PSYCHOMETRIKA—VOL. 62, NO. 3, 331–347
SEPTEMBER 1997

STOCHASTIC ORDERING USING THE LATENT TRAIT AND THE SUM SCORE IN POLYTOMOUS IRT MODELS

BAS T. HEMKER AND KLAAS SIJTSMA

UTRECHT UNIVERSITY

IVO W. MOLENAAR

UNIVERSITY OF GRONINGEN

BRIAN W. JUNKER

CARNEGIE MELLON UNIVERSITY

In a restricted class of item response theory (IRT) models for polytomous items the unweighted total score has monotone likelihood ratio (MLR) in the latent trait θ. MLR implies two stochastic ordering (SO) properties, denoted SOM and SOL, which are both weaker than MLR, but very useful for measurement with IRT models. Therefore, these SO properties are investigated for a broader class of IRT models for which the MLR property does not hold.

In this study, first a taxonomy is given for nonparametric and parametric models for polytomous items based on the hierarchical relationship between the models. Next, it is investigated which models have the MLR property and which have the SO properties. It is shown that all models in the taxonomy possess the SOM property. However, counterexamples illustrate that many models do not, in general, possess the even more useful SOL property.

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

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 dichotomous or polytomous items. Let the score on item i be denoted X_i. The total score on the test, X_+, is the unweighted sum of the L item scores X_i. In testing generally, and in item response theory (IRT) in particular, the total score X_+, which is observable, is often used as a proxy for the unobservable latent trait value θ. In particular the ordering of subjects by X_+ is usually assumed to approximate the ordering of subjects by θ. It is thus desirable to identify IRT models in which a higher total score corresponds to a higher expected latent trait value.

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

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

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), P.O. Box 1034, 6801 MG Arnhem, THE NETHERLANDS.

0033-3123/97/0900-95415$00.75/0

is a nondecreasing function of θ.

Grayson (1988) also used 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 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 given by Huynh (1994). Because of its widespread use, and its fundamental role in binary IRT models, we will concentrate for the remainder of the paper on the total score X_+, although in certain settings other nondecreasing item summaries may be of interest (see, e.g., Rosenbaum, 1984, 1985).

It can easily be shown that the MLR property is symmetric in its arguments, which means that MLR of X_+ in θ is equivalent to MLR of θ in X_+. MLR is a technical property that implies two stochastic ordering (SO) properties (Lehmann, 1959, p. 74) that are easier to interpret in an IRT context. These SO properties are both weaker than the MLR property, in the sense that neither SO property implies the MLR property (Lehmann, 1959, sec. 3.3; see also, Junker, 1993; Rosenbaum, 1985).

First, MLR implies that X_+ is stochastically ordered by θ. That is, for any two respondents a and b with $\theta_a < \theta_b$,

$$P(X_+ \geq x_+ | \theta_a) \leq P(X_+ \geq x_+ | \theta_b). \qquad \text{(SOM)}$$

This first SO property (stochastic ordering of a manifest variable by θ, to be abbreviated SOM, in this case of X_+ by θ) takes the ordering on θ as a starting point. It implies that a higher latent trait value results in a higher expected total score (see Lehmann, 1986, p. 85, Lemma 2(i); which pertains to the stronger MLR property).

The second SO property concerns the stochastic ordering of θ by X_+. For any constant value s of θ, and for all $0 \leq C < K \leq L$,

$$P(\theta > s | X_+ = C) \leq P(\theta > s | X_+ = K). \qquad \text{(SOL)}$$

This second SO property (stochastic ordering of the latent trait, to be abbreviated SOL, in this case by X_+), which takes the ordering of X_+ as a starting point, is probably of more interest to the practical use of tests than SOM of X_+, because only the ordering on X_+ can be observed and inferences with respect to θ may be drawn on the basis of X_+. SOL by X_+ is evidently what is required for making mastery decisions based on cutoffs for the total score X_+; it also follows from SOL by X_+ that a higher total score results in a higher expected latent trait value (Lehmann, 1986, p. 85, Lemma 2(i)).

Many models for binary items begin with the three assumptions of latent trait UD, LI, and nondecreasing IRFs. The class of models that possess these three properties is called "strictly unidimensional" by Stout (1990) and Junker (1993). Mokken's (1971) formulation of monotone homogeneity was one of the earliest to explicitly consider all models satisfying just these three assumptions. Recently, variations have been studied extensively by Ellis and van den Wollenberg (1993), who characterize a "stochastic subject" version of strict unidimensionality, and by Holland (1981), Rosenbaum (1984), and Junker (1993), who consider "random sampling" versions of strict unidimensionality (the terms "stochastic subject" and "random sampling" were introduced by Holland (1990) to denote two ways of justifying these modeling assumptions in psychological/statistical terms). The MLR result of Grayson (1988) and Huynh (1994) applies to all strictly unidimensional models. Parametric examples which thus have the MLR property include the normal ogive models and the logistic models for binary items (e.g., Lord, 1980).

A general class of strictly unidimensional, monotone IRT models for polytomous items can also be defined (Hemker, Sijtsma, & Molenaar, 1995; Junker, 1991; Molenaar, 1982, 1997). Holland and Rosenbaum (1986), Ellis and Junker (1995) and Junker and Ellis

(1995) have studied theoretical model-fit issues for these and other monotone latent variable models. However only recently has inference for θ in this class of models been considered. Hemker, Sijtsma, Molenaar, and Junker (1996) show that the MLR result of Grayson and Huynh does not apply to this general class; the least restrictive model considered by Hemker, et al. that possesses MLR of X_+ by θ, is the Partial Credit Model (PCM; Masters, 1982) or a trivial generalization of it. For less restrictive models for polytomous items, counterexamples were found that showed that these models do not have the MLR property (Hemker, et al., 1996).

In this paper, we investigate the weaker SO properties for a broader class of unidimensional polytomous IRT models for which the stronger MLR property does not hold. These SO properties and the MLR property will be related to a taxonomy for nonparametric and parametric IRT models for polytomous items, based on the hierarchical relationships between the various models. First, this taxonomy is presented. Next, it is shown which of the models in the taxonomy have the MLR property and which models have one or both SO properties.

A Taxonomy of Unidimensional Polytomous IRT Models

A taxonomy of IRT models was given by Thissen and Steinberg (1986). Unidimensional parametric IRT models for polytomous items were organized as members of three distinct classes: divide-by-total models, difference models and left-side added multiple category models. This last class of models, which describe multiple-choice responses with guessing, is not considered in our taxonomy. Another difference is that we added two nonparametric models to our taxonomy. This clarifies the fact that all models from the first two classes can be integrated into one class of polytomous IRT models. Recently, Mellenbergh (1995) provided an alternative classification of parametric polytomous IRT models, mainly based on the definitions of the conditional probabilities of choosing a particular answer category.

We assume that all items have the same number of answer categories in the models we consider. Generalization of our discussion to models in which items have different numbers of answer categories is straightforward but would lead to more cumbersome notation.

Divide-by-Total Models

Probably the best known member of the class of divide-by-total models (Thissen & Steinberg, 1986) is the PCM (Masters, 1982). Let each of the L items have $m + 1$ ordered answer categories which are scored $X_i = 0, \ldots, m$, respectively. Masters' PCM assumes the parametric form

$$P(X_i = j \mid \theta; X_i = j \text{ or } X_i = j - 1) = \frac{\exp(\theta - \delta_{ij})}{1 + \exp(\theta - \delta_{ij})}, \tag{2}$$

where δ_{ij} is the difficulty of step j of item i (Masters, 1982). We shall call the conditional probability $P(X_i = j \mid \theta; X_i = j \text{ or } X_i = j - 1)$ of responding in category j rather than category $j - 1$, given θ, the partial credit item step response function (partial credit ISRF) for step j of item i.

To have a more compact notation the conditional probability $P(X_i = x \mid \theta)$ is denoted as π_{ix}, omitting the argument θ. From (2) it follows (Masters, 1982) that in the PCM π_{ix} is

$$\pi_{ix} = \frac{\exp\left[\sum_{j=1}^{x} (\theta - \delta_{ij})\right]}{\sum_{k=0}^{m} \exp\left[\sum_{j=1}^{k} (\theta - \delta_{ij})\right]}, \tag{3}$$

where for notational convenience $\Sigma_{j=1}^{0} (\theta - \delta_{ij}) \equiv 0$ in case of $x = 0$. Note that in the denominator of (3) we have the sum of the numerators across all answer categories, which explains the qualifier divide-by-total. We will further discuss this terminology after we have introduced the class of difference models (Thissen & Steinberg, 1986). In the Rating Scale Model (RSM; Andrich, 1978), $\delta_{ij} = \delta_i + \tau_j$ where δ_i is the location of item i on θ and τ_j is the location of the j-th step of each item relative to that item's location on θ. Note that the RSM is a special case of the PCM (Masters, 1982).

A more flexible model than (3) can be defined by inserting a positive discrimination parameter α_{ij}. The resulting model may be called the two-parameter Partial Credit Model (2p-PCM; Hemker et al., 1996), in which

$$\pi_{ix} = \frac{\exp\left[\sum_{j=1}^{x} \alpha_{ij}(\theta - \delta_{ij})\right]}{\sum_{k=0}^{m} \exp\left[\sum_{j=1}^{k} \alpha_{ij}(\theta - \delta_{ij})\right]}. \tag{4}$$

Note that this definition of π_{ix} yields a model that is identical to the Nominal Response Model (NRM; Bock, 1972) if nominal response categories are assumed, that is, if α_{ij} is not restricted to be positive (Muraki, 1992; Samejima, 1972).

Special cases of the 2p-PCM which are generalizations of the original PCM (Masters, 1982) can easily be defined. If the discrimination parameter is held constant across the item steps of the same item ($\alpha_{ij} = \alpha_i$) the generalized PCM (g-PCM; Muraki, 1992) is obtained. Note that this model has also been referred to as Thissen and Steinberg's Ordinal Model (TSOM; Maydeu-Olivares, Drasgow, & Mead, 1994). Using a similar line of reasoning, the 2p-PCM with the same discrimination parameter for item step j across all items ($\alpha_{ij} = \alpha_j$) can be defined (Hemker, et al. 1996). To discriminate between these three 2p-PCMs they are denoted 2p(ij)-, 2p(i)-, and 2p(j)-PCM, respectively. The term between brackets clarifies whether α varies across items i *and* item steps j, or only across items i or item steps j, respectively. If α is constant across both items and item steps ($\alpha_{ij} = \alpha$), the one-parameter PCM is obtained. Note that this model is a trivial generalization of the original PCM in which $\alpha_{ij} = \alpha = 1$. Thus, it is not distinguished from the PCM in this study.

Difference Models

Perhaps the best known model from the class of difference models (Thissen & Steinberg, 1986) is the Graded Response Model (GRM; Samejima, 1969; see also Masters, 1982). Samejima's GRM assumes the well known logistic form

$$P(X_i \geq j|\theta) = \frac{\exp\left[\alpha_i(\theta - \lambda_{ij})\right]}{1 + \exp\left[\alpha_i(\theta - \lambda_{ij})\right]}, \tag{5}$$

for all $j = 1, \ldots, m$, where λ_{ij} is the threshold parameter, with $\lambda_{i1} \leq \lambda_{i2} \leq \cdots \leq \lambda_{im}$, and $\alpha_i > 0$, for all i. For $x = 0$ and $x = m$, by definition $P(X_i \geq 0|\theta) = 1$ and

$P(X_i \geq m + 1|\theta) = 0$, respectively. We shall call the conditional probability $P(X_i \geq j|\theta)$ of responding in category j or higher, the graded response item step response function (graded response ISRF) for step j of item i.

The probability of having item score x is given by the difference

$$\pi_{ix} = P(X_i \geq x|\theta) - P(X_i \geq x + 1|\theta). \tag{6}$$

The terminology of difference models was derived (Thissen & Steinberg, 1986) from (6) in which the difference between two adjacent model ISRFs is used to obtain π_{ix}. It is important to note that the ISRF $P(X_i \geq x|\theta)$ of a difference model is a *simple parametric* function, for example, a logistic function. Also, note that divide-by-total models, such as the PCM, do not have a simple parametric form for $P(X_i \geq x|\theta)$. For this reason they are not considered to be difference models. Similarly, the probability $P(X_i = j|\theta; X_i = j$ or $X_i = j - 1)$ as in (2) of the PCM has a *simple parametric* form which is characteristic of divide-by-total models. Difference models do not have a simple parametric form for $P(X_i = j|\theta; X_i = j$ or $X_i = j - 1)$ and, therefore, they are not considered to be divide-by-total models.

In the following, Samejima's GRM is referred to as the 2p(i)-GRM, which is more parallel to our PCM naming conventions. Note that the 2p(ij)-GRM and the 2p(j)-GRM do not exist because α cannot vary over item steps, for otherwise the ISRFs in (5) would cross for different values of x_i, and this is evidently impossible (Samejima, 1969, 1972; Thissen & Steinberg, 1986). If discrimination parameters are assumed to be the same for all items ($\alpha_i = \alpha$) a special case of the 2p(i)-GRM is obtained. This is the one-parameter GRM (1p-GRM), which is obviously also a difference model.

Nonparametric Models

Two nonparametric models that are based on the parametric models discussed here are defined in our taxonomy: the nonparametric Partial Credit Model (np-PCM), and the nonparametric Graded Response Model (np-GRM). Both nonparametric models are defined by three assumptions: UD, LI, and ISRFs that are nondecreasing in the latent trait θ. The two models, however, differ in the definition of the ISRFs, analogous to the difference in the definition of ISRFs between the divide-by-total models and the difference models. Both nonparametric models serve the purpose of uniting the parametric classes of divide-by-total models and difference models in a more comprehensive hierarchical framework. In addition but not pursued here, both models can be seen as alternative models for describing the data; see for example Hemker, Sijtsma, and Molenaar (1995) who discuss the np-GRM as a data analysis method. The np-PCM, however, is new.

The np-PCM is defined by assuming that the partial credit ISRFs

$$P(X_i = x|\theta; X_i = x \text{ or } X_i = x - 1) = \frac{\pi_{ix}}{\pi_{ix} + \pi_{i,x-1}} \tag{7}$$

are nondecreasing in θ for all i and all $x = 1, \ldots, m$. In the 2p(ij)-PCM (see (4)) the ISRF is defined as

$$\frac{\pi_{ix}}{\pi_{ix} + \pi_{i,x-1}} = \frac{\exp\left[\alpha_{ix}(\theta - \delta_{ix})\right]}{1 + \exp\left[\alpha_{ix}(\theta - \delta_{ix})\right]}.$$

Because the discrimination parameter is positive, this function is nondecreasing in θ, so the 2p(ij)-PCM is indeed a special case of the np-PCM. Because all other parametric divide-by-total models are special cases of the 2p(ij)-PCM, all models from this class are special cases of the np-PCM.

The np-GRM is defined by assuming that the graded response ISRFs

$$P(X_i \geq x|\theta) = \sum_{j=x}^{m} \pi_{ij}$$

are nondecreasing in θ for all i and all $x = 1, \ldots, m$. Note that the definition of the np-GRM is identical to the definition of the Mokken model of monotone homogeneity for polytomous items (Hemker, Sijtsma, & Molenaar, 1995; Molenaar, 1982, 1997). The abbreviation np-GRM is used here because it better fits in the nomenclature of this study. The np-GRM is called "strictly unidimensional" by Junker (1991), who uses it as the starting point for an investigation of essential unidimensionality (see also Stout, 1987, 1990) for polytomous items. Ellis and Junker (1995) and Junker and Ellis (1995) give characterizations of an infinite item pool formulation of the np-GRM, in terms of conditional association (Holland & Rosenbaum, 1986) and a vanishing conditional dependence condition. The 2p(i)-GRM is a special case of the np-GRM because in the 2p(i)-GRM, $P(X_i \geq x|\theta)$, given by (5), is nondecreasing in θ.

Less obvious than the results that the parametric divide-by-total models are special cases of the np-PCM, and the parametric difference models are special cases of the np-GRM, is that the np-PCM is a special case of the np-GRM (Theorem 2; to be discussed below). This relation follows directly from the MLR and SO properties of the two models that will be discussed in the next section. As a result, all divide-by-total models are also special cases of the np-GRM. It can also be shown that the difference models defined here (i.e., all 2p(i)-models) are special cases of the np-PCM (Theorem 3). The proof of this result also uses the MLR and SO properties of both models and is given after the introduction of these properties.

Summary

All polytomous models discussed thus far can be organized in a taxonomy that emphasizes the hierarchical relations between the models. The most general model, and thus the least restrictive model, is the np-GRM. A special case of this model is the np-PCM. The models from the class of divide-by-total models as well as the models from the class of difference models are special cases of the np-PCM. Finally, because difference models are neither a special case nor a generalization of the divide-by-total models (Thissen & Steinberg, 1986), the taxonomy is complete. This taxonomy of relations between the various models is displayed as a Venn-diagram in Figure 1. Note that this figure only holds for items with at least three answer categories; for dichotomous items the set structure is more simple. For example, the np-GRM and the np-PCM are identical for dichotomous items; that is, if $m = 1$, $P(X_i \geq 1|\theta) = P(X_i = 1|\theta; X_i = 1 \text{ or } X_i = 0)$. This latter equality also implies that the distinction between difference and divide-by-total models no longer exists (Thissen & Steinberg, 1986).

MLR, SOM, and SOL in Polytomous IRT Models

The definition of MLR of X_+ in θ for polytomous models is almost the same as in the dichotomous case (see (1)). The only difference concerns the range of the total score. Because for polytomous items $X_i = 0, \ldots, m$, for the total score $X_+ = 0, \ldots, mL$, and thus $0 \leq C < K \leq mL$.

The least restrictive model for polytomous items in our taxonomy that implies MLR is the PCM (Hemker et al., 1996). Thus, MLR also holds for the RSM (Andrich, 1978). Counterexamples were found (Hemker et al., 1996) for the models from the divide-by-total class in which α_{ij} varied over items or item steps or both, and for all models from the class

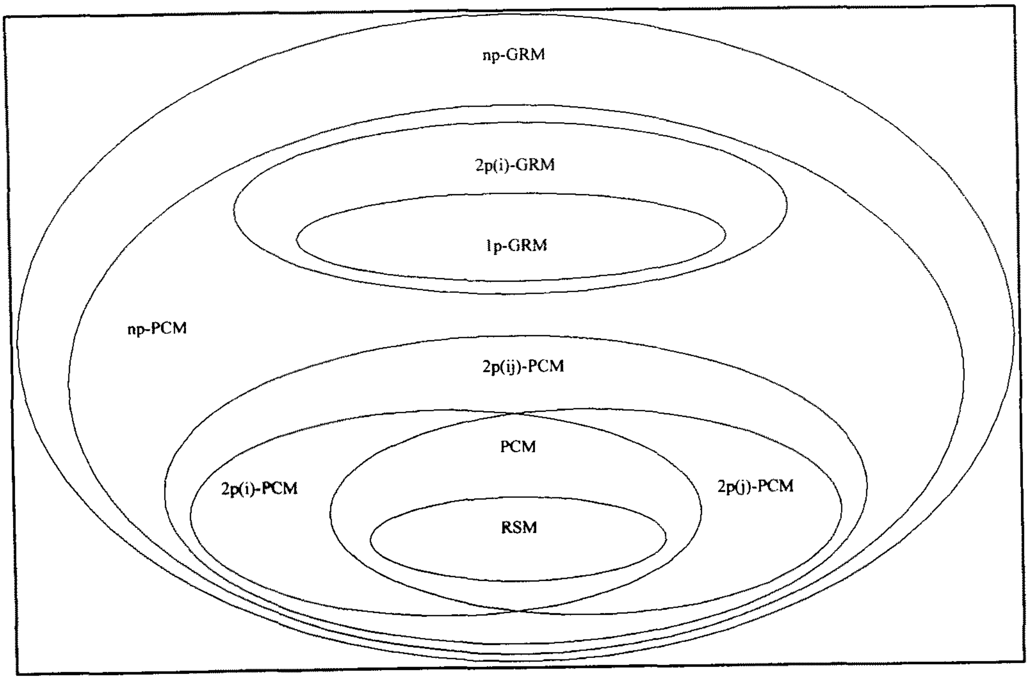

FIGURE 1.
Venn-diagram displaying the taxonomy of relations between the different models. Note: The models in the taxonomy are Samejima's Graded Response Model (2p(i)-GRM), the nonparametric GRM (np-GRM), the one-parameter GRM (1p-GRM), the Partial Credit Model (PCM), the nonparametric PCM (np-PCM), the two-parameter PCMs [the 2p(ij)-PCM), the 2p(i)-PCM, and the 2p(j)-PCM], and the Rating Scale Model (RSM).

of difference models. Obviously, the nonparametric models do not imply MLR of X_+ in θ because they are generalizations of the parametric models that do not have this property.

It can thus be concluded that for polytomous items the class of IRT models that have the MLR property is smaller and subject to more restrictions than the class of models with MLR for dichotomous items. However, it can not be concluded that the PCM is the only model for polytomous items that implies the SO properties on total score level, because SOM of X_+, SOL by X_+, or both do not imply MLR of X_+ (Junker, 1993).

The np-GRM and SO

In the np-GRM, $P(X_i \geq x|\theta)$ is nondecreasing in the latent trait θ for all x and all i. This assumption is identical to SOM of X_i, that is, for any two respondents a and b with $\theta_a < \theta_b$,

$$P(X_i \geq x|\theta_a) \leq P(X_i \geq x|\theta_b), \tag{8}$$

for all $i = 1, \ldots, L$ and all $x = 0, 1, \ldots, m$.

Theorem 1. The np-GRM has the property of SOM of X_+.

Proof. Theorem 1 follows from a general result cited in Holland and Rosenbaum (1986; Lemma 2). Let $\mathbf{X} = (X_1, \ldots, X_L)$; from this result it follows that if the np-GRM holds, then for any bounded function g(**X**) that is nondecreasing in each coordinate, the conditional expectation $E[\text{g}(\mathbf{X})|\theta]$ is nondecreasing in θ. Let

$$g(X_1, \ldots, X_L) = \begin{cases} 1 & \text{if } X_+ \geq x_+; \\ 0 & \text{if } X_+ < x_+, \end{cases}$$

which is nondecreasing in all coordinates of $\mathbf{X}$, then $E[g(X_1, \ldots, X_L)|\theta]$ equals $P(X_+ \geq x_+|\theta)$. Thus, the np-GRM implies SOM of X_+. An alternative, more elementary proof that uses (8) is sketched in the Appendix along with an illustrative example. □

Analogous to the SOM property for item scores X_i in (8), the SOL property can be defined for item scores. For any constant value s of θ, and for all $0 \leq c < k \leq m$,

$$P(\theta > s|X_i = c) \leq P(\theta > s|X_i = k).$$

The np-GRM, however, does not imply SOL by X_i. This is shown next (Example 1) by extending the counterexample that the np-GRM does not imply MLR given by Junker (1993, Example 4.1).

Example 1: The np-GRM does not imply SOL by X_i. Let $0 \leq \theta \leq 1$ and consider an item i with three answer categories ($X_i = 0, 1, 2$) that satisfies the np-GRM, with

$$P(X_i \geq 1|\theta) = \begin{cases} 3\theta, & \text{if } 0 \leq \theta \leq \frac{1}{4}; \\ \frac{2}{3} + \frac{1}{3}\theta, & \text{if } \frac{1}{4} < \theta \leq 1. \end{cases}$$

$$P(X_i \geq 2|\theta) = \begin{cases} 2\theta, & \text{if } 0 \leq \theta \leq \frac{1}{4}; \\ \frac{1}{4} + \theta, & \text{if } \frac{1}{4} < \theta \leq \frac{1}{2}; \\ \frac{1}{2} + \frac{1}{2}\theta, & \text{if } \frac{1}{2} < \theta \leq 1. \end{cases}$$

Consider for this example a three-point distribution of θ, $P(\theta = 1/4) = P(\theta = 1/2) = .25$ and $P(\theta = 1) = .5$; then $P(\theta > 1/4|X_i = 0) = .40$ and $P(\theta > 1/4|X_i = 1) = .25$. Therefore, SOL by X_i does not hold.

It can also be shown that the np-GRM does not imply SOL by X_+. A counterexample, however, will be given for a more restricted model and thus by implication a counterexample has been found for the np-GRM. For reasons of an efficient presentation of other results, this counterexample is postponed to Example 2, following Theorem 3 below.

Because the np-GRM has the property of SOM of X_+, and because the np-GRM is the most general polytomous IRT model in our taxonomy, all models in our taxonomy have this property. However, the np-GRM does not have the SOL property for either X_+ or X_i, which leaves the np-PCM as the most general candidate that may have SOL by X_i or by X_+.

The np-PCM, MLR, and SO

We will prove the next proposition for the np-PCM.

Proposition. The np-PCM has the property of MLR of X_i.

Proof. The assumption that characterizes the np-PCM, in addition to UD and LI, is that $\pi_{ix}/(\pi_{ix} + \pi_{i,x-1})$ is nondecreasing in θ, for all i and $x = 1, \ldots, m$ (see (7)). This holds if and only if

$$\frac{1}{1 + \dfrac{\pi_{i,x-1}}{\pi_{ix}}}$$

is nondecreasing in θ for all i and x, which holds if and only if $\pi_{ix}/\pi_{i,x-1}$ is nondecreasing in θ for all i and x. Monotonicity of the latter ratio expresses MLR for $X_i = x$ versus $X_i = x - 1$, and holds for any x. By multiplying similar ratios, we obtain, for all $0 \leq c < k \leq m$, and all i, that

$$\frac{\pi_{ik}}{\pi_{ic}} = \frac{P(X_i = k|\theta)}{P(X_i = c|\theta)} \tag{9}$$

is nondecreasing in θ, which is equivalent to MLR of X_i, for all i. □

This result is used to prove the next theorem.

Theorem 2. The np-PCM is a proper special case of the np-GRM.

Proof. The np-PCM and the np-GRM have the first two assumptions (UD and LI) in common. The third assumption of the np-PCM $[\pi_{ix}/(\pi_{ix} + \pi_{i,x-1})$ nondecreasing$]$ is equivalent to MLR on the item score level (Equation (9)). The third assumption of the np-GRM ($\Sigma_{j=x}^{m} \pi_{ij}$ nondecreasing) is equivalent to SOM of X_i (Equation (8)). Because MLR of X_i (np-PCM) implies SOM of X_i (np-GRM) (Lehmann, 1959, p. 74) but not vice versa (Junker, 1993, Example 4.1), the np-PCM is a proper special case of the np-GRM. □

Next we will show that the 2p(i)-GRM is a special case of the np-PCM. Because the first two assumptions of the np-PCM and the 2p(i)-GRM (UD and LI) are the same, it is sufficient to show that the third assumption of the 2p(i)-GRM implies MLR of X_i (the third assumption of the np-PCM), but not vice versa.

Theorem 3. The 2p(i)-GRM is a special case of the np-PCM.

Proof. The third assumption of the 2p(i)-GRM is that $P(X_i \geq x|\theta) = \exp[\alpha_i(\theta - \lambda_{ix})]/\{1 + \exp[\alpha_i(\theta - \lambda_{ix})]\}$ (Equation (5)) is nondecreasing in θ for all $x = 1, \ldots, m$, and $i = 1, \ldots, L$, with $\lambda_{i1} \leq \lambda_{i2} \leq \cdots \leq \lambda_{im}$, for all i. Note that $P(X_i \geq 0|\theta) = 1$ and that $P(X_i \geq m + 1|\theta) = 0$. Let $f_x = \exp[\alpha_i(\theta - \lambda_{ix})]$ for notational convenience, then in the 2p(i)-GRM

$$\pi_{ix} = \frac{f_x}{1 + f_x} - \frac{f_{x+1}}{1 + f_{x+1}},$$

for $x = 1, \ldots, m$, with $f_{m+1} = 0$ (Equation (6)). Note that the first derivative of f_x with respect to θ is equal to $\alpha_i f_x$. The first derivative of π_{ix} can thus be written as

$$\frac{\alpha_i\{f_x[1 + f_{x+1}]^2 - f_{x+1}[1 + f_x]^2\}}{[1 + f_x]^2\,[1 + f_{x+1}]^2},$$

which is identical to

$$\frac{\alpha_i[\mathrm{f}_x - \mathrm{f}_{x+1}][1 - \mathrm{f}_x\mathrm{f}_{x+1}]}{[1 + \mathrm{f}_x]^2\,[1 + \mathrm{f}_{x+1}]^2}.$$

This means that for all x and i, $[\log \pi_{ix}]'$ is equal to

$$\frac{\pi'_{ix}}{\pi_{ix}} = \frac{\alpha_i[1 - \mathrm{f}_x \cdot \mathrm{f}_{x+1}]}{[1 + \mathrm{f}_x][1 + \mathrm{f}_{x+1}]}$$

$$= \alpha_i\left[1 - \frac{\mathrm{f}_x}{1 + \mathrm{f}_x} - \frac{\mathrm{f}_{x+1}}{1 + \mathrm{f}_{x+1}}\right],$$

with $\mathrm{f}_0/(1 + \mathrm{f}_0) = 1$, by definition. As a result, it holds that for all x and i

$$\frac{\pi'_{ix}}{\pi_{ix}} - \frac{\pi'_{i,x-1}}{\pi_{i,x-1}} = \alpha_i\left[\frac{\mathrm{f}_{x-1}}{1 + \mathrm{f}_{x-1}} - \frac{\mathrm{f}_x}{1 + \mathrm{f}_x} + \frac{\mathrm{f}_x}{1 + \mathrm{f}_x} - \frac{\mathrm{f}_{x+1}}{1 - \mathrm{f}_{x+1}}\right]$$

$$= \alpha_i(\pi_{i,x-1} + \pi_{ix}) \geq 0.$$

This means that if the 2p(i)-GRM holds $\pi'_{ix}\pi_{i,x-1} \geq \pi'_{i,x-1}\pi_{ix}$, and thus $\pi'_{ix}\pi_{i,x-1} - \pi'_{i,x-1}\pi_{ix} \geq 0$ for all x and i; this implies that $\pi_{ix}/\pi_{i,x-1}$ is nondecreasing in θ for all x and i. This result is sufficient to prove that (9) is nondecreasing, which means that MLR of X_i in θ holds. Therefore, the third assumption of the 2p(i)-GRM implies MLR of X_i in θ, which is the third assumption of the np-PCM. Thus, the 2p(i)-GRM implies the np-PCM. □

The reverse relation which says that the np-PCM implies the 2p(i)-GRM does not hold. This follows from the result that the PCM does not imply the 2p(i)-GRM, or vice versa (Thissen & Steinberg, 1986), and because the PCM is a special case of the np-PCM. Thus the np-PCM and the 2p(i)-GRM are not equivalent.

Because we have established that the 2p(ij)-PCM and the 2p(i)-GRM are special cases of the np-PCM, it follows that all parametric models from our taxonomy are special cases of the np-PCM. Therefore, by implication all these parametric models have MLR of X_i in θ and, consequently, SOM of X_i and SOL by X_i. We have seen that SOM of X_i implies SOM of X_+. However, SOL by X_i does not imply SOL by X_+ in general. This can be shown by means of a counterexample. Note that the definition of SOL by X_+ is equivalent to Equation (SOL) with $0 \leq C < K \leq mL$.

Example 2: The 1p-GRM does not imply SOL by X_+. Consider two items ($i = 1, 2$), each with four answer categories ($X_i = 0, 1, 2, 3$). Let $\alpha_1 = \alpha_2 = 1$, $\lambda_{11} = \log 49/51$, $\lambda_{12} = 0$, and $\lambda_{13} = \log 51/49$; and $\lambda_{21} = \log 33/67$, $\lambda_{22} = \log 33/17$, and $\lambda_{23} = \log 99$. Assume that the latent trait θ has a standard normal distribution. Then one obtains by numerical integration $P(\theta > 0|X_+ = 2) = .536$ and $P(\theta > 0|X_+ = 3) = .464$. Figure 2 and Table 1 show $P(\theta > s|X_+)$ for all total scores and for $s = -3, -2, \ldots, 3$.

This counterexample not only implies that the 1p-GRM does not imply SOL by X_+, but also that all models that are generalizations of the 1p-GRM do not imply SOL by X_+. These include the 2p(i)-GRM, the np-GRM and the np-PCM. This counterexample also shows that SOL by X_i does not imply SOL by X_+. This leaves the 2p(ij)-PCM as possibly the least restrictive of the models we have considered with SOL by X_+. However, for these models counterexamples can also be found.

Example 3: The 2p(i)-PCM does not imply SOL by X_+. Consider two items ($i = 1, 2$), each with three answer categories ($X_i = 0, 1, 2$). Let $\alpha_1 = 2$ and $\alpha_2 = .5$. Let $\delta_{11} = \delta_{12}$

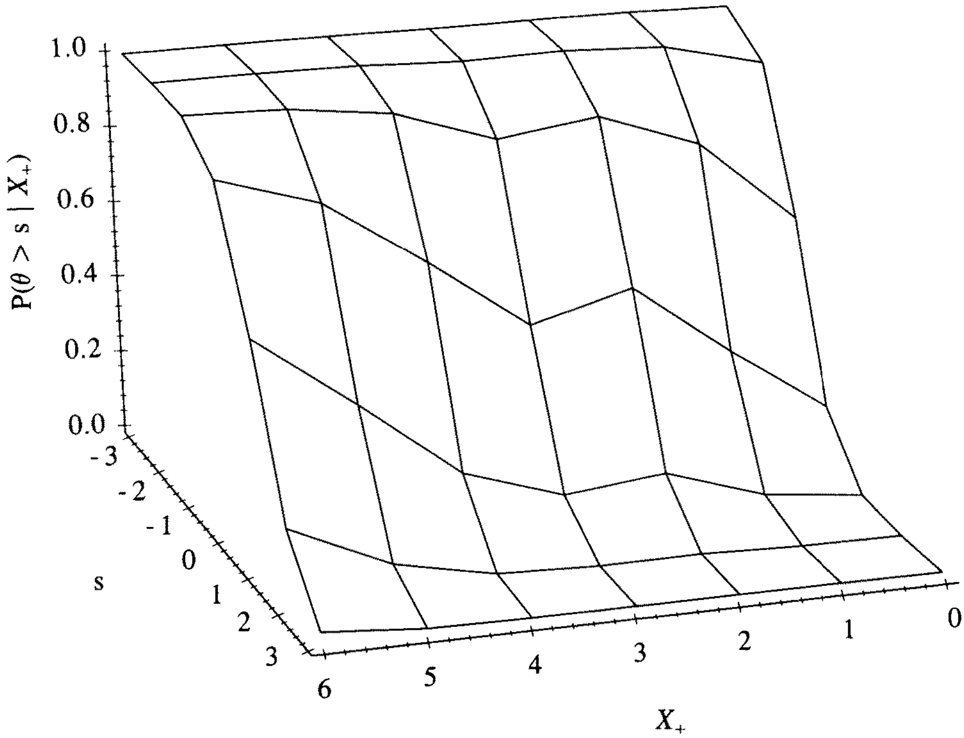

FIGURE 2.
$P(\theta > s|X_+)$ for all total scores and for $s = -3, -2, \ldots, 3$, in case of two items satisfying the 1p-GRM, each with four answer categories with the following parameter vector: $\alpha = 1$; $\lambda_{11} = \log 49/51$, $\lambda_{12} = 0$, and $\lambda_{13} = \log 51/49$; and $\lambda_{21} = \log 33/67$, $\lambda_{22} = \log 33/17$, and $\lambda_{23} = \log 99$. The latent trait θ has a standard normal distribution.

$= \delta_{21} = 0$, and $\delta_{22} = -2 \log 98$. Assume that the latent trait θ has a standard normal distribution. Then $P(\theta > 0|X_+ = 1) = .121$ and $P(\theta > 0|X_+ = 2) = .120$. Figure 3 and Table 2 show $P(\theta > s|X_+)$ for all total scores in this case and for $s = -3, -2, \ldots, 3$.

Because the 2p(i)-PCM is a special case of the 2p(ij)-PCM, Example 3 also shows that the 2p(ij)-PCM does not imply SOL by X_+. A similar counterexample can be found for the 2p(j)-PCM. This means that the PCM is not only the least restrictive model that has MLR of X_+ (Hemker, et al. 1996), but also the least restrictive of the models we have considered that has SOL by X_+.

This study thus shows that the PCM is the least restrictive model considered in this study that allows the ordering of subjects by means of their total score in *all* cases. Note, however, that the counterexamples that show that the less restrictive models do not imply SOL by X_+ are based on extreme parameter vectors. It is obvious that many examples with less extreme and, therefore, more practical parameter vectors, can be found that show that the less restrictive models can have SOL by X_+. Since the property of SOL by X_+ depends on the parameter setups for these less restrictive models, it is incumbent on the user of polytomous IRT models to check that SOL by X_+ holds in the fitted model before asserting that higher total scores correspond to higher expected θ values, using total score cutoffs for mastery decisions, etcetera.

Discussion

Nonparametric graded response models (np-GRM) provide a natural, large nonparametric class of IRT models for polytomous items with ordered response categories. This

TABLE 1

Values of $P(\theta > s \mid X_+ = x_+)$ for Example 2

X_+	s: -3	-2	-1	0	1	2	3
0	.994	.920	.590	.165	.015	.000	.000
1	.999	.982	.808	.342	.046	.002	.000
2	1.000	**.993**	**.900**	**.536**	**.131**	**.010**	.000
3	1.000	**.989**	**.865**	**.464**	**.102**	**.008**	.000
4	1.000	.998	.953	.653	.188	.017	.000
5	1.000	1.000	.984	.829	.397	.073	.004
6	1.000	1.000	.993	.911	.594	.195	.026

Note: boldface indicates where $P(\theta > s \mid X_+ = x_+)$ decreases in X_+. Calculations were accurate up to 40 decimals.

class is characterized by the three assumptions of latent trait UD, LI, and monotonicity of the ISRFs, $P(X_i \geq x|\theta)$. It has been considered under many names, including "strictly unidimensional IRT models" (Stout, 1990; also Junker, 1991, 1993), the "Mokken model of monotone homogeneity for polytomous items" (Hemker, Sijtsma, & Molenaar, 1997; Molenaar, 1982, 1997), and "monotone unidimensional latent variable models" (Ellis & Junker, 1995; Holland & Rosenbaum, 1986; Junker & Ellis, 1995); We have shown that all commonly-considered parametric and nonparametric models for polytomous items with ordered response categories, including the RSM of Andrich (1978), one-parameter, two-parameter, and nonparametric PCMs (Masters, 1982; Muraki, 1992; Hemker, et al., 1996), as well as one- and two-parameter GRMs (Samejima, 1969; Hemker, Sijtsma, Molenaar, & Junker, 1996), can be organized into a hierarchical taxonomy within the class of np-GRMs (Figure 1). All models in our taxonomy enjoy the property of SOM of X_+, that is, stochastic ordering of the total score X_+ by the latent trait θ; this follows directly from monotonicity of the ISRFs (Theorem 1).

The class of np-PCMs replaces the assumption of monotonicity of $P(X_i \geq x|\theta)$ with the assumption of monotonicity of $P(X_i = x|\theta; X_i = x \text{ or } x - 1)$. This is equivalent to a monotone likelihood ratio property for individual items (Proposition), from which it follows that the np-PCM is a special case of the np-GRM (Theorem 2). Interestingly, all of the parametric models that we considered, including parametric PCMs and even parametric GRMs, can be shown to be special cases of the np-PCM (Theorem 3). Counterexamples show that none of these classes are equivalent. Other relationships among these models can be seen in Figure 1.

For inference on θ, a more useful stochastic ordering property is SOL by X_+: stochastic ordering of θ by the total score X_+. Of the models we considered, only the PCM

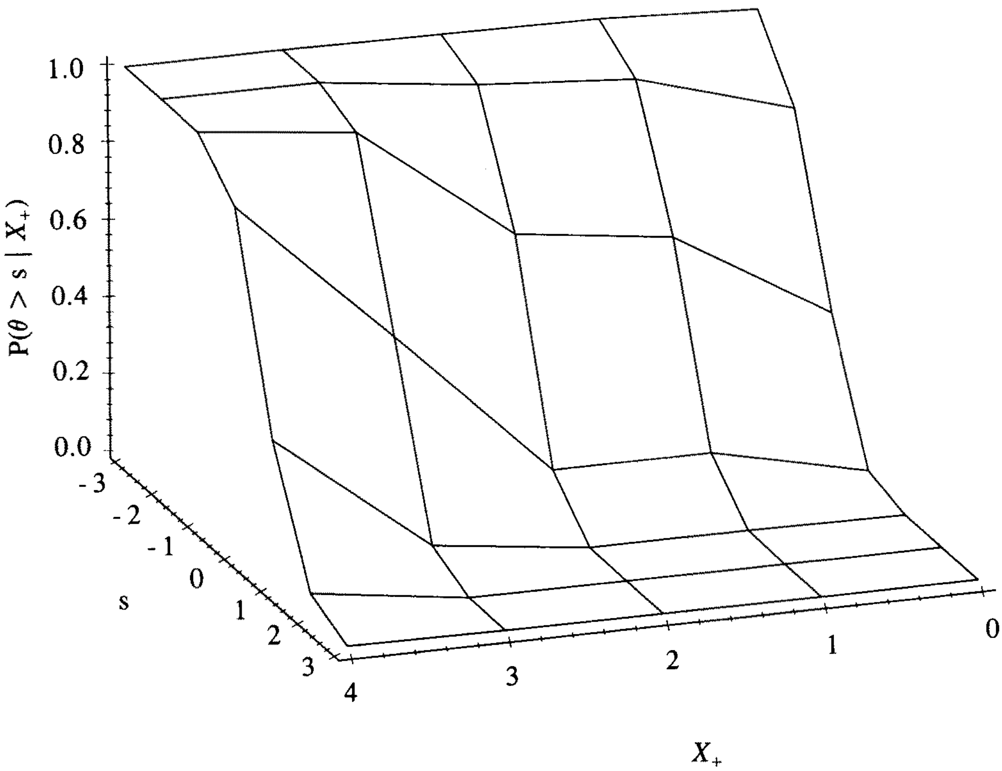

FIGURE 3.
$P(\theta > s|X_+)$ for all total scores in this case and for $s = -3, -2, \ldots, 3$, in case of two items satisfying the 2p(i)-PCM, each with three answer categories with the following parameter vector: $\alpha_1 = 2$ and $\alpha_2 = .5$; $\delta_{11} = \delta_{12} = \delta_{21} = 0$, and $\delta_{22} = -2 \log 98$. The latent trait θ has a standard normal distribution.

and models that are special cases of this model, such as the RSM, enjoy SOL by X_+. Counterexamples show that in all the other parametric models we considered, as well as in the two nonparametric classes of np-PCM and np-GRM, SOL by X_+ does not hold in all cases. However, many examples can be found that suggest that SOL by X_+ holds for many realistic sets of parameter values in the parametric models we considered. Thus it is incumbent on the user of parametric polytomous IRT models to check that SOL by X_+ holds in the fitted model before asserting that higher total scores correspond to higher expected θ values, using total score cutoffs for mastery decisions, etcetera. Two next steps in future research may be a general characterization of SOL models, and the search for methods for investigating the validity of SOL in empirical research.

Appendix

Another proof of Theorem 1 can be given. This direct proof uses (8) to show that SOM of the total score X_+ holds in the np-GRM, which is the same as showing that the derivative to θ of $P(X_+ \geq x_+|\theta)$ is nonnegative (see also Equation (SOM)) if the derivative to θ of $P(X_i \geq x|\theta)$ is nonnegative. The full proof is a lengthy combinatorial argument. However the sketch we present here is relatively simple. The idea of the full proof can be gained from the sketch and the example which follows. A complete proof for the general case can be obtained on request from the first author.

The following is an outline of the proof: It has to be shown that the first derivative of the probability $P(X_+ \geq x_+|\theta)$ with respect to θ is nonnegative for each value x_+. Deriv-

TABLE 2

Values of P(θ > s | $X_+ = x_+$) for Example 3

X_+	s -3	-2	-1	0	1	2	3
0	.975	.801	.357	.031	.000	.000	.000
1	.994	.920	.595	**.121**	**.004**	.000	.000
2	.997	.950	.647	**.120**	**.003**	.000	.000
3	1.000	.999	.954	.509	.052	.002	.000
4	1.000	1.000	.998	.889	.369	.056	.003

Note: boldface indicates where P(θ > s | $X_+ = x_+$) decreases in X_+. Calculations were accurate up to 40 decimals.

atives are denoted by means of a prime. The minimum and maximum values of X_+ will be considered first. For $X_+ = 0$, the probability $P(X_+ \geq 0|\theta) = 1$; therefore its derivative equals 0, which does not contradict nondecreasingness of $P(X_+ \geq x_+|\theta)$ in θ. For $X_+ = mL$, the probability

$$P(X_+ \geq mL|\theta) = P(X_+ = mL|\theta)$$
$$= \prod_{i=1}^{L} P(X_i = m|\theta)$$
$$= \prod_{i=1}^{L} P(X_i \geq m|\theta).$$

Each probability in the last product is nondecreasing in θ by (8).

For $0 < X_+ < mL$, it can be checked that the derivative of $P(X_+ \geq x_+|\theta)$ can always be expressed as a sum of positive products where each product consists of one derivative $P'(X_j \geq x_j|\theta)$ which is nonnegative by (8), and $L - 1$ probabilities of the form π_{ix}, $i \neq j$:

$$P'(X_j \geq x|\theta) \prod_{i \neq j}^{L} \pi_{ix}. \tag{10}$$

The following line of reasoning clarifies how terms as in (10) are obtained. First, note that

$$P(X_+ \geq x_+|\theta) = \sum_{t=x_+}^{mL} P(X_+ = t|\theta). \tag{11}$$

Taking the first derivative means that each probability on the right has to be differentiated with respect to θ. Before differentiating note that each probability $P(X_+ = t|\theta)$ can be written as the sum of products of the L probabilities of individual item scores that add up to $X_+ = t$. A product that is based on such an array is an element indexed a_t from the set A_t that contains all these products. Thus we can write

$$P(X_+ = t|\theta) = \sum_{a_t \in A_t} \left[\prod_{i=1}^{L} \pi_{ix} \right]_{a_t}. \tag{12}$$

Combining (11) and (12) thus yields

$$P(X_+ \geq x_+|\theta) = \sum_{t=x_+}^{mL} \sum_{a_t \in A_t} \left[\prod_{i=1}^{L} \pi_{ix} \right]_{a_t}. \tag{13}$$

Taking the derivative of such a sum of products is done by means of the product rule which is independently applied to each of the products. Let $L = 2$ and $m = 2$, so that $X_i = 0, 1, 2$; and $X_+ = 0, 1, 2, 3, 4$. Then, as an example,

$$P(X_+ = 3|\theta) = \pi_{11}\pi_{22} + \pi_{12}\pi_{21},$$

with

$$P'(X_+ = 3|\theta) = [\pi'_{11}\pi_{22} + \pi_{11}\pi'_{22}] + [\pi'_{12}\pi_{21} + \pi_{12}\pi'_{21}]. \tag{14}$$

The product rule for differentiation has to be applied to each probability on the right in (13). In our example, the *total* sequence of products that is obtained in this way can be rearranged and factored such that only a sum of positive products of the form

$$P'(X_j \geq x|\theta)\pi_{ix} \tag{15}$$

is obtained (see (12)) and no terms are left. It is *crucial* to note that all the derivatives of probabilities such as in (14) are used to form the derivative such as in (10) and (15). Note, in particular, that to do this successfully it has to be recognized that

$$P'(X_i \geq x|\theta) = [\pi_{ix} + \cdots + \pi_{im}]'$$
$$= \pi'_{ix} + \cdots + \pi'_{im}.$$

This completes the outline of the proof. □

A full example for $L = 2$ and $m = 2$

For the special case that $L = 2$ and $m = 2$ we show explicitly that if $P'(X_i \geq x|\theta) \geq 0$ for all i and x, then $P'(X_+ \geq x_+|\theta) \geq 0$ for all x_+. Note that $X_i = 0, 1, 2$; and $X_+ = 0, 1, 2, 3, 4$. Let $\pi_{ix} = P(X_i = x|\theta)$ for notational convenience.

$$\begin{aligned} P'(X_+ \geq 4|\theta) &= P'(X_+ = 4|\theta) \\ &= \pi'_{12}\pi_{22} + \pi'_{22}\pi_{12} \\ &= P'(X_1 \geq 2|\theta)\pi_{22} + P'(X_2 \geq 2|\theta)\pi_{12} \\ &\geq 0 \end{aligned}$$

$$\begin{aligned} P'(X_+ \geq 3|\theta) &= P'(X_+ = 3|\theta) + P'(X_+ \geq 4|\theta) \\ &= \pi'_{11}\pi_{22} + \pi'_{22}\pi_{11} + \pi'_{12}\pi_{21} + \pi'_{21}\pi_{12} + \pi'_{12}\pi_{22} + \pi'_{22}\pi_{12}. \end{aligned}$$

Grouping the terms with the same π_{ix} leads to

$$P'(X_1 \geq 1|\theta)\pi_{22} + P'(X_2 \geq 1|\theta)\pi_{12} + P'(X_1 \geq 2|\theta)\pi_{21} + P'(X_2 \geq 2|\theta)\pi_{11} \geq 0$$

Analogously,

$$
\begin{aligned}
P'(X_+ \geq 2|\theta) &= P'(X_+ = 2|\theta) + P'(X_+ \geq 3|\theta) \\
&= \pi'_{10}\pi_{22} + \pi'_{22}\pi_{10} + \pi'_{11}\pi_{21} + \pi'_{21}\pi_{11} + \pi'_{12}\pi_{20} \\
&\quad + \pi'_{20}\pi_{12} + P'(X_+ \geq 3|\theta) \\
&= P'(X_1 \geq 0|\theta)\pi_{22} + P'(X_2 \geq 0|\theta)\pi_{12} + P'(X_1 \geq 1|\theta)\pi_{21} \\
&\quad + P'(X_2 \geq 1|\theta)\pi_{11} + P'(X_1 \geq 2|\theta)\pi_{20} + P'(X_2 \geq 2|\theta)\pi_{10} \\
&\geq 0
\end{aligned}
$$

$$
\begin{aligned}
P'(X_+ \geq 1|\theta) &= P'(X_+ = 1|\theta) + P'(X_+ \geq 2|\theta) \\
&= \pi'_{10}\pi_{21} + \pi'_{21}\pi_{10} + \pi'_{11}\pi_{20} + \pi'_{20}\pi_{11} + P'(X_+ \geq 2|\theta) \\
&= P'(X_1 \geq 0|\theta)\pi_{22} + P'(X_2 \geq 0|\theta)\pi_{12} + P'(X_1 \geq 0|\theta)\pi_{21} \\
&\quad + P'(X_2 \geq 0|\theta)\pi_{11} + P'(X_1 \geq 1|\theta)\pi_{20} + P'(X_2 \geq 1|\theta)\pi_{10} \\
&\geq 0
\end{aligned}
$$

$$P'(X_+ \geq 0|\theta) = [1]' = 0$$

References

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., & Junker, B. W. (1995). *Tail-measurability in monotone latent variable models*. Manuscript submitted for publication.

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.

Hemker, B. T., Sijtsma, K., & Molenaar, I. W. (1995). Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. *Applied Psychological Measurement, 19*, 337–352.

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*, 679–693.

Holland, P. W. (1981). When are item response models consistent with observed data? *Psychometrika, 46*, 79–92.

Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika, 55*, 577–601.

Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics, 14*, 1523–1543.

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.

Junker, B. W., & Ellis, J. L. (1995). *A characterization of monotone unidimensional latent variable models* (CMU Statistics Department Tech. Rep. #614). Manuscript submitted for publication.

Lehmann, E. L. (1959, 1986). *Testing statistical hypotheses*. New York: Wiley.

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.

Maydeu-Olivares, A., Drasgow, F., & Mead, A. D. (1994). Distinguishing among parametric item response models for polychotomous ordered data. *Applied Psychological Measurement, 18*, 245–256.

Mellenbergh, G. J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement, 19*, 91–100.

Mokken, R. J. (1971). *A theory and procedure of scale analysis*. New York/Berlin: De Gruyter.
Molenaar, I. W. (1982). Mokken scaling revisited. *Kwantitatieve Methoden, 3* (No. 8), 145–164.
Molenaar, I. W. (1997). Nonparametric models for polytomous responses. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 369–380). New York: Springer.
Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement, 16*, 159–176.
Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika, 49*, 425–435.
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.
Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph*, No 18.
Stout, W. F. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*, 589–617.
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.

Manuscript received 8/25/95
Final version received 4/9/96