Relationships and Properties of Polytomous Item Response Theory Models

L. Andries van der Ark
Tilburg University

Relationships among twenty polytomous item response theory (IRT) models (parametric and nonparametric) and eight measurement properties relevant to polytomous IRT are described. Three tables are provided to assist in choice of an appropriate model, and examples demonstrate the use of the models in test construction. *Index terms: invariant item ordering, monotone likelihood ratio, monotonicity, nonparametric item response theory, parametric item response theory, polytomous items, stochastic ordering.*

Tests or questionnaires are frequently used to obtain data in the behavioral sciences and in research in education, marketing, and medicine. Often tests with polytomous items are preferred over those with dichotomous items. This preference might be based on (1) the fact that fewer polytomous items are typically needed to obtain the same degree of reliability, (2) some traits are more easily measured on rating scales, and/or (3) item responses for certain kinds of variables are better expressed on an ordinal scale.

Many psychometric models are available for analyzing tests with polytomous items. When choosing a model, it is important to consider its measurement properties. Within item response theory (IRT), much literature (e.g., Akkermans, 1999; Hemker, Sijtsma, Molenaar, & Junker, 1996, 1997; Samejima, 1995; Sijtsma & Hemker, 1998, 2000) is available on the measurement properties of IRT models for items with polytomous responses. Much of this literature focuses on the relationships among unidimensional polytomous IRT models and their measurement properties—particularly, whether a certain model implies a certain property.

This paper also discusses relationships among unidimensional polytomous IRT models and their measurement properties. However, no sharp distinction is made between IRT models and measurement properties. Instead, each polytomous IRT model and measurement property is defined as a set of assumptions, which makes it possible to directly compare models and properties.

Types of Relationships

Consider two sets of assumptions, A and B. There are five logical relationships between A and B:

1. A implies B (A $\Rightarrow$ B). That is, A is a special case of B (A is a subset of B).
2. A is implied by B (A $\Leftarrow$ B). That is, A is a generalization of B (B is a subset of A).
3. A and B are disjoint (A $\otimes$ B). That is, if the assumptions of B hold, those of A do not hold, and vice versa.
4. A and B are unrelated (A $\circ$ B). That is, if the assumptions of A hold, those of B might or might not hold, and vice versa.
5. A and B are identical (A $\Leftrightarrow$ B).

Applied Psychological Measurement, Vol. 25 No. 3, September 2001, 273–282

These symbols (O, $\otimes$, $\Leftarrow$, $\Rightarrow$, and $\Leftrightarrow$) are used here in a conceptual framework in which a large part of the literature on the measurement properties of polytomous IRT models is summarized and expanded.

Some of the findings in this paper are illustrated with real data based on the responses of 825 examinees to a test measuring coping strategies in the event of industrial odor annoyance. The test consisted of five items with four answer categories each (Cavalini, 1992; see also Molenaar & Sijtsma, 2000). All analyses were conducted with the computer program ℓEM (Vermunt, 1997).

Polytomous IRT Models

Assume a test of J polytomous locally independent items, indexed by $j = 1, 2, \ldots, J$. Each item has $m+1$ ordered answer categories, indicated by $x = 0, 1, 2, \ldots, m$. The difference between two adjacent answer categories can be thought of as an item step (Molenaar, 1983) yielding m item steps for $m + 1$ categories. Let X_j be a random variable for the discrete response to item j, and let $X_+ = \sum_{j=1}^{J} X_j$ be the unweighted total score based on the sum of the item category weights associated with the responses selected by an examinee. Assume that a unidimensional latent trait, θ, establishes the probability of $X_j = x$, $P(X_j = x|\theta)$.

Classification of Polytomous IRT Models

All polytomous IRT models discussed here can be placed into three classes (Molenaar, 1983; see also Agresti, 1990, pp. 318–322; Hemker, 1996, Chap. 6; Mellenbergh, 1995)—graded response models (GRMs), sequential models (SMs), and partial credit models (PCMs). The logit of passing the xth item step is defined differently in each class. GRMs assume that the cumulative logits,

$$\text{logit}[P(X_j \geq x|\theta)] , \quad x = 1, 2, \ldots, m , \tag{1}$$

are nondecreasing functions of θ. This assumption is known as monotonicity. SMs assume that the continuation ratio logits,

$$\text{logit}\left[\frac{P(X_j \geq x|\theta)}{P(X_j \geq x-1|\theta)}\right] , \quad x = 1, 2, \ldots, m , \tag{2}$$

are nondecreasing functions of θ. PCMs assume that the adjacent category logits,

$$\text{logit}\left[\frac{P(X_j = x|\theta)}{P(X_j = x-1 \vee X_j = x|\theta)}\right] , \quad x = 1, 2, \ldots, m , \tag{3}$$

are nondecreasing functions of θ.

Parametric IRT (PIRT) models. In addition to being classified based on logits, IRT models can also be classified as parametric or nonparametric. The PIRT models discussed here assume local independence, unidimensionality, and logits linear in θ. The linear function for the logit of item step x of item j is

$$\text{logit (Item Step}_{j,x}) = \alpha_{jx}(\theta - \beta_{jx}) , \quad \alpha_{jx} > 0 , \tag{4}$$

where α_{jx} is the slope parameter and β_{jx} is the location parameter. Restrictions on α_{jx} yield four model types:

1. Models with unrestricted slope parameters. These are denoted 2P(jx). For example, a PCM (Equation 3) with logits equal to $\alpha_{jx}(\theta - \beta_{jx})$ is indicated by 2P(jx)-PCM.

2. Models with $\alpha_{jx} = \alpha_j$, for all x (i.e., slope parameters constant over categories, but varying over items). These are denoted 2P(j). Thus, a PCM with logits equal to $\alpha_j(\theta - \beta_{jx})$ is indicated by 2P(j)-PCM.
3. Models with $\alpha_{jx} = \alpha_x$, for all j (i.e., slope parameters constant over items, but varying over categories). These are denoted 2P(x). Therefore, a PCM with logits equal to $\alpha_x(\theta - \beta_{jx})$ is indicated by 2P(x)-PCM.
4. Models with $\alpha_{jx} = 1$, for all x and j. These are denoted 1P. For example, a PCM with logits equal to $\theta - \beta_{jx}$ is indicated by 1P-PCM.

β_{jx} can be restricted by a rating formulation (Andrich, 1978) indicated by -R, which divides β_{jx} into an item component (δ_j) and a step component (τ_x), such that $\beta_{jx} = \delta_j + \tau_x$, with $\Sigma_n \tau_n = 0$. Thus, a PCM with logits equal to $\theta - \delta_j - \tau_x$ is indicated by 1P-PCM-R.

Using this notation:

2P(j)-GRM is the GRM (Samejima, 1969),
1P-PCM is the PCM (Masters, 1982),
1P-PCM-R is the rating scale model (Andrich, 1978),
2P(j)-PCM is the generalized PCM (Muraki, 1992),
1P-SM is the SM (Tutz, 1990), and
1P-SM-R is the sequential rating scale model (Tutz, 1990).

The acceleration model (AM; Samejima, 1995) also is discussed. The AM is a nonlogistic SM and assumes local independence, unidimensionality, and

$$\frac{P(X > x|\theta)}{P(X \geq x-1|\theta)} = \left\{\frac{\exp[D\alpha_{jx}(\theta - \beta_{jx})]}{1 + \exp[D\alpha_{jx}(\theta - \beta_{jx})]}\right\}^{\xi_j}, \tag{5}$$

where D is a scaling parameter and $\xi_j \geq 0$ is the acceleration parameter. Note that the AM is a logistic model for $\xi = 1$.

For GRMs, α_{jx} cannot change over the categories of the same item, so 2P(jx)-GRM and 2P(x)-GRM do not exist (e.g., Mellenbergh, 1995). Moreover, a GRM's β_{jx} must satisfy the ordering $\beta_{j1} < \beta_{j2} < \ldots < \beta_{jm}$ (Samejima, 1969, p. 23).

Nonparametric IRT (NIRT) models. Three NIRT models (indicated by NP-) assume local independence, unidimensionality, and nondecreasing logits in θ. These are the NP-GRM (Equation 1), NP-SM (Equation 2), and NP-PCM (Equation 3).

Three additional NIRT models—the double monotonicity model (DMM; Molenaar, 1997), the isotonic ordinal probabilistic model (ISOP; Scheiblechner, 1995; see also Junker, 1998), and the strong double monotonicity model (SDMM; Sijtsma & Hemker, 1998)—have additional assumptions; these models are all GRMs. The DMM assumes that the cumulative logits are nonintersecting. The ISOP assumes that the cumulative logits are invariantly ordered for all items; that is, logit$[P(X_i \geq x|\theta)] \leq$ logit$[P(X_j \geq x|\theta)]$ for all θ and x. The SDMM assumes that the cumulative logits are nonintersecting and invariantly ordered.

Relationships Among Models

Table 1 shows the relationships among twenty polytomous IRT models. These relationships should be read as follows: NP-SM is above 1P-PCM on the vertical axis. Therefore, the NP-SM should be read from the diagonal axis. The relationship $\Rightarrow$ between these two models is indicated at the intersection of the column below NP-SM on the diagonal axis and the row for the 1P-PCM on the vertical axis: hence, 1P-PCM $\Rightarrow$ NP-SM.

In Table 1, polytomous PIRT models from different classes are disjoint ($\otimes$). This means that item parameters that pertain to the logit of one class of models cannot be written as a function of

Table 1
Relationships Among 20 Polytomous IRT Models

	NP-GRM																			
NP-GRM	⇔	ISOP																		
ISOP	⇒	⇔	DMM																	
DMM	⇒	O	⇔	SDMM																
SDMM	⇒	⇒	⇒	⇔	NP-PCM															
NP-PCM	⇒	O	O	O	⇔	NP-SM														
NP-SM	⇒	O	O	O	⇐	⇔	2P(j)-GRM													
2P(j)-GRM	⇒	O	O	O	⇒	⇒	⇔	1P-GRM												
1P-GRM	⇒	O	⇒	O	⇒	⇒	⇒	⇔	1P-GRM-R											
1P-GRM-R	⇒	⇒	⇒	⇒	⇒	⇒	⇒	⇒	⇔	2P(jx)-PCM										
2P(jx)-PCM	⇒	O	O	O	⇒	⇒	⊗	⊗	⊗	⇔	2P(j)-PCM									
2P(j)-PCM	⇒	O	O	O	⇒	⇒	⊗	⊗	⊗	⇒	⇔	2P(x)-PCM								
2P(x)-PCM	⇒	O	O	O	⇒	⇒	⊗	⊗	⊗	⇒	O	⇔	1P-PCM							
1P-PCM	⇒	O	O	O	⇒	⇒	⊗	⊗	⊗	⇒	⇒	⇒	⇔	1P-PCM-R						
1P-PCM-R	⇒	⇒	O	O	⇒	⇒	⊗	⊗	⊗	⇒	⇒	⇒	⇒	⇔	2P(jx)-SM					
2P(jx)-SM	⇒	O	O	O	O	⇒	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⇔	2P(j)-SM				
2P(j)-SM	⇒	O	O	O	⇒	⇒	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⇒	⇔	2P(x)-SM			
2P(x)-SM	⇒	O	O	O	O	⇒	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⇒	O	⇔	1P-SM		
1P-SM	⇒	O	O	O	⇒	⇒	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⇒	⇒	⇒	⇔	1P-SM-R	
1P-SM-R	⇒	⇒	O	O	⇒	⇒	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⇒	⇒	⇒	⇒	⇔	AM
AM	⇒	O	O	O	O	⇒	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⇐	⇐	⇐	⇐	⇐	⇔

those of the other two classes (Mellenbergh, 1995; see also Thissen & Steinberg, 1986, for PCMs and GRMs).

In some tests, failing the xth item step yields an item score of $x - 1$. Passing the xth item step means that the $x + 1$th item step might be tried, yielding an item score of at least x. Such sequentially scored items can be modeled by a parametric SM (for an example, see Hemker, van der Ark, & Sijtsma, in press). In these situations, location parameters can be interpreted as difficulty parameters for each item step. That is, examinees with θ values less than β_{jx} have probabilities of less than .5 of passing the item step; examinees with $\theta > \beta_{jx}$ have a probability of passing greater than .5.

Software for estimating SMs is not widely available [although WINMIRA (Von Davier, 1996) can estimate the 1P-SM]. A test constructor might have to analyze data with a different model (e.g., a PCM) using a program such as ℓEM (Vermunt, 1997). However, parametric PCMs and SMs are disjoint: PCM location parameter estimates cannot be expressed as those of SMs. Therefore, PCM parameter estimates cannot be interpreted within the sequential scoring context of a test.

Hemker (1996, Chap. 6) proved that NIRT models—NP-GRM, NP-SM, and NP-PCM—are hierarchically nested (i.e., NP-PCM $\Rightarrow$ NP-SM $\Rightarrow$ NP-GRM). Other NIRT models (DMM, ISOP, and SDMM) are special cases of the NP-GRM. Table 1 shows that the 1P-GRM implies the DMM, and that the ISOP, DMM, and SDMM are unrelated to most PIRT models discussed here. Although this result is new, other relationships between PIRT and NIRT models were proven by Hemker et al. (1997), Hemker et al. (in press), and Sijtsma & Hemker (1998).

Figure 1 shows the hierarchical order of the polytomous IRT models. A solid arrow indicates an implication, and a dotted two-way arrow indicates that two models are unrelated—they neither imply nor exclude the other. The 1P-PCM-R, 1P-SM-R, and 1P-GRM-R are the most restrictive of the models. The NP-GRM is the most general model. When searching for a suitable model to analyze data, start with a restrictive model and if it does not fit, try the next model in the hierarchical order.

Figure 1
Hierarchical Structure of Twenty Polytomous IRT Models

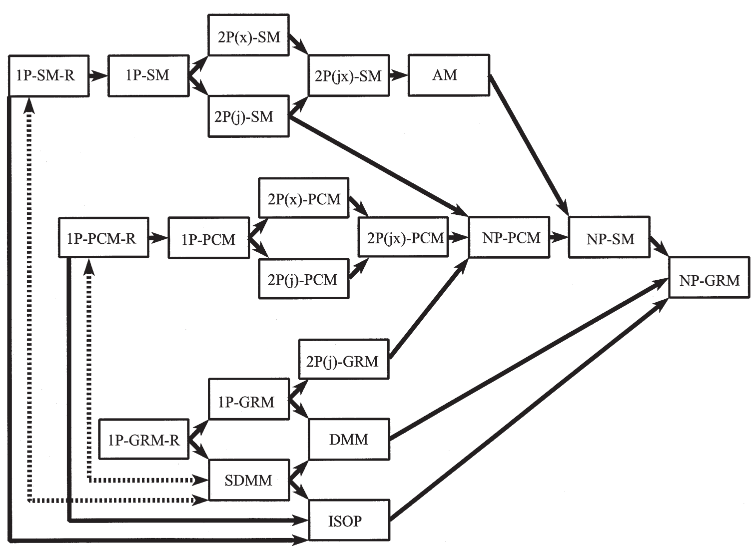

Table 2 shows the result of fitting several logically related models to the Cavalini (1992) data using ℓEM (Vermunt, 1997). To test model-data fit, the likelihood ratio statistic

$$L^2 = 2\sum_i \log\left(\frac{n_i}{\hat{m}_i}\right) \tag{6}$$

was used, where n_i is the frequency of the ith response pattern, and $\hat{m}_i$ is the estimated frequency of the ith response pattern (see Vermunt, 2001, for a discussion of L^2 in an NIRT context). PIRT models were estimated by marginal maximum likelihood, assuming a normal distribution for θ with seven quadrature points. NIRT models were considered latent class models with seven ordered classes (thus, NIRT inequality restraints were satisfied) and estimated by maximum likelihood. Table 2 shows that the NP-PCM was the most restrictive model with an acceptable fit. Whether each model was identifiable also is indicated in Table 2. For unidentifiable models, the implied measurement properties held, but parameter estimates could not be interpreted (see discussion below).

Table 2
Fit for the 1P-PCM-R and Its Generalizations

Model	L^2	df	p	Identified
1P-PCM-R	1888.3	1015	.000	Yes
1P-PCM	1620.1	1007	.000	Yes
2P(x)-PCM	1608.5	1005	.000	Yes
2P(j)-PCM	1267.0	1003	.000	Yes
2P(jx)-PCM	1250.3	993	.000	Yes
NP-PCM	966.2	983	.643	No
NP-GRM	824.1	960	.999	No

Measurement Properties

In IRT, stochastic ordering properties relate the examinee ordering on a manifest variable, Y, and on θ. Two manifest variables were considered: the item score, X_j, and the unweighted total score, X_+.

Monotonicity. For monotonicity, $P(X_j \geq x|\theta)$ (or, equivalently, Equation 1) is nondecreasing in θ. Note that local independence and unidimensionality are assumed. Therefore, monotonicity always is assumed in combination with them.

Monotone likelihood ratio (MLR). MLR (Lehmann, 1986, p. 78–86; see also Hemker et al., 1996) holds if

$$\frac{P(Y = K|\theta)}{P(Y = C|\theta)} \tag{7}$$

is a nondecreasing function of θ for all C and K, where $C < K$. There are two versions of MLR.

1. MLR of the item score (MLR-X_j). Equation 7 holds when $Y \equiv X_j$.
2. MLR of the total score (MLR-X_+). Equation 7 holds when $Y \equiv X_+$.

Stochastic ordering of the manifest variable (SOM). In SOM (Hemker et al., 1997), the order of the examinees on the latent variable gives a stochastically correct ordering of the examinees on the manifest variable. That is,

$$P(Y \geq x|\theta_A) \leq P(Y \geq x|\theta_B)\ , \tag{8}$$

for all x and $\theta_A < \theta_B$. SOM has two versions:

1. SOM of the item score (SOM-X_j). Equation 8 holds for $Y \equiv X_j$.
2. SOM of the total score (SOM-X_+). Equation 8 holds for $Y \equiv X_+$.

Stochastic ordering of the latent trait (SOL). In SOL (Hemker et al., 1997), the order of the examinees on the manifest variable gives a stochastically correct ordering of the examinees on the latent variable. That is,

$$P(\theta \geq s|Y = C) \leq P(\theta \geq s|Y = K) \ , \tag{9}$$

for all s, C, and K, where $C < K$.

SOL allows inferences to be drawn about the unknown latent trait. Two versions also exist for SOL:

1. SOL of the item score (SOL-X_j). Equation 9 holds for $Y \equiv X_j$.
2. SOL of the total score (SOL-X_+). Equation 9 holds for $Y \equiv X_+$.

Invariant item ordering (IIO). IIO pertains to the ordering of items rather than of examinees: all items have the same order of difficulty for all examinees. Items have IIO if they can be ordered and numbered accordingly such that (Sijtsma & Hemker, 1998)

$$\varepsilon(X_1|\theta) \leq \varepsilon(X_2|\theta) \leq \ldots \leq \varepsilon(X_J|\theta) \ , \tag{10}$$

for all θ. [For IIO applications and methods of investigation, see Sijtsma & Junker (1996; dichotomous items) and Sijtsma & Hemker (1998; polytomous items).]

Relationships Among Properties

Table 3 shows the relationships among measurement properties. IIO is unrelated to all other measurement properties: the presence of any other measurement property does not guarantee or preclude an IIO. As indicated above, MLR-X_+ implies SOM-X_+ and SOL-X_+ (Lehmann, 1986). MLR-X_j implies SOM-X_j and SOL-X_j, which follows for $J = 1$. SOM-X_j implies SOM- X_+ (Hemker et al., 1997, Theorem 1), so MLR-X_j also implies SOM-X_+. The remaining properties are unrelated.

Table 3
Relationships Among Measurement Properties
(M = Monotonicity)

	M							
M	⇔	MLR-X_j						
MLR-X_j	⇒	⇔	MLR-X_+					
MLR-X_+	O	O	⇔	SOM-X_j				
SOM-X_j	⇔	⇐	O	⇔	SOM-X_+			
SOM-X_+	⇐	⇐	⇐	⇐	⇔	SOL-X_j		
SOL-X_j	O	⇐	O	O	O	⇔	SOL-X_+	
SOL-X_+	O	O	⇐	O	O	O	⇔	IIO
IIO	O	O	O	O	O	O	O	⇔

It is surprising that MLR-X_+ and MLR-X_j, and SOL-X_+ and SOL-X_j, are unrelated. Most IRT models do not assume SOL-X_+ or MLR-X_+. Therefore, for most IRT models, SOL-X_+ and MLR-X_+ hold only for some fitted models (i.e., models with specific values for the item parameters) describing data for a particular set of J items, but not for other sets. All parametric models, however, imply MLR-X_j. Because MLR-X_+ and MLR-X_j are unrelated, MLR-X_+ might not hold for these fitted models when an item is added or removed.

Table 4 shows the item parameter estimates of the 2P(j)-GRM for five items used to compute the likelihood ratio (Equation 7) of X_+ for 2,000 values of θ between $[-10, 10]$. Note that MLR-X_j

is implied by the 2P(j)-GRM. In Table 4, the five items have MLR-X_+. However, when Item 5 was deleted and X_+^* of the other four items recomputed, the likelihood ratio $P(X_+^* = 3|\theta)/P(X_+^* = 2|\theta)$ decreased for $\theta < -4.58$. That is, MLR-X_+ did not hold for the remaining four items. Thus, if an IRT model does not imply MLR-X_+, each time an item is added or removed from the test it is necessary to determine whether the model has MLR-X_+, even when all items have MLR-X_j.

Table 4
Parameter Estimates From the 2P(j)-GRM

Item	$\hat{\alpha}_j$	$\hat{\beta}_{j1}$	$\hat{\beta}_{j2}$	$\hat{\beta}_{j3}$
1	.873	−2.628	.532	.653
2	3.633	−.481	.120	.601
3	.444	−4.743	−.865	1.191
4	6.459	−.199	.092	.424
5	.268	−.481	5.164	8.972

Relationships Among Models and Measurement Properties

Table 5 shows the relationships between polytomous IRT models and measurement properties. All models imply SOM-X_j, SOM-X_+, and monotonicity. Monotonicity (or SOM-X_j) is equivalent to the NP-GRM because all models and properties imply local independence and unidimensionality, so monotonicity is the only additional property.

Table 5
Relationships Between Polytomous IRT Models and
Measurement Properties (M = Monotonicity)

	Property							
Model	M	MLR-X_j	MLR-X_+	SOM-X_j	SOM-X_+	SOL-X_j	SOL-X_+	IIO
NP-GRM	⇔	⇐	○	⇔	⇒	○	○	○
ISOP	⇒	○	○	⇒	⇒	○	○	⇒
DMM	⇒	○	○	⇒	⇒	○	○	○
SDMM	⇒	○	○	⇒	⇒	○	○	⇒
NP-PCM	⇒	⇔	○	⇒	⇒	⇒	○	○
NP-SM	⇒	⇐	○	⇒	⇒	○	○	○
2P(j)-GRM	⇒	⇒	○	⇒	⇒	⇒	○	○
1P-GRM	⇒	⇒	○	⇒	⇒	⇒	○	○
1P-GRM-R	⇒	⇒	○	⇒	⇒	⇒	○	⇒
2P(jx)PCM	⇒	⇒	○	⇒	⇒	⇒	○	○
2P(j)-PCM	⇒	⇒	○	⇒	⇒	⇒	○	○
2P(x)-PCM	⇒	⇒	○	⇒	⇒	⇒	○	○
1P-PCM	⇒	⇒	⇒	⇒	⇒	⇒	⇒	○
1P-PCM-R	⇒	⇒	⇒	⇒	⇒	⇒	⇒	⇒
2P(jx)-SM	⇒	○	○	⇒	⇒	○	○	○
2P(j)-SM	⇒	⇒	○	⇒	⇒	⇒	○	○
2P(x)-SM	⇒	○	○	⇒	⇒	○	○	○
1P-SM	⇒	⇒	○	⇒	⇒	⇒	○	○
1P-SM-R	⇒	⇒	○	⇒	⇒	⇒	○	⇒
AM	⇒	○	○	⇒	⇒	○	○	○

Hemker et al. (1996, 1997) showed that only the 1P-PCM and 1P-PCM-R imply MLR-X_+ and SOL-X_+. Hemker et al. (1997, Proposition, p. 338) proved the equivalence of MLR-X_j and the NP-PCM.

Thus, special cases of the NP-PCM imply MLR-X_j and SOL-X_j. It was shown here, however, that IRT models that are not special cases of the NP-PCM are unrelated to MLR-X_j and SOL-X_j. IIO is only implied by (1) parametric rating scale models discussed here and (2) NIRT models that assume invariantly ordered logits (the ISOP and SDMM).

Table 5 permits identification of the least-restrictive (i.e., most desirable) models to satisfy a set of measurement properties. These models are the 1P-PCM (for SOL-X_+ and MLR-X_+), 1P-PCM-R (for all properties), ISOP (for IIO), NP-PCM (for MLR-X_j and SOL-X_j), and NP-GRM (for monotonicity, SOM-X_+ and SOM-X_j).

Discussion

A practical researcher might select a model with all desirable properties. This type of model usually does not fit well. For example, the 1P-PCM-R (Table 2) had a very poor fit. Less-restrictive models had a better fit, but usually do not imply the desired measurement properties. When a model is unrelated to a particular measurement property, it sometimes can be determined whether the property holds for a particular fitted model. Item parameter estimates can be used to compute, for example, MLR (Equation 7), SOL (Equation 9) or IIO (Equation 10) for various values of θ. However, a condition for using this approach is that the model is identifiable. In Table 2, the NP-PCM is the most-restrictive model with an acceptable fit. Table 3 shows that SOL-X_j, monotonicity, SOM-X_j, and SOM-X_+ are implied by the NP-PCM, but that it does not imply IIO, MLR-X_+, and SOL-X_+. Because the fitted NP-PCM was unidentifiable (Table 2)—yielding meaningless parameter estimates—the latter properties could not be investigated. This could be an argument for choosing the 2P(jx)-PCM over the NP-PCM. Sometimes, neither the most-restrictive nor the least-restrictive model is adequate. In these cases, the choice depends on the willingness to pay for (1) identifiability and measurement properties or (2) fit.

IRT models that are unrelated to certain measurement properties might still be of interest. Often, a model has a property in most practical situations, and only exceptionally does the property not hold. (See Sijtsma & van der Ark, 2001, for an example involving the NP-GRM and SOL-X_+.)

Some important measurement properties were not included and should be incorporated into this framework. These include ordinal modal points and unique maximum condition. For ordinal modal points, the modes of $P(X_j = x|\theta)$ (for $x = 0, 1, 2, \ldots, m$) are consecutively ordered in θ (Samejima, 1995). For the unique maximum condition, the estimates of $P(X_j = x|\theta)$ have a unique maximum (Samejima, 1972, 1995). Other properties cannot easily be incorporated into the present framework because they cannot be defined unambiguously (e.g., the psychological reality of a model; Samejima, 1995) or because their definition depends on the model, making a comparison across models difficult (e.g., additivity and scale reversibility; Hemker, 1996, Chap. 6; Jansen & Roskam, 1986; Samejima, 1995). For NIRT models, parameter estimates often can be obtained by fitting an ordered latent class model (see also Vermunt, 2001, who extensively discusses this topic).

References

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

Akkermans, W. (1999). Polytomous item scores and Guttman dependence. *British Journal of Mathematical and Statistical Psychology, 52,* 39–62.

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

Cavalini, P. M. (1992). *It's an ill wind that brings no good: Studies on odour annoyance and the dispersion of odour concentrations from industries.* Unpublished doctoral dissertation. University of Groningen, Groningen, The Netherlands.

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

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.

Hemker, B. T., van der Ark, L. A., & Sijtsma, K. (in press). On measurement properties of continuation ratio models. *Psychometrika.*

Jansen, P. G. W., & Roskam, E. E. (1986). Latent trait models and dichotomization of graded responses. *Psychometrika, 51,* 69–91.

Junker, B. W. (1998). Some remarks on Scheiblechner's treatment of ISOP models. *Psychometrika, 63,* 73–85.

Lehmann, E. L. (1986). *Testing statistical hypotheses.* (2nd ed.). New York: Wiley.

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

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

Molenaar, I. W. (1983). *Item steps* (Heymans Bulletins HB-83-630-EX). Groningen, The Netherlands: University of Groningen.

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.

Molenaar, I. W., & Sijtsma, K. (2000). *MSP5 for Windows* [Software manual]. Groningen, The Netherlands: ProGAMMA.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159–177.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph No. 17.*

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

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

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 of IRT models 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.

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.

van der Ark, L. A. (1999). *A reference card for the relationships between IRT models for ordered polytomous items and some relevant properties* (WORC Paper No. 99.10.02). Tilburg, The Netherlands: Tilburg University, WORC.

Vermunt, J. K. (1997). *ℓEM: A general program for the analysis of categorical data.* Unpublished manuscript, Tilburg University, Tilburg, The Netherlands. Available from http://www.kub.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html.

Vermunt, J. K. (2001). The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models. *Applied Psychological Measurement, 25,* 283–294.

Von Davier, M. (1996). *WINMIRA V1.68: A program system for analyses with the Rasch model, with the latent class analysis and with the mixed Rasch model* [Software manual]. Kiel, Germany: IPN Institute for Science Education.

Author's Note

Mathematical proofs for the relationships among the models can be found in van der Ark (1999) or on the Internet at http://come.to/nirt. Elaborated versions of the real-data examples also can be found at http://come.to/nirt.

Acknowledgments

This research was supported by the Netherlands Research Council, Grant No. 400.20.011. Thanks are due to Klaas Sijtsma, Brian Junker, and Ivo Molenaar for comments on earlier versions of this paper.

Author's Address

Send requests for reprints or further information to L. Andries van der Ark, Department of Methodology and Statistics, Faculty of Social and Behavioral Sciences, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands. Email: a.vdark@kub.nl.