Educational and Psychological Measurement
71(4) 684–711

DOI: 10.1177/0013164410378690
http://epm.sagepub.com

Target Rotations and Assessing the Impact of Model Violations on the Parameters of Unidimensional Item Response Theory Models

Steven Reise[1], Tyler Moore[1], and Alberto Maydeu-Olivares[2]

Abstract

Reise, Cook, and Moore proposed a "comparison modeling" approach to assess the distortion in item parameter estimates when a unidimensional item response theory (IRT) model is imposed on multidimensional data. Central to their approach is the comparison of item slope parameter estimates from a unidimensional IRT model (a restricted model), with the item slope parameter estimates from the general factor in an exploratory bifactor IRT model (the unrestricted comparison model). In turn, these authors suggested that the unrestricted comparison bifactor model be derived from a target factor rotation. The goal of this study was to provide further empirical support for the use of target rotations as a method for deriving a comparison model. Specifically, we conducted Monte Carlo analyses exploring (a) the use of the Schmid–Leiman orthogonalization to specify a viable initial target matrix and (b) the recovery of true bifactor pattern matrices using target rotations as implemented in Mplus. Results suggest that to the degree that item response data conform to independent cluster structure, target rotations can be used productively to establish a plausible comparison model.

Keywords

item response theory, bifactor model, ordinal factor analysis, unidimensionality, multidimensionality, Schmid–Leiman transformation, target rotations

[1]University of California–Los Angeles, Los Angeles, CA, USA
[2]University of Barcelona, Barcelona, Spain

Corresponding Author:
Steven Reise, Department of Psychology, University of California–Los Angeles, 1285 Franz Hall, Los Angeles, CA 90091, USA
Email: reise@psych.ucla.edu

Most applications of item response theory (IRT) modeling have focused on unidimensional models (i.e., models that assume a single latent trait). Yet if more than one latent trait is required to achieve local independence, then using a unidimensional model will result in biased IRT item parameters. There is a large literature that has explored the robustness of unidimensional IRT model parameter estimates to unidimensionality violations (e.g., Drasgow & Parsons, 1983; Reckase, 1979). One conclusion arising from this literature is that if factor analysis or a related method is applied to the data and a ''strong'' common factor, or multiple highly correlated factors are obtained, then unidimensional IRT item parameter estimates are reasonably robust, that is, they are not seriously distorted.

Drawing from the IRT robustness literature, many researchers have adopted the idea that it is safe to apply unidimensional IRT models when data have a ''strong'' common dimension. Consequently, researchers considering the application of unidimensional IRT models rely almost exclusively on rule-of-thumb guidelines (e.g., ratio of first to second eigenvalues greater than 3) and post hoc model analysis (e.g., small residuals after fitting a single factor, or acceptable ''fit'' to a one-factor model; McDonald & Mok, 1995) to judge whether item response data are ''unidimensional enough'' for application of IRT.

Reise, Cook, and Moore (2010), on the other hand, argued that no matter how reasonable, intuitive, or statistically rigorous, no rule-of-thumb index can provide an estimate of the amount of bias that exists in the item parameter estimates obtained by fitting a unidimensional IRT model when multidimensionality is present. Thus, as a complement to rule-of-thumb guidelines and post hoc model checking, they suggested fitting a multidimensional and a unidimensional IRT model and then comparing the resulting item slope parameters (see also Ip, 2010). For this comparison to be meaningful, the multidimensional IRT model must consist of an overall general dimension (to capture what is in common among the items) plus some additional group dimensions (to capture local dependencies). Furthermore, the comparison should be performed using an intercept/slope parameterization, not a threshold/factor loading parameterization, for reasons that will be discussed shortly.

Now, to obtain a multidimensional IRT solution with an overall dimension and specific additional dimensions (i.e., a hierarchical solution) requires some work. One first needs to find a plausible multidimensional model for the data at hand. Then, the resulting solution is transformed into a hierarchical solution using a Schmid–Leiman transformation (SL; Schmid & Leiman, 1957). However, as we shall show, the resulting SL solution may yield biased estimates of the parameters on the general dimension and should not be compared with the parameters estimated in the unidimensional model. This is because the SL solution leads to a constrained hierarchical solution. There are restrictions among the estimated parameters (see Yung, Thissen, & McLeod, 1999). Thus, an additional step is needed. Using the SL solution (i.e., the constrained hierarchical solution) as target, a target rotation needs to be performed on the parameters of the multidimensional solution. The resulting solution is an unconstrained hierarchical model whose parameters on the general dimension are to be compared with those obtained in the unidimensional solution. Notice that if in the unconstrained solution each item depends on just the overall factor and a group factor (regardless of the number of group factors) this is a bifactor model (Gibbons & Hedeker, 1992) and throughout this article we shall use the term *bifactor* and *hierarchical model* interchangeably.

Obviously, finding the best multidimensional IRT model may be rather computer intensive and one may question the utility of so much computational effort simply to investigate the robustness of IRT parameter estimates. However, the multidimensional IRT model can be estimated using limited information methods (from tetrachoric/polychoric correlations) and as a result, the full procedure may take mere seconds. Note that when estimating the model in this fashion, thresholds and factor loadings are estimated. Ultimately, these need to be transformed to intercepts and slopes for comparison to the unidimensional results, as described below.

Reise et al. (2010) have summarized the steps needed to investigate the robustness of unidimensional IRT parameter estimates to local independence violations as follows:

1. Find the best fitting and plausible multidimensional factor model for the data at hand estimating the model using tetrachoric/polychoric correlations.
2. Apply an SL transformation to the factor loadings obtained in Step 1 to obtain a restricted hierarchical solution.
3. Using the solution obtained in Step 2 as target, obtain an unrestricted hierarchical solution using target rotations (Browne, 2001).
4. Transform the factor loadings obtained in Step 3 to IRT slopes. Compare them with the slopes obtained for a unidimensional solution. It is preferable that for this comparison the parameters of the unidimensional solution are also obtained using limited information methods (i.e., estimating the model using tetrachoric/polychoric correlations) to avoid possible biases in the comparison induced by the use of different estimation methods.

Reise et al. (2010) suggested that a finding of little or no difference in IRT slopes argues for the application of a unidimensional IRT model whereas a finding of large differences argues for (a) fitting a multidimensional IRT model—a bifactor IRT model (Gibbons & Hedeker, 1992), a unidimensional model with local dependencies modeled through correlated residuals (Ip, 2010), or an independent clusters solution (McDonald, 1999, 2000); (b) revising the scale by deleting problematic items from the analyses (e.g., items that load on more than one group factor in a hierarchical solution model); or (c) abandoning any latent variable model (i.e., concluding the scale lacks any identifiable and interpretable latent structure).

In this study, our primary goal is to elaborate on and provide empirical support for the approach suggested in Reise et al. (2010). To do so, first we provide an example to illustrate their proposal. Two sections follow the numerical example: First, we review the relations between ordinal factor analysis (OFA) and IRT models to justify our use of OFA for studying the parameters of IRT models. Second, Monte Carlo simulation is used to explore (a) the viability of using a SL transformation to suggest an initial bifactor target matrix (Steps 2 and 3) and (b) assuming that the target pattern matrix is specified correctly, the ability of targeted bifactor pattern rotations to recover true population parameters.

Ordinal Factor Analysis and Item Response Theory

In this section, we review the different parameterizations that may be used in two-parameter IRT models. These are well known and have received considerable treatment in the psychometric literature (e.g., Ackerman, 2005; Bartholomew & Knott, 1999; Forero & Maydeu-Olivares, 2009; Knol & Berger, 1991; McDonald, 2000; McDonald & Mok, 1995; McLeod, Swygert, & Thissen, 2001; Takane, & de Leeuw, 1987; Wirth & Edwards, 2007). However, as we shall see, in the case of bifactor models equivalent bifactor models can appear very different in terms of the implied data structure just because of the choice of parameterization.

Item Response Theory Models

IRT models specify an item response curve (IRC) to characterize the nonlinear relation between individual differences on a continuous latent variable (θ) and the probability of endorsing a dichotomous item x_i. Equation (1) shows the two-parameter normal-ogive model:

$$P(x_i = 1|\theta) = E(x_i|\theta) = \int_{-\infty}^{z_i} \frac{1}{\sqrt{2\pi}} \exp(-Z_i^2/2)dt. \tag{1}$$

In Equation (1), z is a normal deviate equal to $\alpha(\theta - \beta)$, where α is an item discrimination parameter and β is an item location parameter corresponding to where on the latent trait scale the probability of endorsing the item is .50. Stated in different terms, Equation (1) is a cumulative normal distribution with a mean equal to β and standard deviation $\sigma = 1/\alpha$. Thus, $z = (\theta - \mu)/\sigma$ is simply the z-score derived from a particular normal distribution which is used to determine the cumulative proportion of cases that fall at or below the normal deviate.

A more commonly applied IRT model that yields a very similar fit is the two-parameter logistic model (2PL) shown in Equation (2).

$$P(x_i = 1|\theta) = E(x_i|\theta) = \frac{\exp(1.7z_i)}{1 + \exp(1.7z_i)} = \frac{\exp(1.7\alpha_i(\theta - \beta_i))}{1 + \exp(1.7\alpha_i(\theta - \beta_i))}, \tag{2}$$

where 1.7 is a scaling factor that makes the item slope parameter comparable to its value under the normal-ogive parameterization (Equation 1). Including the 1.7 is critical because only the normal-ogive parameterization can be easily linked with ordinal factor analytic terms.

Equations (1) and (2) express the two-parameter model using its most commonly found parameterization. However, this is not the parameterization that is used by most IRT software packages, and this parameterization is not readily generalizable to the multidimensional case. Instead it is better to write

$$z_i = \alpha_i\theta - \alpha_i\beta_i \tag{3}$$

and then define an intercept term $\gamma = -\alpha\beta$ and rewrite the deviate as shown in Equation (4):

$$z_i = \alpha_i\theta + \gamma_i. \tag{4}$$

Obviously, because the slope is always positive, items with negative locations (easy items) will have positive intercepts and items with positive locations (difficult items) will have negative intercepts. Thus γ reflects the relative ''easiness'' of the item.

Equation (4) has a straightforward generalization to the multivariate case, where the number of dimensions is $p = 1, \ldots, P$ (possibly) correlated factors.

$$z_i = \sum \alpha_{ip}\theta_p + \gamma_i. \tag{5}$$

In Equation (5), γ is now a multidimensional intercept parameter representing the deviate corresponding to the expected proportion endorsing an item for individuals who are at the mean on all latent traits. In the simple case of a bifactor IRT model (e.g., Gibbons & Hedeker, 1992) where each item loads on a single general factor (G), and at most on one additional orthogonal group factor ($p =$ relevant group factor), Equation (5) becomes

$$z_i = \alpha_{iG}(\theta_G) + \alpha_{ip'}(\theta_{p'}) + \gamma_i. \tag{6}$$

Ordinal Factor Analytic Models

Applying the usual (linear) factor analysis model to dichotomous or polytomous items is problematic for a variety of reasons (see Bernstein & Teng, 1989; Wirth & Edwards, 2007). However, the model can be easily modified to accommodate ordinal responses by adding a threshold process to the linear factor model for continuously distributed responses (Christoffersson, 1975; Maydeu-Olivares, 2006; Muthén, 1979). For example, consider the binary case and n items. The ordinal factor analysis model assumes that the observed 0 or 1 item response is a discrete realization of a continuous and normally distributed latent response process ($x*$) underlying the items. More specifically, consider the fully standardized model:

$$E(x_i^*|\theta_1 \ldots \theta_P) = x_i^* = \lambda_{i1}\theta_1 + \lambda_{i2}\theta_2 + \lambda_{i3}\theta_3 + \cdots + \lambda_{iP}\theta_P + \varepsilon_i. \tag{7}$$

In this model, $x*$ is not an observed 0 or 1 item score but rather is a continuous and normally distributed response propensity variable assumed to underlie item performance. In addition, λ is a factor loading (correlation), which, in the unidimensional case, can be shown to be analogous to an item-test biserial correlation. To develop the model further, an item threshold parameter (τ) needs to be specified such that $x = 1$, if $x* \geq t_i$ and $x_i = 0$, if $x* < t_i$. That is, an individual will endorse an item only if his or her response propensity is above the item's threshold.

The model in Equation (7) assumes that the errors are multivariate normal and uncorrelated with each other and the latent traits. The implied correlation matrix is thus

$$\Sigma = \Lambda\Phi\Lambda' + \Psi, \tag{8}$$

where, Σ is the implied correlation matrix for the ''hypothetical'' latent propensities (i.e., it is a tetrachoric correlation matrix), Λ is an $n \times p$ matrix of factor loadings, Φ is a $p \times p$ matrix of factor correlations, and Ψ is a $p \times p$ diagonal matrix of residual variances such that $\Psi = \mathbf{I} - \mathrm{diag}(\Lambda\Phi\Lambda')$.

It turns out that the OFA model just described and the normal-ogive model in Equation (1) are equivalent models. Indeed, we can write the normal deviate for the multidimensional or bifactor cases:

$$z_i = \frac{\lambda_{i1}\theta_1 + \lambda_{i2}\theta_2 + \cdots + \lambda_{iP}\theta_P - \tau_i}{\sqrt{1 - \sum \lambda_i^2}} \quad \text{or} \quad z_i = \frac{\lambda_{iG}\theta_G + \lambda_{ip'}\theta_{p'} - \tau_i}{\sqrt{1 - \sum \lambda_i^2}}. \tag{9}$$

Clearly, in either case the normal deviate is formed as a weighted linear combination of the factor loadings, which is then scaled relative to the square root of an item's unique variance (the residual standard deviation of $x*$). Finally, when either of the normal deviates in Equation (9) is substituted into Equation (1), the model becomes a normal-ogive OFA model.

Transformations

The equations to translate between an OFA and the two-parameter normal-ogive IRT model are cited in dozens of readily available sources (e.g., Ackerman, 2005; Knol & Berger, 1991; McDonald, 2000, McLeod et al., 2001), and are obvious through inspection of Equation (9). Assuming that the latent factors are uncorrelated, the translations between FA loadings and IRT slopes (normal-ogive metric) are

$$\lambda_{ip} = \frac{\alpha_{ip}}{\sqrt{1 + \sum_{p=1}^{P} \alpha_{ip}^2}}, \quad \alpha_{ip} = \frac{\lambda_{ip}}{\sqrt{\psi_i}} = \frac{\lambda_{ip}}{\sqrt{1 - \sum_{p=1}^{P} \lambda_{ip}^2}}. \tag{10}$$

The denominator in the latter equation is readily interpretable as the square root of the item's uniqueness, ψ_i (one minus the communality). Thus, it is immediately apparent that a slope α on a particular dimension is a function of both an item's loading on that dimension, λ, and its uniqueness, ψ, within the context of a specific model (Muthén & Christoffersson, 1981, p. 411); items with more common variance in the OFA model will have relatively higher slopes in the IRT parameterization. This is exactly why the interpretation of equivalent multidimensional IRT and OFA models is challenging, as we will show shortly.

Note that the denominator of either of the above equations becomes considerably more complicated in the case of multiple correlated dimensions. For example, in the case of two correlated factors, the denominator for the IRT slope equation is

$$\sqrt{\psi_i} = \sqrt{1 - \sum_{p=1}^{P} \lambda_{ip}^2 - 2\lambda_1\lambda_2\phi_{12}}. \tag{11}$$

Finally, the translation between IRT item intercepts and factor analytic thresholds follows a similar logic as Equation (10).

$$\tau_i = \frac{-\gamma_i}{\sqrt{1 + \sum_{p=1}^{P} \alpha_{ip}^2}}, \quad \gamma_i = \frac{-\tau_i}{\sqrt{1 - \sum_{p=1}^{P} \lambda_{ip}^2}}. \tag{12}$$

Interpreting Multidimensional Solutions in OFA and IRT

When data are strictly unidimensional the relations between OFA and IRT are simple because the item parameter estimates leave the researcher with much the same impression of the data structure and psychometric functioning of the items. For example, in the first column in the top half of Table 1 are factor loadings for 15 items. This matrix mimics a hypothetical test with three content clusters of five items each, where the relation between the item and common latent trait is equal within cluster but not between. In the next column are the transformed IRT parameters. Clearly, the same message about the data structure is conveyed—Items 1 through 5 are the best, Items 11 through 15 are the worst, and the relations between item and latent trait are the same within cluster but not between.

Such symmetry of translation is, generally speaking, not possible when data have a multidimensional structure and the items vary in communality. In the top middle portion of Table 1 we display the loading matrix for a bifactor model with one general and three orthogonal group factors. The items are all equally related to the general trait, but the content clusters differ in their loadings on the group factors. In the final set of columns are shown the equivalent transformed IRT model parameters. The items do not have equal relations to the general trait, but rather the subset of items with the smallest uniqueness appear to be the best items in terms of measuring the general factor. In short, the factor analytic parameterization suggests a different psychometric interpretation than under the IRT parameterization.

In the bottom portion of Table 1 we reversed the process by starting with an IRT model with equal slopes on the general factor and varying the group factor slopes. Specifically, the item slopes on the general factor are 0.75, and group slopes are 0.58, 0.32, and 0.10. The true slopes in the bottom portion of Table 1 mimic the true factor values in the top half of Table 1 in that these are the IRT slopes equal to the factor loadings ignoring the multidimensionality. That is, the true slopes are simply

Table 1. Demonstration of OFA and IRT Transformations

	Unidimensional		Multidimensional							
	FA	IRT	FA				IRT			
Item	λ	α	λ_g	λ_1	λ_2	λ_3	α_g	α_1	α_2	α_3
1	0.78	1.17	0.60	0.50			0.96	0.80		
2	0.78	1.17	0.60	0.50			0.96	0.80		
3	0.78	1.17	0.60	0.50			0.96	0.80		
4	0.78	1.17	0.60	0.50			0.96	0.80		
5	0.78	1.17	0.60	0.50			0.96	0.80		
6	0.67	0.90	0.60		0.30		0.81		0.40	
7	0.67	0.90	0.60		0.30		0.81		0.40	
8	0.67	0.90	0.60		0.30		0.81		0.40	
9	0.67	0.90	0.60		0.30		0.81		0.40	
10	0.67	0.90	0.60		0.30		0.81		0.40	
11	0.61	0.77	0.60			0.10	0.76			0.13
12	0.61	0.77	0.60			0.10	0.76			0.13
13	0.61	0.77	0.60			0.10	0.76			0.13
14	0.61	0.77	0.60			0.10	0.76			0.13
15	0.61	0.77	0.60			0.10	0.76			0.13
	Unidimensional		**Multidimensional**							
	IRT	**FA**	**IRT**				**FA**			
Item	λ	α	α_g	α_1	α_2	α_3	λ_g	λ_1	λ_2	λ_3
1	1.17	0.78	0.75	0.58			0.54	0.42		
2	1.17	0.78	0.75	0.58			0.54	0.42		
3	1.17	0.78	0.75	0.58			0.54	0.42		
4	1.17	0.78	0.75	0.58			0.54	0.42		
5	1.17	0.78	0.75	0.58			0.54	0.42		
6	0.90	0.67	0.75		0.32		0.58		0.25	
7	0.90	0.67	0.75		0.32		0.58		0.25	
8	0.90	0.67	0.75		0.32		0.58		0.25	
9	0.90	0.67	0.75		0.32		0.58		0.25	
10	0.90	0.61	0.75		0.32		0.58		0.25	
11	0.77	0.61	0.75			0.10	0.60			0.08
12	0.77	0.61	0.75			0.10	0.60			0.08
13	0.77	0.61	0.75			0.10	0.60			0.08
14	0.77	0.61	0.75			0.10	0.60			0.08
15	0.77	0.61	0.75			0.10	0.60			0.08

OFA = ordinal factor analysis; IRT = item response theory.

the true loadings divided by the square root of 1 minus the loading squared. Clearly, the right hand side in the lower half of Table 1 demonstrates that when IRT parameters are translated to OFA terms, the items no longer load on the general factor equally.

The results in Table 1 demonstrate that although the IRT and OFA models are equivalent in implied correlational structure, in the multidimensional case there is a nonsymmetry to the interpretation owing to the critical role of an item's uniqueness. This nonsymmetry of interpretation must be considered when a researcher is using the results of one parameterization to understand what the equivalent model would look like under the alternative parameterization. Fortunately, programs for estimating multidimensional models such as TESTFACT (Bock et al., 2002) and NOHARM (Fraser, 1988; Fraser & McDonald, 1988) routinely provide results in both IRT and OFA terms. Nevertheless, this nonsymmetry of interpretation in multidimensional models can cause confusion when interpreting the results of research into the effects of model violations on item parameter estimates.

Parameter Estimation in OFA and IRT

The above section outlined the theoretical relations between OFA and IRT parameters; however, in practice, such parameters need to be estimated. Generally speaking, there are two strategies for estimating these parameters. One strategy, used in OFA, estimates the parameters using only bivariate information, in several stages, generally through the use of tetrachoric (or polychoric) correlations (see, e.g., Muthén, 1993). The other strategy, used in IRT estimates the parameters using full information maximum likelihood estimation (Bock & Aitkin, 1981). Thus, even when the same model is estimated, as in the case of the normal-ogive model, OFA and IRT use slightly different approaches. In OFA, the threshold and factor loading parameterization is used, and parameters are estimated using bivariate information. In IRT, the intercept and slope parameterization is used, and parameters are estimated using full information maximum likelihood. Full information maximum likelihood is theoretically superior to bivariate estimation methods, but the latter are considerably faster for multidimensional models. Another advantage of the bivariate information methods is that a goodness-of-fit test of the model to the tetrachoric correlations is available (see Muthén, 1993; Maydeu-Olivares, 2006)—and an associated root mean square error of approximation. In contrast, comparable procedures for goodness-of-fit assessment in the case of full information maximum likelihood estimation (e.g., Maydeu-Olivares & Joe, 2005, 2006) are not yet available in standard software.

Reise et al. (2010) argued that the simplest bivariate estimation method, common factoring via unweighted least squares estimation of a tetrachoric matrix is a good enough approach for the purposes of determining the robustness of unidimensional modeling of multidimensional data. This view is supported by research such as Forero and Maydeu-Olivares (2009), who performed an exhaustive Monte Carlo comparison of full and bivariate estimation strategies and concluded that for samples with more than 500 observations the differences in parameter accuracy between both methods were negligible (and even slightly favoring unweighted least squares). In the next section, we consider the development and specification of a comparison bifactor model using the SL and target rotations.

The Schmid–Leiman and Target Bifactor Rotations

As described above, in the comparison modeling approach, estimating the degree to which model violations affect unidimensional IRT item parameter estimates involves the comparison of an exploratory bifactor model (i.e., the unrestricted model) with a highly restricted model (i.e., a unidimensional model). Although the model comparison idea is simple enough, the realization in practice is challenging because of the fact that commonly used analytic rotations are geared toward identifying simple structure solutions (see Browne, 2001 for detailed discussion). When the true data structure is bifactor in the population, exploratory analytic rotations such as promax or oblimin will fail to identify the correct underlying data structure.

It must be recognized that in practice, to use exploratory multidimensional models to judge the effects of forcing data into a restricted model, the comparison model must be plausible and reasonably reflect the population loading pattern. That is, it must approximate what the true factor loadings or IRT slopes would be without the imposed restrictions; one cannot meaningfully compare a wrong bifactor model with a more wrong unidimensional model. The challenge is, given the limitations of analytic rotations, how can researchers identify an acceptable exploratory method that can identify a bifactor structure?

As noted in the steps above, Reise et al. (2010) suggested that an SL orthogonalization (Schmid & Leiman, 1957) be used as a basis for specifying a bifactor target pattern, and then a target pattern rotation (Browne, 2001) be used to estimate an unrestricted comparison model. This begs two questions. First, assuming that in the population the data structure is bifactor, under what conditions can the SL orthogonalization identify a good target matrix? By good target matrix, we mean a matrix that specifies all near-zero loadings to specified (0), and all nontrivial loadings to unspecified (?). Second, given a correct target matrix, how well can target rotations recover true population parameters? In the following two sections, we use Monte Carlo simulation to address each of these issues in turn.

The Schmid–Leiman: Getting the Target Bifactor Pattern Right

Target pattern rotations require an initial target matrix of specified (0) and unspecified (?) elements. ‘‘This target matrix reflects partial knowledge as to what the factor pattern should be’’ (Browne, 2001, p. 124). In addition, the target matrix forms the basis for a rotation that minimizes the sum of squared differences between the target and the rotated factor pattern.

Although target rotations potentially have a self-correcting element (i.e., just because a loading is specified as zero does not mean the loading estimate must be zero and vice versa), the robustness of misspecifying the elements of a target matrix is currently unknown. Thus, at this point, target rotations (Browne, 2001) can only be expected to function optimally when the target pattern is correctly specified.

It follows that an important goal of comparison modeling is to identify a bifactor structure so that a researcher can form a target matrix such that items with nontrivial (e.g., $\geq .30$) loadings on the group factors in the population are unspecified (?), and items with trivial or near-zero loadings are specified (0). Note that it is assumed all items have nontrivial loadings on the general factor. Of course, there are many way to specify an initial target matrix. For example, a target can be generated from theory, or from exploratory data analyses such as factor analysis, cluster analyses, or multidimensional scaling.

In conjunction with these methods, Reise et al. (2010) suggested that the SL orthogonalization be used as an exploratory tool to determine the number of identified group factors in a data matrix and to determine the pattern of trivial and nontrivial loadings on the group factors. Here, we define a nontrivial loading as being greater than or equal to .15 in an SL (Schmid & Leiman, 1957) solution (as opposed to a population value) for reasons that will be made clear shortly. In this section, we consider the ability of the SL to identify an appropriate target matrix.

The SL is an orthogonalization of a higher order factor model, which itself is a product of a common factor analysis with oblique rotation (McDonald, 2000). For example, starting with a tetrachoric correlation matrix, a researcher: (a) extracts (e.g., unweighted least squares) a specified number of primary factors and rotates the factors obliquely (e.g., oblimin) and (b) factor analyzes the primary factor correlation matrix to determine the relation between each primary factor and a general factor. Given this higher order representation, the SL is easily obtained. Specifically, an item's loading on the general factor is simply its loading on the primary factor multiplied by the loading of the primary factor on the general factor. An item's loading on a group factor is its loading on the primary factor multiplied by the square root of the disturbance (i.e., that part of the primary factor that is not explained by the general factor).

The Achilles' heel of the SL procedure is that because of the well known proportionality constraints (Yung et al., 1999), the SL estimates of population loadings are unbiased under only a narrow range of conditions. The proportionality constraints emerge because, as described above, for each item, the group and general loadings in the SL are functions of common elements (i.e., the loading of the primary on the general and the unique variance of the primary). As a consequence, a researcher cannot rely on the SL to derive an acceptable comparison model because of the bias in factor loadings that emerge under some conditions.

To demonstrate this, in the first set of columns in the top portion of Table 2, we specified a population bifactor structure with equal loadings on the general and equal loadings within each group factor. In the bottom portion is a SL of the implied correlation matrix, extracting three factors using unweighted least squares and rotating using oblimin. The SCHMID routine included in the PSYCH package (Revelle, 2009) of the R software program (R Development Core Team, 2008) was used for all analyses. Obviously, in this example, the SL perfectly reflects the population structure because the loadings within the group factors are proportional to the general factor. If real data were to always mimic this perfectly proportional structure, the SL

Table 2. The Schmid–Leiman Orthogonalization Under Three Conditions

	True population structure											
	IC: proportional				IC: not proportional				IC basis			
Item	Gen	G1	G2	G3	Gen	G1	G2	G3	Gen	G1	G2	G3
1	.50	.60			.40	.27			.50	.60	.50	
2	.50	.60			.40	.68			.50	.60		
3	.50	.60			.40	.65			.50	.60		
4	.50	.60			.40	.21			.50	.60		
5	.50	.60			.40	.65			.50	.60		
6	.50		.50		.40		.43		.50		.50	.50
7	.50		.50		.40		.51		.50		.50	
8	.50		.50		.40		.29		.50		.50	
9	.50		.50		.40		.27		.50		.50	
10	.50		.50		.40		.19		.50		.50	
11	.50			.40	.40			.61	.50	.50		.40
12	.50			.40	.40			.13	.50			.40
13	.50			.40	.40			.46	.50			.40
14	.50			.40	.40			.64	.50			.40
15	.50			.40	.40			.53	.50			.40
	Schmid–Leiman											
Item	Gen	G1	G2	G3	Gen	G1	G2	G3	Gen	G1	G2	G3
1	.50	.60	.00	.00	.33	.32	.09	.05	**.65**	**.49**	**.36**	**.11**
2	.50	.60	.00	.00	.41	.67	.01	.01	.49	.61	.01	.02
3	.50	.60	.00	.00	.40	.65	.00	.00	.49	.61	.01	.02
4	.50	.60	.00	.00	.31	.27	.10	.06	.49	.61	.01	.02
5	.50	.60	.00	.00	.40	.65	.00	.00	.49	.61	.01	.02
6	.50	.00	.50	.00	.41	.01	.41	.01	**.61**	**.05**	**.40**	**.38**
7	.50	.00	.50	.00	.43	.02	.45	.02	.50	.01	.50	.03
8	.50	.00	.50	.00	.37	.02	.33	.02	.50	.01	.50	.03
9	.50	.00	.50	.00	.36	.02	.32	.03	.50	.01	.50	.03
10	.50	.00	.50	.00	.33	.04	.27	.05	.50	.01	.50	.03
11	.50	.00	.00	.40	.41	.01	.01	.60	**.56**	**.45**	**.05**	**.33**
12	.50	.00	.00	.40	.30	.07	.12	.20	.42	.05	.04	.48
13	.50	.00	.00	.40	.38	.02	.03	.48	.42	.05	.04	.48
14	.50	.00	.00	.40	.42	.01	.02	.63	.42	.05	.04	.48
15	.50	.00	.00	.40	.39	.01	.01	.54	.42	.05	.04	.48

IC = independent cluster. The boldfaced numbers are the loadings for the three items with cross-loadings in the population.

would always yield the correct comparison matrix (given enough data) and there would be no need for target bifactor rotations.

In the top portion of Table 2 in the middle set of columns we specified an extreme condition where loadings on the general are all .40 but loadings within each group factor were selected from a random uniform distribution ranging from 0.10 to 0.70. Note that this structure is perfect independent cluster (IC; McDonald, 1999; 2000), that is, there are no items with cross-loadings on the group factors (which is equivalent to saying there are no items with cross-loadings in the oblimin solution), and each orthogonal group factor has three items to identify it. This particular true loading pattern, although probably unrealistic in practice, was selected because of its clear violations of proportionality.

After converting these loadings into a correlation matrix, in the bottom portion of Table 2 in the middle set of columns are shown the SL loadings for this independent clusters data. Consider first group factor one (Items 1 through 5) where for three of the five items, the true group factor loadings are higher than the true general factor loadings, and for the remaining two items, the group factor loadings are way below the loadings for the general. For the latter two items, in the SL the loadings on the general factor are underestimated, and the loadings on the group factor are overestimated. In addition, small positive loadings occur for these items on Group Factors 2 and 3. For the items in Group Factor 2, Items 6 and 7 have overestimated general factor loadings and underestimated group factor loadings. The reverse occurs for Items 8 through 10.

Such findings are easily anticipated from inspection of the original communalities and inspection of the loadings in an oblique rotated factor pattern. However, the important lesson is that although the SL can reasonably reproduce the correct pattern of trivial and nontrivial loadings even under these extreme conditions, the specific factor loadings in the SL are not proper estimates of their corresponding population values. This result can be entirely attributed to the proportionality constraints, and implies that the SL loadings should not, generally speaking, be used as a comparison matrix.

Finally, in the third set of columns we have reproduced the population loading pattern from the first demonstration but added large (.50) cross-loadings for Items 1, 6, and 11. The data are no longer independent cluster but rather have an independent cluster basis (ICB; each group factor is identified by three items with simple loadings, but there are also items with cross-loadings). In the bottom portion we see the effects of cross-loadings on the SL results. Specifically, a cross-loading raises an item's communality and thus causes an item's loading on the general factor to increase. This overestimation is proportional to the communality of the item. In turn, loadings on the group factors decrease relative to their true population values. The critical lesson learned from this demonstration is that in the SL, one must proceed very cautiously in interpreting the size of the loadings, and they should not be taken seriously as estimates of population parameters.

The above demonstrations do not necessarily spoil our ability to use the SL to identify an appropriate target matrix. In specifying a target pattern a researcher does not have to get the loadings right but only the pattern of trivial and nontrivial loadings. However, the fact that loadings in the SL are not good estimates of their population values must be considered in judging how to specify an initial target in real data situations. A topic we now address.

As noted, we are interested in identifying any group factor loading that is greater than or equal to .30 in the population (we always assume items on the general meet this criterion). To accomplish this, we use the SL results to build a target pattern by (a) specifying that all items are unspecified (?) on the general factor (if preliminary analyses suggest otherwise, we delete the item) and on the group factors and (b) if the SL loading is greater than or equal to .15 the element is unspecified, otherwise it is specified. The reason we selected this liberal criterion is that we are most concerned with not missing any cross-loadings if they exist in the population; if an item has a sizable loading on the general and group factor, as well as a cross-loading, its communality will tend to be high and its group factor loadings underestimated in the SL. To allow for this potential underestimation we selected the liberal .15 criterion.

This decision rule begs the following question: How well does it work in practice? Even without data, it is obvious that the largest problems will occur with loadings that are around .30 in the population. Depending on sample size, and the distortions of the SL, such values may be underestimated in real data and thus incorrectly specified. The other, less serious problem is that of loadings that are truly zero in the population, but overestimated in the SL using real data. This would cause the error of not specifying an element when it truly should be specified. Because of limited research on the effects of incorrect targets in target rotations, it is unclear which of these potential errors is more serious. Given the self-correcting nature of target rotations mentioned previously, neither may be serious if a researcher allows an iterative method of defining a target rotation. Such iteration strategies are beyond the present scope, however.

In what follows, we demonstrate the fitting of targeted rotations to multidimensional data in order to evaluate unidimensional and restricted bifactor IRT models. First, we evaluate the ability of the SL to recover a good target matrix, and second, we evaluate the target rotations themselves (assuming a correct target). In both cases, we specify a true population factor loading matrix, and then we convert that matrix into a true population tetrachoric matrix Σ using the relation given in Equation (13):

$$\Sigma = \Lambda\Phi\Lambda' + \Psi. \tag{13}$$

This tetrachoric matrix is then used to generate dichotomous data. A sample tetrachoric matrix is estimated from the simulated dichotomous data and submitted to the analyses detailed in later sections. In all analyses, we specified a structure with 15 items, and OFA threshold (or IRT intercept) parameters are fixed to zero for all items. This was done to control for the possibility that the estimation of loadings or slopes is affected by the value of the thresholds or intercepts, respectively.

Simulations

To explore the efficacy of the SL to suggest an initial target, we conducted two sets of Monte Carlo analyses. In the first, we specified true population bifactor structures that were independent cluster—each item loads on the general and one and only one group factor. For all analyses, we specified 15 items with 5 items loading on each of three group factors. We defined three levels of loadings: low = .3 to .5, medium = .4 to .6, and high = .5 to .7. The design was then completely crossed with general and group factor loadings randomly selected from a uniform distribution with boundaries set to either low, medium or high values (defined above). For example, conditions ran from general factor loadings high and group factor loadings low, to, general factor loadings low and group factor loadings high. Note that because we knew the true population pattern matrix, we automatically knew the correct target matrix. Note also that the true loadings specified in the leftmost columns of Tables 3 and 4 are means of the uniform distribution from which the simulated loadings were sampled. For example, loadings sampled from a uniform distribution ranging from 0.4 to 0.6 (''medium,'' as defined above) would be indicated in the tables as 0.5 (the mean of said distribution).

Based on the true loading matrices simulated above, we computed the implied population correlation matrix. This process was then repeated 1,000 times for three sample size conditions of 250, 500, and 1,000. For each simulated data set, we then estimated the tetrachoric correlation matrix, and conducted a SL orthogonalization extracting three group factors using unweighted least squares and rotating to an oblimin solution. For each SL result, we specified a target matrix using the .15 criterion mentioned above. The outcome of interest is simply the number of times each element of the target matrix was correctly specified.

There are two types of mistakes possible: (a) a loading that is zero in the population is estimated above .15 and is thus left erroneously unspecified (?) and (b) a loading that is .30 and above in the population is underestimated in the data and gets erroneously specified (0). The results under IC conditions are very easy to summarize. When the data are IC cluster and sample size is 500 and above, errors in specifying the cells of the target matrix occur less than 1% of the time. Although we simulated a total of six conditions (strong general, strong groups; strong general, medium groups, etc.), each with sample sizes of 250, 500, and 1,000, we show only the four most extreme conditions. Table 3 shows the percentages of times errors occurred in each cell of the target matrix under conditions of a strong general trait with strong group factors (top portion) and a weak general trait with weak group factors (bottom portion). Note that the proportion errors for the general factor are not shown, because those are always unspecified (correctly) in the comparison modeling procedure. As expected, strong general and group factor loadings are almost always discovered and specified correctly (for the target) by the SL, even in small sample sizes. In sample sizes above 500, it is fair to say the SL will always recover a correctly specified target in this strong general/strong groups condition.

The condition involving a weak general factor combined with weak group factors (bottom portion of Table 3) is far more likely to cause problems for the SL when trying to recover the true target. This is because low loadings are more vulnerable to underestimation below 0.15, especially in small samples. Estimation below this threshold results in the target element erroneously being specified (0) rather than unspecified

Table 3. Effects of Sample Size on the Ability of the Schmid–Leiman to Specify a Correct Target

Demonstration A: Strong general, strong groups

True loadings				Frequency of incorrect specification (%)								
				$N = 250$			$N = 500$			$N = 1,000$		
Gen	G1	G2	G3	G1	G2	G3	G1	G2	G3	G1	G2	G3
0.60	0.60			0	3	1	0	0	0	0	0	0
0.60	0.60			0	0	2	0	0	0	0	0	0
0.60	0.60			0	1	2	0	0	0	0	0	0
0.60	0.60			0	1	1	0	0	0	0	0	0
0.60	0.60			0	3	2	0	0	0	0	0	0
0.60		0.60		2	0	1	0	0	1	0	0	0
0.60		0.60		1	0	2	0	0	0	0	0	0
0.60		0.60		1	0	0	0	0	0	0	0	0
0.60		0.60		3	0	3	0	0	0	0	0	0
0.60		0.60		1	0	2	0	0	0	0	0	0
0.60			0.60	2	0	0	1	0	0	0	0	0
0.60			0.60	0	3	0	0	0	0	0	0	0
0.60			0.60	1	1	0	0	0	0	0	0	0
0.60			0.60	1	1	0	0	0	0	0	0	0
0.60			0.60	4	4	0	0	0	0	0	0	0

Demonstration B: Weak general, weak groups

True loadings				Frequency of incorrect specification (%)								
				$N = 250$			$N = 500$			$N = 1,000$		
Gen	G1	G2	G3	G1	G2	G3	G1	G2	G3	G1	G2	G3
0.40	0.40			3	17	15	0	4	2	0	0	0
0.40	0.40			1	16	18	1	4	1	0	1	1
0.40	0.40			2	24	14	2	5	3	0	0	0
0.40	0.40			2	22	13	1	3	2	0	0	2
0.40	0.40			1	18	17	2	10	6	0	3	0
0.40		0.40		25	6	15	6	0	5	0	0	1
0.40		0.40		16	4	19	2	0	3	0	0	0
0.40		0.40		20	5	17	4	1	6	1	0	2
0.40		0.40		19	2	9	4	0	6	1	0	0
0.40		0.40		14	2	21	6	0	9	0	0	0
0.40			0.40	23	12	1	5	5	0	0	0	0
0.40			0.40	23	10	1	3	6	0	0	3	0
0.40			0.40	19	15	0	2	4	0	2	1	0
0.40			0.40	13	15	2	7	5	1	0	1	0
0.40			0.40	20	24	3	6	1	0	1	0	0

(?). Even given this problematic condition of weak general and group factors, however, sufficient sample size (500-1,000) can correct the problem almost entirely.

Table 4 shows the equivalent results for conditions in which the general factor is strong and the group factors are weak (top portion) and vice versa (bottom portion). Results in the top portion of Table 4 suggest that the existence of a strong general factor improves the SL's ability to recover the true target, but perhaps not enough to justify sample sizes below 500. Whereas a condition involving strong general *and* group factors (top portion of Table 3) appears robust to small samples, the same cannot necessarily be said of the condition shown in the top portion of Table 4. The bottom portion of Table 4 contains the results when the top portion is reversed. Interestingly, the existence of strong group factors (and a weak general factor) is even more conducive to correct target specification by the SL than the reverse condition (strong general, weak groups). Indeed, given strong group factors, one can expect the SL to specify a correct target 90% of the time with a sample size of 250. Conditions involving medium general and group factor strengths are not shown, because these can be interpolated from the four extreme conditions in Tables 3 and 4.

Target Bifactor Rotations: How Accurate?

Given the results above, we now turn to consideration of the accuracy of targeted rotations under different conditions, assuming that the target matrix is correctly specified. As noted, rotations to a target structure are not new (e.g., Tucker, 1940), but the rotation of a factor pattern to a partially specified target matrix (Browne, 1972a, 1972b, 2001) is only recently gaining increased attention due to the availability of software packages to implement targeted and other types of nonstandard rotation methods (e.g., Mplus—Asparouhov & Muthén, 2008; comprehensive exploratory factor analysis [CEFA]—Browne, Cudeck, Tateneni, & Mels, 2004). The Mplus program allows the user to conduct Monte Carlo simulations by specifying a true population loading matrix, and target pattern matrix of specified (0) and unspecified (?) elements. The weighted least squares estimator was used.

''This target matrix reflects partial knowledge as to what the factor pattern should be'' (Browne, 2001, p. 124). The target matrix forms the basis for a rotation that minimizes the sum of squared differences between the target and the rotated factor pattern. It is important to recognize that a specified element of a target pattern matrix is not the same as a fixed element in structural equation modeling. A fixed element in structural equation modeling is not allowed to vary at all from its specified value. If the said value is unreasonable, it will manifest in poor fit statistics (comparative fit index, root mean square error of approximation, etc.). In targeted rotations, however, a ''fixed/specified'' element is free to vary from its ''specified'' value if a different value will lead to better *overall* fit of the target to the estimated loading matrix. As mentioned above, in targeted rotation, ''fit'' is measured by the sum of squared deviations of the pattern matrix from the estimated matrix.

Above, we focused on the first relevant question in evaluating the effectiveness of the comparison modeling approach: How often does the SL orthogonalization succeed

Table 4. Effects of Sample Size on the Ability of the Schmid–Leiman to Specify a Correct Target

Demonstration C: Strong general, weak groups

True loadings				Frequency of incorrect specification (%)								
				$N = 250$			$N = 500$			$N = 1{,}000$		
Gen	G1	G2	G3	G1	G2	G3	G1	G2	G3	G1	G2	G3
0.60	0.40			0	13	7	0	4	4	0	0	0
0.60	0.40			2	13	20	0	4	2	0	0	0
0.60	0.40			0	13	9	0	4	4	0	0	0
0.60	0.40			0	15	9	0	2	0	0	0	0
0.60	0.40			0	4	11	0	2	4	0	0	0
0.60		0.40		7	0	13	0	0	4	0	0	0
0.60		0.40		20	0	7	0	0	2	0	0	0
0.60		0.40		15	0	7	4	0	2	0	0	0
0.60		0.40		17	0	11	2	0	4	0	0	2
0.60		0.40		11	2	7	2	0	7	0	0	0
0.60			0.40	13	15	0	2	2	0	2	0	0
0.60			0.40	17	13	0	2	0	0	0	0	0
0.60			0.40	9	22	0	0	2	0	0	0	0
0.60			0.40	22	15	0	2	2	0	0	0	0
0.60			0.40	13	13	0	0	2	0	0	0	0

Demonstration D: Weak general, strong groups

True loadings				Frequency of incorrect specification (%)								
				$N = 250$			$N = 500$			$N = 1{,}000$		
Gen	G1	G2	G3	G1	G2	G3	G1	G2	G3	G1	G2	G3
0.40	0.60			2	4	4	0	0	0	0	0	0
0.40	0.60			2	4	7	0	0	0	0	0	0
0.40	0.60			2	2	2	0	0	0	0	0	0
0.40	0.60			2	4	9	0	0	0	0	1	0
0.40	0.60			2	9	2	0	0	2	0	0	0
0.40		0.60		4	0	2	0	0	0	0	0	0
0.40		0.60		2	0	7	0	0	2	0	0	0
0.40		0.60		7	0	4	0	0	0	0	0	0
0.40		0.60		2	0	2	0	0	2	0	0	0
0.40		0.60		2	0	7	2	0	0	0	0	0
0.40			0.60	4	4	0	0	0	0	0	0	0
0.40			0.60	2	4	0	0	0	0	1	0	0
0.40			0.60	4	7	0	0	0	0	0	0	0
0.40			0.60	2	9	0	0	5	0	0	0	0
0.40			0.60	2	11	0	0	0	0	0	0	0

in pinpointing a good target matrix? We now turn to the second question: Assuming the SL pinpoints a good target, how effective is the target rotation itself in recovering the population parameters?

Whereas population parameters in the above simulations were sampled from a uniform distribution, population parameters in the present simulations were fixed at a single number. This was necessary to compare true and estimated parameters on a continuous scale, whereas in the previous simulations, we were interested only in counts of success/failure. The fixed population parameters used in the present simulations were specified simply as the means of the analogous uniform distribution from the previous simulations. For example, loadings drawn from a 0.3-0.5 uniform distribution in the above simulations are defined as 0.4 presently.

The ''Monte Carlo'' procedure in Mplus requires typical inputs: number of simulated samples, number of persons per sample, population parameters from which to generate observations, and the desired analysis to be run on said observations. Population parameters were specified as variations on bifactor independent cluster structure directly analogous to the structures simulated in the above Monte Carlo analyses. Weak/moderate/strong group and general factors were completely crossed with three sample sizes (250, 500, and 1,000). Additionally, we simulated some extreme conditions, which are indicated below. The target matrix in each condition was specified perfectly—that is, if a particular loading is zero in the population, it was specified as zero. Otherwise, Mplus was told to estimate it.

Table 5 (Demonstration A) presents true and estimated loadings when the general and group factors are equal and of moderate strength. To save space, only sample sizes of 250 and 1,000 are shown. At a sample size of 250, loadings were recovered reasonably well, with an average absolute error of 0.020. Exceptions to this include Item 4's loading on the first group factor, with an error of 0.08, and an apparent, small positive bias on the group factors. The larger sample size (1,000) offered surprisingly little improvement. Average absolute error (AAE) decreased to 0.014, an improvement of only 0.006. Additionally, the group factors remained positively biased, and the worst error (Item 6, Group 3) was not much better than when $N = 250$.

Demonstration B (bottom portion of Table 5) presents true and estimated loadings when the general factor is strong but the group factors are weak. An immediately observable phenomenon is that the general factor (which is stronger than in Demonstration A) is estimated more accurately, with AAEs of 0.009 ($N = 250$) and 0.004 ($N = 1{,}000$). This improved accuracy over Demonstration A extends to the group factors as well. Even though Item 14's loading on Group 3 is off by 0.10 when $N = 250$, the AAE was lower on the group factors in Demonstration A (despite their being weaker).

Table 6 contains two other independent cluster examples: a weak general factor with strong group factors (Demonstration C), and a weak general factor with unbalanced group factors (Demonstration D). Demonstration C is analogous to Demonstration B (Table 5), only reversed, such that the group factors actually account for more item variance than does the general factor. Interestingly, even though the communalities do not differ between Demonstrations B and C, the latter proved more challenging for the targeted rotation to estimate. At sample sizes of 250 and 1,000, the AAE

Table 5. True and Estimated Loadings by Monte Carlo Simulation in Mplus

Demonstration A: Simple structure (moderate general)[a]

	Population (true)				Estimated, $N = 250$				Estimated, $N = 1{,}000$			
Item	Gen	G1	G2	G3	Gen	G1	G2	G3	Gen	G1	G2	G3
1	0.50	0.50			0.48	0.51	0.01	0.02	0.50	0.49	−0.01	−0.01
2	0.50	0.50			0.47	0.51	0.04	0.02	0.46	0.52	0.02	0.02
3	0.50	0.50			0.50	0.49	-0.01	0.00	0.48	0.54	0.00	0.01
4	0.50	0.50			0.47	0.58	0.02	0.02	0.48	0.51	0.01	0.02
5	0.50	0.50			0.49	0.50	0.01	0.02	0.47	0.55	0.01	0.02
6	0.50		0.50		0.48	0.02	0.50	0.03	0.50	0.01	0.51	0.01
7	0.50		0.50		0.49	0.01	0.51	0.00	0.52	0.00	0.47	0.00
8	0.50		0.50		0.47	0.02	0.53	0.02	0.52	−0.01	0.51	0.00
9	0.50		0.50		0.48	0.02	0.52	0.02	0.51	0.01	0.50	0.01
10	0.50		0.50		0.50	0.01	0.48	0.01	0.50	0.02	0.51	0.01
11	0.50			0.50	0.46	0.02	0.01	0.53	0.48	0.01	0.02	0.50
12	0.50			0.50	0.45	0.02	0.03	0.55	0.48	0.02	0.01	0.56
13	0.50			0.50	0.47	0.03	0.04	0.54	0.49	0.01	0.00	0.52
14	0.50			0.50	0.50	0.01	−0.01	0.49	0.48	0.01	0.01	0.52
15	0.50			0.50	0.47	0.01	0.02	0.52	0.49	0.01	0.01	0.50

Demonstration B: Simple structure (Strong general, weak groups)[a]

	Population (true)				Estimated, $N = 250$				Estimated, $N = 1{,}000$			
Item	Gen	G1	G2	G3	Gen	G1	G2	G3	Gen	G1	G2	G3
1	0.60	0.40			0.60	0.43	0.02	0.02	0.60	0.39	0.00	−0.01
2	0.60	0.40			0.60	0.40	0.01	−0.01	0.60	0.39	0.00	0.00
3	0.60	0.40			0.60	0.40	0.01	0.02	0.59	0.41	0.02	0.01
4	0.60	0.40			0.59	0.46	0.00	0.01	0.59	0.44	0.01	0.01
5	0.60	0.40			0.60	0.37	0.00	0.01	0.59	0.42	0.00	−0.01
6	0.60		0.40		0.59	0.02	0.40	0.01	0.60	0.00	0.39	−0.01
7	0.60		0.40		0.59	−0.01	0.41	0.00	0.59	0.01	0.40	0.01
8	0.60		0.40		0.59	0.00	0.48	0.02	0.60	0.01	0.40	0.02
9	0.60		0.40		0.59	0.01	0.42	0.01	0.59	0.01	0.41	0.01
10	0.60		0.40		0.59	0.02	0.42	0.02	0.60	0.02	0.41	0.00
11	0.60			0.40	0.59	0.00	0.00	0.40	0.60	0.01	0.01	0.43
12	0.60			0.40	0.58	0.02	0.02	0.44	0.60	0.00	0.00	0.40
13	0.60			0.40	0.60	−0.02	0.01	0.39	0.60	0.01	0.00	0.40
14	0.60			0.40	0.58	0.03	0.02	0.50	0.60	0.00	0.01	0.39
15	0.60			0.40	0.59	0.01	0.00	0.44	0.60	−0.01	0.01	0.42

a. Although 1,000 replications were specified in Mplus, not all replications converged. Thus, the actual number of replications is less than 1,000.

was 0.027 and 0.017, respectively (compared with 0.009 and 0.004 in Demonstration B). The source of this larger error in Demonstration C is due mainly to consistent underestimation of the general factor loadings. Whereas the strong general factor in

Table 6. True and Estimated Loadings by Monte Carlo Simulation in Mplus

Demonstration C: Simple structure (weak general, strong groups)[a]

	Population (true)				Estimated, $N = 250$				Estimated, $N = 1{,}000$			
Item	Gen	G1	G2	G3	Gen	G1	G2	G3	Gen	G1	G2	G3
1	0.40	0.60			0.40	0.60	0.01	0.00	0.36	0.64	0.02	0.02
2	0.40	0.60			0.36	0.61	0.04	0.01	0.38	0.59	0.00	0.01
3	0.40	0.60			0.37	0.63	0.02	0.02	0.37	0.64	0.01	0.01
4	0.40	0.60			0.36	0.64	0.02	0.03	0.36	0.60	0.02	0.02
5	0.40	0.60			0.38	0.58	0.03	0.03	0.38	0.66	0.01	0.01
6	0.40		0.60		0.36	0.02	0.65	0.03	0.39	0.01	0.64	0.01
7	0.40		0.60		0.34	0.02	0.60	0.02	0.39	0.01	0.58	0.01
8	0.40		0.60		0.35	0.02	0.61	0.03	0.39	0.01	0.64	0.02
9	0.40		0.60		0.36	0.02	0.60	0.02	0.40	0.01	0.58	0.01
10	0.40		0.60		0.34	0.04	0.62	0.03	0.39	0.02	0.58	0.01
11	0.40			0.60	0.36	0.02	0.02	0.61	0.40	0.01	0.00	0.56
12	0.40			0.60	0.37	0.01	0.02	0.62	0.39	0.01	0.01	0.59
13	0.40			0.60	0.37	0.02	0.03	0.60	0.38	0.02	0.02	0.60
14	0.40			0.60	0.35	0.03	0.02	0.65	0.38	0.02	0.02	0.63
15	0.40			0.60	0.35	0.02	0.03	0.62	0.40	0.00	0.01	0.63

Demonstration D: Simple structure (moderate general, unbalanced groups)[a]

	Population (true)				Estimated, $N = 250$				Estimated, $N = 1{,}000$			
Item	Gen	G1	G2	G3	Gen	G1	G2	G3	Gen	G1	G2	G3
1	0.40	0.60			0.44	0.55	−0.02	−0.01	0.38	0.60	0.00	0.00
2	0.40	0.60			0.36	0.62	0.03	0.03	0.35	0.62	0.02	0.02
3	0.40	0.60			0.37	0.62	0.02	0.03	0.38	0.61	0.00	0.00
4	0.40	0.60			0.36	0.71	0.02	0.03	0.36	0.62	0.02	0.02
5	0.40	0.60			0.39	0.57	0.02	0.02	0.35	0.64	0.02	0.02
6	0.40		0.50		0.39	0.01	0.50	0.03	0.39	0.01	0.49	0.01
7	0.40		0.50		0.37	0.01	0.53	0.02	0.41	0.01	0.53	0.00
8	0.40		0.50		0.35	0.02	0.51	0.03	0.40	0.01	0.47	0.02
9	0.40		0.50		0.36	0.03	0.51	0.03	0.40	0.01	0.51	0.02
10	0.40		0.50		0.36	0.02	0.52	0.02	0.42	0.01	0.50	−0.01
11	0.40			0.40	0.36	0.01	0.03	0.40	0.38	0.02	0.02	0.46
12	0.40			0.40	0.34	0.02	0.03	0.47	0.39	0.01	0.00	0.40
13	0.40			0.40	0.33	0.03	0.06	0.47	0.38	0.03	0.02	0.46
14	0.40			0.40	0.34	0.04	0.02	0.52	0.40	0.01	0.00	0.42
15	0.40			0.40	0.34	0.03	0.03	0.43	0.38	0.01	0.02	0.38

a. Although 1,000 replications were specified in Mplus, not all replications converged. Thus, the actual number of replications is less than 1,000.

Demonstration B was estimated almost perfectly by the target rotation, the weaker general factor in Demonstration C appears to be vulnerable to slight underestimation. As sample size increases, however, this underestimation approaches zero.

Demonstration D (also in Table 6) is the most complicated of the independent cluster examples shown presently, because the general factor is weak and the group factors vary from explaining 36% of item variance (G1) to only 16% (G3). It is also the most extreme example of sample size effects, with AAE dropping from 0.033 ($N = 250$) to 0.016 ($N = 1{,}000$). This is due mainly to an apparent vulnerability of the small sample size to produce a few very large errors. Item 14's loading on G3 and Item 4's loading on G1 are both misestimated by more than 0.10. Further simulations (not shown) confirmed that errors of this size are indeed common with this factor structure at small sample sizes. A second noticeable phenomenon in Demonstration D is a pattern of increasing underestimation of the general factor as the group factors decrease in size, with general factor loadings underestimated (on average) by 0.02 on G1, 0.04 on G2, and 0.06 on G3. This would indicate a misalignment of the general latent vector, obviously caused by the imbalanced group factors.

Table 7 displays the true and estimated factor loadings for more problematic conditions involving violations of independent cluster structure. In Demonstration E, the general and group factor loadings are moderate and equal, but Item 3 cross-loads on the third group factor. Cross-loadings present problems for a few reasons, but one of the most important is that they indicate the existence of factors unaccounted for by the hypothesized structure. Discovery of cross-loadings is crucial to proper item parameter estimation, and should lead researchers to consider extracting additional factors or to remove the problematic item(s) altogether.

As in Demonstrations A to C, the effect of sample size in Demonstration E (Table 7) appears negligible, decreasing average absolute error from 0.024 ($N = 250$) to 0.021 ($N = 1{,}000$). Additionally, even though we assume here that the target matrix is perfect, the targeted rotation still does not estimate the cross-loading (0.50, Item 3) perfectly. Specifically, it tends to overestimate the general factor loading and underestimate both the group loading and the cross-loading. These over- and underestimations, which average 0.09, are large enough to cause problems for item parameter estimation. Nonetheless, the SL (not shown) performed even worse, over- and underestimating the cross-loadings almost twice as badly. This lends support to our argument that performing a target rotation after performing an SL will lead to more accurate loading estimation, even though the target is not perfect. Furthermore, with the exception of Item 3, the target rotation recovered the true factor loadings with accuracy comparable with Demonstrations A through D.

Finally, the population factor structure in Demonstration F (also in Table 7) is bordering on incoherent. Given the severity of independent clusters (three unbalanced sets of two cross-loadings each), a researcher would be unlikely to overlook such a problem, and would probably deem the hypothesized factor structure unreasonable. We include Demonstration F here to test the limits of our proposed method. As expected, the average absolute error was greater for Demonstration F than for any of the previous demonstrations. The effect of sample size is also large, with AAE decreasing from 0.039 ($N = 250$) to 0.027 ($N = 1{,}000$). Remarkably, however, given sufficient sample size (1,000), the *largest* absolute error in Demonstration F is only 0.07. Indeed, whereas the SL orthogonalization failed miserably with such a complex

Table 7. True and Estimated Loadings by Monte Carlo Simulation in Mplus

Demonstration E: IC basis, moderate general, balanced groups, one cross-loading[a]

	Population (true)				Estimated, $N = 250$				Estimated, $N = 1{,}000$			
Item	Gen	G1	G2	G3	Gen	G1	G2	G3	Gen	G1	G2	G3
1	0.50	0.50			0.49	0.53	0.03	0.01	0.49	0.52	0.01	0.00
2	0.50	0.50			0.47	0.54	0.04	0.02	0.47	0.51	0.03	0.02
3	0.50	0.50		0.50	0.58	0.40	−0.06	0.42	0.60	0.43	−0.08	0.37
4	0.50	0.50			0.46	0.54	0.03	0.04	0.47	0.51	0.03	0.02
5	0.50	0.50			0.47	0.51	0.04	0.02	0.47	0.55	0.02	0.01
6	0.50		0.50		0.48	0.01	0.51	0.00	0.47	0.01	0.57	0.02
7	0.50		0.50		0.47	0.00	0.51	0.00	0.49	0.00	0.51	−0.01
8	0.50		0.50		0.46	0.03	0.54	0.02	0.49	0.00	0.50	0.01
9	0.50		0.50		0.46	0.02	0.53	0.04	0.48	0.01	0.50	0.01
10	0.50		0.50		0.49	0.02	0.48	0.01	0.49	0.02	0.50	0.00
11	0.50			0.50	0.47	0.03	0.05	0.48	0.49	0.01	0.02	0.51
12	0.50			0.50	0.50	0.00	0.01	0.50	0.49	0.01	0.02	0.54
13	0.50			0.50	0.52	0.01	0.01	0.48	0.50	0.01	0.02	0.48
14	0.50			0.50	0.49	0.03	0.02	0.52	0.50	0.00	0.01	0.54
15	0.50			0.50	0.50	0.01	0.01	0.50	0.50	0.01	0.02	0.49

Demonstration F: IC basis, weak general, unbalanced groups, six unbalanced cross-loadings[a]

	Population (true)				Estimated, $N = 250$				Estimated, $N = 1{,}000$			
Item	Gen	G1	G2	G3	Gen	G1	G2	G3	Gen	G1	G2	G3
1	0.40	0.60		0.20	0.45	0.55	0.01	0.14	0.42	0.59	0.00	0.16
2	0.40	0.60		0.20	0.40	0.56	0.02	0.17	0.40	0.60	0.01	0.18
3	0.40	0.60			0.36	0.64	0.03	0.02	0.37	0.62	0.03	0.00
4	0.40	0.60			0.35	0.66	0.02	0.00	0.35	0.63	0.03	0.01
5	0.40	0.60			0.39	0.58	0.02	0.02	0.36	0.61	0.02	0.01
6	0.40		0.50		0.32	0.04	0.58	0.02	0.33	0.03	0.53	0.02
7	0.40		0.50		0.34	0.01	0.53	0.02	0.35	0.02	0.51	0.02
8	0.40		0.50	0.40	0.43	−0.02	0.45	0.35	0.44	−0.01	0.48	0.35
9	0.40		0.50	0.40	0.43	0.00	0.46	0.33	0.45	−0.01	0.48	0.33
10	0.40		0.50		0.30	0.05	0.63	0.03	0.36	0.02	0.56	0.00
11	0.40			0.40	0.36	0.02	0.03	0.41	0.40	0.01	0.03	0.37
12	0.40			0.40	0.35	0.03	0.04	0.45	0.41	0.01	0.01	0.38
13	0.40			0.40	0.37	0.03	0.04	0.41	0.40	0.01	0.02	0.38
14	0.40	0.60		0.40	0.45	0.55	−0.02	0.32	0.46	0.57	−0.02	0.33
15	0.40	0.60		0.40	0.46	0.57	−0.03	0.32	0.46	0.56	−0.02	0.33

IC = independent cluster. a. Although 1,000 replications were specified in Mplus, not all replications converged. Thus, the actual number of replications is less than 1,000.

structure (not shown), the target rotation recovered 90% of the loadings within 0.06 of their true values.

Discussion

Parameters can only be estimated correctly to the degree that the data are appropriate for a specific model. In an IRT context, when considering a highly restrictive unidimensional model, item response data must have a structure that is consistent with this restriction to accurately estimate population parameters. From an applied standpoint, accurate item parameter estimation is critical because it underlies all applications of IRT models, including (a) scaling individual differences, (b) performing scale linking, (c) developing and administering computerized adaptive tests, (d) analyzing the psychometric functioning of items and scales, and (e) identifying differential item functioning.

Nevertheless, it is generally recognized that even well constructed psychological measures are unlikely to produce item response data that are perfectly consistent with the restrictions of a unidimensional model. In fact, it is arguable that many psychological measures produce item response matrices that are consistent with both a single general common factor (i.e., ''strong'' common factor) and multidimensional models caused by the presence of two or more content parcels. Precisely because the assessment of ''unidimensionality'' is so challenging, applied researchers rely heavily on rule-of-thumb guidelines to judge when a restricted model may be safely applied to a given dataset.

In this report, we reviewed and provided further empirical support for a proposed ''comparison modeling'' approach to evaluating the viability of particular restricted IRT model applications (Reise et al., 2010). Specifically, in comparison modeling, a researcher identifies an exploratory multidimensional factor rotation that is assumed to better represent the true structure of the data. In turn, a researcher then compares the item slopes with those estimated under a more restricted model. As stated by Browne (2001, p. 113) in the context of structural equation modeling, ''The discovery of misspecified loadings, however, is more direct through rotation of the factor matrix than through the examination of model modification indices.'' We believe that a similar logic applies to IRT model fitting as well. That is, we believe that it is better to thoroughly understand the multidimensional structure of the data, and to identify potential modeling problems, prior to fitting an IRT model as opposed to either relying on a rule of thumb (e.g., ratio of first to second eigenvalue greater than 3), or fitting an IRT model and then searching for inadequacies (e.g., inspection of correlated residuals or fit indices).

As illustrated by the present research, at their best (an adequate sample size and a multidimensional structure that is independent cluster), an SL orthogonalization can yield an adequate initial target pattern. In turn, assuming the target is correct, a target rotation can provide, in either a factor analytic or IRT metric, an accurate unrestricted bifactor comparison model. Finally, this bifactor comparison model can provide a direct index of the degree of item parameter bias caused by forcing multidimensional data into unidimensional models. Of course, we note that no targeted rotation procedure can be expected to work under all the possible conditions that confront a researcher. Moreover, the application of targeted rotations is much more complicated when working with real data, as opposed to the hypothetical examples provided here. Thus, in the next section, we briefly consider challenges and limitations of targeted rotations and describe directions for future research.

Limitations and Future Directions

The most significant challenge of comparison IRT modeling is to identify a correct exploratory factor model to compare with a more restricted model. The ability to accomplish this goal depends on (a) the true structure of the data (e.g., the severity and form of multidimensionality); (b) the ability of the target rotation method to recover this true structure (e.g., what particular function should be minimized?, should weighted or unweighted least squares be used?), which depends partially on (c) how the researcher specifies an initial target pattern (e.g., based on theory or a priori data analysis?). We will focus the following discussion on this latter task—identifying the dimensionality of a target matrix and specifying the values of an initial target.

To specify a target matrix, the first challenge is the historically important problem of identifying the dimensionality of the target matrix. Recall that to test a restricted model (e.g., unidimensionality), the researcher needs to compare that with an exploratory multidimensional model. Drawing partly from experience and partly from the present results, we suggest that in many situations, an unweighted least squares extraction followed by an oblique rotation can provide a lot of insight into the questions of how many factors should be extracted as well as the question of how the target should be specified. As shown in the above simulations, when the data are multidimensional because of having at least three "clusters" of items, and the items have an independent cluster structure, the SL is capable of adequately recovering the configuration of meaningful loadings. However, although not explicitly illustrated by the present simulations, it is easy to demonstrate that the SL can run into serious problems when the data depart drastically from independent cluster structure (i.e., many cross-loadings in an oblique rotated solution). On the other hand, the degree to which any model should be applied to such data is highly questionable.

In closing, we note that although it is critical to correctly specify the dimensionality of the initial target matrix, it may not be so critical to correctly specify the particular pattern of specified and unspecified elements in that matrix. We have observed that if a wrong element is specified (e.g., a zero is placed for a loading that is truly nonzero), Mplus will identify this problem by estimating a nonzero rotated factor loading. In turn, the researcher may then respecify the target matrix, and reestimate the rotation. The opposite is also true in many circumstances. That is, if a researcher mistakenly specified a question mark for an element that should be zero, Mplus can identify this problem. The conditions under which a researcher can iterate to the correct target in this manner remains to be demonstrated in future studies. Finally, although not detailed here, we warn that despite the present positive results, bifactor structures are often challenging to estimate even in simulated data. The reasons are very complicated and beyond the present scope. We hope to address these issue more thoroughly in future work.

Summary

We described the use of exploratory bifactor IRT models (estimated as exploratory factor models) to judge the effects of forcing multidimensional data into a unidimensional model. We propose that the comparison modeling approach allows researchers to evaluate the item parameter bias caused by forcing data into a restricted model. At the very least, exploratory modeling forces the researcher into considering alternative multidimensional models and thus to seriously considering the structure of their data and its impact on item parameter estimates. For this reason, we suggest that research that proposes a unidimensional IRT application should be required to also report the tetrachoric or polychoric correlation matrix so that other researchers can make their own judgments.

Declaration of Conflicting Interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

This work was supported by the Consortium for Neuropsychiatric Phenomics (NIH Roadmap for Medical Research grants UL1-DE019580 (PI: Robert Bilder), and RL1DA024853 (PI: Edythe London). Additional research support was obtained through the National Institutes of Health the NIH Roadmap for Medical Research Grant (AR052177; PI: David Cella); a pre-doctoral advanced quantitative methodology training grant (#R305B080016) awarded to UCLA by the Institute of Education Sciences of the U.S. Department of Education; and, a NCI grant 4R44CA137841-03 (PI: Patrick Mair) for IRT software development for health outcomes and behavioral cancer research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

References

Ackerman, T. A. (2005). Multidimensional item response theory modeling. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics* (pp. 3-26). Mahwah, NJ: Lawrence Erlbaum.

Asparouhov, T., & Muthén, B. (2008). *Exploratory structural equation modeling* (Tech. Rep.). Retrieved from http://www.statmodel.com

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis*. London, England: Arnold.

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence of multidimensionality. *Psychological Bulletin, 105*, 467-477.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443-458.

Bock, R. D., Gibbons, R., Shilling, S. G., Muraki, E., Wilson, D. T., & Wood, R. (2002). *TEST-FACT 4*. Chicago, IL: Scientific Software.

Browne, M. W. (1972a). Orthogonal rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology, 25*, 115-120.

Browne, M. W. (1972b). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology, 25*, 207-212.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 35*, 111-150.

Browne, M. W., Cudeck, R., Tateneni, K., & Mels, G. (2004). *CEFA: Comprehensive exploratory factor analysis, Version 2.00* [Computer software and manual]. Retrieved from http://quantrm2.psy.ohio-state.edu/browne/

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40*, 5-32.

Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement, 7*, 189-199.

Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded models for rating data: Limited versus full information methods. *Psychological Methods, 14*, 275-299.

Fraser, C. (1988). NOHARM II. *A Fortran program for fitting unidimensional and multidimensional normal ogive models of latent trait theory*. Armidale, New South Wales, Australia: The University of New England, Center for Behavioral Studies.

Fraser, C., & McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research, 23*, 267-269.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika, 57*, 423-436.

Knol, D. L., & Berger, M. P. F. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research, 26*, 457-477.

Ip, E. H. (2010). Empirically indistinguishable multidimensional IRT and locally dependent unidimensional item response models. *British Journal of Mathematical and Statistical Psychology, 63*, 395-416.

Maydeu-Olivares, A. (2006). Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika, 71*, 57-77.

Maydeu-Olivares, A., & Joe, H. (2005). Limited and full information estimation and testing in 2^n contingency tables: A unified framework. *Journal of the American Statistical Association, 100*, 1009-1020.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71*, 713-732.

McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Lawrence Erlbaum.

McDonald, R. P. (2000). A basis for multidimensional item response theory. *Applied Psychological Measurement, 24*, 99-114.

McDonald, R. P., & Mok, M. M. C. (1995). Goodness of fit in item response models. *Multivariate Behavioral Research, 30*, 23-40.

McLeod, L. D., Swygert, K. A., & Thissen, D. (2001). Factor analysis for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 189-216). Mahwah, NJ: Lawrence Erlbaum.

Muthén, B. (1979). A structural probit model with latent variables. *Journal of American Statistical Association, 74*, 807-811.

Muthén, B. (1993). Goodness of fit with categorical and other nonnormal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equations models* (pp. 205-234). Newbury Park, CA: Sage.

Muthén, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46*, 407-419.

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*, 207-230.

Revelle, W. (2009). *psych: Procedures for psychological, psychometric, and personality research. R package version 1.0-68*. Retrieved from http://personality-project.org/r/psych.manual.pdf

Reise, S. P., Cook, K. F., & Moore, T. M. (2010). *A comparison modeling approach for evaluating the impact of multidimensionality on unidimensional item response theory model parameters*. Manuscript submitted for publication.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*, 53-61.

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*, 393-408.

Tucker, L. R. (1940). A rotational method based on the mean principal axis of a subgroup of tests. *Psychological Bulletin, 3*, 289-294.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58-79.

Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64*, 113-128.