# Latent and manifest monotonicity in item response models

Junker, B.W.; Sijtsma, K.

*Published in:*
Applied Psychological Measurement

*Publication date:*
2000

*Document Version*
Publisher's PDF, also known as Version of record



*Citation for published version (APA):*
Junker, B. W., & Sijtsma, K. (2000). Latent and manifest monotonicity in item response models. *Applied Psychological Measurement*, *24*(1), 65-81.

Applied Psychological Measurement

http://apm.sagepub.com

# Latent and Manifest Monotonicity in Item Response Models

Brian W. Junker and Klaas Sijtsma

*Applied Psychological Measurement* 2000; 24; 65

DOI: 10.1177/01466216000241004

# Latent and Manifest Monotonicity in Item Response Models

**Brian W. Junker, Carnegie Mellon University**

**Klaas Sijtsma, Tilburg University**

The monotonicity of item response functions (IRF) is a central feature of most parametric and nonparametric item response models. Monotonicity allows items to be interpreted as measuring a trait, and it allows for a general theory of nonparametric inference for traits. This theory is based on monotone likelihood ratio and stochastic ordering properties. Thus, confirming the monotonicity assumption is essential to applications of nonparametric item response models. The results of two methods of evaluating monotonicity are presented: regressing individual item scores on the total test score and on the "rest" score, which is obtained by omitting the selected item from the total test score. It was found that the item-total regressions of some familiar dichotomous item response models with monotone IRFs exhibited nonmonotonicities that persist as the test length increased. However, item-rest regressions never exhibited nonmonotonicities under the nonparametric monotone unidimensional item response model. The implications of these results for exploratory analysis of dichotomous item response data and the application of these results to polytomous item response data are discussed. *Index terms: elementary symmetric functions, essential unidimensionality, latent monotonicity, manifest monotonicity, monotone homogeneity, nonparametric item response models, strict unidimensionality.*

Most item response theory (IRT) models for dichotomous item scores ($X_1, X_2, \ldots, X_J$, taking values in {0,1}) assume that the probability of correctly responding to an item given a latent trait $\theta$ ($P_j(\theta) = P[X_j = 1|\theta]$) is a monotonic, nondecreasing function of $\theta$. Moreover, Hemker, Sijtsma, Molenaar, & Junker (1997) showed that for all graded response and partial-credit IRT models for polytomous items, the item step response functions (ISRFs) $P^*_{js}(\theta) = P[X_j > s|\theta]$ are also nondecreasing in $\theta$ for each $j$ and $s$, where $s$ is an integer item score on a polytomous item.

Monotonicity plays a central role in most nonparametric and parametric formulations of IRT because it captures the intuitive idea that the items measure $\theta$; higher $\theta$s indicate a higher probability of answering an item correctly. The general nonparametric model discussed here has been studied before under many different names (e.g., Ellis & Junker, 1997; Hemker et al., 1996, 1997; Holland & Rosenbaum, 1986; Junker, 1991, 1993; Junker & Ellis, 1997; Mokken, 1971; van der Linden & Hambleton, 1997); here it is referred to as the nonparametric unidimensional monotone item response theory (UMIRT) model. It is defined as

$$P(X_1 = x_1, \ldots, X_J = x_J) = \int P(X_1 = x_1, \ldots, X_J = x_J|\theta)dF(\theta)\ , \tag{1}$$

where

$x_1, x_2, \ldots, x_J$ are the observed values of the item response variables $X_1, X_2, \ldots, X_J$,
$P_{jx}(\theta) = P[X_j = x|\theta]$ are the associated item category response functions, and
$dF(\theta)$ is an arbitrary distribution function.

*Applied Psychological Measurement,* Vol. 24 No. 1, March 2000, 65–81

The UMIRT model assumes unidimensionality ($\theta \in \mathbb{R}$), local independence (LI; $P[X_1 = x_1, x_2, \ldots, x_J = X_J|\theta] = \prod_{j=1}^{J} P_{jx_j}(\theta)$, and monotonicity. For the latter, the ISRFs $P^*_{js}(\theta) = \sum_{x=(s+1)}^{m_j} P_{jx}(\theta)$ are assumed to be nondecreasing in $\theta$ for each $j$ and $s$. Throughout most of this paper $X_j \in \{0, 1\}$, so that $P^*_{j0}(\theta) = P[X_j = 1|\theta] = P_j(\theta)$, which is the usual item response function (IRF) of dichotomous IRT.

Grayson (1988) showed that under the UMIRT model for dichotomous items, the likelihood $P[S = s|\theta]$ for the total score based on a test with $J$ items, $S = \sum_1^J X_j$, has a monotone likelihood ratio property (see also Huynh, 1994). Two stochastic ordering properties follow from this: stochastic ordering of the manifest score $S$ (SOM; $P[S > s|\theta]$ is nondecreasing in $\theta$ for each $s$), and stochastic ordering of the latent trait $\theta$ (SOL; $P[\theta > t|S = s]$ is nondecreasing in $s$ for each $t$).

These properties, which lead to a nonparametric theory of inference for $\theta$, were studied in detail by Hemker et al. (1996, 1997). They showed that SOM holds for any monotone unidimensional IRT model (Equation 1) and that SOL is surprisingly restrictive when items are polytomous. In the nonparametric estimation of IRFs (e.g., Ramsay, 1991; Ramsay & Abrahamowicz, 1989) the shape of the estimated IRF can reveal information on exactly how the IRF deviates from the expected monotonicity. For example, decreasingness at the high end of the scale may suggest that the item has a flaw that distracts high $\theta$ examinees.

Thus, IRF monotonicity must be evaluated as a modeling assumption for data. For dichotomous items, two functions have been used for this purpose—the regression of the item score $X_j$ on the total score $S$ ($P[X_j = 1|S = s]$), and the regression of the item score $X_j$ on the *rest score* $S_{(-j)} = S - X_j$ ($P[X_j = 1|S_{(-j)} = s]$; Junker, 1993). However, using the first function (the item-total regression) can be problematic. A primary purpose of this paper was to demonstrate some familiar situations in which the item-total regression can lead to false rejection of IRF monotonicity.

## Background

Omnibus tests of model fit for specific parametric forms of Equation 1 can be ambiguous about the cause of misfit, e.g., lack of IRF monotonicity or lack of fit with a particular parametric form. A more informative alternative would be to investigate the monotonicity of the empirical regression function $P[X_j = 1|\hat{\theta}]$, where $\hat{\theta}$ is an estimator of $\theta$ that does not depend on the parametric form of the model. For such investigations, it is not important that $\hat{\theta}$ be efficient as a point estimator of $\theta$, but rather that it order examinees as $\theta$ would.

For example, Stout (1990, Theorem 3.2) showed that the total score $S$ is ordinally consistent for $\theta$. That is, there are monotone transformations $f_J(\theta)$ such that $|S - f_J(\theta)|$ becomes small with high probability as $J$ increases. Moreover, Clarke & Ghosh (1995) showed that the conditional distribution of $\theta$, given $S = s$, becomes tighter as $J$ increases. These results suggest that as $J$ increases, $S$ and $\theta$ should be similarly ordered, and hence $S$ is a good candidate for an ordinal $\hat{\theta}$; (non)monotonicities of $P[X_j = 1|S = s]$ should correspond to (non)monotonicities of $P_j(\theta) = P[X_j = 1|\theta]$.

On this basis, some authors have advocated confirming the monotonicity of the item-total regression function as a way to evaluate the UMIRT model (Anastasi, 1988, p. 220; Ramsay, 1991; Sijtsma, 1988); others have proposed constraining existing models so that this condition is satisfied (Croon, 1991; Scheiblechner, 1998). More recently Thissen & Orlando (1997; see also Orlando & Thissen, 1997) proposed item-fit indices and graphical displays based on Friendly's (1994) mosaic plots: $P[X_j = 1|S = s]$ is plotted as a step function and the joint probabilities $P[X_j = 1, S = s]$ are plotted as rectangular areas under the step function.

However, there is a strong tradition in classical test theory of omitting the studied item from an item-total correlation (Lord & Novick, 1968, p. 331; see also Cureton, 1966; Wolf, 1967). The uncorrected point-biserial correlation is usually expected to be inflated because $X_j$ should have a stronger linear relation with the total score $S$ than with the rest score $S_{(-j)}$. The item-total regression has sometimes been replaced with the item-rest regression in studies of IRF shape and model fit. For example, Lord (1965) examined item-rest regressions to determine IRF shapes, and Wainer (1983) and Wainer, Wadkins, & Rogers (1984) used item-rest regressions to explore methods of identifying incorrectly keyed items.

The practical difference between item-rest and item-total regressions seems to have been confused [e.g., compare Lord (1965) with Lord & Novick (1961, pp. 363–364) and Lord (1980, pp. 27–28)]. However, Junker (1993) compared these two regressions and reported Snijders' example, in which three dichotomous items satisfied the UMIRT model and one of the item-total regressions dramatically failed to be monotone in $s$. Junker showed that the item-rest regression is guaranteed to be monotone nondecreasing in $s$ when the UMIRT model holds. Thus, in some cases conditioning on $S$ is inappropriate; $P[X_j = 1|S = s]$ can be artificially nonmonotone even when all $P_j(\theta)$ are nondecreasing. However, conditioning on $S_{(-j)}$ always fixes the problem.

## Manifest Monotonicity

The definition of manifest monotonicity (MM; Junker, 1993, p. 1371) can be adapted to a general score $R$; MM holds for item $j$ and manifest score $R$ if $P[X_j = 1|R = r]$ is nondecreasing in $r$, where $r$ is a realization of $R$: $r = 0, 1, \ldots, J - 1$. Here, the focus is on the total score ($R = S \equiv \sum_1^J X_j$) and the rest score ($R = S_{(-j)} \equiv S - X_j$).

Junker (1993; Proposition 4.1) gave a direct proof of MM for binary items in the UMIRT model, when $R = S_{(-j)}$, using

$$P[X_J = 1|S_{(-j)}] = E\left\{P[X_J = 1|S_{(-j)}, \Theta]|S_{(-j)}\right\} = E[P_j(\Theta)|S_{(-j)}] \ . \tag{2}$$

The first equality follows by general properties of conditional expectations. The second equality follows from the LI assumption that $X_j$ and $S_{(-j)}$ are conditionally independent, given $\theta$. $E[P_j(\theta)|S_{(-j)} = s]$ is nondecreasing in $s$ by SOL, which clearly holds for $S_{(-j)}$ as well as $S$ (Lehmann, 1955; see also Stout, 1990, Lemma 3.1).

Junker (1998) outlined an alternative method that reproduced the above results and accounted better for the Rasch (1960/80) model. The argument that $P[X_j = 1|R = r - 1] \le P[X_j = 1|R = r]$ can be organized as follows:

$$\left.\begin{aligned} P(X_j = 1|R = r - 1) \quad &= \quad \int P(X_j = 1|R = r - 1, \theta)dF(\theta|R = r - 1) \\ &\stackrel{?}{\le} \quad \int P(X_j = 1|R = r, \theta)dF(\theta|R = r - 1) \\ &\stackrel{?}{\le} \quad \int P(X_j = 1|R = r, \theta)dF(\theta|R = r) \\ &= \quad P(X_j = 1|R = r) \ , \end{aligned}\right\} \tag{3}$$

where $dF(\theta|R = r)$ is the conditional distribution of $\theta$ given $R = r$, and the first and last equalities are always true by general properties of conditional expectations.

To establish the first inequality marked by "?" in Equation 3, it is sufficient to show that

$$P(X_j = 1|R = r, \theta) \text{ is nondecreasing in } r \text{ for each fixed } \theta \ . \tag{4}$$

When $R = S_{(-j)}$, LI implies $P[X_j = 1|R = r, \theta] = P[X_j = 1|\theta]$, which is constant in $r$, a special case of Equation 4. When $R = S$, Equation 4 still holds by Scheiblechner's argument (1995; Theorem 4).

To establish the second inequality marked by "?" in Equation 3, it is sufficient to show that

$$\theta \text{ is stochastically ordered by } R \tag{5}$$

and

$$P(X_j = 1|R = r, \theta) \text{ is nondecreasing in } \theta \text{ for each fixed } r \, . \tag{6}$$

The stochastic ordering argument following Equation 2 can also be applied here. For $R = S$ or $R = S_{(-j)}$, Equation 5 is SOL. Equation 6 is more difficult; for $R = S_{(-j)}$, it always holds: under LI, $P[X_j = 1|S_{(-j)}, \theta] = P[X_j = 1|\theta] = P_j(\theta)$, which is assumed to be nondecreasing. For $R = S$, Equation 6 does not need to hold, in general (see the examples below). However, in the special case of the Rasch model, $R = S$ is sufficient for $\theta$ so that $P[X_j = 1|R = r, \theta] = P[X_j = 1|R = r]$, which is constant in $\theta$. This is a special case of Equation 6.

Thus, when the UMIRT model holds for binary items, MM is implied by Equations 4, 5, and 6. In particular, MM holds for item $X_j$ and rest score $S_{(-j)}$ for each $j$ $(j = 1, 2, \ldots, J)$. For the binary Rasch model, MM also holds for all $X_j$ and the total score $S$. Equations 4 and 5 always hold for the total score $S$, so that only violations of Equation 6 can lead to violations of MM for $S$.

## Elementary Symmetric Functions

To more carefully study the behavior of $P[X_j = 1|S = s]$, the definition of the elementary symmetric functions used in Rasch models can be extended to general nonparametric item response models for dichotomous items. The conditional odds of answering item $j$ correctly, given $\theta$, is $\epsilon_j(\theta) = P_j(\theta)/[1 - P_j(\theta)]$, where $P_j(\theta)$ is the IRF for the $j$th item. By slightly extending theory about symmetric functions for the traditional Rasch model (e.g., Fischer, 1974, p. 226; Scheiblechner, 1995), the elementary symmetric function for total score $s$, latent variable $\theta$, and the vector of conditional odds $\boldsymbol{\varepsilon}(\theta) = [\varepsilon_1(\theta), \varepsilon_2(\theta), \ldots, \varepsilon_J(\theta)]$ is

$$\gamma_s[\boldsymbol{\varepsilon}(\theta)] = \sum_{S=s} \prod_{j=1}^{J} \varepsilon_j(\theta)^{x_j} = \sum_{S=s} \prod_{j=1}^{J} \left[ \frac{P_j(\theta)}{1 - P_j(\theta)} \right]^{x_j} \, , \tag{7}$$

where the summation is over all score patterns $x_1, x_2, \ldots, x_J$ such that $S = s$. Note that this is exactly the symmetric function $\gamma_s$ of the Rasch model, but it is evaluated at $\varepsilon_j(\theta) = P_j(\theta)/[1 - P_j(\theta)]$ instead of at the exponentiated Rasch item difficulties $\varepsilon_j = \exp(-b_j)$.

Thus,

$$P(S = s|\theta) = \gamma_s[\boldsymbol{\varepsilon}(\theta)] \cdot \prod_{j=1}^{J} [1 - P_j(\theta)] \, , \tag{8}$$

and the conditional distribution of $\theta$, given $S = s$, is

$$dF(\theta|S = s) = \frac{P(S = s|\theta)dF(\theta)}{\int P(S = s|t)dF(t)} = \frac{\gamma_s[\boldsymbol{\varepsilon}(\theta)] \cdot \prod_{j=1}^{J} [1 - P_j(\theta)]dF(\theta)}{P(S = s)} \, . \tag{9}$$

By first integrating Equation 8 with respect to the latent distribution $dF(\theta)$, then multiplying and dividing on the right by the marginal probability $P[S = 0]$, and then applying Equation 9 with $s = 0$,

$$\begin{aligned} P(S=s) &= \int \gamma_s[\boldsymbol{\varepsilon}(\theta)] \cdot \prod_{j=1}^{J}[1-P_j(\theta)]dF(\theta) \\ &= \int \gamma_s[\boldsymbol{\varepsilon}(\theta)]dF(\theta|S=0) \cdot P(S=0) \\ &= E\left\{\gamma_s[\boldsymbol{\varepsilon}(\theta)]|S=0\right\} \cdot P(S=0) \ , \end{aligned} \tag{10}$$

where the expected value is with respect to the posterior distribution of $\theta$, given $S = 0$. Equation 10 extends Holland's (1990) *Dutch identity* from the Rasch model to arbitrary dichotomous IRT models satisfying LI.

Note also that $\gamma_s[\boldsymbol{\varepsilon}(\theta)]$ satisfies standard identities for the elementary symmetric functions (e.g., Molenaar, 1995, pp. 44–45). For example,

$$\frac{d}{d\varepsilon_j(\theta)}\gamma_s[\boldsymbol{\varepsilon}(\theta)] = \gamma_{s-1}[\varepsilon_1(\theta), \ \ldots \ , \varepsilon_{j-1}(\theta), \varepsilon_{j+1}(\theta), \ \ldots \ , \varepsilon_J(\theta)] \equiv \gamma_{s-1}^{(j)}[\boldsymbol{\varepsilon}(\theta)] \ , \tag{11}$$

and

$$\gamma_s[\boldsymbol{\varepsilon}(\theta)] = \varepsilon_j(\theta)\gamma_{s-1}^{(j)}[\boldsymbol{\varepsilon}(\theta)] + \gamma_s^{(j)}[\boldsymbol{\varepsilon}(\theta)] \ , \tag{12}$$

for each $j$.

To investigate the behavior of $P[X_j = 1|S = s, \theta]$, an additional identity is needed. From Equation 8,

$$P(X_j = 1, S = s|\theta) = P[X_j = 1, S_{(-j)} = s-1|\theta] = \varepsilon_j(\theta)\gamma_{s-1}^{(j)}[\boldsymbol{\varepsilon}(\theta)]\prod_{j=1}^{J}[1-P_j(\theta)] \ . \tag{13}$$

By applying the definition of conditional probability and then Equations 13, 8, and 12 (in that order), the following identity is obtained:

$$\begin{aligned} P(X_j = 1|S = s, \theta) &= \frac{P(X_j = 1, S = s|\theta)}{P(S = s|\theta)} \\ &= \frac{\varepsilon_j(\theta)\gamma_{s-1}^{(j)}[\boldsymbol{\varepsilon}(\theta)]\prod_{j=1}^{J}[1-P_j(\theta)]}{\gamma_s[\boldsymbol{\varepsilon}(\theta)]\prod_{j=1}^{J}[1-P_j(\theta)]} \qquad (14) \\ &= \frac{\varepsilon_j(\theta)\gamma_{s-1}^{(j)}[\boldsymbol{\varepsilon}(\theta)]}{\varepsilon_j(\theta)\gamma_{s-1}^{(j)}[\boldsymbol{\varepsilon}(\theta)] + \gamma_s^{(j)}[\boldsymbol{\varepsilon}(\theta)]} \\ &= \left[1 + \frac{\gamma_s^{(j)}[\boldsymbol{\varepsilon}(\theta)]}{\varepsilon_j(\theta)\gamma_{s-1}^{(j)}[\boldsymbol{\varepsilon}(\theta)]}\right]^{-1} \ . \qquad (15) \end{aligned}$$

One version of Equation 14 appears in the literature on testing the Rasch model (e.g., Molenaar, 1983, Equation 3.1; van den Wollenberg, 1982, Equation 11). The further development of Equation 15 allows conditions to be specified under which MM fails.

## Examples and Counterexamples

### Application to IRT Models

*The Guttman model.* The Guttman (1950) model is the simplest IRT model, and it is clear that $P[X_j = 1|S = s]$ is monotone in $s$ in this model. Suppose that items $X_1, \ldots, X_J$ have nondecreasing IRFs that step from 0 to 1,

$$P_j(\theta) = \begin{cases} 0, & \theta \le b_j \\ 1, & b_j < \theta \end{cases} , \tag{16}$$

that the items are ordered so that $b_1 < b_2 < \cdots < b_J$, as well, and that $dF(\theta)$ is a $\theta$ distribution for which $P[S = s] > 0$ for all $s = 0, 1, \ldots, J$. Thus, $P[X_j = 1, S = s] = P[S = s]$ when $j \le s$, and 0 otherwise. Hence,

$$P(X_j = 1|S = s) = \begin{cases} 1, & \text{if } j \le s \\ 0, & \text{if } s < j \end{cases} , \tag{17}$$

so that the item-total regression is also a nondecreasing step function.

*The Rasch model.* For the Rasch (1960/1980) model, $\varepsilon_j(\theta) = \exp(\theta - b_j)$, and the following can be calculated as in the standard Rasch model:

$$\gamma_s[\boldsymbol{\varepsilon}(\theta)] = \exp(s\theta) \cdot \gamma_s[\exp(-b_1), \ldots, \exp(-b_J)] . \tag{18}$$

For the Rasch model, Equation 14 becomes

$$\begin{aligned} P(X_1 = 1|S = s, \theta) &= \frac{\varepsilon_1(\theta)\gamma_{s-1}^{(1)}[\boldsymbol{\varepsilon}(\theta)]}{\gamma_s[\boldsymbol{\varepsilon}(\theta)]} \\ &= \frac{\exp(\theta - b_1)\exp[(s-1)\theta]\gamma_s^{(1)}[\exp(-b_1), \ldots, \exp(-b_J)]}{\exp(s\theta)\gamma_s^{(1)}[\exp(-b_1), \ldots, \exp(-b_J)]} \\ &= \frac{\exp(-b_1)\gamma_{s-1}^{(1)}[\exp(-b_1), \ldots, \exp(-b_J)]}{\gamma_s^{(1)}[\exp(-b_1), \ldots, \exp(-b_J)]} . \end{aligned} \tag{19}$$

This is independent of $\theta$, as it should be due to the sufficiency of $S$ for $\theta$ in the Rasch model. Thus, Equation 6 is explicitly established for $R = S$, so MM holds for $S$.

*The two-parameter logistic model.* For the two-parameter logistic model (2PLM; e.g., Lord, 1980), $\varepsilon_j(\theta) = \exp[a_j(\theta - b_j)] = \exp(a_j\theta - \beta_j)$, where $\beta_j = a_j b_j$. There is no simple, general factorization of $\gamma_s[\boldsymbol{\varepsilon}(\theta)]$ for the 2PLM as in Equation 18, so a special case is considered here. Suppose $a_1$ is fixed, and $a_j \equiv a_2$ for all $j$. In this special case, Equation 15 becomes

$$\begin{aligned} P(X_1 = 1|S = s, \theta) &= \left\{1 + \frac{\gamma_s^{(1)}[\boldsymbol{\varepsilon}(\theta)]}{\varepsilon_1(\theta)\gamma_{s-1}^{(1)}[\boldsymbol{\varepsilon}(\theta)]}\right\}^{-1} \\ &= \left\{1 + \frac{\exp(sa_2\theta)\gamma_s^{(1)}[\exp(-\beta_1), \ldots, \exp(-\beta_J)]}{\exp(a_1\theta - \beta_1)\exp[(s-1)a_2\theta]\gamma_{s-1}^{(1)}[\exp(-\beta_1), \ldots, \exp(-\beta_J)]}\right\}^{-1} \\ &= \left\{1 + \exp[(a_2 - a_1)\theta]\exp(\beta_1)\frac{\gamma_s^{(1)}[\exp(-\beta_1), \ldots, \exp(-\beta_J)]}{\gamma_{s-1}^{(1)}[\exp(-\beta_1), \ldots, \exp(-\beta_J)]}\right\}^{-1} . \end{aligned} \tag{20}$$

Note that Equation 20 is itself a 2PLM IRF, which can be increasing or decreasing depending on the sign of $a_2 - a_1$. This leads to the following result, which violates Equation 6: In the special case of the 2PLM in which $a_1$ is fixed and $a_j \equiv a_2$ for all $j = 2, 3, \ldots, J$, $P[X_1 = 1|S = s, \theta]$ is decreasing in $\theta$ for each fixed $s$ whenever $a_2 > a_1$.

For $S$, the second expression in Equation 3 is always true. Thus, the above violation of the third expression must be so great that it dominates the calculation if MM is violated for $S$. It is shown below that this is not only possible, but that the violations can persist and become worse as $J$ increases.

The nature of the violations produced below is opposite of what might be expected based on previous experience with corrected point-biserial correlations. Whereas the linear relation between $X_j$ and $S_{(-j)}$ might be expected to be less strong than that between $X_j$ and $S$, $E[X_j|S_{(-j)} = s]$ is guaranteed to be monotone in $s$, whereas $E[X_j|S = s]$ might not be.

**Some Examples**

*Snijders' example.* Three binary response variables are considered, with a two-point distribution for $\theta$, $P(\theta = \theta_0) = P(\theta = \theta_1) = .5$, where $\theta_0 < \theta_1$. Let

$$P_j(\theta_0) \equiv \delta,\ j = 1, 2, 3; \quad P_1(\theta_1) = \frac{1}{2}; \quad \text{and } P_2(\theta_1) = P_3(\theta_1) = 1 - \delta\ . \tag{21}$$

It follows that, as $\delta \to 0$, $P[X_1 = 1|S = s]$ tends toward 0, .25, 0, and 1 for $s = 0, 1, 2$, and 3, respectively, whereas $P[X_2 = 1|S = s]$ and $P[X_3 = 1|S = s]$ both tend toward 0, 1/3, 1, and 1 for $s$ = 0, 1, 2, and 3, respectively. Thus, MM using the total score $R$ holds for $X_2$ and $X_3$, but fails for $X_1$.

Figure 1 shows the behavior of $P[X_1 = 1|R = r]$ for the total score $R = S$ and the rest score $R = S_{(-j)}$, for $J = 3$ and $\delta = .1$. Figure 1a shows the latent structure of Snijders' model with $\theta_0$ and $\theta_1$ fixed at 0 and 1, respectively. The IRFs, which are linearly interpolated between the discrete values given in Equation 21, are graphed above the horizontal axis (note that $X_2$ and $X_3$ have the same IRF). The three latent distributions—$P[\theta] = .5$ for $\theta = 0$ or 1 (outline only), $P[\theta|S = 1]$, and $P[\theta|S = 2]$—are shown as histograms with class intervals centered at $\theta_0$ and $\theta_1$ below the horizontal axis, with $\theta_0 = 0$ and $\theta_1 = 1$.

Figure 1b shows the total score distribution below the horizontal axis and the curves for $E[X_1 = 1|S = s]$ and $E[X_1 = 1|S_{(-j)} = s]$ above the horizontal axis. When $\delta$ is almost zero, the values of $P[X_1 = 1|S = s]$ are near their limiting values of 0, .25, 0, and 1 for $s$ = 0, 1, 2, and 3, respectively. However, the lack of monotonicity in $P[X_1 = 1|S = s]$ can still be seen. The item-rest correlation ($r_{t(-1)}$) was .39 and the item-total correlation ($r_t$) was .69.

By Equation 15, if this example is extended from 3 to $J$ items by replicating $P_2(\theta)$, then

$$P(X_1|S = s, \theta_0) = \left(1 + \frac{J - s}{s}\right)^{-1} \tag{22}$$

and

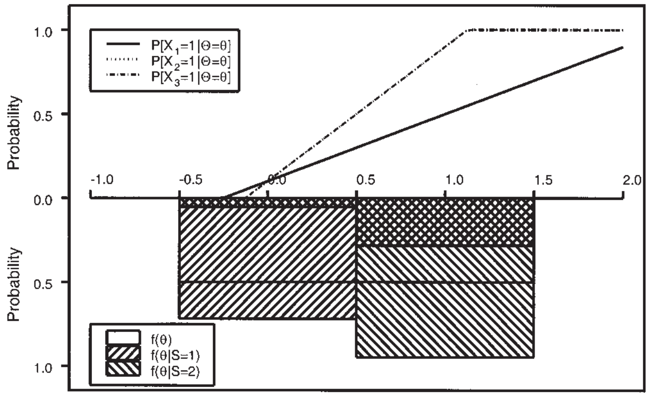
$$P(X_1|S = s, \theta_1) = \left[1 + \left(\frac{1-\delta}{\delta}\right)\left(\frac{J-s}{s}\right)\right]^{-1}\ . \tag{23}$$

This violates Equation 6 as expected; $(1 - \delta)/\delta > 1$ for $0 < \delta < .5$. The nonmonotonicity in $P[X_1 = 1|S = s]$ increases as $J$ increases.

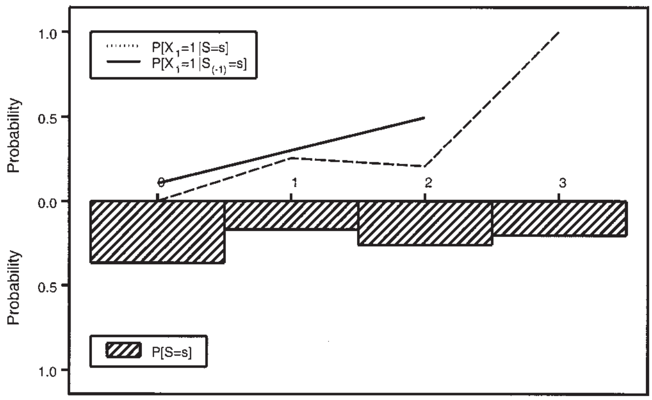
*An extreme 2PLM example.* In the 2PLM,

$$P_j(\theta) = \frac{\exp[a_j(\theta - b_j)]}{1 + \exp[a_j(\theta - b_j)]}\ . \tag{24}$$

**Figure 1**
Violation of MM for $S$ in Snijders' Model
a. Latent Structure




b. Manifest Structure ($r_{t(-1)} = .39$; $r_t = .69$)




In this example, the differences in discrimination were made to be very large. This highlighted the effect of having different discriminations for all items. If $a_1 = .01$ and $a_j = 9$ for all $j > 1$, and if the difficulty of all items is $b_j \equiv 0$ for all $j$, then $X_1$ has a nearly constant probability of being answered correctly or incorrectly, and all other items are nearly perfect Guttman items. The $\theta$ distribution was created as a discrete distribution between $-2$ and 2, which was derived by discretizing a standard normal distribution. Figures 2 and 3 illustrate the behavior of the IRFs and the $\theta$ distributions, and the manifest curves and $S$ distributions, for $J$ = 3, 6, 8, and 10 items. Figures 2a, 2c, 3a, and 3c show that the latent structure of $P_1(\theta)$ was essentially constant (.5), and that $P_j(\theta)$ for $j > 1$ are identical. Below the horizontal axis in each figure are the latent distributions $f(\theta)$, $f(\theta|S = 1)$, and $f(\theta|S = J - 1)$. Note that $f(\theta)$ is represented by a normal curve for clarity [$f(\theta)$ was actually a discrete distribution with $\theta = -2, -1, 0, 1, 2$, and with $f(\theta)$ proportional to $(1/\sqrt{2\pi}) \times \exp(-\theta^2/2)$ for these five values].

Figures 2b, 2d, 3b, and 3d show the manifest structure of the model. Below the horizontal axis is the distribution of $S$, and above the axis are the curves for $P[X_1 = 1|S = s]$ (thin nonmonotone line) and $P[X_1 = 1|S_{(-j)} = s]$. As $J$ increases, it becomes clear why the decreasingness of

**Figure 2**
Violation of MM for $S$ With Extreme Differences in Discrimination (3 and 6 2PLM Items)

a. Latent Structure, 3 Items — b. Manifest Structure, 3 Items ($r_{t(-1)} = 0.0$; $r_t = .49$)

c. Latent Structure, 6 Items — d. Manifest Structure, 6 Items ($r_{t(-1)} = 0.0$; $r_t = .24$)

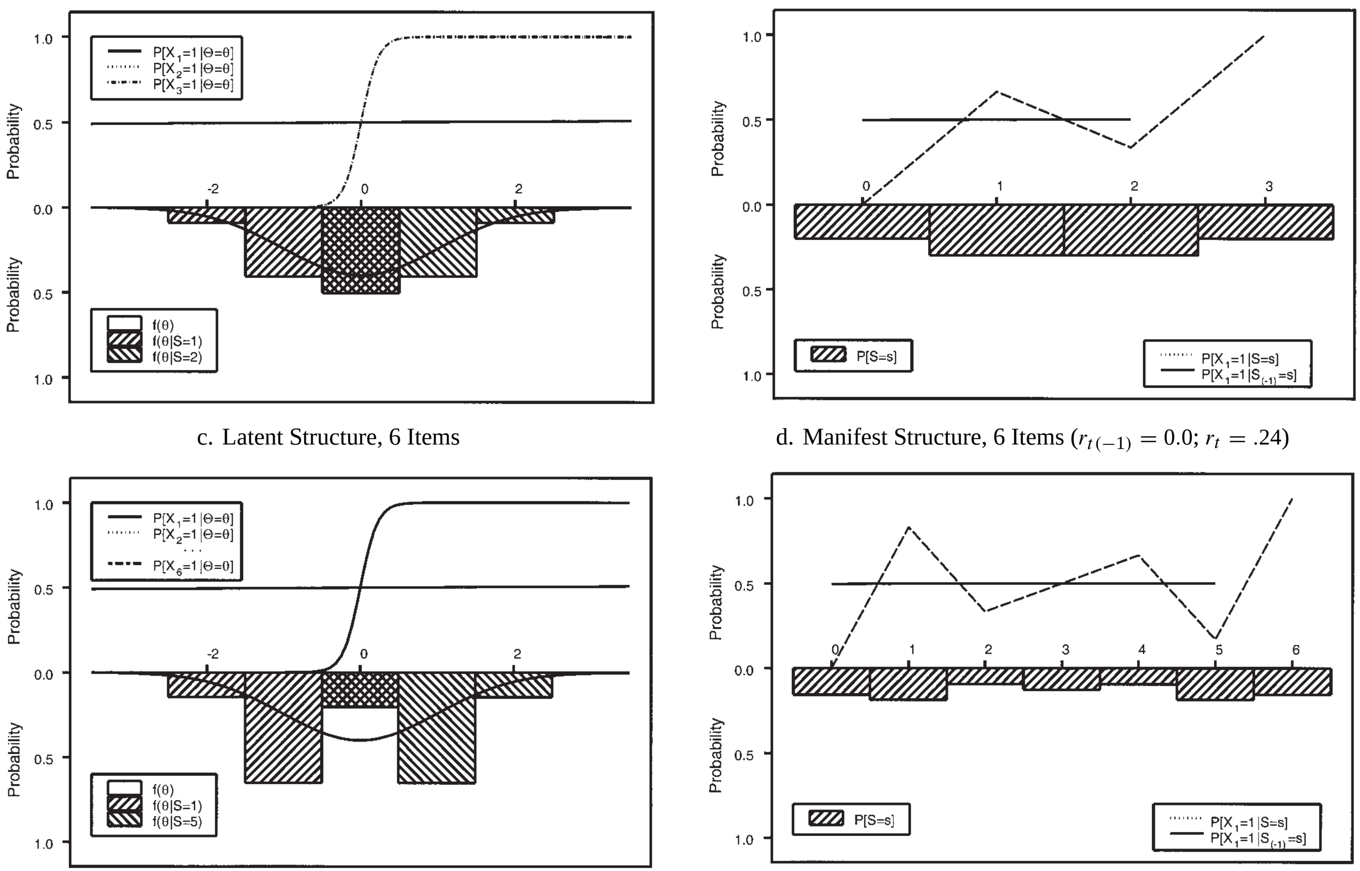
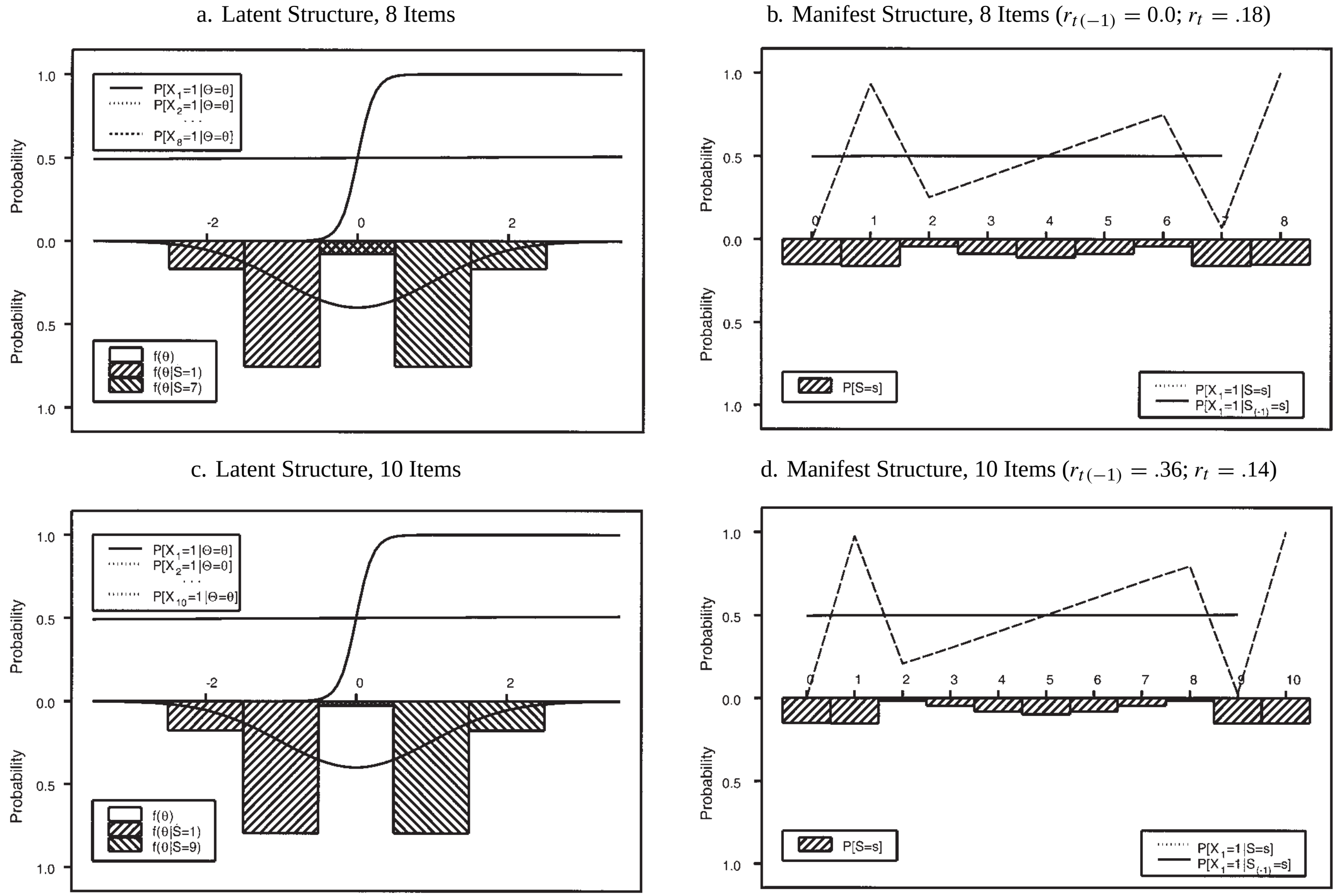


**Figure 3**
Violation of MM for $S$ With Extreme Differences in Discrimination (8 and 10 2PLM Items)



$P[X_1 = 1|S = s, \theta]$ in $\theta$ for each $s$ under the 2PLM with $a_1$ fixed and $a_j = a_2$ for all $j > 2$ can lead to nonmonotonicity in $P[X_1 = 1|S = s]$ and why the nonmonotonicities can increase. For example, Figure 3a shows that when $\theta$ was much less than 0, $P_j(\theta) \approx 0$ for all items except $j = 1$. For this item, $P_1(\theta)$ was still approximately .5. Thus, when $S = 1$ and $\theta$ is low, it is very likely that $X_1 = 1$, i.e., $P[X_1 = 1|S = 1, \theta] \approx 1$. However, when $\theta$ is much greater than 0, $P_j(\theta) \approx 1$ for $j > 1$, and $P_1(\theta)$ is still approximately .5. Thus, when $S = J - 1$, it is very likely that all item responses except $X_1$ equal 1, i.e., $P[X_1 = 1|S = J - 1, \theta] \approx 0$. Moreover, the conditional distribution $f(\theta|S = 1)$ concentrates where $P[X_1 = 1|S = 1, \theta] \approx 1$, and $f(\theta|S = J - 1)$ concentrates where $P[X_1 = 1|S = J - 1, \theta] \approx 0$. This leads to the manifest values $P[X_1 = 1|S = 1] \approx 1$ and $P[X_1 = 1|S = J - 1] \approx 0$ in Figure 3b.

The observed score distribution $P[S = s]$ in Figure 3b shows that a large proportion of examinees was actually located at these points of nonmonotonicity of the item-total regression. By comparing Figures 2c through 3d, it can be seen that the conditional distributions $f(\theta|S = 1)$ and $f(\theta|S = J-1)$ became more separated as $J$ increased, which increased nonmonotonicity. Although the curve for $P[X_1 = 1|S_{(-j)} = s]$ was always monotone, it can be much flatter overall than $P[X_1 = 1|S = s]$. This is especially true near the modes of the total score distribution, which leads to a lower $r_{t(-1)}$ than $r_t$.

*A less extreme 2PLM example.* This case was defined with $a_1 = 1$ and $a_j = 3$ for all $j > 1$. The item difficulties were $b_1 = 0$, and $b_j$, $j > 1$, uniformly spaced between $-1$ and 1. The $\theta$ distribution was the same as in the previous example. The latent and manifest structures for this example, using the same numbers of items, are shown in Figures 4 and 5.

For smaller $J$, (Figure 4) there were no violations of nondecreasingness for $P[X_1 = 1|S = s]$. However, when $J$ increased, the conditional distributions $f(\theta|S = 1)$ and $f(\theta|S = J - 1)$ pushed outward into a region where $P[X_1 = 1|S = 1, \theta]$ and $P[X_1 = 1|S = J - 1, \theta]$ strongly violated nondecreasingness. It could be argued that, as $J$ increases, the conditional distributions will continue to push out past the region where $P[X_1 = 1|S = 1, \theta]$ and $P[X_1 = 1|S = J - 1, \theta]$ exhibit strong reverse monotonicity. Once past this region, these quantities would again be comparable in size. In this case, conditioning on $S = k$ and $S = j - k$ for suitably selected $k$ could yield the same reverse monotonicity in $P[X_1 = 1|S = k, \theta]$ and $P[X_1 = 1|S = J - k, \theta]$, as well as $P[X_1 = 1|S = k]$ and $P[X_1 = 1|S = J - k]$. The item-rest and item-total correlations again gave few hints about the monotonicity of the corresponding regressions.

*Noncrossing IRFs.* Snijders' example resulted in a specific instance in which $P[X_1 = 1|S = s]$ failed to be monotone, although the IRFs did not cross. This behavior can be replicated in logistic models as well. Assume that $f(\theta)$ is distributed the same as above, and the IRFs are the 2PLM, where $a_1 = .25$ and $b_1 = 2$, and for $j > 1$, $a_j = 2$, where $b_j$ is equally spaced between $-3.1$ and $-2.9$. These curves do not intersect within the range of $\theta$s to which $f(\theta)$ assigns positive probability.

Figure 6a verifies that the IRFs do not cross. Figure 6b shows a nonmonotonicity in $P[X_1 = 1|S = s]$ to the left of the major mode of the $S$ distribution. The large disparity between the item-rest correlation and the item-total correlation in this example is entirely due to the interaction of a large increase in the item-total relationship from $s = 9$ to $s = 10$ and the fact that most of the total score distribution is concentrated on these two values of $s$.

## Discussion

### Dichotomous Items

For dichotomous items, there are three problems with using the item-total regression under the UMIRT model. First, MM for the total score might fail in realistic situations, e.g., Figures 4 and 5. The degree of decreasingness in the item-total regression seems to depend on the difference

**Figure 4**
No Violation of MM for $S$ With Less Extreme Item Parameters (3 and 6 2PLM Items)

a. Latent Structure, 3 Items

b. Manifest Structure, 3 Items ($r_{t(-1)} = .31$; $r_t = .76$)

c. Latent Structure, 6 Items

d. Manifest Structure, 6 Items ($r_{t(-1)} = .37$; $r_t = .59$)

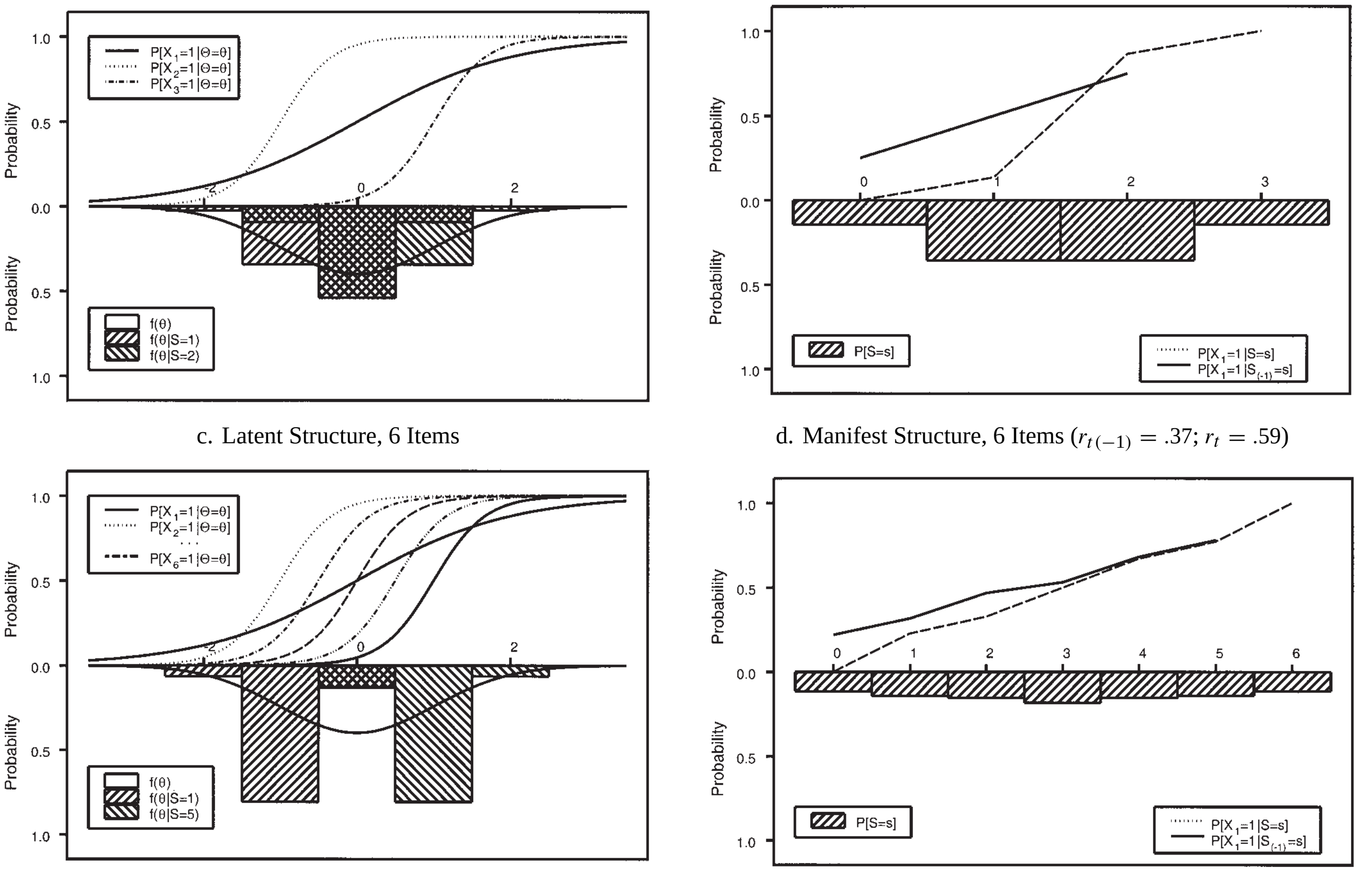
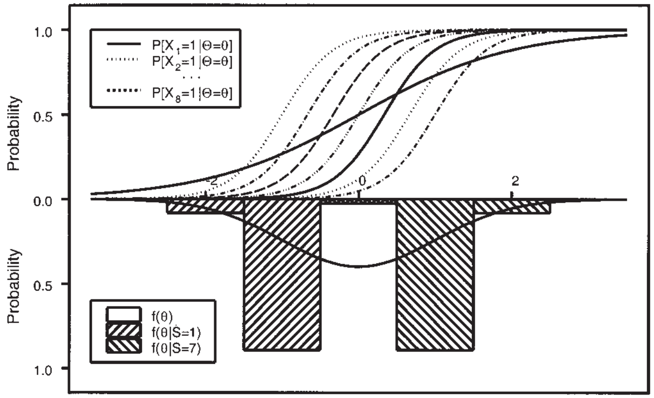


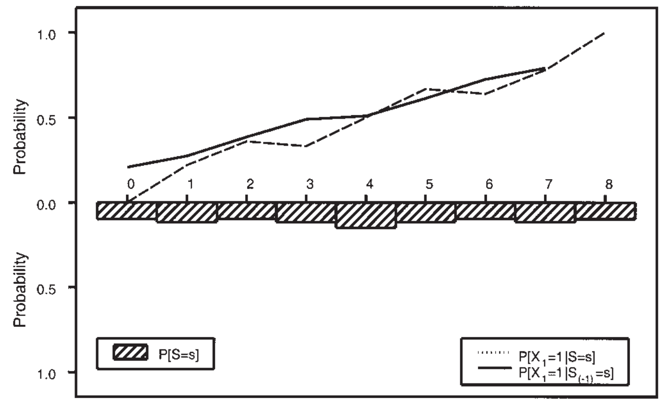
**Figure 5**

Violation of MM for $S$ with Less Extreme Item Parameters (8 and 10 2PLM Items)

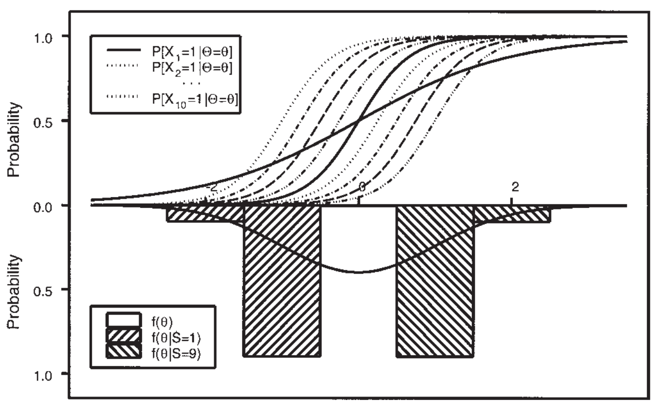
a. Latent Structure, 8 Items




b. Manifest Structure, 8 Items ($r_{t(-1)} = .38$; $r_t = .54$)




c. Latent Structure, 10 Items




d. Manifest Structure, 10 Items ($r_{t(-1)} = .38$; $r_t = .52$)

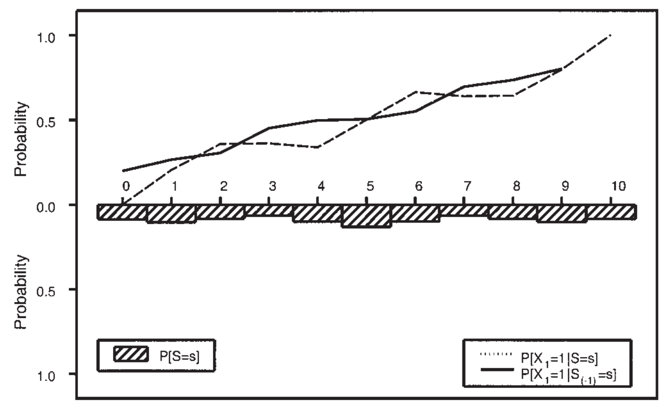
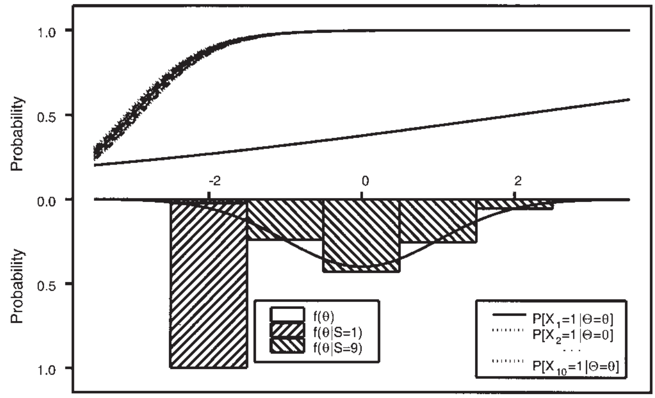
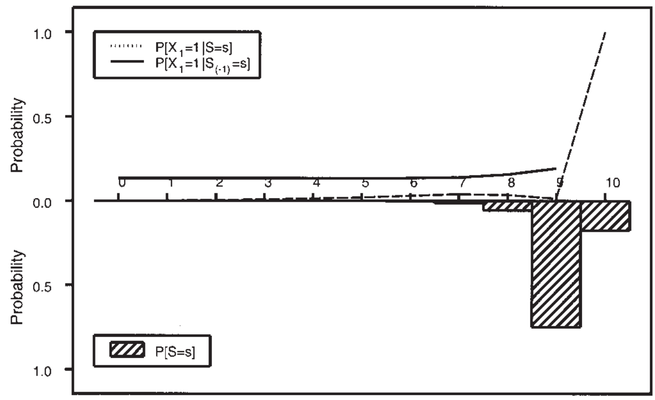


**Figure 6**
Violation of MM for $S$ for 10 2PLM Items With Noncrossing IRFs in the Range $\theta \in [-2,2]$

a. Latent Structure




b. Manifest Structure ($r_{t(-1)} = .03$; $r_t = .71$)




between the slope of the selected item and the slopes of the other items; flatter slopes for the selected item lead to greater nonmonotonicities.

Second, the nonmonotonicities can persist as the number of test items $(J)$ increases. C. Lewis (personal communication, July 8, 1998) conjectured that the graph of the item-total regression converges pointwise to the graph of the item true-score regression (where true score is expected total score), as $J$ increases. If Lewis's conjecture is correct, the examples presented here show that the convergence is not likely to be uniform in the total or proportion-correct scores. Examples not reported here suggest that the situation does not improve when the five-point approximation to a normal $\theta$ distribution is replaced by a more nearly continuous approximation, e.g., with 200 or 500 quadrature points.

Third, the item-total regression can give very misleading results. In particular, items can be rejected although they have properties that do not contradict the IRT model used to build the test. Such false negatives can be avoided by using item-rest regressions. (There is no evidence yet regarding the possibility of false positives for the item-total regression.)

Despite these problems, item-total regressions continue to be frequently used for evaluating the monotonicity of $P_j(\theta)$. The present results indicate that item-rest regressions should be used instead. Not only are they guaranteed to be nondecreasing when all the IRFs are nondecreasing, but they will also be sensitive to some violations of the model in $P_1(\theta)$ if all items except $X_1$ are known to satisfy the UMIRT model. This sensitivity provides some justification for procedures such as those of Wainer (1983) and Wainer, Wadkins & Rogers (1984), who studied the probabilities of distractors of a defective item conditional on number correct for the remaining items.

## Polytomous Items

B. T. Hemker (personal communication, July 7, 1997) concluded that the nonparametric graded response model (np-GRM) does not imply nondecreasing item-rest regressions. For $0 \leq \theta \leq 1$ and two items ($X_1$ and $X_2$) with three answer categories (0, 1, and 2) and identical ISRFs,

$$P(X_j \geq 1|\theta) = \begin{cases} 3\theta, & 0 \leq \theta \leq \frac{1}{4} \\ \frac{2}{3} + \frac{1}{3}\theta, & \frac{1}{4} \leq \theta \leq 1 \end{cases} \tag{25}$$

and

$$P(X_j \geq 2|\theta) = \begin{cases} 2\theta, & 0 \leq \theta \leq \frac{1}{4} \\ \frac{1}{4} + \theta, & \frac{1}{4} < \theta \leq \frac{1}{2} \\ \frac{1}{2} + \frac{1}{2}\theta, & \frac{1}{2} \leq \theta \leq 1\,. \end{cases} \tag{26}$$

These ISRFs are nondecreasing in $\theta$; hence, this is a NGRM. Let $\theta$ have the discrete distribution on $\{.25, .5, 1\}$ with $P[\theta = .25] = P[\theta = .5] = .25$, and $P[\theta = 1] = .5$.

Two different definitions of item-rest regression can be considered for the np-GRM. Manifest ISRFs are $P[X_j \geq x|S_{(-j)} = s]$ where

$$S_{(-j)} = \sum_{i=1}^{J} X_i - X_j\ , \tag{27}$$

and the simpler manifest IRFs are $E[X_j|S_{(-j)} = s]$.

For the manifest ISRF $P[X_1 \geq x|S_{(-1)} = s] = P[X_1 \geq x|X_2 = s]$,

$$P(X_1 \geq 1|X_2 = s) = \begin{cases} 0.7833, & s = 0 \\ 0.7708, & s = 1 \\ 0.9231, & s = 2 \end{cases} \tag{28}$$

and

$$P(X_1 \geq 2|X_2 = s) = \begin{cases} 0.6000, & s = 0 \\ 0.5625, & s = 1 \\ 0.8654, & s = 2 \end{cases}\ . \tag{29}$$

Both manifest ISRFs fail to be nondecreasing in $s$.

For the manifest IRFs, $E[X_1|S_{(-1)} = s] = E[X_1|X_2 = s]$ and $E[X_1|X_2] = 1(P[X_1 \geq 1|X_2] - P[X_1 \geq 2|X_2]) + 2P[X_1 \geq 2|X_2] = P[X_1 \geq 1|X_2] + P[X_1 \geq 2|X_2]$, so that

$$E(X_1|X_2 = 0) = 0.7833 + 0.6000 = 1.3833\ , \tag{30}$$

$$E(X_1|X_2 = 1) = 0.7708 + 0.5625 = 1.3333\ , \tag{31}$$

and

$$E(X_1|X_2 = 2) = 0.9321 + 0.8654 = 1.7975\ , \tag{32}$$

so that $E[X_1|X_2 = s]$ also fails to be nondecreasing in $s$.

Part of the problem is that SOL does not hold for many UMIRT models for polytomous responses (Hemker et al., 1997). Indeed, the example showed that $P[\theta \geq .5|X_2 = s]$ was .4, .25, and .85 for $s$ = 0, 1, and 2, respectively.

Thus, for polytomous item response models the first $\overset{?}{\leq}$ in an adaptation of Equation 3 for polytomous items (not presented here) also fails, in general. It is not known whether there is a simple and general unidimensional summary $R$ of polytomous item responses, one for which $P[\theta c|R = r]$ can be nondecreasing in $r$ when Equation 1 holds.

It is always possible, however, to reduce polytomous test data with ordered categories to dichotomous test data. Each polytomously scored $X_j$ is replaced with a dichotomous variable $Y_{jc_j}$ that is equal to 1 when $X_j > c_j$ and 0 otherwise. For every choice of the sequence of thresholds $c_j$ $(j = 1, 2, \ldots, J)$, this will produce a set of dichotomous item responses that satisfies Equation 1 (as long as the polytomous responses also satisfied it; e.g., Junker, 1991; Junker & Ellis, 1997; Samejima, 1969; Scheiblechner, 1995). Thus for each such dichotomization, the monotonicity of $P[Y_{jc_j} = 1|S_{(-j,\mathbf{c})} = s]$ can be examined, where $S_{(-j,\mathbf{c})}$ is the rest score corresponding to the sequence $\mathbf{c} = (c_1, c_2, \ldots, c_J)$ and item $j$. This leads to a large number of monotonicity conditions to analyze, but many should be stochastically dependent on one another. Thus, a careful search strategy could quickly find cases in which monotonicity might be violated.

## References

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Clarke, B. S., & Ghosh, J. K. (1995). Posterior convergence given the mean. *Annals of Statistics, 23,* 2116–2144.

Croon, M. A. (1991). Investigating Mokken scalability of dichotomous items by means of ordinal latent class analysis. *British Journal of Mathematical and Statistical Psychology, 44,* 315–331.

Cureton, E. E. (1966). Corrected item-test correlations. *Psychometrika, 31,* 93–96.

Ellis, J. L., & Junker, B. W. (1997). Tail-measurability in monotone latent variable models. *Psychometrika, 62,* 495–523.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests* [Introduction to the theory of psychological tests]. Bern: Huber.

Friendly, M. (1994). Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association, 89,* 190–200.

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

Guttman, L. (1950). The basis of scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Measurement and prediction* (pp. 60–90). Princeton NJ: Princeton University Press.

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.

Holland, P. W. (1990). The Dutch Identity: A new tool for the study of item response models. *Psychometrika, 55,* 5–18.

Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent trait models. *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. *Annals of Statistics, 21,* 1359–1378.

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

Junker, B. W., & Ellis, J. L. (1997). A characterization of monotone unidimensional latent variable models. *Annals of Statistics, 25,* 1327–1343.

Lehmann, E. L. (1955). Ordered families of distributions. *Annals of Mathematical Statistics, 26,* 399–419.

Lord, F. M. (1965). An empirical study of item-test regression. *Psychometrika, 30,* 373–376.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale NJ: Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Wesley.

Mokken, R. J. (1971). *A theory and procedure of scale analysis.* The Hague, The Netherlands: Mouton.

Molenaar, I. W. (1983). Some improved diagnostics for failure of the Rasch model. *Psychometrika, 48,* 49–72.

Molenaar, I. W. (1995). Estimation of item parameters. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments and applications* (pp. 39–51). New York: Springer.

Orlando, M., & Thissen, D. (1997, July). *New item fit indices for dichotomous item response theory models.* Paper presented at the annual meeting of the Psychometric Society, Gatlinburg TN.

Ramsay, J. O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika, 56,* 611–630.

Ramsay, J. O., & Abrahamowicz, M. (1989). Binomial regression with monotone splines: A psychometric application. *Journal of the American Statistical Association, 84,* 906–915.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* (Copenhagen, Danish Institute for Educational Research). Expanded edition (1980), with foreword and afterword by B. D. Wright. Chicago: University of Chicago Press.

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

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

Scheiblechner, H. (1998). Corrections of theorems in Scheiblechner's treatment of ISOP models and comments on Junker's remarks. *Psychometrika, 63,* 87–91.

Sijtsma, K. (1988). *Contributions to Mokken's nonparametric item response theory.* Amsterdam: Free University Press.

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., & Orlando, M. (1997, July). Graphical fit displays for dichotomous item response theory models. Paper presented at the annual meeting of the Psychometric Society, Gatlinburg TN.

van den Wollenberg, A. L. (1982). Two new test statistics for the Rasch model. *Psychometrika, 47,* 123–140.

van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory.* New York: Springer.

Wainer, H. (1983). Pyramid power: Searching for an error in test scoring with 830,000 helpers. *American Statistician, 37,* 87–91.

Wainer, H., Wadkins, J. R. J., & Rogers, A. (1984). Was there one distractor too many? *Journal of Educational Statistics, 9,* 5–24.

Wolf, R. (1967). Evaluation of several formulae for correction of item-total correlations in item analysis. *Journal of Educational Measurement, 4,* 21–26.

## Acknowledgments

*This research was supported in part by the National Institutes of Health, Grant CA54852, and the National Science Foundation, Grants DMS-94.04438 and DMS-97.05032. An earlier version of this paper was presented at the June, 1997 meeting of the Psychometric Society, Gatlinburg, Tennessee.*

## Author's Address

Send requests for reprints or further information to Brian W. Junker, Department of Statistics, Carnegie Mellon University, Pittsburgh PA 15213, U.S.A. Email: brian@stat.cmu.edu.