C = 0$.

Counterexamples of MLR for $K < mL$ and $C > 0$. In the first counterexample, the discrimination parameter varies over items but not over item steps: $\alpha_{ij} = \alpha_i$. Thus, the generalized PCM (g-PCM; Muraki, 1992) is obtained. In this model, the left-hand side of (5) equals

$$\sum_{u=1}^{R_K} \sum_{v=1}^{R_C} \left(\sum_{i=1}^{L} [(x_{i(u)} - x_{i(v)})\alpha_i] \prod_{i=1}^{L} \pi_{ix(u)} * \pi_{ix(v)} \right). \tag{8}$$

Say, there are two items, indexed $i = 1, 2$, each with three answer categories, $x_i = 0, 1, 2$, that satisfy the 2p-PCM with item-specific discrimination parameters (i.e., the g-PCM). Let $K = 2$ and $C = 1$, then $R_K = 3$ [(2, 0), (1, 1), and (0, 2)] and $R_C = 2$ [(1, 0) and (0, 1)]. Let $\alpha_1 = 2$ and $\alpha_2 = 0.5$, and let $\delta_{11} = \delta_{12} = \delta_{21} = 0$ and $\delta_{22} = -2 \log 98 = -9.17$. Figure 1 shows the likelihood ratios $g(K + 1, K; \theta)$ for $K = 1, 2$.

For $L = 2$ and $m = 2$ nondecreasingness of $g(1, 0; \theta)$ and $g(4, 3; \theta)$ follows from the results concerning maximum K or minimum C and, therefore, does not provide any information about MLR. Furthermore, it is easily checked that $g(3, 1; \theta)$ is the product of $g(2, 1; \theta)$ and $g(3, 2; \theta)$. In general, if K and C differ by at least 2, $g(K, C; \theta)$ is the product of $K - C$ functions $g(K^*, K^* - 1; \theta)$,

$$g(K, C; \theta) = \prod_{K^*=C+1}^{K} g(K^*, K^* - 1; \theta).$$

We examined many examples and it was always found that if K and C differ by at least 2 the corresponding likelihood ratio was monotonely increasing; therefore, Figure 1 and later figures do not show such MLR functions. In Figure 1, $g(2, 1; \theta)$ decreases across approximately 1.5 units on the θ scale and $g(3, 2; \theta)$ decreases more to the right across approximately 1.5 units. Consequently, both curves reveal that for particular ranges of θ the ratio of the probability of a total score $K + 1$ and the probability of a total score K becomes smaller rather than larger as predicted by the MLR property.

Note that the vertical scale of the function $g(K, C; \theta)$ in Figure 1 has been compressed so that at first sight the derivative might appear smaller than it actually is. To give an example of the actual values of the derivative of $g(K, C; \theta)$ consider, for example, $\theta = 0$ for which $\pi_{10} = \pi_{11} = \pi_{12} = 0.33$, and $\pi_{20} = \pi_{21} = 0.01$ and $\pi_{22} = 0.98$. Substitution of these probabilities, the item scores, and the discrimination parameters in the quotient rule formula for the first derivative of $g(K, C; \theta)$, of which (8) is the numerator, yields $g'(2, 1; 0) = -10.25$ after some cumbersome algebra.

The second counterexample concerns the 2p-PCM with item step-specific discrimination parameters: $\alpha_{ij} = \alpha_j$. In this case, the left-hand side of (5) equals

$$\sum_{u=1}^{R_K} \sum_{v=1}^{R_C} \left(\sum_{i=1}^{L} \left[\sum_{j=1}^{x_{i(u)}} \alpha_j - \sum_{j=1}^{x_{i(v)}} \alpha_j \right] \prod_{i=1}^{L} \pi_{ix(u)} * \pi_{ix(v)} \right).$$

Note that this model is substantially different from the model with item-specific discrimination parameters. Suppose, there are two items, indexed $i = 1, 2$, each with four answer categories, indexed $x_i = 0, 1, 2, 3$. Using these restrictions, for $K = 3$ and $C = 2$, it follows that $R_K = 4$ [(3, 0), (2, 1), (1, 2), and (0, 3)] and $R_C = 3$ [(2, 0), (1, 1), and (0, 2)]. Furthermore, assume that $\delta_{11} = \delta_{12} = \delta_{13} = \delta_{21} = \delta_{22} = 0$ and $\delta_{23} = -\log 97 = -4.57$. Let $\alpha_1 = 5$ and $\alpha_2 = \alpha_3 = 1$. Figure 2 gives the functions $g(K + 1, K; \theta)$, except for $K = 0$ and $K = 5$. For reasons mentioned earlier score differences larger than 1 are neglected. Figure 2 reveals that only for $K = 2$ the corresponding MLR function is locally decreasing across the interval between approximately $\theta = 0$ and $\theta = 1$.

To have an impression of the actual numerical values of the derivative which are obscured in Figure 2 by the compression of the vertical axis, consider $\theta = 0$ for which $\pi_{1x} = 0.25$ for all x_i ($x_i = 0, 1, 2, 3$), $\pi_{20} = \pi_{21} = \pi_{22} = 0.01$, and $\pi_{23} = 0.97$. Using these results, $g'(3, 2; 0) = -8.44$.

The model with item-specific discrimination parameters and the model with item step-specific discrimination parameters are both special cases of the 2p-PCM in which α_{ij} varies over items and over item steps. It is obvious that the two counterexamples show that this more general model does not imply MLR.

The Graded Response Model

Another parametric IRT model for polytomous items is the GRM (Samejima, 1969; see also Masters, 1982). Let $X_i = x_i$; $x_i = 1, \ldots, m$, and define discrimination parameters α_i, and threshold parameters λ_{ix}. Then the GRM is defined as

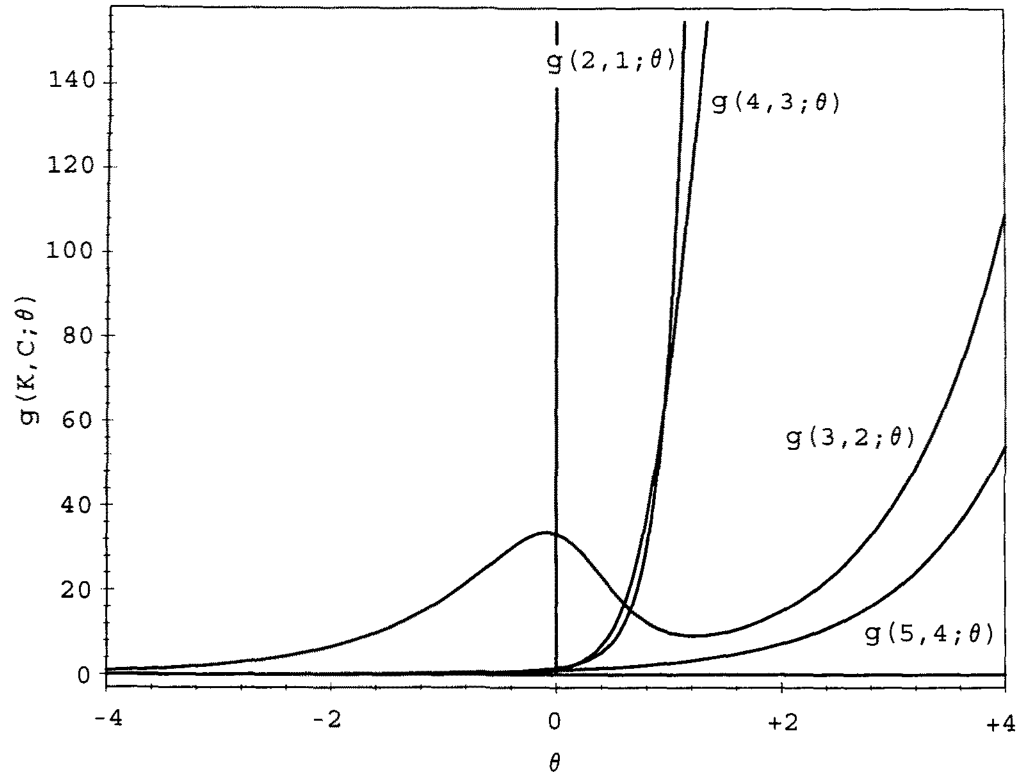

FIGURE 2.
MLR functions based on two items from the 2p-PCM with equal discrimination parameters across items per item step ($\alpha_{ij} = \alpha_j$). Precise parameter values in text.

$$P(X_i \geq x_i | \theta) = \frac{\exp[\alpha_i(\theta - \lambda_{ix})]}{1 + \exp[\alpha_i(\theta - \lambda_{ix})]}. \tag{9}$$

In this model, the discrimination parameters do not differ over item steps (Samejima, 1969, 1972; see also Thissen & Steinberg, 1986). Note that $P(X_i \geq 0|\theta) = 1$ and $P(X_i \geq m + 1|\theta) = 0$. The probability of having item score x_i is given by

$$\pi_{ix} = P(X_i \geq x_i|\theta) - P(X_i \geq x_i + 1|\theta).$$

It can be shown (Appendix 1, (A1)) that in the GRM, the derivative of log π_{ix}, which is a part of (5), equals

$$\frac{\pi'_{ix}}{\pi_{ix}} = \alpha_i[1 - P(X_i \geq x_i|\theta) - P(X_i \geq x_i + 1|\theta)]$$

$$= \alpha_i\left[1 - \pi_{ix} - 2\sum_{k=x+1}^{m} \pi_{ik}\right].$$

This means that for $x > y$

$$\frac{\pi'_{ix}}{\pi_{ix}} - \frac{\pi'_{iy}}{\pi_{iy}} = \alpha_i\left[\sum_{k=y}^{x-1} (\pi_{ik} + \pi_{i,k+1})\right]; \tag{10}$$

see also (5) and (Appendix 1, (A2)).

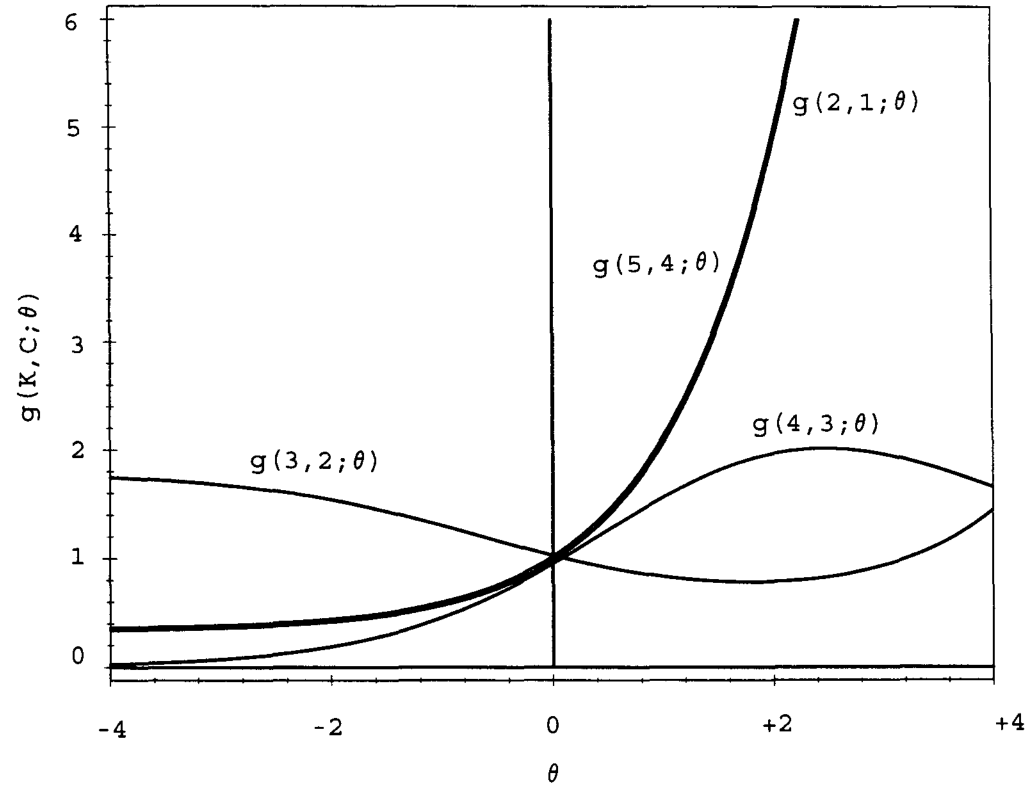

FIGURE 3.
MLR functions based on two items from the GRM. Precise parameter values in text.

Analogously to the 2p-PCM it can be shown for the GRM that MLR holds for $K = mL$ and C free ($K > C$) and $C = 0$ and K free ($K > C$) and counterexamples disproving MLR can be given for other values of K and C. The proof uses (10). If $K = mL$ then $x = m$ for all items, and $m \geq y$, and it follows readily that (10) and thus the left-hand side of (5) are nonnegative. For $C = 0$ a similar line of reasoning yields the same result.

The counterexample concerns two items ($L = 2$), each with four answer categories ($m = 3$). Let $K = 3$ and $C = 2$, then $R_K = 4$ [(3, 0), (2, 1), (1, 2), and (0, 3)] and $R_C = 3$ [(2, 0), (1, 1), and (0, 2)]. Let $\alpha_1 = \alpha_2 = 1$, and let $\lambda_{11} = \log (49/51) = -0.04$, $\lambda_{12} = 0$, $\lambda_{13} = \log (51/49) = 0.04$, $\lambda_{21} = \log (33/67) = -0.71$, $\lambda_{22} = \log (33/17) = 0.66$, and $\lambda_{23} = \log 99 = 4.60$. Figure 3 shows the corresponding functions $g(K + 1, K; \theta)$ for appropriate values of K. Of these functions, the two functions for $K = 2$ and $K = 3$ decrease slowly across relatively large intervals on the scale of θ.

As an illustration of the value of the derivative we consider $\theta = 0$ for which $\pi_{10} = \pi_{13} = 0.49$ and $\pi_{11} = \pi_{12} = 0.01$, and $\pi_{20} = \pi_{21} = \pi_{22} = 0.33$ and $\pi_{23} = 0.01$. Using these results, $g'(3, 2; 0) = -.25$.

Note that in contrast to the 2p-PCM, a constant value of the discrimination parameter does not imply MLR in the GRM. Obviously, the more general case in which α_i varies over items can also result in negative values of $g'(K, C; \theta)$.

Nonparametric models

Since Grayson's (1988) results hold in a nonparametric framework—unidimensionality, local independence, and IRFs that are nondecreasing in θ—for dichotomous items

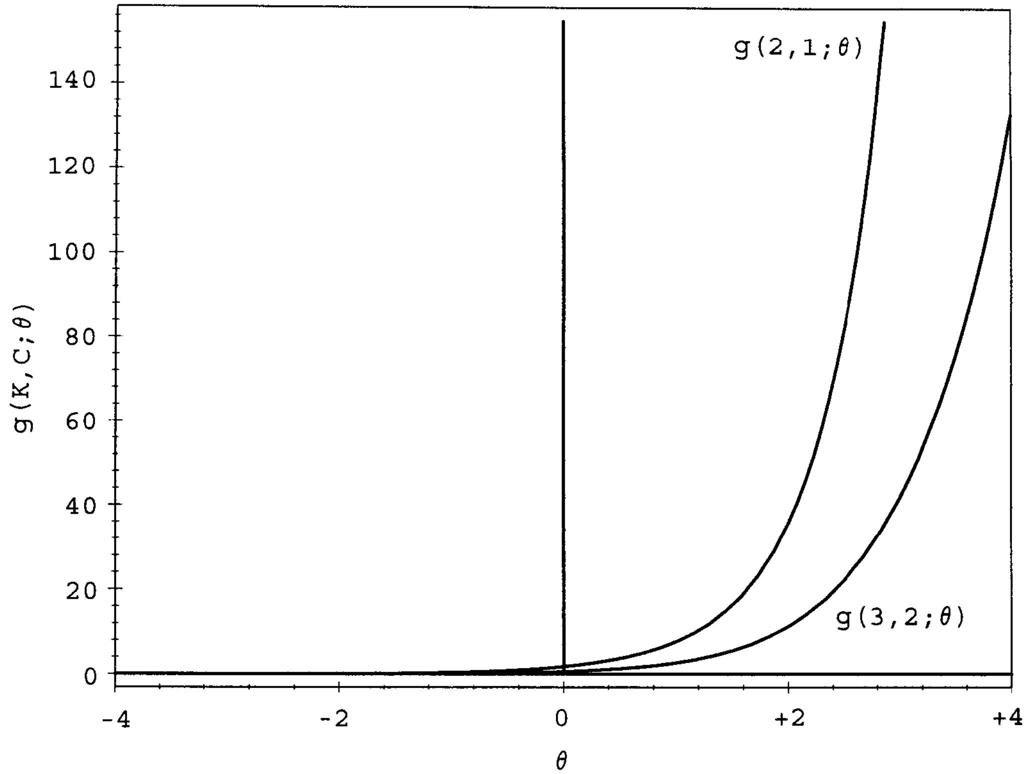

FIGURE 4.
MLR functions based on two items from the 2p-PCM with $\alpha_{ij} = \alpha_i$ (i.e., the g-PCM). The parameters are: $\alpha_{1.} = 2$, $\alpha_{2.} = 1$, $\delta_{11} = -.5$, $\delta_{12} = .5$, $\delta_{21} = -1$, $\delta_{22} = 1$.

with several parametric IRT models as a special case, we will next discuss MLR results for two nonparametric models for polytomous items. At least two nonparametric models can be defined that are based on the parametric models discussed here: the nonparametric Partial Credit Model (np-PCM), and the nonparametric Graded Response Model (np-GRM). Both nonparametric models are defined by three assumptions: unidimensionality, local independence and item step response functions (ISRFs) that are nondecreasing in θ. The two models, however, differ in the definition of the ISRFs.

In the np-PCM an ISRF is defined by

$$P(Y_{ij} = 1|\theta) = \frac{\pi_{ij}}{\pi_{ij} + \pi_{i,j-1}},$$

where $P(Y_{ij} = 1|\theta)$ is the ISRF of step j in item i. The original PCM and the 2p-PCM are both special cases of the np-PCM, since in the PCM [Equation (6)],

$$P(Y_{ij} = 1|\theta) = \frac{\pi_{ij}}{\pi_{ij} + \pi_{i,j-1}} = \frac{\exp(\theta - \delta_{ij})}{1 + \exp(\theta - \delta_{ij})},$$

and in the 2p-PCM [Equation (7)]

$$P(Y_{ij} = 1|\theta) = \frac{\pi_{ij}}{\pi_{ij} + \pi_{i,j-1}} = \frac{\exp[\alpha_{ij}(\theta - \delta_{ij})]}{1 + \exp[\alpha_{ij}(\theta - \delta_{ij})]}.$$

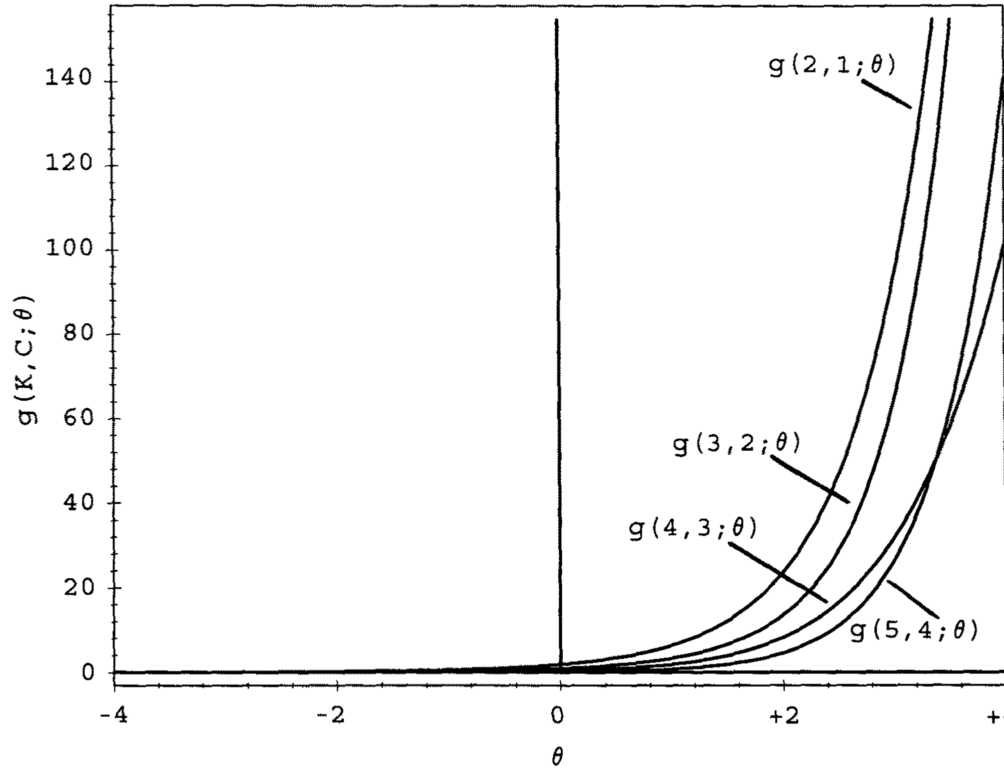

FIGURE 5.
MLR functions based on two items from the 2p-PCM with $\alpha_{ij} = \alpha_j$. The parameters are: $\alpha_{.1} = 1$, $\alpha_{.2} = 1.5$, $\alpha_{.3} = 2$, $\delta_{11} = -1$, $\delta_{12} = 0$, $\delta_{13} = 1$, $\delta_{21} = -.5$, $\delta_{22} = .5$, $\delta_{23} = 2$.

Obviously, both functions are nondecreasing in θ. Note that in the 2p-PCM, $P(Y_{ij} = 1|\theta)$ is nondecreasing because α_{ij} is positive, and that this does not depend on whether α_{ij} varies or not. Since it is known that there are special cases of the np-PCM that do not possess the MLR property, it can be concluded that the np-PCM does not imply MLR.

In the np-GRM an ISRF is defined as the probability of having at least an item score x_i, $P(X_i \geq x_i|\theta)$. Note that this definition of the np-GRM is identical to the definition of the Mokken model of monotone homogeneity for polytomous items (Molenaar, 1982, 1986, in press).

The definition of $P(X_i \geq x_i|\theta)$ in Samejima's GRM is given by (9). This is a nondecreasing function in θ, which implies that Samejima's GRM is a special case of the np-GRM. Because it was shown that Samejima's GRM does not possess the MLR property, it can be concluded that the np-GRM does not imply MLR.

Discussion

The only model for polytomous items investigated here that implies MLR is the parametric PCM in which $\alpha_{ij} = \alpha$. Of course, MLR also holds for models that are special cases of this model, like the original PCM (Masters, 1982) and the rating scale model (Andrich, 1978). Counterexamples were found for the 2p-PCM with varying α_{ij}, and the GRM.

Note that in the two counterexamples involving the 2p-PCM the ordering of the location parameters within one of the two items considered was reversed in comparison with the ordering of the item steps. The location parameter of the highest step was large

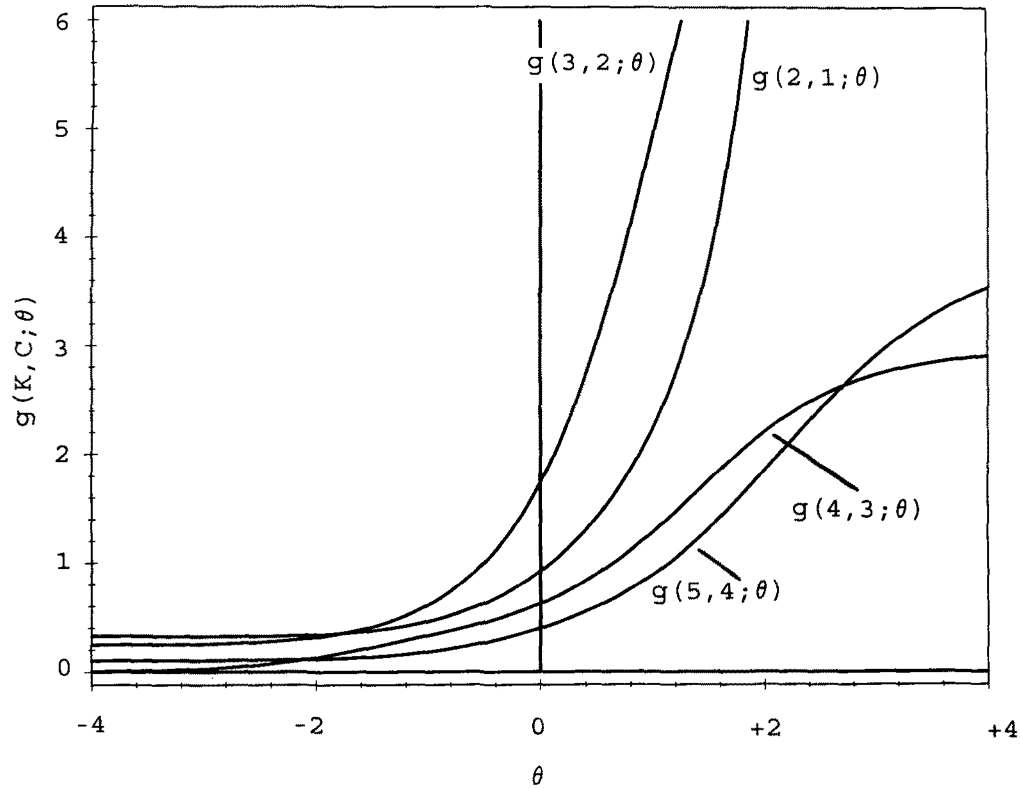

FIGURE 6.
MLR functions based on two items from the GRM. The parameters are: $\alpha_{1.} = 2$, $\alpha_{2.} = 1$, $\lambda_{11} = \log(1/3) = -1.10$, $\lambda_{12} = 0$, $\lambda_{13} = \log 3 = 1.10$, $\lambda_{21} = \log(1/2) = -.69$, $\lambda_{22} = \log 2 = .69$, $\lambda_{23} = \log 8 = 2.08$.

negative whereas the location parameters of the other steps were 0. Note that in the GRM a conflict like this is impossible by definition (Masters, 1982). In the counterexample with respect to the GRM the locations were widely spaced for one item and closely for the other item. In Figures 4, 5, and 6 for each of the models an example is given, again for $L = 2$ and for m varying across the examples, but for considerably less extreme and thus more realistic values of the item parameters. Many other choices for the 2p-PCM yielded monotonically increasing MLR functions. For the GRM even realistic choices would sometimes yield locally decreasing MLR functions, however. A simulation study in which all the relevant characteristics are systematically manipulated may provide more precise information.

The two nonparametric models may also result in negative values for the derivative of $g(K, C; \theta)$. Additional restrictions on the models might result in nonparametric models that yield MLR. However, simple additional restrictions that do not considerably limit the practical usefulness of the models have not yet been found.

It can be concluded that for polytomous items a less general class of IRT models has the MLR property than for dichotomous items. However, it can not be concluded that the PCM is the only model that implies the property of SO, because SO does not imply MLR (Junker, 1993). To find out which polytomous models have the SO property, further research is needed.

Appendix 1

Let item i satisfy the GRM with discrimination parameter α_i and threshold parameters λ_{ix} ($x = 1, \ldots, m$), then in (5)

$$\frac{\pi'_{ix}}{\pi_{ix}} = \alpha_i\left[1 - \pi_{ix} - 2\sum_{k=x+1}^{m}\pi_{ik}\right]. \tag{A1}$$

Next, let item score x_i be larger than item score y_i, and subtract the ratio involving the smaller score y_i from the ratio involving x_i, then in (5)

$$\frac{\pi'_{ix}}{\pi_{ix}} - \frac{\pi'_{iy}}{\pi_{iy}} = \alpha_i\sum_{j=y}^{x-1}\{\pi_{ij} + \pi_{i,j+1}\}. \tag{A2}$$

Proof of (A1). In the GRM

$$P(X_i \geq x_i|\theta) = \frac{\exp[\alpha_i(\theta - \lambda_{ix})]}{1 + \exp[\alpha_i(\theta - \lambda_{ix})]},$$

for all $x_i = 1, \ldots, m$, and with $P(X_i \geq 0|\theta) = 1$ and $P(X_i \geq m + 1|\theta) = 0$. Item score probability π_{ix} equals $P(X_i \geq x_i|\theta) - P(X_i \geq x_i + 1|\theta)$. This means that in this model, π_{ix} equals

$$\pi_{ix} = \frac{\exp[\alpha_i(\theta - \lambda_{ix})]}{1 + \exp[\alpha_i(\theta - \lambda_{ix})]} - \frac{\exp[\alpha_i(\theta - \lambda_{i,x+1})]}{1 + \exp[\alpha_i(\theta - \lambda_{i,x+1})]}$$

$$= \frac{\exp[\alpha_i(\theta - \lambda_{ix})] - \exp[\alpha_i(\theta - \lambda_{i,x+1})]}{\{1 + \exp[\alpha_i(\theta - \lambda_{ix})]\} * \{1 + \exp[\alpha_i(\theta - \lambda_{i,x+1})]\}}$$

for all $x_i = 1, \ldots, m$. For notational convenience, $f_{ix}(\theta) = \exp[\alpha_i(\theta - \lambda_{ix})] - \exp[\alpha_i(\theta - \lambda_{i,x+1})]$, and $g_{ix}(\theta) = \{1 + \exp[\alpha_i(\theta - \lambda_{ix})]\} * \{1 + \exp[\alpha_i(\theta - \lambda_{i,x+1})]\}$. Using this short-hand notation, we find that $[\log \pi_{ix}]'$ equals

$$\left[\log\left(\frac{f_{ix}(\theta)}{g_{ix}(\theta)}\right)\right]' = \frac{f'_{ix}(\theta)}{f_{ix}(\theta)} - \frac{g'_{ix}(\theta)}{g_{ix}(\theta)}.$$

Let $\exp[\alpha_i(\theta - \lambda_{ix})] = \exp(\kappa_{ix})$ for notational convenience, then $\{\exp[\alpha_i(\theta - \lambda_{ix})]\}' = \alpha_i * \exp(\kappa_{ix})$, and thus

$$f'_{ix}(\theta) = \alpha_i * [\exp(\kappa_{ix}) - \exp(\kappa_{i,x+1})]; \text{ and}$$

$$g'_{ix}(\theta) = [1 + \exp(\kappa_{ix}) + \exp(\kappa_{i,x+1}) + \exp(\kappa_{ix})\exp(\kappa_{i,x+1})]'$$

$$= \alpha_i * [\exp(\kappa_{ix}) + \exp(\kappa_{i,x+1}) + 2\exp(\kappa_{ix})\exp(\kappa_{i,x+1})].$$

This implies that

$$[\log \pi_{ix}]' = \alpha_i - \alpha_i * \left(\frac{\exp(\kappa_{ix}) + \exp(\kappa_{i,x+1}) + 2\exp(\kappa_{ix})\exp(\kappa_{i,x+1})}{[1 + \exp(\kappa_{ix})] * [1 + \exp(\kappa_{i,x+1})]}\right).$$

This means that

$$\frac{\pi'_{ix}}{\pi_{ix}} = \alpha_i * \left(1 - \frac{\exp[\alpha_i(\theta - \lambda_{ix})]}{1 + \exp[\alpha_i(\theta - \lambda_{ix})]} - \frac{\exp[\alpha_i(\theta - \lambda_{i,x+1})]}{1 + \exp[\alpha_i(\theta - \lambda_{i,x+1})]}\right)$$

which equals

$$\frac{\pi'_{ix}}{\pi_{ix}} = \alpha_i[1 - P(X_i \geq x_i|\theta) - P(X_i \geq x_i + 1|\theta)].$$

It can be proven that for $x_i = 0$ and $x_i = m$ this equation also holds. In terms of item score probabilities we thus have for all item scores

$$\frac{\pi'_{ix}}{\pi_{ix}} = \alpha_i\left[1 - \pi_{ix} - 2 \sum_{k=x_i+1}^{m} \pi_{ik}\right]. \qquad \square$$

Proof of (A2). Writing y rather than x, we have

$$\frac{\pi'_{ix}}{\pi_{iy}} = \alpha_i\left[1 - \pi_{iy} - 2 \sum_{k=y+1}^{m} \pi_{ik}\right].$$

Thus,

$$\frac{\pi'_{y}}{\pi_{ix}} - \frac{\pi'_{iy}}{\pi_{iy}} = \alpha_i\left[-\pi_{ix} - 2 \sum_{k=x+1}^{m} \pi_{ik} + \pi_{iy} + 2 \sum_{k=y+1}^{m} \pi_{ik}\right].$$

If $x > y$ this is equal to

$$\frac{\pi'_{ix}}{\pi_{ix}} - \frac{\pi'_{iy}}{\pi_{iy}} = \alpha_i\left[-\pi_{ix} + \pi_{iy} + 2 \sum_{k=y+1}^{x} \pi_{ik}\right],$$

which can be written as

$$\frac{\pi'_{ix}}{\pi_{ix}} - \frac{\pi'_{iy}}{\pi_{iy}} = \alpha_i \sum_{k=y}^{x-1} (\pi_{ik} + \pi_{i,k+1}). \qquad \square$$

References

Andersen, E. B. (1980). *Discrete statistical models with social science applications*. Amsterdam: North Holland.
Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika, 43*, 561–573.
Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29–51.
Ellis, J. L., & van den Wollenberg, A. L. (1993). Local homogeneity in latent trait models. A characterization of the homogeneous monotone IRT model. *Psychometrika, 58*, 417–429.
Grayson, D. A. (1988). Two-group classification in latent trait theory: Scores with monotone likelihood ratio. *Psychometrika, 53*, 383–392.
Huynh, H. (1994). A new proof for monotone likelihood ratio for the sum of independent bernoulli random variables. *Psychometrika, 59*, 77–79.
Junker, B. W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika, 56*, 255–278.
Junker, B. W. (1993). Conditional association, essential independence and monotone unidimensional item response models. *The Annals of Statistics, 21*, 1359–1378.
Lehmann, E. L. (1959). *Testing statistical hypotheses*. New York: Wiley.
Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 149–158.
Lord, F. M. (1980). *Applications 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.
Meijer, R. R., Sijtsma, K., & Smid, N. G. (1990). Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement, 14*, 283–298.
Mokken, R. J. (1971). *A theory and procedure of scale analysis*. New York/Berlin: De Gruyter.

Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement, 6,* 417–430.

Molenaar, I. W. (1982). Mokken scaling revisited. *Kwantitatieve Methoden, 3*(8), 145–164.

Molenaar, I. W. (1986). Een vingeroefening in item response theorie voor drie geordende antwoordcategorieën [An exercise in item response theory for three ordered response categories]. In G. F. Pikkemaat & J. J. A. Moors (Eds.), *Liber Amicorum Jaap Muilwijk* (pp. 39–57). Groningen, The Netherlands: Econometrisch Instituut.

Molenaar, I. W. (in press). Nonparametric models for polytomous responses. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern psychometrics*. New York: Springer.

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

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen & Lydiche.

Rosenbaum, P. R. (1985). Comparing distributions of item responses for two groups. *British Journal of Mathematical and Statistical Psychology, 38,* 206–215.

Samejima, F. (1969). Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monograph No. 17, 34*(4, Pt. 2).

Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph No. 18, 37*(4, Pt. 2).

Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika, 55,* 293–325.

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

Verhelst, N. D., & Glas, C. A. W. (1995). The one parameter logistic model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models. Foundations, recent developments, and applications* (pp. 215–237). New York: Springer-Verlag.

Manuscript received 2/17/95
Final version received 6/12/95