**Supplementary Online Content: Technical Appendix**

# 1 Item Factor Analysis Based on Item Response Theory

IRT-based item factor analysis makes use of all information in the original categorical responses and does not depend on pairwise indices of association such as tetrachoric or polychoric correlation coefficients. For that reason it is referred to as *full information* item factor analysis. It works directly with item response models giving the probability of the observed categorical responses as a function of latent variables descriptive of the respondents and parameters descriptive of the individual items. It differs from the classical formulation in its scaling, however, because it does not assume that the response process has unit standard deviation and zero mean; rather it assumes that the *residual* term has unit standard deviation and zero mean. The latter assumption implies that the response processes have zero mean and standard deviation equal to

$$\sigma_{y_j} = \sqrt{1+\Sigma_v^d \alpha_{jv}^2}\ .$$

Inasmuch as the scale of the model affects the relative size of the factor loadings and thresholds, we rewrite the model for dichotomous responses in a form in which the factor loadings are replaced by factor slopes, $a_{jv}$, and the threshold is absorbed in the intercept, $c_j$:

$$y_j = \sum_{v=1}^{d} a_{jv}\theta_v + c_j + \varepsilon_j\ .$$

To convert factor slopes into loadings we divide by the above standard deviation and similarly convert the intercepts to thresholds:

$$\alpha_{jv} = a_{jv}/\sigma_{y_j} \quad \text{and} \quad \gamma_j = -c_j/\sigma_{y_j}\ .$$

Conversely, to convert to factor analysis units, we change the standard deviation of the residual from 1 to

$$\sigma_{\varepsilon_j}^{*} = \sqrt{1-\Sigma_v^d \alpha_{jv}^2}\ ,$$

and change the scale of the slopes and intercept accordingly:

$$a_{jv} = \alpha_{jv}/\sigma^*_{\varepsilon_j} \text{ and } c_j = -\gamma_j/\sigma^*_{\varepsilon_j} .$$

For polytomous responses, the model generalizes as:

$$z_j = \sum_{v=1}^{d} a_{jv}\theta_v ,$$

$$P_{jh}(\theta) = \Phi(z_j + c_{jh}) - \Phi(z_j + c_{j,h-1}) ,$$

where $\Phi(z_j + c_{j0}) = 0$ and $\Phi(z_j + c_{jm_j}) = 1 - \Phi(z_j + c_{j,m_j-1})$ as previously. In the context of item factor analysis, this is the multidimensional generalization of the *graded* model introduced by Samejima (1969).

## 2 Confirmatory Item Factor Analysis

In confirmatory factor analysis, indeterminacy of rotation is resolved by assigning arbitrary fixed values to certain loadings of each factor during maximum likelihood estimation. An important example of confirmatory item factor analysis is the bifactor pattern for general and group factors, which applies to tests and scales with item content drawn from several well-defined sub-areas of the domain in question. To analyze these kinds of structures for dichotomously scored item responses, Gibbons & Hedeker (1992) developed full-information item bifactor analysis for binary item responses, and Gibbons extended it to the polytomous case (Gibbons et.al., 2007). To illustrate, consider a set of $n$ test items for which a $d$-factor solution exists with one general factor and $d-1$ group or method-related factors. The bifactor solution constrains each item $j$ to a non-zero loading $\alpha_{j1}$ on the primary dimension and a second loading ($\alpha_{jv}, v = 2,\ldots,d$) on not more than one of the $d-1$ group factors. For four items, the bifactor pattern matrix might be

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & 0 \\ \alpha_{21} & \alpha_{22} & 0 \\ \alpha_{31} & 0 & \alpha_{33} \\ \alpha_{41} & 0 & \alpha_{43} \\ & & \end{bmatrix}$$

This structure, which Holzinger & Swineford (1937) termed the ``bifactor" pattern, also appears in the inter-battery factor analysis of Tucker (1958) and is one of the confirmatory factor analysis models considered by Jöreskog (1969). In the latter case, the model is restricted to test scores assumed to be continuously distributed. However, the bifactor pattern might also arise at the item level (Muthén, 1989). Gibbons & Hedeker (1992) showed that paragraph comprehension tests, where the primary dimension represents the targeted process skill and additional factors describe content area knowledge within paragraphs, were described well by the bifactor model. In this context, they showed that items were conditionally independent between paragraphs, but conditionally dependent within paragraphs.

The bifactor restriction leads to a major simplification of likelihood equations that (1) permits analysis of models with large numbers of group factors since the integration always simplifies to a two-dimensional problem, (2) permits conditional dependence among identified subsets of items, and (3) in many cases, provides more parsimonious factor solutions than an unrestricted full-information item factor analysis.

# 3 The Bifactor Model

In the bifactor case, the graded response model is

$$z_{jh}(\theta) = \sum_{v=1}^{d} a_{jv}\theta_v + c_{jh} , \tag{1}$$

where only one of the $v = 2, \ldots, d$ values of $a_{jv}$ is non-zero in addition to $a_{j1}$. Assuming independence of the $\theta$, in the unrestricted case, the multidimensional model above would require a $d$-fold integral in order to compute the unconditional probability for response pattern $\mathbf{u}$, i.e.,

$$P(\mathbf{u} = \mathbf{u}_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L_i(\theta) g(\theta_1) g(\theta_2) \ldots g(\theta_d) d\theta_1 d\theta_2 \ldots d\theta_d , \tag{2}$$

for which numerical approximation is limited as previously described. Gibbons & Hedeker (1992) showed that for the binary response model, the bifactor restriction always results in a two-dimensional integral regardless of the number of dimensions, one for $\theta_1$ and the other for $\theta_v, v > 1$. The reduction formula is due to Stuart (1958), who showed that if $n$ variables follow a standardized multivariate normal distribution where the correlation $\rho_{ij} = \sum_{v=1}^{d} \alpha_{iv}\alpha_{jv}$ and $\alpha_{iv}$ is nonzero for only one $v$, then the probability that respective variables are simultaneously less than $\gamma_j$ is given by,

$$P = \prod_{v=1}^{d} \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{n} \left[ \Phi\left( \frac{\gamma_j - \alpha_{jv}\theta}{\sqrt{1-\alpha_{jv}^2}} \right) \right]^{u_{jv}} \right\} g(\theta) d\theta, \tag{3}$$

where $\gamma_j = -c_j / y_j$, $\alpha_{jv} = a_{jv} / y_j$, $y_j = (1 + a_{j1}^2 + a_{jv}^2)^{1/2}$, $u_{jv} = 1$ denotes a nonzero loading of item $j$ on dimension $v$ ($v = 1, \ldots, d$), and $u_{jv} = 0$ otherwise. Note that for item $j$, $u_{jv} = 1$ for only one $d$. Note also that $\gamma_j$ and $\alpha_{jv}$ used by Stuart (1958) are equivalent to the item threshold and factor loading, and are related to the more traditional IRT parameterization as described above.

This result follows from the fact that if each variate is related only to a single dimension, then the $d$ dimensions are independent and the joint probability is the product of $d$ unidimensional probabilities. In this context, the result applies only to the $d-1$ content dimensions ( i.e., $v = 2, \ldots, d$ ). If a primary dimension exists, it will not be independent of the other $d-1$ dimensions, since each item now loads on each of two dimensions. Gibbons & Hedeker (1992) derived the necessary two-dimensional generalization of Stuart's (1958) original result as

$$P = \int_{-\infty}^{\infty} \left\{ \prod_{v=2}^{d} \int_{-\infty}^{\infty} \left[ \prod_{j=1}^{n} \left( \Phi \left[ \frac{\gamma_j - \alpha_{j1}\theta_1 - \alpha_{jv}\theta_v}{\sqrt{1 - \alpha_{j1}^2 - \alpha_{jv}^2}} \right] \right)^{u_{jv}} \right] g(\theta_v) d\theta_v \right\} g(\theta_1) d\theta_1 , \tag{4}$$

For the graded response model, the probability of a value less than the category threshold $\gamma_{jh} = -c_{jh} / y_j$ can be obtained by substituting $\gamma_{jh}$ for $\gamma_j$ in the previous equation. Let $\delta_{ijh} = 1$ if person $i$ responds positively to item $j$ in category $h$ and $\delta_{ijh} = 0$ otherwise. The unconditional probability of a particular response pattern $\mathbf{u}_i$ is then

$$P(\mathbf{u} = \mathbf{u}_i) = \int_{-\infty}^{\infty} \left\{ \prod_{v=2}^{d} \int_{-\infty}^{\infty} \left[ \prod_{j=1}^{n} \prod_{h=1}^{m_j} \left[ \Phi_{jh}(\theta_1, \theta_v) - \Phi_{jh-1}(\theta_1, \theta_v) \right)^{\delta_{ijh} \cdot u_{jv}} \right] g(\theta_v) d\theta_v \right\} g(\theta_1) d\theta_1 , \tag{5}$$

which can be approximated to any degree of practical accuracy using two-dimensional Gauss-Hermite quadrature, since for both the binary and graded bifactor response models, the dimensionality of the integral is 2 regardless of the number of subdomains ( i.e., $d-1$) that comprised the scale.

# 4 Parameter Estimation

Gibbons & Hedeker (1992) showed how parameters of the item bifactor model for binary responses can be estimated by maximum marginal likelihood using a variation of the EM algorithm described by Bock & Aitkin (1981). For the graded case, the likelihood equations are derived as follows.

Denoting the $v$th subset of the components of $\theta$ as $\theta_v^* = \begin{bmatrix} \theta_1 \\ \theta_v \end{bmatrix}$, let

$$
\begin{aligned}
P_i &= P(\mathbf{u}=\mathbf{u}_i) \\
&= \int_{\theta_1}\left\{\prod_{v=2}^{d}\int_{\theta_v}\left[\prod_{j=1}^{n}\prod_{h=1}^{m_j}\left(\Phi_{jh}(\theta_v^*)-\Phi_{jh-1}(\theta_v^*)\right)^{\delta_{ijh}\cdot u_{jv}}\right]g(\theta_v)d\theta_v\right\}g(\theta_1)d\theta_1 \\
&= \int_{\theta_1}\left\{\prod_{v=2}^{d}\int_{\theta_v}L_{iv}(\theta_v^*)g(\theta_v)d\theta_v\right\}g(\theta_1)d\theta_1\,, \qquad (6)
\end{aligned}
$$

where

$$
L_{iv}(\theta_v^*)=\prod_{j=1}^{n}\prod_{h=1}^{m_j}\left(\Phi_{jh}(\theta_v^*)-\Phi_{jh-1}(\theta_v^*)\right)^{\delta_{ijh}\cdot u_{jv}}\,.
$$

Then the log-likelihood is

$$
\log L=\sum_{i=1}^{s}r_i\,\log P_i, \qquad (7)
$$

where $s$ denotes number of unique response patterns, and $r_i$ the frequency of pattern $i$. As the number of items gets large, $s$ typically is the number of respondents and $r_i=1$. Complete details of the likelihood equations and their solution are provided in Gibbons et.al. (2007).

# 5 Trait Estimation

In practice, the ultimate objective is to estimate the trait level of person $i$ on the primary trait the instrument was designed to measure. For the bifactor model, the goal is to estimate the latent variable $\theta_1$ for person $i$. A good choice for this purpose (Bock & Aitkin, 1981) is the expected *a posteriori* (EAP) value (Bayes estimate) of $\theta_1$, given the observed response vector $\mathbf{u}_i$ and levels of the other subdimensions $\theta_2\ldots\theta_d$. The Bayesian estimate of $\theta_1$ for person $i$ is:

$$
\hat{\theta}_{1i}=E(\theta_{1i}\mid\mathbf{u}_i,\theta_{2i}\ldots\theta_{di})=\frac{1}{P_i}\int_{\theta_1}\theta_{1i}\left\{\prod_{v=2}^{d}\int_{\theta_v}L_{iv}(\theta_v^*)g(\theta_v)d\theta_v\right\}g(\theta_1)d\theta_1\,. \qquad (8)
$$

Similarly, the posterior variance of $\hat{\theta}_{1i}$, which may be used to express the precision of the EAP estimator, is given by

$$
V(\theta_{1i}\mid\mathbf{u}_i,\theta_{2i}\ldots\theta_{di})=\frac{1}{P_i}\int_{\theta_1}(\theta_{1i}-\hat{\theta}_{1i})^2\left\{\prod_{v=2}^{d}\int_{\theta_v}L_{iv}(\theta_v^*)g(\theta_v)d\theta_v\right\}g(\theta_1)d\theta_1\,. \qquad (9)
$$

These quantities can be evaluated using Gauss-Hermite quadrature as previously described.

In some applications, we are also interested in estimating a person's location on the secondary domains of interest as well. For the $v$th sub-domain, the EAP estimate and its variance can be written as:

$$\hat{\theta}_{vi} = E(\theta_{vi} \mid \mathbf{u}_i, \theta_{1i}) = \frac{1}{P_i} \int_{\theta_v} \theta_{vi} \left\{ \int_{\theta_1} L_{iv}(\theta_v^*) g(\theta_1) d\theta_1 \right\} g(\theta_v) d\theta_v , \tag{10}$$

and

$$V(\theta_{vi} \mid \mathbf{u}_i, \theta_{1i}) = \frac{1}{P_i} \int_{\theta_v} (\theta_{vi} - \hat{\theta}_{vi})^2 \left\{ \int_{\theta_1} L_{iv}(\theta_v^*) g(\theta_1) d\theta_1 \right\} g(\theta_v) d\theta_v . \tag{11}$$

# 6 Item Information Functions

Suppose there are $i = 1,2,\ldots N$ examinees, and $j = 1,2,\ldots n$ items. Let the probability of a response in category $h = 1,2,\ldots m_j$ to graded response item $j$ for examinee $i$ with factor $\theta$ be denoted by $P_{ijh}(\theta)$. We call $P_{ijh}(\theta)$ a category probability. $P_{ijh}(\theta)$ is given by the difference between two adjacent boundaries.

$$P_{ijh}(\theta) = P(x_{ij} = h \mid \theta) = P_{ijh}^*(\theta) - P_{ijh-1}^*(\theta) \tag{12}$$

where $P_{ijh}^*(\theta)$ is the boundary probability. The boundary probability $P_{ijh}^*$ and category probability $P_{ijh}(\theta)$ are related as:

$$\begin{aligned}
&P_{ij0}^* = 0, \\
&P_{ij1}^* = P_{ij1}(\theta), \\
&P_{ij2}^* = P_{ij1}(\theta) + P_{ij2}(\theta), \\
&\ldots \\
&P_{ijh}^* = P_{ij1}(\theta) + P_{ij2}(\theta) + \cdots + P_{ijh}(\theta), \\
&\ldots \\
&P_{ijm_j}^* = P_{ij1}(\theta) + P_{ij2}(\theta) + \cdots + P_{ijm_j-1}(\theta) + P_{ijm_j}(\theta) = 1.
\end{aligned}$$

In general,

$$P_{ijh}^* = \sum_{g=1}^{h} P_{ijg}(\theta). \tag{13}$$

Let $P_{ij}(\theta)$ be given by the following multinomial probability model:

$$P_{ij}(\theta) = P(x_{ij} \mid \theta) = \prod_{h=1}^{m_j} P_{ijh}^{1_{\{x_{ij}=h\}}} \tag{14}$$

where

$$1_{\{x_{ij}=h\}} = \begin{cases} 1 & \mathit{if}\ \ x_{ij} = h \\ 0 & \mathit{otherwise.} \end{cases}$$

For simplicity, we drop the index $i$ in the subsequent notation.

For the unidimensional model, Samejima (1969) defined the item information function (IIF) as:

$$I_j(\theta) = \sum_{h=1}^{m_j} \frac{1}{P_{jh}(\theta)} \left(\frac{\partial P_{jh}(\theta)}{\partial \theta}\right)^2$$

$$= \sum_{h=1}^{m_j} \frac{(P_{jh}^{*\prime}(\theta) - P_{jh-1}^{*\prime}(\theta))^2}{P_{jh}^{*}(\theta) - P_{jh-1}^{*}(\theta)} \tag{15}$$

The definition of information for the multidimensional case is the same as that given in the previous equation for the unidimensional case. However, in the multidimensional case, information corresponds to the composite of factors in the $\theta$ space. At each point in the $\theta$ space, the shape of multidimensional item response surface differs on the direction of the movement from the point. For the multidimensional model with a graded response item, Yao and Schwarz (2006) defined IIF as:

$$I_j(\theta) = \sum_{h=1}^{m_j} \frac{(\nabla_\alpha P_{jh}(\theta))^2}{P_{jh}(\theta))} \tag{16}$$

where

$$\theta : p\ \ x\ 1\ \mathit{vector}$$

$$\nabla_\alpha P_{jh}(\theta) = \left(\frac{\partial P_{jh}}{\partial \theta}\right)\left(\cos \alpha_j\right)$$

$$\left(\frac{\partial P_{jh}}{\partial \theta}\right)' = \left(\frac{\partial P_{jh}}{\partial \theta_1}, \ldots, \frac{\partial P_{jh}}{\partial \theta_p}\right)$$

$$\left(\cos \alpha_j\right)' = \left(\cos \alpha_{j1}, \ldots, \cos \alpha_{jp}\right).$$

$\alpha$ is the vector of angles with the coordinate axes that define the direction taken from the $\theta$ point, and $\nabla_\alpha$ is the directional derivatives in the direction $\alpha$.

In the bifactor model, the IIF expression can be written as:

$$I_j(\theta) = \sum_{h=1}^{m_j} \left( \frac{1}{P_{jh}(\theta)} \right) \left( \left(\frac{\partial P_{jh}}{\partial \theta_1}, \frac{\partial P_{jh}}{\partial \theta_2}\right) \cdot (\cos\alpha_{j1}, \cos\alpha_{j2})' \right)^2 \tag{17}$$

where $\theta_1$ and $\theta_2$ are primary and secondary factors, respectively. Under the normal ogive model, the boundary probability is given by:

$$\begin{aligned} P^*_{jh}(\theta) &= \Phi(z_{jh}) \\ &= \int_{-\infty}^{z_{jh}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \end{aligned}$$

where

$$z_{jh} = a_{j1}\theta_1 + a_{j2}\theta_2 + c_{jh}.$$

Then, the IIF has a specific form:

$$\begin{aligned} I_j(\theta) &= \sum_{h=1}^{m_j} \left( \frac{1}{P_{jh}} \right) \left[ (a_{j1}\phi(z_{jh}) - a_{j1}\phi(z_{jh-1}), a_{j2}\phi(z_{jh}) - a_{j2}\phi(z_{jh-1})) \cdot (\cos\alpha_{j1}, \cos\alpha_{j2})' \right]^2 \\ &= \left[\sum_{h=1}^{m_j} \frac{\left[\phi(z_{jh}) - \phi(z_{jh-1})\right]^2}{\Phi(z_{jh}) - \Phi(z_{jh-1})}\right] [a_{j1}\cos\alpha_{j1} + a_{j2}\cos\alpha_{j2}]^2 \end{aligned} \tag{18}$$

where

$$\phi(z_{jh}) = \frac{\partial \Phi(z_{jh})}{\partial z_{jh}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_{jh}^2}{2}}.$$

Given $\theta$ and $a_{j1}, a_{j2}$ and $c_{jh}$, the first term of the right hand side of (18)

$$\sum_{h=1}^{m_j} \frac{\left(\phi(z_{jh}) - \phi(z_{jh-1})\right)^2}{\Phi(z_{jh}) - \Phi(z_{jh-1})}$$

is constant; so at a fixed point of $\theta$, $I_j(\theta)$ depends only on the direction $\alpha$. Two special but important cases are

$$I_j(\theta) = \sum_{h=1}^{m_j} \frac{\left(a_{j1}[\phi(z_{jh}) - \phi(z_{jh-1})]\right)^2}{\Phi(z_{jh}) - \Phi(z_{jh-1})} \tag{19}$$

which is the item information associated with a change in a primary factor, but no change in a secondary factor, and

$$I_j(\theta) = \sum_{h=1}^{m_j} \frac{(a_{j2}[\phi(z_{jh}) - \phi(z_{jh-1})])^2}{\Phi(z_{jh}) - \Phi(z_{jh-1})} \tag{20}$$

which is the item information associated with a change in a secondary factor, but no change in a primary factor. It should be noted that the previous two equations depend on both the primary $\theta_1$ and $a_{j1}$, and secondary $\theta_2$ and $a_{j2}$ factors and loadings.

When we are interested in estimating the IIF for $\theta_1$ in the presence of other sub-domains, the sub-domains can be integrated out of the objective function. Suppose that for the purpose of computerized adaptive testing (CAT), $\theta_1$ is our focus; however, $\theta_2$ is also present in a bifactor model. In this case, we are interested in obtaining $I_j(\theta_1)$, which is a function only of $\theta_1$. To get $I_j(\theta_1)$, we integrate the previous bifactor IIF expression with the conditional distribution $h(\theta_2 \mid \theta_1)$ of $\theta_2$ and obtain

$$I_j(\theta_1) = \sum_{h=1}^{m_j} \int \frac{\left[\phi(z_{jh}) - \phi(z_{jh-1})\right]^2}{\Phi(z_{jh}) - \Phi(z_{jh-1})} h(\theta_2 \mid \theta_1) d\theta_2, \tag{21}$$

which provides an estimate of the information associated with $\theta_1$ averaged over the $\theta_2$ distribution. It is this expression that we have used as the basis for selecting items with maximal information in the CAT-DI.

# 7 Computerized Adaptive Testing

The bifactor model is extremely useful for CAT of multidimensional data. The conditional dependencies produced by the sub-domains can be directly incorporated in trait estimation and item information functions as shown in the previous two sections, leading to improved estimates of uncertainty and elimination of pre-mature termination of the CAT and potential bias in the estimated trait score. After each item administration, the primary ability estimate and posterior standard deviation (PSD) are re-computed, and based on the estimate of $\theta_1$, the item with maximal information is selected as the next item to be administered. This process continues until the PSD is less than a threshold value (*e.g.,* 0.3). Once the primary dimension has been estimated via CAT, sub-domain scores can also be estimated by adding items from the sub-domain that have not been previously administered in evaluating the primary domain, until the sub-domain score is estimated with similar precision.

When the trait score is at a boundary ( i.e., either the low or high extreme of the trait distribution), it may take a large number of items to reach the intended PSD (standard error) convergence criterion (e.g., se=0.3). In such extreme cases, we generally do not require such high levels of precision, since we know that the subject either does not suffer from the condition of interest or is among the most severly impaired. A simple solution to this problem is to add a second termination condition based on item information at the current estimate of the trait score and if there is less information than the threshold, the CAT terminates. The choice of the threshold is application specific and can be selected based on simulated CAT. A good value will affect only a small percentage of cases (*e.g.,* <20%) and only be used in extreme (i.e., high or low) cases.

For large item banks, there may be items that are too similar to be administered within a given session. In these cases, we can declare these as "enemy items" and not administer the other members of the list of enemy items when one of the members has been administered. The idea of enemy items can be extended to the longitudinal case to insure that the same respondent is not repeatedly administered the same items on adjacent testing sessions.

CAT will often result in a subset of the entire item bank being used exclusively, because these items have the highest loadings on primary and sub-domains. Often the difference between the loadings of items that are selected by the CAT versus those that are not, are quite small and the items have similar information. To insure that the majority of the items in the item bank are administered, we can add a probabilistic component in which a selected item is only administered if a uniform random number exceeds a threshold. Typically a threshold of 0.5 works well, but again, the exact choice can be based on simulated adaptive testing, in which the largest set of unique items are used without compromising the other characteristics of the measurement process ( i.e., average number of items administered and correlation with the total bank score).

# References

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443-459.

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

Gibbons, R.D., Bock, R.D., Hedeker, D., Weiss, D., Bhaumik, D.K., Kupfer, D., Frank, E., Grochocinski, V., & Stover, A. (2007). Full-Information Item Bifactor Analysis of Graded Response Data. *Applied Psychological Measurement*, *31*, 4-19.

Holzinger, K. J., & Swineford, F. (1937). The bifactor method. *Psychometrika, 2*, 41-54.

Joreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183-202.

Muthen, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557-585.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *17*.

A. (1958). Equally correlated variates and the multinormal integral. *Journal of the Royal Statistical Society Series B*, *20*, 373-378.

Tucker, L. R. (1958). An inter-battery method of factor analysis. *Psychometrika*, *23*, 111-136.

Yao, L., and Schwarz, R. D. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. *Applied Psychological Measurement, 30*, 469-492.