Measurement and Research Department Reports 2000–6

On Measurement Properties of Continuation Ratio Models

B.T. Hemker

L.A. van der Ark

K. Sijtsma

Citogroep

Measurement and Research Department Reports 2000–6

On Measurement Properties of Continuation Ratio Models

B.T. Hemker

L.A. van der Ark

K. Sijtsma

Citogroep
Arnhem, december 2000

Abstract

Three classes of polytomous IRT models are distinguished. These classes are the adjacent category models, the cumulative probability models, and the continuation ratio models. So far, the latter class has received relatively little attention. The class of continuation ratio models includes logistic models, such as the sequential model (Tutz, 1990), and non-logistic models, such as the acceleration model (Samejima, 1995) and the nonparametric sequential model (Hemker, 1996). Four measurement properties are discussed. These are monotone likelihood ratio of the total score, stochastic ordering of the latent trait by the total score, stochastic ordering of the total score by the latent trait, and invariant item ordering. These properties have been investigated previously for the adjacent category models and the cumulative probability models, and for the continuation ratio models this is done here. It is shown that stochastic ordering of the total score by the latent trait is implied by all continuation ratio models, while monotone likelihood ratio of the total score and stochastic ordering on the latent trait by the total score are not implied by any of the continuation ratio models. Only the sequential rating scale model implies the property of invariant item ordering. Also, we present a Venn-diagram showing the relationships between all known polytomous IRT models from all three classes.

Key words: acceleration model, adjacent category models, continuation ratio models, cumulative probability models, hierarchical relationships between IRT models, invariant item ordering, monotone likelihood ratio, polytomous IRT models, sequential model, stochastic ordering.

General Introduction

Sequential Scoring of Polytomous Items

In the social and behavioral sciences data collected by means of items in tests and questionnaires are often ordered scores, where a higher score indicates a higher position on a latent trait such as arithmetic ability, introversion, or attitude towards capital punishment. Examples of items with ordered scores are used in the "NT2-profiel toets " (CITO, 1999), an ability test for Dutch as a foreign language. We discuss such an item and its sequential scoring rule here because it appears to be well suited for the class of continuation ratio IRT models (CRMs; Agresti, 1990, pp. 319-321; Mellenbergh, 1995; Molenaar, 1983) that is central to this paper. Each item of the "NT2-profiel toets" consists of a spoken Dutch text that ends with a question about this text, for example (see also Hemker, 2001),

[translated from Dutch:] *Suppose, you work at an office. You have to fax a letter for your boss. You have no experience with the fax machine. You know a colleague who is able to use the fax machine. What do you ask your colleague?*

An examinee has to give a verbal response (in Dutch). Examinees are tested individually by an examiner, who scores each item. The item is scored as follows. In the first step, the content of the answer is assessed. If the response is incorrect with respect to content (e.g., “Can I use this fax machine?”), the first step is failed and the result is an item score of 0. Only if the examinee's response is correct or almost correct (i.e., a request for help or for an explanation of the operation of the fax machine) the first step is passed and the examiner proceeds with the second step. In the second step the examinee's use of grammar is assessed. If the examinee makes more than just a few insignificant grammatical errors, the second step is failed and the result is an item score of 1. Only if the examinee's response contains no more than a few unimportant grammatical errors the second step is passed and the examiner proceeds with the third step. In the third step the pronunciation of the response is assessed. If the examiner thinks that the average Dutchman will not be able to understand the response easily, the third step is failed and the result is an item score of 2. If the examiner thinks that the average Dutchman can understand the response without too much difficulty, the third is passed and the result is an item score of 3.

Classes of Polytomous Item Response Models

Continuation Ratio Models

The class of CRMs to be discussed here may be suited particularly for modeling data obtained through a sequential scoring rule as illustrated by the example. CRMs usually have logistic response functions. Hemker (1996, chapter 6) extended the class of CRMs to also include nonparametric response functions of which logistic functions are special cases. Before discussing the general form of CRMs, we first introduce some notation. Let the latent trait be denoted by θ, the random variable for the score on item j by X_j, and realizations by $x = 0, \ldots, m$. Furthermore, all models discussed here assume a unidimensional θ and locally independent item scores. First, we define the conditional probability of passing an item step as

$$M_{jx}(\theta) = P(X_j \geq x \mid X_j \geq x-1; \theta) = \frac{P(X_j \geq x \mid \theta)}{P(X_j \geq x-1 \mid \theta)}. \quad (1)$$

Equation 1 implies that if $x = 0$ then $M_{jx}(\theta) = 1$ for all θ. Equation 1 is the item step response function (ISRF). The conditional probability of obtaining an item score of *x*, $P(X_j = x \mid \theta)$, is decomposed into a product of *x* terms $M_{jx}(\theta)$ and one term $1 - M_{j,x+1}(\theta)$ as

$$P(X_j = x \mid \theta) = \prod_{y=0}^{x} M_{jy}(\theta)\left[1 - M_{j,x+1}(\theta)\right] \quad (2)$$

(Samejima, 1972, chapter 4). Equation 2 is the category characteristic curve (CCC). Thus, CRMs formalize sequential scoring by writing the CCC as a product of *x* ISRFs for the *x* subtasks that were successfully solved and the conditional probability of failing subtask *x*+1 given that the previous subtasks were mastered. Thus, it is assumed that the steps are executed in a fixed sequence. Tutz (1990) discussed two parametric CRMs and characterized both as sequential models.

Adjacent Category Models

If the order in which the steps are presented to the respondent is not fixed, then two other classes of models for ordered item scores might be used. These two classes use alternative definitions of the ISRF (e.g., Mellenbergh, 1995; Molenaar, 1983). One class of models is known as adjacent category models (ACMs). The ISRF of models from this class is defined as

$$F_{jx}(\theta) = \frac{P(X_j = x \mid \theta)}{P(X_j = x \mid \theta) + P(X_j = x-1 \mid \theta)}. \quad (3)$$

It may be noted that the ISRF of ACMs (Equation 3) and the ISRF of CRMs (Equation 1) are related by

$$F_{jx}(\theta) = \frac{M_{jx}(\theta) - M_{jx}(\theta) M_{j,x+1}(\theta)}{1 - M_{jx}(\theta) M_{j,x+1}(\theta)}.$$

Thissen and Steinberg (1986) called parametric models from the class of ACMs divide-by-total models and Andrich (1995) called these models Rasch models. Some well-known divide-by-total models are the rating scale model (Andersen, 1977; Andrich, 1978) and the generalized partial credit model (Muraki, 1992). The best known of these parametric ACMs is the partial credit model (Masters, 1982), defined by

$$F_{jx}(\theta) = \frac{\exp(\theta - \delta_{jx})}{1 + \exp(\theta - \delta_{jx})}, \quad (4)$$

where δ_{jx} is a location parameter. Hemker, Sijtsma, Molenaar and Junker (1996) introduced a more general class including a nonparametric model. They called this model the nonparametric partial credit model, defined by $F_{jx}(\theta)$ (Equation 3) nondecreasing in θ.

Cumulative Probability Models

The third class of models is known as cumulative probability models (CPMs). The ISRF of models from this class is defined as

$$G_{jx}(\theta) = P(X_j \geq x \mid \theta) \; . \tag{5}$$

It may be noted that the ISRF of CPMs (Equation 4) and the ISRF of CRMs (Equation 1) are related by

$$G_{jx}(\theta) = \prod_{y=1}^{x} M_{jy}(\theta) . \tag{6}$$

Thissen and Steinberg (1986) called parametric CPMs difference models, because the CCC is obtained by the difference of two adjacent ISRFs. Andrich (1995) called these models Thurstone models. A well-known CPM is the homogeneous case of the graded response model (Samejima, 1969; also, see Samejima, 1997), defined as

$$G_{jx}(\theta) = \frac{\exp[\alpha_j(\theta - \lambda_{jx})]}{1 + \exp[\alpha_j(\theta - \lambda_{jx})]} ,$$

where α_j denotes the slope parameter and λ_{jx} a location parameter, different from δ_{jx} in Equation 4 (see Masters, 1988, for a discussion of the interpretations of δ_{jx} and λ_{jx}). When it is assumed that the ISRF in Equation 5 is nondecreasing in 2, without defining the ISRF parametrically, the nonparametric graded response model is obtained (for example, Hemker et al., 1996).

Table 1 summarizes the terminology used to identify the three classes of polytomous IRT models. Van Engelenburg (1997, chapter 2) argued that with each of the three classes of polytomous IRT models corresponds a particular type of task, and Akkermans (1998, chapter 3) argued that with each class corresponds a particular type of scoring rule.

Table 1:

An Overview of the Terminology Used to Identify Classes of IRT Models

	Definition	ISRF	
Terminology	$\frac{P(X_j = x \mid \theta)}{P(X_j = x \vee x-1 \mid \theta)}$	$P(X_j \geq x \mid \theta)$	$\frac{P(X_j \geq x \mid \theta)}{P(X_j \geq x-1 \mid \theta)}$
Molenaar (1983); parametric and nonparametric models	Adjacent Category Models (ACMs)	Cumulative Probability Models (CPMs)	Continuation Ratio Models (CRMs)
Thissen and Steinberg (1986); Parametric models only	Divide-by-Total Models	Difference Models	
Andrich (1995); Parametric models only	Rasch Models	Thurstone Models	
Tutz (1990); Parametric models only			Sequential Models

Motivation of this Study

Thissen and Steinberg (1986) discussed a taxonomy for divide-by-total models and difference models. This taxonomy also included models with guessing parameters that are not relevant for this study and, consequently, are left out of consideration. The taxonomy only pertained to parametric models. Hemker et al. (1997) discussed a taxonomy that basically extended the taxonomy of Thissen and Steinberg to include nonparametric models. Moreover, the formal relationships between all models were described by means of a Venn-diagram, based on stochastic ordering (SO) relations between the latent trait *2* and the unweighted sum of J item scores, denoted X_+. Sijtsma and Hemker (1998) discussed the same classes of models with respect to the item ordering property known as invariant item ordering (Sijtsma & Junker, 1996).

A missing link in this research is the class of CRMs. Both classes of ACMs and CPMs have been investigated thoroughly (Andersen, 1977, 1997; Andrich, 1978, 1995; Glas, 1989;

Kelderman & Rijkes, 1994; Masters, 1982; Masters & Wright, 1997; Muraki, 1990, 1992; Samejima, 1969, 1972; Verhelst, Glas, & Verstralen, 1995) and models from these classes have been applied to many practical data analysis problems (some recent applications include Alexander & Murphy, 1998; Cooke, Michie, Hart, & Hare, 1999; Gumpel, Wilson, & Shalev, 1998; Maurer, Raju, & Collins, 1998; and Sijtsma & Verweij, 1999). Although potentially useful, the class of CRMs thus far has received relatively little attention (the exceptions are Samejima, 1995; Tutz, 1990, 1997; and Verhelst, Glas, & De Vries, 1997). CRMs are attracting more attention nowadays, given the recent studies by Hemker (1996), Van Engelenburg (1997) and Akkermans (1998), who compared CRMs with other polytomous IRT models. Thus, it seems reasonable to better incorporate the class of CRMs into the polytomous IRT framework. A contribution to this is given in this paper, where we investigate likelihood ratio and SO properties between the latent trait *2* and the sum score X_+, and also the invariant item ordering property. Insight into these relationships contributes to a better understanding of the relationships of CRMs to ACMs and CPMs and, moreover, gives indications of the practical usefulness of CRMs.

Introduction to Continuation Ratio Models

We discuss the most general model from the class of CRMs, and then we discuss several special cases. The most general model is the nonparametric sequential model (Hemker, 1996, chapter 6), which assumes an order-restricted ISRF, without parametrically defining it. The nonparametric sequential model assumes a unidimensional *2*, locally independent item scores, and a nondecreasing ISRF, given by Equation 1. Several special cases have been proposed.

Samejima (1995) assumed a semi-parametric ISRF, $M_{jx}(\theta) = [\Psi_{jx}(\theta)]^{\xi_j}$, where $\xi_j \geq 0$ is the acceleration parameter. The function $\Psi_{jx}(\theta) = P(X_j \geq x \mid X_j \geq x-1; \theta)$ is nonparametric and is assumed to be strictly increasing with 0 and 1 as its horizontal asymptotes. The acceleration model is the parametric version of $M_{jx}(\theta) = [\Psi_{jx}(\theta)]^{\xi_j}$. Let α_{jx} denote the discrimination parameter and β_{jx} the location parameter associated with category x of item j, and let D be a scaling constant, usually equal to 1.7 to scale the logistic function to the normal-ogive. The acceleration model assumes that

$$M_{jx}(\theta) = \left\{ \frac{\exp[D\alpha_{jx}(\theta - \beta_{jx})]}{1 + \exp[D\alpha_{jx}(\theta - \beta_{jx})]} \right\}^{\xi_j}.$$

It may be noted that the acceleration model is not a logistic model for $\xi_j \neq 1$. The acceleration parameter contributes to the steepness of the complete ISRFs, whereas α_{jx} influences the steepness of a logistic curve in its inflection point: $\xi_j > 1$ "pushes down" the entire curve and $\xi_j < 1$ "lifts up" the entire curve, where both effects add to the effect of α_{jx} on the slope of an ISRF in the inflection point. Figure 1a gives a graphic example of the acceleration model, and shows two items each having two ISRFs (solid and dashed curves for different items; parameter values are given in Appendix B) with three answer categories. Figure 1a shows that $M_{jx}(\theta)$ is not symmetric in its inflection point.

For $\xi_j = 1$, the 2-parameter sequential model with parameters for each (j, x) combination, abbreviated 2p(jx)-sequential model, is obtained. This is a logistic model defined by

$$M_{jx}(\theta) = \frac{\exp[\alpha_{jx}(\theta - \beta_{jx})]}{1 + \exp[\alpha_{jx}(\theta - \beta_{jx})]}.$$

A special case of this model can be obtained by fixing α_{jx} across answer categories, so that $\alpha_{jx} = \alpha_j$. The resulting model is the 2p(j)-sequential model. Another possibility is to fix α_{jx} across items, so that $\alpha_{jx} = \alpha_x$. This results in the 2p(x)-sequential model. In Figure 1b, Figure 1c and Figure 1d, we give graphic examples of the 2p(jx)-sequential model, the 2p(j)-sequential model and the 2p(x)- sequential model, respectively (parameter values in Appendix B).

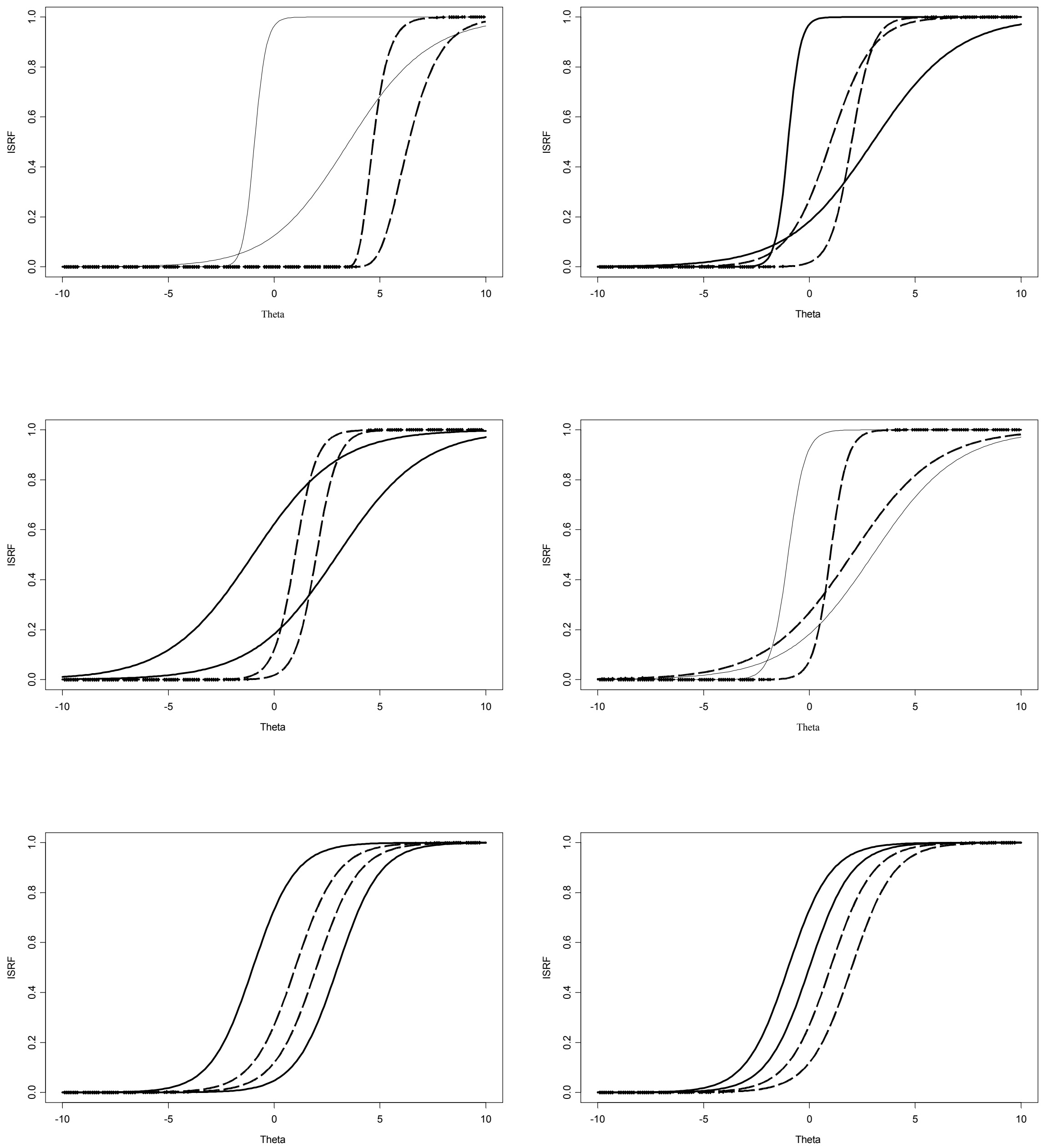


Figure 1: The ISRFs of Six Parametric CRMs for Two Items (Solid and Dashed Curve) with Three Answer Categories. Figure 1a is the Acceleration Model; Figure 1b is the 2p(jx)-Sequential Model; Figure 1c is the 2p(j)- Sequential Model; Figure 1d is the 2p(x)- Sequential Model; Figure 1e is the Sequential Rasch Model; Figure 1f is the Sequential Rating Scale Model.

In the sequential Rasch model (Tutz, 1990) or, equivalently, the 1p-sequential model, the ISRF $M_{jx}(\theta)$ is further constrained by fixing $\alpha_{jx} = 1$, so that

$$M_{jx}(\theta) = \frac{\exp(\theta - \beta_{jx})}{1 + \exp(\theta - \beta_{jx})}. \qquad (7)$$

Alternatively, we may write

$$\operatorname{logit}\left[M_{jx}(\theta)\right] = \log\left[\frac{M_{jx}(\theta)}{1 + M_{jx}(\theta)}\right] = \theta - \beta_{jx}.$$

De Vries (1988) and Verhelst, Glas, and de Vries (1997) introduced the sequential model to analyze partial credit as an alternative to Masters' partial credit model. Their model is equivalent to the sequential Rasch model (Equation 7).

A special case of the 1p-sequential model is the sequential rating scale model (Tutz, 1990), in which the location parameter β_{jx} is split up into an item location parameter ε_j and a step location parameter τ_x, with $\sum_x \tau_x = 0$. The sequential rating scale model is the most restricted CRM proposed. Graphic examples of the sequential Rasch model and the sequential rating scale model are given in Figure 1e and Figure 1f, respectively. It may be noted that in the sequential Rasch model, the sequential rating scale model, and the 2p-sequential models the logit of $M_{jx}(\theta)$ is a linear function of the model parameters (Mellenbergh, 1995; Molenaar, 1983). This is not true in the acceleration model. Figure 2 shows the relationship between the various CRMs. The arrows in Figure 2 should be read as logical symbols for an implication.

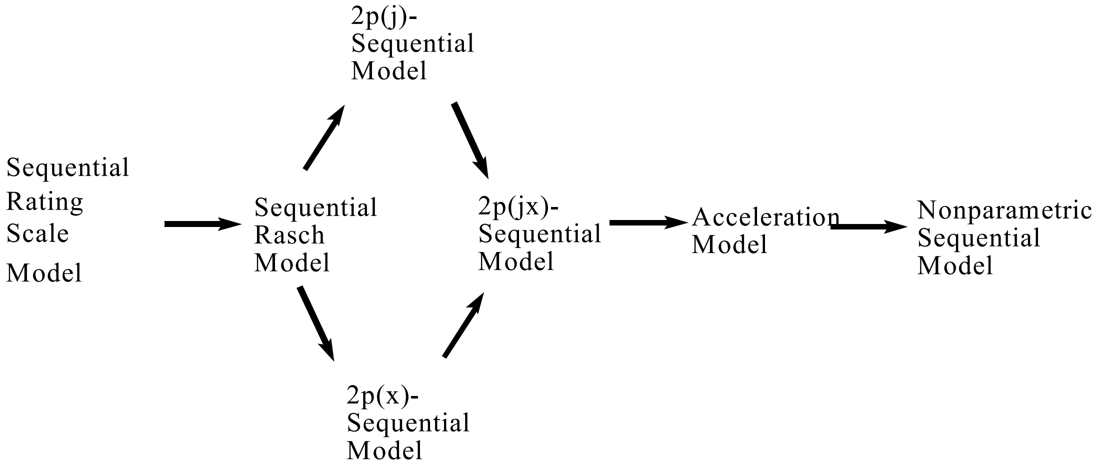


Figure 2: Hierarchical Relationships Within the Class of CRMs.

Measurement Properties for Persons and Items

Measurement Properties for Persons

Motivation for Using Total Score

We assume J polytomous items with $m+1$ ordered answer categories each and a simple scoring rule for each item, that is, $X_j = 0, \ldots, m$, for all j. The unweighted total score is

$$X_+ = \sum_{j=1}^{J} X_j,\ X_+ = 0, \ldots, mJ .$$

Samejima (1996) criticized the use of X_+ for estimating θ, because the amount of test information based on any aggregation of the response patterns, such as X_+, cannot exceed the amount of test information obtained from the response patterns, unless X_+ is a sufficient statistic for θ (Samejima, 1969, chapter 6). Sijtsma and Hemker (2000) extensively discussed the *practical* usefulness of X_+ as opposed to the *theoretical* usefulness of θ, for example, as discussed by Samejima. They argue that X_+ is better suited than θ for communicating test results to measurement practitioners and laymen, because X_+ has an interpretation closely related to solving problems correct or incorrect (dichotomous items) or the number of points earned (polytomous items), whereas θ has a complicated interpretation in terms of logits (see Mellenbergh, 1995). On the contrary, for test practitioners X_+ is quick and simple, and allows immediate feedback to testees. Also, Sijtsma and Hemker (2000) note that nothing prevents psychometricians and test constructors to use IRT for test construction and the information function for measurement evaluation of the estimated θ on the one hand, and test practitioners, including teachers, to score performance on those same tests by means of summary scores such as X_+ on the other hand. The use of X_+ is further corroborated by a theoretical result of Junker (1991), who showed in the context of the nonparametric graded response model [Equation 5, response probability $G_{jx}(\theta)$ nondecreasing in θ] that for infinitely many polytomous items X_+ consistently estimates θ. In this paper we investigate for CRMs whether X_+ can be used for ordering respondents on θ in an SO sense, which is also useful in a nonparametric IRT context where numerical estimates of θ are not available.

We agree with Samejima (1996) that for the evaluation of measurement precision X_+ is not the optimal statistic, but we also believe that X_+ may be an adequate summary test score for ordering persons on θ in a nonparametric context and for communication purposes in a general IRT context. Also, Hemker et al. (1997) used measurement properties based on X_+ to study the relationships between all known ACMs and CPMs. This paper completes this investigation by presenting a Venn-diagram displaying the relationships between all known polytomous IRT models from the classes of ACMs, CPMs, and also CRMs.

Monotone Likelihood Ratio

The first measurement property we consider is *monotone likelihood ratio* (MLR). For polytomous items, MLR of X_+ in θ means that for $0 \le C < K \le mJ$,

$$g(K, C; \theta) = \frac{P(X_+ = K|\theta)}{P(X_+ = C|\theta)} \qquad \text{(MLR)}$$

is a nondecreasing function of random variable θ (Lehmann, 1959). It can be shown that the MLR property is symmetric in its arguments. By writing the ratio in Equation (MLR) twice, conditioning once on θ_a and once on θ_b , with $\theta_a < \theta_b$, so that

$$\frac{P(X_+ = K | \theta = \theta_a)}{P(X_+ = C | \theta = \theta_a)} \leq \frac{P(X_+ = K | \theta = \theta_b)}{P(X_+ = C | \theta = \theta_b)},$$

then rearranging probabilities, and applying Bayes' Theorem, eventually we have that

$$\frac{P(\theta = \theta_b \mid X_+ = C)}{P(\theta = \theta_a \mid X_+ = C)} \leq \frac{P(\theta = \theta_b \mid X_+ = K)}{P(\theta = \theta_a \mid X_+ = K)}.$$

This result means that MLR of X_+ in θ is equivalent to MLR of θ in X_+. MLR is a technical property that implies two SO properties (Lehmann, 1959, p. 74) that can be interpreted conveniently 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, Section 3.3; see also, Junker, 1993; Rosenbaum, 1985). In addition, the SO properties do not imply each other.

Stochastic Ordering Properties

First, MLR implies the stochastic ordering of the manifest variable X_+ by θ (abbreviated SOM). That is, for any two respondents *a* and *b* with $\theta_a < \theta_b$, and for any x_+,

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

SOM takes the ordering on θ as a starting point, and implies that a higher θ results in a higher expected total score [see Lehmann, 1986, p. 85, Lemma 2(i); which pertains to the MLR property].

Second, MLR implies the stochastic ordering of the latent trait θ by X_+ (abbreviated SOL). This means that for any constant value *s* of θ , and for all $0 \leq C < K \leq mJ$,

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

SOL takes the ordering on X_+ as a starting point, and implies that a higher X_+ results in a higher expected θ [Lehmann, 1986, p. 85, Lemma 2(i)]. In practice, SOL is of more interest than SOM, because only the ordering on X_+ can be observed and inferences about θ are based on X_+. For example, SOL is required for making mastery decisions based on cutoffs for the total score X_+.

Grayson (1988; also see Huynh, 1994) showed that, given unidimensionality, local independence, and monotonicity, MLR holds for tests consisting of dichotomously scored items. By implication, SOM and SOL also hold under these assumptions. For the classes of well known ACMs and CPMs, Hemker, et al. (1996) showed that MLR holds only for the partial credit model (and its special cases), but for none of the other well known polytomous models. In addition, Hemker et al. (1997) showed that SOL also holds only for the partial credit model, but that SOM holds for each of the well known parametric and nonparametric ACMs and CPMs. For the class of CRMs, the properties of MLR, SOM, and SOL have not been investigated thus far.

A Measurement Property for Items: Invariant Item Ordering

Let $E(X_j \mid \theta)$ denote the conditional expected score of item *j*. This conditional expectation is the item response function (IRF), both for dichotomous and polytomous items (Chang & Mazzeo, 1994). Unlike for dichotomous items, for polytomous items the IRF is not a probability, but a function ranging from 0 to *m*. Invariant item ordering (IIO; Sijtsma & Junker, 1996; Sijtsma & Hemker, 1998) means that the items have the same ordering by $E(X_j \mid \theta)$, except for possible ties, for all values of θ. In general, *J* items have an IIO (Sijtsma & Hemker, 1998; Definition) if they can be ordered and numbered such that

$$E(X_1 \mid \theta) \leq E(X_2 \mid \theta) \leq \ldots \leq E(X_J \mid \theta) \text{ ; for all } \theta. \qquad \text{(IIO)}$$

Within meaningful subgroups, such as age groups, items may also be ordered using $E(X_j)$, $j = 1, \ldots, J$, which is the mean item score across the distribution of θ in a particular subgroup. If an IIO holds, that is, an item ordering that is the same for all θ s, then the items also have the same ordering with respect to $E(X_j)$ between different subgroups.

IIO is a useful property when the application of a test assumes that items have the same ordering for different θ s. For example, in intelligence testing using a conventional test format (i.e., not an adaptive test format) items are often ordered from easy to difficult to facilitate the use of starting and stopping rules for individuals (e.g., the Amsterdam Child Intelligence Test; Bleichrodt, Drenth, Zaal, & Resing, 1985) in the following way. The youngest age group starts with the easiest item and an individual child stops when he/she failed at, for example, three consecutive items (the next items are more difficult and it is assumed that the child will also fail at those items). The next age groups skips, say, the first five items, which are assumed to be too easy for them, and starts at item 6. For each individual child, the same stopping rule applies. The third age group starts at, say, item 16, and so on. Obviously, this test administration procedure uses the assumption that for the whole population the items have an IIO.

Other applications where an IIO is relevant are the following. Several person fit detection methods are based on the difficulty ordering of the items, and applications to individuals all use the same item difficulty ordering. Also, items may reflect a developmental sequence that is assumed to hold for each individual, and the difficulty ordering that results from the developmental ordering by implication also holds at the individual level. Finally, when items are assumed to be unbiased the ordering according to difficulty should be the same in different meaningful subgroups, for example, defined by gender, ethnicity, and social economic status.

For dichotomous and polytomous items, all IRT models having nonintersecting IRFs imply an IIO (Sijtsma & Junker, 1996; Sijtsma & Hemker, 1998). For dichotomous items, the Rasch (1960) model and the double monotonicity model (Mokken & Lewis, 1982) are well known examples. For polytomous items, the ISRFs of different items need not be nonintersecting to obtain nonintersecting IRFs. Sijtsma and Hemker (1998) showed that in the ACM class the rating scale model (Andrich, 1978) implies an IIO, and in the CPM class the rating scale version of the graded response model with equal ISRF slopes [a special case of Muraki's (1990) model], the strong double monotonicity model (Sijtsma & Hemker, 1998), and the isotonic ordinal probabilistic model (ISOP; Scheiblechner, 1995) each imply an IIO.

Measurement Properties of the Continuation Ratio Models

First, we show that CRMs do *not* imply MLR. Next, we show that all CRMs imply SOM, but that none of the CRMs imply SOL. Finally, we show that the sequential rating scale model implies an IIO when all items have the same number of answer categories. We will derive all results assuming that the number of answer categories is fixed over items, which is realistic in most applications. Also, this is the assumption followed in previous research on MLR (Hemker et al., 1996), SOM and SOL (Hemker et al., 1997), and IIO (Sijtsma & Hemker, 1998).

Monotone Likelihood Ratio

Example 1 (below) shows that the sequential rating scale model (Equation 7, with $\beta_{jx} = \tau_x + \varepsilon_j$ substituted) does not imply MLR. Since the sequential rating scale model is a special case of all other CRMs (see Figure 2), it follows that none of these more general models implies MLR.

EXAMPLE 1. *The sequential rating scale model does not imply MLR*. Consider two items $(J = 2; j = 1, 2)$, each with five answer categories ($m = 4$). Let the item locations be $\varepsilon_1 = 0$ and $\varepsilon_2 = 1$, and let the category locations be $\tau_1 = -.99$, $\tau_2 = .98$, $\tau_3 = -1.00$, and $\tau_4 = 1.01$. This means that $\beta_{11} = -.99$, $\beta_{12} = .98$, $\beta_{13} = -1.00$, and $\beta_{14} = 1.01$; and $\beta_{21} = .01$, $\beta_{22} = 1.98$, $\beta_{23} = .00$, and $\beta_{24} = 2.01$. Figure 3 shows the corresponding functions $g(C+1, C; \theta) = \dfrac{P(X_+ = C+1 \mid \theta)}{P(X_+ = C \mid \theta)}$ (see Equation MLR) for $0 \le C \le 7$. The likelihood ratio function that decreases from $\theta \approx 1.47$ to infinity, is $g(C+1=6, C=5; \theta)$. This function shows that the sequential rating scale model does not imply MLR.

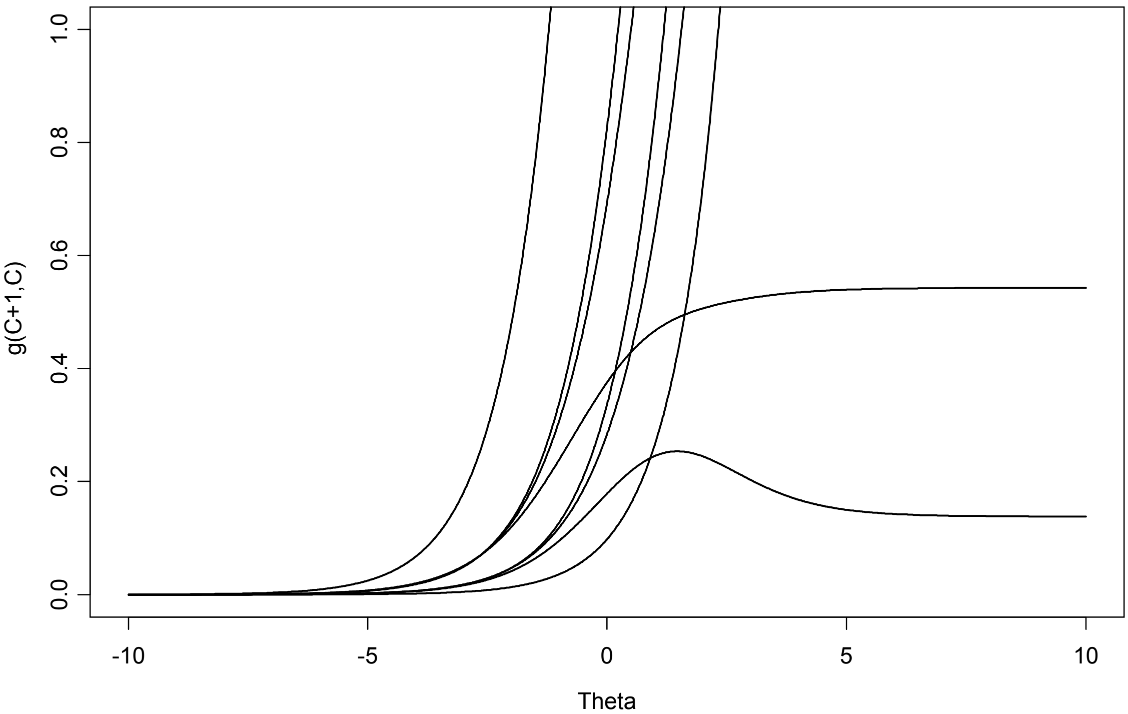


Figure 3: Graphic Display of Eight Curves Representing $\dfrac{P(X_+ = C+1 \mid \theta)}{P(X_+ = C \mid \theta)}$ for $C = 0, \ldots, 7$, Obtained From a Sequential Rating Scale Model for Two Items With Five Ordered Answer Categories.

For many other choices of the location parameters than the values in Example 1, the likelihood ratio $g(C+1, C; \theta)$ is often found to be nondecreasing for all C. For the special cases of maximum total score $X_+ = mJ$ and minimum total score $X_+ = 0$, CRMs even imply MLR mathematically (proof in Appendix C). Another special case is MLR of item score X_j. Hemker et al. (1997; Proposition) showed that MLR of item score X_j is equivalent to nondecreasingness of the ISRF of the ACM class (Equation 3). Additionally, Hemker (1996, chapter 6) showed that parametric CRMs with $\alpha_{jx} \geq \alpha_{j,x+1}$ imply that the ISRFs of the ACM class are nondecreasing. Thus, the 2p(j)-sequential model and its special cases imply MLR when $X_+ = X_j$.

Stochastic Ordering

Since MLR is a sufficient, but not a necessary condition for the properties of SOM and SOL, models that do not have MLR may have one or both SO properties. First, we show that all CRMs imply SOM. Next, we show that none of the CRMs imply SOL.

THEOREM 1. *All CRMs imply SOM.*

PROOF: The proof consists of two parts. First, we prove that all CRMs discussed here imply SOM of X_j. It may be noted that unidimensional θ, local independence, and SOM of X_j together define the nonparametric graded response model [Hemker, et al., 1996; see Equation 5, where the conditional probability $G_{jx}(\theta)$ is assumed to be nondecreasing]. Since all CRMs assume unidimensionality and local independence, and we prove that these models imply SOM of X_j, it follows logically that all CRMs imply the nonparametric graded response model. Second, we prove that the nonparametric graded response model implies SOM. The first part of the proof is given here (also, see Hemker, 1996, chapter 6), and the second part was proven in Hemker et al. (1997, Theorem 1).

Let $\theta_a < \theta_b$. In the nonparametric sequential model the ISRF (Equation 2) is nondecreasing and, therefore,

$$\frac{P(X_j \geq x \mid \theta_a)}{P(X_j \geq x-1 \mid \theta_a)} \leq \frac{P(X_j \geq x \mid \theta_b)}{P(X_j \geq x-1 \mid \theta_b)}, \text{ for all } x \text{ and all } j.$$

It follows that

$$\prod_{y=1}^{x} \frac{P(X_j \geq y \mid \theta_a)}{P(X_j \geq y-1 \mid \theta_a)} \leq \prod_{y=1}^{x} \frac{P(X_j \geq y \mid \theta_b)}{P(X_j \geq y-1 \mid \theta_b)} \Leftrightarrow$$

$$\frac{P(X_j \geq x \mid \theta_a)}{P(X_j \geq 0 \mid \theta_a)} \leq \frac{P(X_j \geq x \mid \theta_b)}{P(X_j \geq 0 \mid \theta_b)}, \text{ for all } x \text{ and all } j. \tag{8}$$

Since the denominators in Equation 8 equal 1, we have that

$$P(X_j \geq x \mid \theta_a) \leq P(X_j \geq x \mid \theta_b),$$

which is equivalent to SOM of the item score X_j. Since the nonparametric sequential model is the least restrictive model in the CRM class all special cases imply SOM.

Next, we investigate SOL. Example 2 (below) gives an example of a sequential rating scale model that does not imply SOL. Because the sequential rating scale model is the most restrictive model in the CRM class, it follows that none of the CRMs imply SOL.

EXAMPLE 2. *The sequential rating scale model does not imply SOL*. This counterexample uses the same parameter values as Example 1. Furthermore, let θ be a discrete latent trait with $P(\theta=0)=0.5$ and $P(\theta=1)=0.5$, then $P(\theta \geq 1 \mid X_+=3) \approx .64$, and $P(\theta \geq 1 \mid X_+=4) \approx .54$. Thus, $P(\theta \geq 1 \mid X_+)$ is *not* nondecreasing in X_+. Consequently, the sequential rating scale model does not imply SOL.

Example 2 remains valid as a counter example of SOL for standard normally distributed θ. The values of $P(\theta > s \mid X_+)$ obtained using numerical integration are given in Table 2, for $X_+ = 4, 5, 6, 7$ and $s = 0,\ 1,\ 2,\ 3$. In Figure 4, $P(\theta > s \mid X_+)$ is depicted for $X_+ = 0, \ldots, 8$ and s ranging from -5 to 5. The left-hand solid curve represents $P(\theta > s \mid X_+ = 0)$, the right-hand solid curve represents $P(\theta > s \mid X_+ = 8)$, and the remaining curves represent the scores ranging from 1 through 7. If SOL holds then the curves are in ascending order according to X_+ and do not intersect. It may be noted, however, that $P(\theta > s \mid X_+ = 5)$ and $P(\theta > s \mid X_+ = 6)$ (third and fourth curve from the right) intersect at $\theta \approx 1.47$; thus, SOL is violated.

Table 2:
Numerical Values Showing That the Sequential Rating Scale Model Does Not Imply SOL. Bold Face Values Indicate Violations of SOL.

X_+	$P(\theta > 0 \mid X_+)$	$P(\theta > 1 \mid X_+)$	$P(\theta > 2 \mid X_+)$	$P(\theta > 3 \mid X_+)$
4	.920	.679	.334	.105
5	.987	.907	.660	.322
6	.991	.912	**.623**	**.265**
7	.999	.983	.845	.492

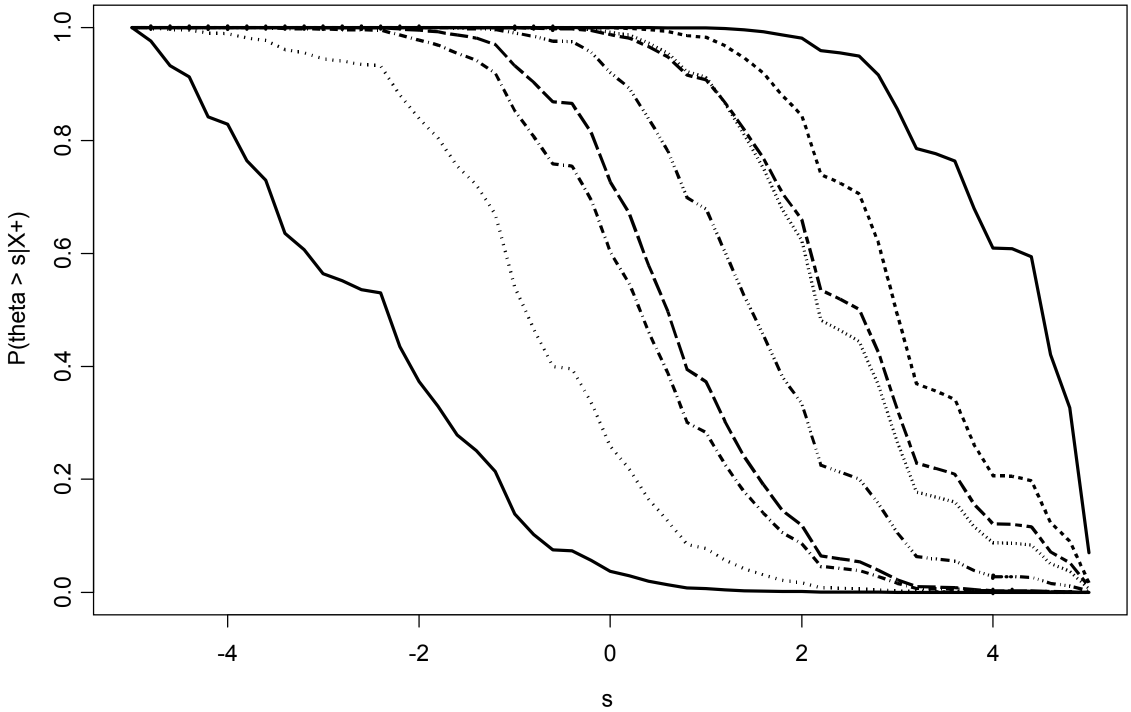


Figure 4: Graphic Display of Nine Curves Representing $P(\theta > s \mid X_+ = K)$ for $K = 0,\dots,8$, Obtained From a Sequential Rating Scale Model for Two Items With Five Ordered Answer Categories.

For most values of X_+ and s there is no problem in the ordering of persons on θ by X_+. In addition, several examples, not provided here, demonstrate that SOL also holds for many values of the item parameters. Example 2 shows, however, that none of the sequential models investigated here implies SOL. SOL is only implied in some special cases. For example, we already showed that MLR holds for all CRMs if $X_+ = mJ$, and that the 2p-sequential model with $\alpha_{jx} \geq \alpha_{j,x+1}$ implies MLR of the item score X_j. We also noted that MLR implies SOL. Consequently, SOL also holds in these special cases. Example 3 (below) shows that, in general, CRMs do not imply SOL of the item score X_j.

EXAMPLE 3. *The 2p(x)-sequential model does not imply SOL of* X_j. Consider an item j with three answer categories. Two ISRFs describe this item: $M_{j1}(\theta)$ and $M_{j2}(\theta)$. Let $\alpha_{j1} = 1$, $\alpha_{j2} = 2$, and $\beta_{j1} = \beta_{j2} = 0$. Thus $\text{logit}[M_{jx}(\theta)] = x\theta$, for all x. Assume a discrete distribution of θ, with $P(\theta = 0) = 0.5$ and $P(\theta = 1) = 0.5$. Then $P(\theta \geq 1 \mid X_j = 0) \approx .52$, $P(\theta \geq 1 \mid X_j = 1) \approx .26$, and $P(\theta \geq 1 \mid X_j = 2) \approx .56$. Thus, $P(\theta \geq 1 \mid X_j)$ is *not* nondecreasing in $X_+ = X_j$. Consequently, SOL does not hold for the 2p(x)-sequential model when $X_+ = X_j$.

Example 3 also implies that the 2p(jx)-sequential model, the acceleration model and the nonparametric sequential model do not imply SOL of the item score X_j.

Invariant Item Ordering

Only the sequential rating scale model implies an IIO. The sequential rating scale model is the most restrictive CRM. First, we prove that the sequential rating scale model implies an IIO. For the sequential Rasch model, Example 4 provides a counterexample, which shows that this model does not imply an IIO. The combination of this result and the hierarchical relationships between the CRMs (see Figure 2) shows that none of the generalizations of the sequential Rasch model imply an IIO.

THEOREM 2. *The sequential rating scale model implies an IIO*

PROOF. Let items i and j have ISRFs according to the sequential rating scale model (Equation 7, with $\beta_{jx} = \varepsilon_j + \tau_x$ substituted). Let the location parameters of the items be ordered $\varepsilon_i \geq \varepsilon_j$, so that $\Delta_{ij} \equiv \varepsilon_i - \varepsilon_j \geq 0$. Because for the CRMs the ISRF (Equation 1) is a nondecreasing function, it follows readily that

$$M_{ix}(\theta) \leq M_{ix}(\theta + \Delta_{ij})\text{; for all } x.$$

From the definition of the sequential rating scale model it follows that for items i and j

$$M_{ix}(\theta + \Delta_{ij}) = M_{jx}(\theta)\text{; for all } x. \tag{9}$$

Equation 9 implies

$$M_{ix}(\theta) \leq M_{jx}(\theta), \text{ for all } x \Rightarrow$$

$$\prod_{k=0}^{x} M_{ix}(\theta) \leq \prod_{k=0}^{x} M_{jx}(\theta) \text{ , for all } x. \quad (10)$$

From Equation 5 and Equation 6 (also see Samejima, 1995) it follows that Equation 10 is identical to

$$P(X_i \geq x \mid \theta) \leq P(X_j \geq x \mid \theta), \text{ for all } x. \quad (11)$$

Next, Equation 11 implies that

$$\sum_{x=1}^{m} P(X_i \geq x \mid \theta) \leq \sum_{x=1}^{m} P(X_j \geq x \mid \theta) \quad . \quad (12)$$

It may be noted that Equation 12 is identical to

$$E(X_i|\theta) \leq E(X_j|\theta).$$

Equation 12 can easily be extended to J items and, therefore, Equation (IIO) holds for all items satisfying the sequential rating scale model.

EXAMPLE 4. *The sequential Rasch model does not imply an IIO*. Consider two items (j = 1, 2), each with three answer categories ($m = 2$). Consider Equation 7 and let the location parameters of the items be $\beta_{11}=-1.5$, $\beta_{12}=2.5$, $\beta_{21}=-.5$, and $\beta_{22}=1$. Figure 5 shows the IRFs for these items. The IRFs intersect at $\theta \approx .4083$. For persons with $\theta < .4083$, item 1 is easier than item 2, and for persons with $\theta > .4083$ the item ordering is reversed.

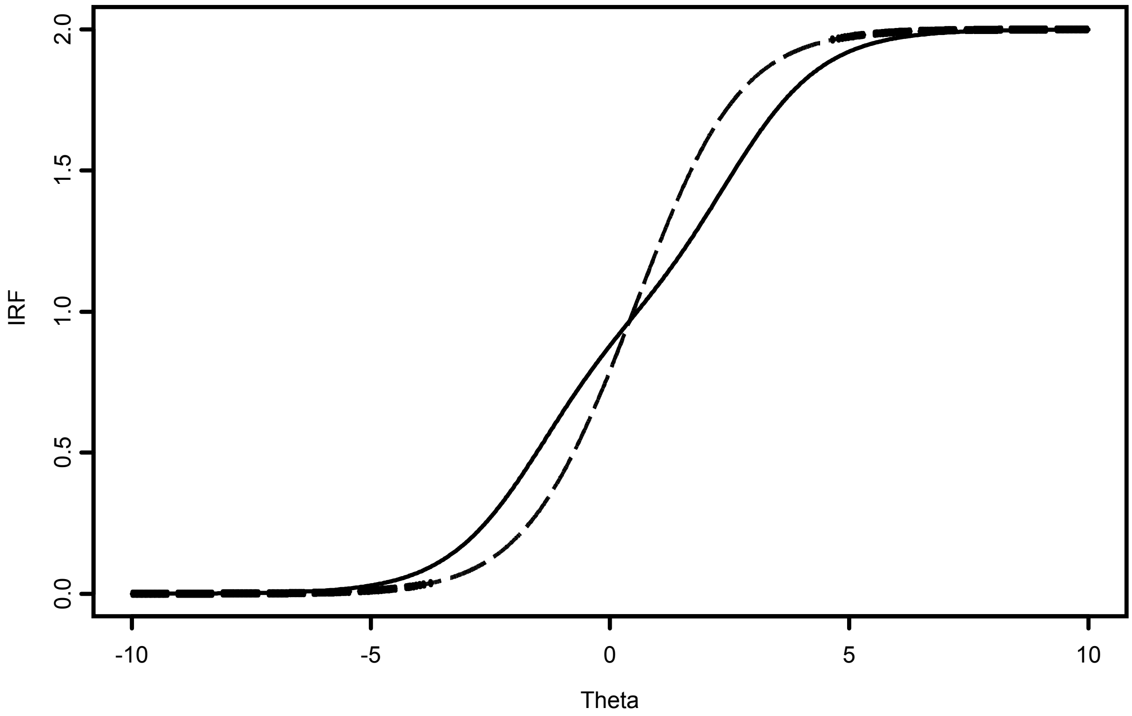


Figure 5: The IRFs (Represented by a Solid and a Dashed Line) of Two Items of the Sequential Rasch Model.

Relationships of Continuation Ratio Models with Other Classes of Polytomous IRT Models

Previous results on formal relationships between all CPMs and ACMs were based on SOL and displayed in a Venn-diagram (Hemker et al., 1997). The results of this paper fit nicely into this framework. Figure 6 extends the Venn-diagram with the relationships between the CRMs, and between the CRMs and the other models. The bold lines indicate the extensions. For a better understanding of Figure 6 we summarize the previous results on the formal relationships.

Molenaar (1983) showed that if the ISRFs of the ACMs, CPMs and CRMs are defined by a logistic function, none of the three types of parametric models can be considered a special case or a generalization of any of the other models. In agreement with this result, Figure 6 shows the three types of parametric models as disjoint clusters of sets, with the outer sets denoted 2p(jx)-PCM, 2p(j)-GRM, and AM, respectively (acronyms explained below Figure 6).

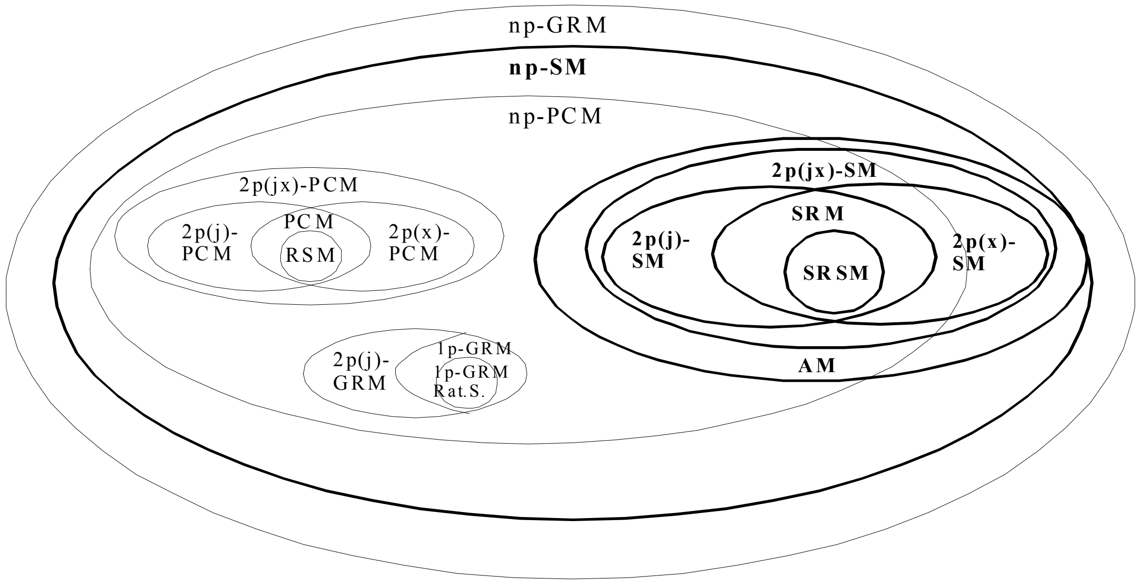
Nonparametric models only restrict the ISRFs to be nondecreasing. When the ISRF in Equation 3 is assumed to be nondecreasing, the nonparametric partial credit model is obtained, and when the ISRF in Equation 5 is assumed to be nondecreasing the nonparametric graded response model is obtained. Hemker (1996, chapter 6) studied the relationship between the nonparametric models of the CRM class, the ACM class and the CPM class. He proved that the nonparametric partial credit model implies the nonparametric sequential model, and that the nonparametric sequential model implies the nonparametric graded response model. In Figure 6, the three outer sets represent this hierarchical relationship.

Hemker et al. (1997) proved that all parametric ACMs and all parametric CPMs are special cases of the nonparametric partial credit model. In Figure 6, the two sets of parametric ACMs [outer set labeled 2p(jx)-PCM] and parametric CRMs (outer set labeled AM) are contained in the set denoted np-SM. Because of this relationship, these two sets of parametric models are also special cases of the nonparametric sequential model and the nonparametric graded response model; see Figure 6. Also, all parametric CRMs are special cases of the nonparametric sequential model (see Figure 2) and, thus, of the nonparametric graded response model (Figure 6).

Finally, Hemker (1996, chapter 6), showed that the 2p(jx)-sequential model is a special case of the nonparametric partial credit model only if $\alpha_{jx} \geq \alpha_{j,x+1}$, for all *j* and *x*. Thus, only the 2p(j)-sequential model and special cases of this model imply nondecreasingness of the ISRF in Equation 3. Therefore, those models are special cases of the nonparametric partial credit model, as can be seen in Figure 6 where only the sets representing these models are contained completely in the set for the np-PCM.

Discussion

This study has yielded two main results. First, we have established which CRMs imply one or more of the measurement properties of monotone likelihood ratio (MLR) of the total score X_+ given the latent trait θ, stochastic ordering of X_+ given θ (SOM), stochastic ordering of θ given X_+ (SOL), and an invariant item ordering (IIO). For polytomous IRT



np-GRM	: nonparametric graded response model
np-SM	: nonparametric sequential model
np-PCM	: nonparametric partial credit model
AM	: acceleration model
2p(jx)-SM	: 2p(jx)-sequential model
2p(j)-SM	: 2p(x)-sequential model
2p(x)-SM	: 2p(j)-sequential model
SRM	: sequential Rasch model
SRSM	: sequential rating scale model
2p(j)-GRM	: graded response model
1p-GRM	: one parameter graded response model
1p-GRM Rat s	: one parameter graded response model with rating scale restrictions
2p(jx)-PCM	: 2p(jx)-partial credit model
2p(j)-PCM	: 2p(j)-partial credit model (generalized partial credit model)
2p(x)-PCM	: 2p(x)-partial credit model
PCM	: partial credit model
RSM	: rating scale model

Figure 6: Venn-Diagram Showing the Relationships of Polytomous IRT Models From the Classes of ACMs, CPMs, and CRMs. Bold Face Notation and Bold Lines Indicate New Results.

models from the classes of adjacent category models (ACMs) and cumulative probability models (CPMs), Hemker et al. (1996) investigated the MLR property. For the same classes of models Hemker et al. (1997) investigated SOM and SOL. This study resulted in a Venn-diagram exhibiting the hierarchical relationships between the models from both classes. Finally, for these two classes of models Sijtsma and Hemker (1998) investigated IIO. The present study thus fills a gap by also investigating these measurement properties for a class of models that was not studied in the previous studies. We now have a complete picture of the measurement properties of MLR, SOM, SOL, and IIO for all polytomous IRT models for ordered items scores that are known to date.

Second, we extended the Venn-diagram for ACMs and CPMs presented by Hemker et al. (1997) with results for CPMs. The resulting Venn-diagram contains the hierarchical relationships between all polytomous IRT models for ordered item scores from each of the three classes of IRT models.

When a model allows for intersecting IRFs, it does not imply an IIO. Because with each intersection of two IRFs the ordering of the $E(X_j \mid \theta)$ s changes, it follows that IRT models with intersecting IRFs imply many different item orderings, which depend on θ. Thus, the question whether some models that do not imply an IIO perhaps might have this property by approximation is not an issue.

The situation is different for the property of SOL, which is the most interesting person ordering property. We have many indications from numerical examples that when a model does not imply SOL, this ordering property still may hold by approximation (Sijtsma & Van der Ark, 2001; Van der Ark, 2000). This means, for example, that when X_+ is used for ordering θ under a model, which does not formally imply SOL, the ordering may be distorted only for two or three adjacent X_+ values. For example, let the scale values run from, say, 0 to 60, decisions be based on a cut-off score of 40, and the distortion of the X_+ ordering occurs only for the values of 21 and 22. Then it could be concluded that the violation of SOL does not really harm an application that uses the cut-off score of 40 as the most relevant scale value.

References

Agresti, A. (1990). *Categorical data analysis.* New York: Wiley.

Akkermans, L.M.W. (1998). *Studies on statistical models for polytomously scored test items.* Unpublished doctoral dissertation, University of Twente, Enschede, The Netherlands.

Alexander, P.A., & Murphy, P.K. (1998). Profiling the difference in students' knowledge, interest, and strategic processing. *Journal of Educational Psychology, 90*, 435-447.

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

Andersen, E.B. (1997). The rating scale model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 67 - 84). New York: Springer.

Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika*, *43*, 561-573.

Andrich, D. (1995). Distinctive and incompatible properties of two common classes of IRT models for graded responses. *Applied Psychological Measurement*, *19*, 101-119.

Bleichrodt, N., Drenth, P. J. D., Zaal, J.N., & Resing, W.C.M. (1985). *Revisie Amsterdamse Kinder-Intelligentie Test (RAKIT)* [Revision of the Amsterdam Child Intelligence Test]. Lisse, The Netherlands: Swets & Zeitlinger.

Chang, H., & Mazzeo, J. (1994). The unique correspondence of the item category response functions in polytomously scored item response models. *Psychometrika, 59*, 391-404.

CITO (1999). *Nederlands als tweede taal (NT2) profiel toets* [Dutch as a foreign language profile test]. Arnhem, The Netherlands: CITO.

Cooke, D.J., Michie, C., Hart, S.D., & Hare, R.D. (1999). Evaluating the screening version of the Hare Psychopathy Checklist Revised (PCI, Sv) - An item response theory analysis. *Psychological Assessment, 11*, 3-13.

De Vries, H. H. (1988). *Het Partial Credit Model en het Sequentiële Rasch Model met stochastisch design* [The partial credit model and the sequential Rasch model with stochastic design]. Unpublished master's thesis, University of Amsterdam.

Glas, C.A.W. (1989). *Contributions to estimating and testing Rasch models*. Unpublished doctoral dissertation, University of Twente, Enschede, The Netherlands.

Grayson, D. A. (1988). Two-group classification in latent trait theory: Scores with monotone likelihood ratio. *Psychometrika, 53*, 383-392.

Gumpel, T., Wilson, M., & Shalev, R. (1998). An item response theory analysis of the Conner's Teachers Rating-Scale. *Journal of Learning Disabilities, 31*, 525-532.

Hemker, B.T. (1996). *Unidimensional IRT models for polytomous items, with results for Mokken scale analysis.* Unpublished doctoral dissertation, Utrecht University, The Netherlands.

Hemker, B.T. (2001). Reversibility revisited and other comparisons of three types of polytomous IRT models. In A. Boomsma, M.A.J. van Duijn, & T.A.B. Snijders (Eds.), *Essays on item response theory* (pp. 277 - 296). New York: Springer.

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.

Hemker, B.T., Sijtsma, K., Molenaar, I.W., & Junker, B.W. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psychometrika, 62*, 331-347.

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.

Kelderman, H., & Rijkes, C.P.M. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika, 59*, 437-450.

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

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

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

Masters, G. N. (1988). Measurement models for ordered response categories. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 11-29). New York: Plenum press.

Masters, G.N., & Wright, B.D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 101 - 121). New York: Springer.

Maurer, T.J., Raju, N.S., & Collins, W.C. (1998). Peer and subordinate performance-appraisal measurement equivalence. *Journal of Applied Psychology, 5*, 693-702.

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

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. (1983). *Item steps* (Heymans Bulletin 83-630-EX). Groningen, The Netherlands: University of Groningen, Department of Statistics and Measurement Theory.

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

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: Danish Institute for Educational Research.

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 Supplement No. 17*.

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

Samejima, F. (1995). Acceleration model in the heterogeneous case of the general graded response model. *Psychometrika, 60*, 549-572.

Samejima, F. (1996, April). *Polychotomous responses and the test score*. Paper presented at National Council on Measurement in Education Meeting, New York.

Samejima, F. (1997, March). *An expansion of the logistic positive exponent family of models to a family of graded response models.* Paper presented at National Council on Measurement in Education Meeting, Chigago.

Scheiblechner, H. (1995). Isotonic ordinal probabilistic models (ISOP). *Psychometrika, 60*, 281-304.

Sijtsma, K., & Hemker, B.T. (1998). Nonparametric polytomous IRT models for invariant item ordering, with results for parametric models. *Psychometrika, 63*, 183-200.

Sijtsma, K., & Hemker, B.T. (2000). A taxonomy for ordering persons and items using simple sum scores. *Journal of Educational and Behavioral Statistics, 25,* 391-415.

Sijtsma, K., & Junker, B.W. (1996). A survey of theory and methods of invariant item ordering. *British Journal of Mathematical and Statistical Psychology, 49*, 79-105.

Sijtsma, K., & Van der Ark, L.A. (2001). Progress in NIRT analysis of polytomous item scores: Dilemmas and practical solutions. In A. Boomsma, M.A.J. van Duijn, & T.A.B. Snijders (Eds.), *Essays on item response theory* (pp. 297 - 318). New York: Springer.

Sijtsma, K., & Verweij, A.C. (1999). Knowledge of solution strategies and IRT modeling of items for transitive reasoning. *Applied Psychological Measurement, 23*, 55-68.

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

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology, 43*, 39-55.

Tutz, G. (1997). Sequential models for ordered responses. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 139 - 152). New York: Springer.

Van der Ark, L.A. (2000). *Practical consequences of stochastic ordering of the latent trait under various polytomous IRT models*. Manuscript submitted for publication.

Van Engelenburg, G. (1997). *On psychometric models for polytomous items with ordered categories within the framework of item response theory.* Unpublished doctoral dissertation, University of Amsterdam.

Verhelst, N. D., Glas, C. A. W., & De Vries, H. H. (1997). A steps model to analyze partial credit. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 123 - 138). New York: Springer.

Verhelst, N. D., Glas, C. A. W., & Verstralen, H.H.F.M. (1995). *OPLM: One Parameter Logistic Model. Computer program and manual*. Arnhem, The Netherlands: CITO.

Appendix A

List of acronyms:

Technical terms:

CCC : category characteristic curve

IRF : item response function

IRT : item response theory

ISRF : item step response function

Item response models:

ACMs : adjacent-category models

CPMs : cumulative probability models

CRMs : continuation ratio models

Technical Properties

IIO : invariant item ordering

MLR : monotone likelihood ratio

SO : stochastic ordering

SOL : stochastic ordering of the latent trait by the total score

SOM : stochastic ordering of the total score by the latent trait

Appendix B

The parameters used to produce the curves in Figure 1 are

Model								
Parameter	j	x	AM	2p(jx)-SM	2p(j)-SM	2p(x)-SM	SRM	SRSM
ξ_j	1		0.2	1.0	1.0	1.0	1.0	1.0
	2		5.0	1.0	1.0	1.0	1.0	1.0
α_{jx}	1	1	3.5	3.5	0.5	2.5	1.0	1.0
	1	2	0.5	0.5	0.5	0.5	1.0	1.0
	2	1	1.0	1.0	2.0	2.5	1.0	1.0
	2	2	2.0	2.0	2.0	0.5	1.0	1.0
β_{jx}	1	1	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
	1	2	3.0	3.0	3.0	3.0	3.0	0.0
	2	1	1.0	1.0	1.0	1.0	1.0	1.0
	2	2	2.0	2.0	2.0	2.0	2.0	2.0

Acronyms:

AM	: acceleration model
2p(j)-SM	: 2p(x)-sequential model
2p(x)-SM	: 2p(j)-sequential model
SRM	: sequential Rasch model
SRSM	: sequential rating scale model

Appendix C

We prove that the nonparametric sequential model implies MLR for the maximum total score $X_+ = mJ$. The nonparametric sequential model assumes unidimensionality, local independence, and $M_{jx}(\theta)$ (Equation 1) nondecreasing in θ and is the least restrictive CRM. By implication, all CRMs imply MLR for the maximum total score $X_+ = mJ$; that is, $g(K{=}mJ, C{<}mJ; \theta)$ is nondecreasing in θ.

In the proof the following notation is used: Let $\pi_{jx}(\theta) \equiv P(X_j = x \mid \theta)$ and let the number of score vectors that yield $X_+ = K$ and $X_+ = C$ be denoted by R_K and R_C, respectively. By convention, K > C. Vectors containing scores on the J items summing to $X_+ = K$ are denoted $\mathbf{X}_{(u)}$, with realizations $\boldsymbol{x}_u$ $(u = 1, \ldots, R_K)$. Similarly, vectors containing scores on the J items summing to $X_+ = C$ are denoted $\mathbf{X}_{(v)}$, with realizations $\boldsymbol{x}_v$ $(v = 1, \ldots, R_C)$. Let the first derivative of a function with respect to θ be denoted by means of a prime. All derivatives in the proof are with respect to θ.

Hemker, et al. (1996) showed that, assuming unidimensionality and local independence, MLR of X_+ holds if the first derivative of the likelihood ratio in Equation (MLR) is nonnegative for all θ, that is

$$\sum_{u=1}^{R_K} \sum_{v=1}^{R_C} \left\{ \left\{ \sum_{j=1}^{J} \left[\frac{\pi'_{jx(u)}(\theta)}{\pi_{jx(u)}(\theta)} - \frac{\pi'_{jx(v)}(\theta)}{\pi_{jx(v)}(\theta)} \right] \right\} \times \prod_{j=1}^{J} \left[\pi_{jx(u)}(\theta) \times \pi_{jx(v)}(\theta) \right] \right\} \geq 0. \qquad \text{(A1)}$$

In Equation (A1) the only part that may result in negative values is $\frac{\pi'_{jx(u)}(\theta)}{\pi_{jx(u)}(\theta)} - \frac{\pi'_{jx(v)}(\theta)}{\pi_{jx(v)}(\theta)}$. Therefore, for the proof of this Theorem it is sufficient to show that for $K = mJ$ this difference is always nonnegative, irrespective of the values of C.

The maximum of X_+ is mJ, and is obtained for $\mathbf{X}_{(u)} = (m, m, \ldots, m)$. It may noted that in this case $R_K = 1$, meaning that $\frac{\pi'_{jx(u)}(\theta)}{\pi_{jx(u)}(\theta)} = \frac{\pi'_{jm}(\theta)}{\pi_{jm}(\theta)}$. Next, it is shown that for any x and any j, $\frac{\pi'_{jm}(\theta)}{\pi_{jm}(\theta)} - \frac{\pi'_{jx}(\theta)}{\pi_{jx}(\theta)}$ is nonnegative.

Note that $\frac{\pi'_{jm}(\theta)}{\pi_{jm}(\theta)} = \ln[\pi_{jm}(\theta)]'$, and in the CRM $\pi_{jm}(\theta) = \prod_{y=0}^{m} M_{jy}(\theta)$ (see Equation 2).

Thus, $\ln[\pi_{jm}(\theta)]' = \ln[\prod_{y=0}^{m} M_{jy}(\theta)]' = \sum_{y=0}^{m} \ln[M_{jy}(\theta)]'$, which means that for any *j*

$$\ln[\pi_{jm}(\theta)]' = \sum_{y=0}^{m} \frac{[M_{jy}(\theta)]'}{M_{jy}(\theta)}. \tag{A2}$$

Similarly, $\frac{\pi'_{jx}(\theta)}{\pi_{jx}(\theta)} = \ln[\pi_{jx}(\theta)]'$. Because in the CRM $\pi_{jx}(\theta) = \prod_{y=0}^{x} M_{jy}(\theta)[1 - M_{j,x+1}(\theta)]$ (see Equation 2), this implies that for any x ($0 \le x \le m$) and any j

$$\ln[\pi_{jx}(\theta)]' = \sum_{y=0}^{x} \frac{[M_{jx}(\theta)]'}{M_{jy}(\theta)} - \frac{[M_{j,x+1}(\theta)]'}{1 - M_{j,x+1}(\theta)}. \tag{A3}$$

From Equations (A2) and (A3) it follows that for any *x* and any *j,*

$$\begin{aligned} \frac{\pi'_{jm}(\theta)}{\pi_{jm}(\theta)} - \frac{\pi'_{jx(v)}(\theta)}{\pi_{jx(v)}(\theta)} &= \sum_{y=0}^{m} \frac{[M_{jy}(\theta)]'}{M_{jy}(\theta)} - \left(\sum_{y=0}^{x} \frac{[M_{jy}(\theta)]'}{M_{jy}(\theta)} - \frac{[M_{j,x+1}(\theta)]'}{1 - M_{j,x+1}(\theta)} \right) \\ &= \sum_{y=x+1}^{m} \frac{[M_{jy}(\theta)]'}{M_{jy}(\theta)} + \frac{[M_{j,x+1}(\theta)]'}{1 - M_{j,x+1}(\theta)}. \end{aligned} \tag{A4}$$

Note that for all x ($0 \le x \le m$) the first derivative of $M_{jx}(\theta)$ is nonnegative in the nonparametric sequential model, because this model assumes that $M_{jx}(\theta)$ is nondecreasing. Also, in this model $0 < M_{jx}(\theta) < 1$, for all *x*. Thus, $M_{jy}(\theta)$ and $[1 - M_{j,x+1}(\theta)]$ are nonnegative. This implies that Equation (A4) is nonnegative for all *x* and *j*. This implies that Equation (A1) holds when $K = mJ$. A similar proof shows that MLR holds for $C = 0$ and $K > 0$.