Applied Psychological Measurement

Full-Information Item Bifactor Analysis of Graded Response Data

Robert D. Gibbons, R. Darrell Bock, Donald Hedeker, David J. Weiss, Eisuke Segawa, Dulal K. Bhaumik, David J. Kupfer, Ellen Frank, Victoria J. Grochocinski and Angela Stover

The online version of this article can be found at:
http://apm.sagepub.com/cgi/content/abstract/31/1/4

Additional services and information for *Applied Psychological Measurement* can be found at:

Email Alerts: http://apm.sagepub.com/cgi/alerts

Subscriptions: http://apm.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations (this article cites 6 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
http://apm.sagepub.com/cgi/content/abstract/31/1/4#BIBL

Full-Information Item Bifactor Analysis of Graded Response Data

Robert D. Gibbons, R. Darrell Bock, and Donald Hedeker,
    University of Illinois at Chicago
David J. Weiss, University of Minnesota
Eisuke Segawa and Dulal K. Bhaumik,
    University of Illinois at Chicago
David J. Kupfer, Ellen Frank, Victoria J. Grochocinski,
    and Angela Stover, Western Psychiatric Institute

A plausible factorial structure for many types of psychological and educational tests exhibits a general factor and one or more group or method factors. This structure can be represented by a bifactor model. The bifactor structure results from the constraint that each item has a nonzero loading on the primary dimension and, at most, one of the group factors. The authors develop estimation procedures for fitting the graded response model when the data follow the bifactor structure. Using maximum marginal likelihood estimation of item parameters, the bifactor restriction leads to a major simplification of the likelihood equations and (a) permits analysis of models with large numbers of group factors, (b) permits conditional dependence within identified subsets of items, and (c) provides more parsimonious factor solutions than an unrestricted full-information item factor analysis in some cases. Analysis of data obtained from 586 chronically mentally ill patients revealed a clear bifactor structure. *Index terms: bi-factor model, maximum marginal likelihood, EM algorithm, item analysis, ordinal data, factor analysis*

Psychological scales and educational tests frequently are developed to measure a particular construct (e.g., depression or emotional well-being) by sampling items from identified subdomains of some larger domain. For example, in mental health services research, outcomes relating to a client's quality of life are the focus of many studies of health care service delivery (e.g., Kaziz et al., 1998). In constructing scales to measure quality of life, investigators (e.g., Lehman, 1988) have selected life satisfaction items from several domains, including satisfaction with work, leisure, family, finance, health, living, safety, and social aspects of life. Although the instrument is designed to measure a single overall concept (i.e., quality of life), a result of the two-stage sampling procedure (i.e., domains within a construct and items within domains) frequently produces rating scales with a multidimensional structure.

Similar item-sampling schemes are also common in the construction of educational achievement tests. If these tests contain open-ended exercises rated on a graded scale, graded response data with an underlying multidimensional pattern are the expected result. For example, the American College Testing Service (ACT) science test (ACT, 2006) involves a series of questions

regarding each of several paragraphs. Each paragraph describes a different scientific domain; therefore, an examinee familiar with that content area will be more likely to respond correctly to the items in that paragraph. This structure results in items within domains (e.g., paragraphs) being more highly related to each other than items between domains. As a result, however, the item responses are no longer independent conditional on the primary dimension that the test was designed to measure. This violation of conditional independence precludes the use of unidimensional models for item responses.

Bock and Aitkin (1981) and Bock, Gibbons, and Muraki (1988) developed full-information item factor analysis for binary responses; however, there has been little progress in multidimensional extensions of the graded response model (see Muraki & Carlson, 1995). In part, this is due to the added computational complexity involved in jointly estimating multiple thresholds and factor loadings for each item. Limited information solutions based on weighted least squares (WLS) or robust WLS are available (Flora & Curran, 2004; Muthén, 1984; Muthén & Satorra, 1995) and have been implemented in the Mplus and LISREL computer programs. Less statistically rigorous approaches in which the graded response categories are assumed to represent a normally distributed continuous response and are analyzed using traditional unweighted least squares factor-analytic models have also been used (Bartholomew, 1980). Bock et al. (1988) have indicated that this can lead to an underestimate of item factor loadings and an overestimate of the number of dimensions when the number of categories is small and the frequency of category use is nonuniform.

Although there are good limited information procedures available for ordinal response data, they are not designed specifically for application to instruments where a primary dimension and several subdimensions are present (i.e., the bifactor case). To obtain a correct estimate of both the general factor score and its standard error, however, the residual association between items within subdomains must be taken into account.

The Bifactor Structure

To analyze these kinds of structures for dichotomously scored item responses, Gibbons and Hedeker (1992) developed full-information item bifactor analysis for binary item responses. To illustrate, consider a set of n test items for which an s-factor solution exists with one general factor and $s-1$ group or method-related factors. The bifactor solution constrains each item j to a nonzero loading α_{j1} on the primary dimension and a second loading ($\alpha_{jk}, k=2,\ldots,s$) on not more than one of the $s-1$ group factors. For four items, the bifactor pattern matrix might be

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

This structure, which Holzinger and Swineford (1937) termed the *bifactor* pattern, also appears in the interbattery 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 and 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. Wilson and Adams (1995) developed a Rasch model for "item bundles," which is a special case of the bifactor model of Gibbons and Hedeker. More recently, Wang and Wilson (2005) extended the "Rasch testlet model" to the graded response case.

The bifactor restriction leads to a major simplification of likelihood equations that (a) permits analysis of models with large numbers of group factors, (b) permits conditional dependence among identified subsets of items, and (c) provides more parsimonious factor solutions than an unrestricted full-information item factor analysis for many scales and tests structured as described here. This article provides the necessary extensions of the original model of Gibbons and Hedeker (1992) to the graded response case and illustrates the model using data from the "Quality of Life Interview for the Chronically Mentally Ill" (Lehman, 1988).

Samejima's Graded Response Model

A typical source of ordered categorical data in behavioral sciences is the response of examinees or observers on some form of rating scale. The scale defines for the respondent a dimension on which he or she is required to make a judgment of quantity, intensity, or degree. In psychological measurement problems, the so-called "Likert" scale (Likert, 1932) is often used to classify the endorsement of an item by a respondent into one of m categories, for example, ranging from strongly disapprove to strongly approve. Typically, $5 \leq m \leq 9$ (see Bock, 1975).

The graded case of the model assumes a normal cumulative distribution function (cdf) for the item response function. The probability of responding in or below category t is

$$p_{jt} = \frac{1}{\sqrt{2\pi}} \int_{-z_{jt}(\theta_i)}^{\infty} \exp(-y^2/2)dy = \Phi_{jt}(\theta), \tag{1}$$

where

$$z_{jt}(\theta) = a_j(\theta - b_{jt}). \tag{2}$$

The slope a_j and thresholds b_{jt} are the so-called "invariant" item parameters (Lord & Novick, 1968), and θ is the underlying ability, disability, attitude, or attribute the scale was designed to measure. This graded response model was originally introduced by Samejima (1969). For computational purposes, it is convenient to use the item intercept parameters, $c_{jt} = -a_j b_{jt}$, and

$$z_{jt}(\theta) = c_{jt} + a_j\theta. \tag{3}$$

Letting $p_{j0} = 0$ and $p_{jm} = 1$, the probability of response to item j in category t is therefore

$$P_{jt} = p_{j,t} - p_{j,t-1} = \Phi_{jt}(\theta) - \Phi_{jt-1}(\theta). \tag{4}$$

Muraki's Rating Scale Model

Muraki (1983, 1990) introduced a rating scale version of the graded response model where

$$z_{jt}(\theta) = c_j + d_t + a_j(\theta) \tag{5}$$

(see also Masters, 1982). Here, c_j is the unique item intercept, and d_t are category parameters, assumed to be constant across all n items in the scale, that represent the psychological distance among points on the rating scale, constrained such that $\sum_t^m d_t = 0$. Let $x_{ijt} = 1$ if person i responds

positively to item j in category t and $x_{ijt} = 0$ otherwise. Assuming conditional independence of the n items, the probability of person i responding with response pattern $\mathbf{w}_i = [w_{i1}, w_{i2}, \ldots, w_{in}]$ conditional on θ is

$$P(\mathbf{w} = \mathbf{w}_i \mid \theta) = \prod_j^n \prod_t^m P_{jt}^{x_{ijt}} = L_i(\theta). \tag{6}$$

For a randomly sampled person from a population with distribution function $g(\theta)$, the unconditional probability is

$$P(\mathbf{w} = \mathbf{w}_i) = \int_{-\infty}^{\infty} L_i(\theta) g(\theta) d\theta, \tag{7}$$

which can be approximated to any practical degree of accuracy using the Gauss-Hermite quadrature as

$$P(\mathbf{w} = \mathbf{w}_i) = \sum_q^Q L_i(X_q) A(X_q), \tag{8}$$

where X_q is the tabled quadrature point, and $A(X_q)$ is the corresponding weight (see Bock & Aitkin, 1981; Stroud & Sechrest, 1966). The distribution function $g(\theta)$ is assumed to be continuous, and Bock and Aitkin (1981) have shown that assuming $g(\theta)$ to be normally distributed has little effect on the estimated parameters. They also show how to obtain a nonparametric estimate of $g(\theta)$ directly from the data.

The major advantages of the rating scale model over Samejima's (1969) original model are that (a) it requires estimation of $(n-1)m$ fewer parameters, (b) the category parameters associated with the points on the rating scale may be separately estimated from the item parameters, and (c) the items can be unidimensionally ordered by the item intercept. The major disadvantages are that (a) items with different numbers of response categories cannot be used and (b) the model assumes a common distance between response categories for all items.

Although the focus of this article is on the rating scale model, the necessary modifications to the estimation procedure are provided so that the bifactor solution for Samejima's (1969) original model can also be obtained.

The Bifactor Model for Graded Response Data

In the bifactor case, the graded response model is

$$z_{jt}(\boldsymbol{\theta}) = c_j + d_t + \sum_{k=1}^{s} a_{jk}(\theta_k), \tag{9}$$

where only one of the $k = 2, \ldots, s$ values of a_{jk} is nonzero in addition to a_{j1}. Assuming independence of the $\boldsymbol{\theta}$, in the unrestricted case, the multidimensional model above would require an s-fold integral to compute the unconditional probability, that is,

$$P(\mathbf{w} = \mathbf{w}_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} L_i(\boldsymbol{\theta}) g(\theta_1) g(\theta_2) \ldots g(\theta_s) d\theta_1 d\theta_2 \ldots d\theta_s, \tag{10}$$

for which numerical approximation is limited to four or five dimensions (see Bock & Aitkin, 1981). Gibbons and 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 θ_1 and the other for θ_k, $k > 1$. The reduction formula is due to Stuart (1958), who showed that if n variables follow a standardized multivariate normal distribution where the correlation $\rho_{ij} = \sum_{k=1}^{s} \alpha_{ik}\alpha_{jk}$ and α_{ik} is nonzero for only one k, then the probability that respective variables are simultaneously less than γ_j is given by

$$P = \prod_{k=1}^{s} \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{n} \left[\Phi\left(\frac{\gamma_j - \alpha_{jk}\theta}{\sqrt{1 - \alpha_{jk}^2}} \right) \right]^{u_{jk}} \right\} g(\theta)dy, \tag{11}$$

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

   Equation (11) follows from the fact that if each variate is related only to a single dimension, then the s dimensions are independent, and the joint probability is the product of s unidimensional probabilities. In this context, the result applies only to the $s - 1$ content dimensions (i.e., $k = 2, \ldots, s$). If a primary dimension exists, it will not be independent of the other $s - 1$ dimensions because each item now loads on each of two dimensions. Gibbons and Hedeker (1992) derived the necessary two-dimensional generalization of Stuart's (1958) original result as

$$P = \int_{-\infty}^{\infty} \left\{ \prod_{k=2}^{s} \int_{-\infty}^{\infty} \left[\prod_{j=1}^{n} \left(\Phi\left[\frac{\gamma_j - \alpha_{j1}\theta_1 - \alpha_{jk}\theta_k}{\sqrt{1 - \alpha_{j1}^2 - \alpha_{jk}^2}} \right] \right)^{u_{jk}} \right] g(\theta_k)d\theta_k \right\} g(\theta_1)d\theta_1. \tag{12}$$

   For the rating scale graded response model, the probability of a value less than the category threshold $\gamma_{jt} = -(c_j + d_t)/y_j$ can be obtained by substituting γ_{jt} for γ_j in the previous equation. The unconditional probability of a particular response pattern $\mathbf{w}_i$ is therefore

$$P(\mathbf{w} = \mathbf{w}_i) = \int_{-\infty}^{\infty} \left\{ \prod_{k=2}^{s} \int_{-\infty}^{\infty} \left[\prod_{j=1}^{n} \prod_{t=1}^{m} [\Phi_{jt}(\theta_1, \theta_k) - \Phi_{jt-1}(\theta_1, \theta_k)]^{u_{jk}x_{ijt}} \right] g(\theta_k)d\theta_k \right\} \atop g(\theta_1)d\theta_1, \tag{13}$$

which can be approximated by

$$\hat{P}_i \cong \sum_{q_1}^{Q} \left\{ \prod_{k=2}^{s} \sum_{q_k}^{Q} \left[\prod_{j=1}^{n} \prod_{t=1}^{m} (\Phi_{jt}(X_{q_1}, X_{q_k}) - \Phi_{jt-1}(X_{q_1}, X_{q_k}))^{u_{jk}x_{ijt}} \right] A(X_{q_k}) \right\} A(X_{q_1}), \tag{14}$$

where

$$\Phi_{jk}(\theta_1, \theta_k) \cong \Phi_{jk}(X_{q_1}, X_{q_k}) = \Phi(c_j + d_t + a_{j1}X_{q_1} + a_{jk}X_{q_k}).$$

Alternatively, Samejima's (1969) original graded response model can be used to compute the above probability as

$$\Phi_{jk}(\theta_1, \theta_k) \cong \Phi_{jk}(X_{q_1}, X_{q_k}) = \Phi(c_{jt} + a_{j1}X_{q_1} + a_{jk}X_{q_k}).$$

Note that for both the binary and graded bifactor response models, the dimensionality of the integral is 2 regardless of the number of subdomains (i.e., $s - 1$) that comprised the scale.

The previously mentioned independence assumption of the θ implies an orthogonal basis for the bifactor model. The orthogonal basis is a reasonable choice in that the secondary factors model the residual association among the items once the unique contribution of the primary dimension has been removed. Furthermore, the orthogonal basis permits evaluation of the likelihood using numerical integration.

Marginal Maximum Likelihood Estimation

Gibbons and 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 and Aitkin (1981). For the graded case, the likelihood equations are derived as follows.

Denoting the kth subset of the components of $\boldsymbol{\theta}$ as $\boldsymbol{\theta}_k^* = \begin{bmatrix} \theta_1 \\ \theta_k \end{bmatrix}$, let

$$
\begin{aligned}
P_i &= P(\mathbf{w} = \mathbf{w}_i) \\
&= \int_{\theta_1} \left\{ \prod_{k=2}^{s} \int_{\theta_k} \left[\prod_{j=1}^{n} \prod_{t=1}^{m} \left(\Phi_{jt}(\boldsymbol{\theta}_k^*) - \Phi_{jt-1}(\boldsymbol{\theta}_k^*) \right)^{u_{jk}x_{ijt}} \right] g(\theta_k) d\theta_k \right\} g(\theta_1) d\theta_1 \\
&= \int_{\theta_1} \left\{ \prod_{k=2}^{s} \int_{\theta_k} L_{ik}(\boldsymbol{\theta}_k^*) g(\theta_k) d\theta_k \right\} g(\theta_1) d\theta_1,
\end{aligned}
\tag{15}
$$

where

$$
L_{ik}(\boldsymbol{\theta}_k^*) = \prod_{j=1}^{n} \prod_{t=1}^{m} \left(\Phi_{jt}(\boldsymbol{\theta}_k^*) - \Phi_{jt-1}(\boldsymbol{\theta}_k^*) \right)^{u_{jk}x_{ijt}}.
$$

Then, the log-likelihood is

$$
\log L = \sum_{i=1}^{S} r_i \log P_i,
\tag{16}
$$

where S denotes the number of unique response patterns, and r_i is 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 the appendix.

Estimating Primary Trait Levels

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 θ_1 for person i. A good choice for this purpose (Bock & Aitkin, 1981) is the expected a posteriori (EAP) value (Bayes estimate) of θ_1, given the observed response vector $\mathbf{w}_i$ and levels of the other subdimensions $\theta_2 \ldots \theta_s$. The Bayesian estimate of θ_1 for person i is

$$
\hat{\theta}_{1i} = E(\theta_{1i} \mid \mathbf{w}_i, \theta_{2i} \ldots \theta_{si}) = \frac{1}{P_i} \int_{\theta_1} \theta_{1i} \left\{ \prod_{k=2}^{s} \int_{\theta_k} L_{ik}(\boldsymbol{\theta}_k^*) g(\theta_k) d\theta_k \right\} g(\theta_1) d\theta_1.
\tag{17}
$$

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

$$V(\theta_{1i} \mid \mathbf{w}_i, \theta_{2i} \ldots \theta_{si}) = \frac{1}{P_i} \int_{\theta_1} (\theta_{1i} - \hat{\theta}_{1i})^2 \left\{ \prod_{k=2}^{s} \int_{\theta_k} L_{ik}(\boldsymbol{\theta}_k^*) g(\theta_k) d\theta_k \right\} g(\theta_1) d\theta_1. \qquad (18)$$

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

In some cases, there may also be interest in obtaining a trait estimate for the subdomains in addition to the primary dimension of interest. For example, in the following quality-of-life example, in addition to obtaining an estimate of a respondent's overall quality of life, one may also be interested in estimating that respondent's quality of health and/or social domains. One solution is to use the estimated factor loadings from the subdomains directly from the bifactor model. It is important to note, however, that the subdomain estimates in the bifactor model describe associations among the residuals between the items within each subdomain, once the primary dimension has been accounted for. As such, the bifactor subdomain factor loadings may underestimate the unconditional subdomain estimates. A reasonable alternative is to break the test into a series of subtests (based on subdomains) and apply a traditional unidimensional IRT model separately to each subtest and obtain a corresponding subdomain trait estimate. The problem of obtaining subdomain trait estimates for bifactor models should be a topic for further research.

Illustration

As an illustration of the bifactor model for graded response data, the "Quality of Life Interview for the Chronically Mentally Ill" (Lehman, 1988) was analyzed based on the item responses of 586 chronically mentally ill patients. The scale consists of seven subdomains (Family, Finance, Health, Leisure, Living, Safety, and Social), each with 4 to 6 items for a total of 34 items. In addition, there is one global life satisfaction item, which was allowed to load on its own subdomain in the event that it had a unique contribution to the residual variation above and beyond its contribution to the primary dimension. Each item is rated on a 7-point scale with the following response categories: 1 = *terrible*; 2 = *unhappy*; 3 = *mostly dissatisfied*; 4 = *mixed, about equally satisfied and dissatisfied*; 5 = *mostly satisfied*; 6 = *pleased*; and 7 = *delighted*.

Item intercepts, primary factor loadings, and factor loadings on the eight subfactors are displayed in Table 1 based on the polytomous rating scale model. Table 1 shows that all items had substantial loading on the primary dimension (Factor 1), indicating that the scale was well designed and that all items were related to overall life satisfaction. The three most discriminating items were "global life satisfaction," $\hat{\lambda}_{11} = .694$; satisfaction with "free time," $\hat{\lambda}_{16,1} = .611$ (Scale 4); and "emotional well-being," $\hat{\lambda}_{15,1} = .609$ (Scale 3). The three least discriminating items were satisfaction with "people in general," $\hat{\lambda}_{35,1} = .385$ (Scale 7); "amount you pay for basic needs," $\hat{\lambda}_{7,1} = .391$ (Scale 2); and "pleasure from TV," $\hat{\lambda}_{21,1} = .414$ (Scale 4). The unique "life as a whole" item loaded heavily on the primary dimension but not at all on the subdomain, indicating that the primary dimension is a good measure of overall life satisfaction.

The item intercepts permit items to be positioned relative to the global life satisfaction item to determine at what point on the scale a person would report global life satisfaction. Table 1 shows that the Health (Scale 3), Living (Scale 5), and Social domains (Scale 7) were typically reported at lower levels of satisfaction than the global item, whereas Financial (Scale 2) and Leisure (Scale 4) items had, on average, higher intercepts than the global satisfaction item. The domains of Family (Scale 1) and Safety (Scale 6) items were located at similar levels to the global item.

Table 1

Nine-dimensional Bifactor Solution for the Lehman (1988) Quality-of-Life Rating
Scale Data ($N = 586$): Item Intercepts and Factor Loadings

Scale	Item	Intercept	Factor 1	2	3	4	5	6	7	8	9
Global	Life as a whole	–0.402	0.694	0.001							
1	Family	–0.768	0.499		0.566						
1	Amount of family contact	–0.349	0.534		0.443						
1	Family with interaction	0.282	0.548		0.518						
1	General family stuff	–0.350	0.597		0.491						
2	Total money you get	0.209	0.435			0.568					
2	Amount pay for basic needs	–0.136	0.391			0.477					
2	Financial well-being	0.319	0.503			0.562					
2	Money for fun	0.242	0.491			0.568					
3	Health in general	–0.482	0.458				0.270				
3	Medical care	–0.701	0.475				0.419				
3	How often see doctor	–0.441	0.441				0.397				
3	Talk to therapist	–0.621	0.478				0.378				
3	Physical condition	–0.582	0.553				0.299				
3	Emotional well-being	–0.284	0.609				0.185				
4	Way spend free time	–0.139	0.611					0.262			
4	Amount of free time	–0.292	0.509					0.342			
4	Chance to enjoy time	–0.552	0.578					0.386			
4	Amount of fun	–0.270	0.597					0.430			
4	Amount of relaxation	–0.306	0.525					0.393			
4	Pleasure from TV	–0.776	0.414					0.163			
5	Living arrangements	–0.435	0.493						0.493		
5	Food	–0.982	0.449						0.468		
5	Privacy	–0.709	0.478						0.610		
5	Amount of freedom	–1.090	0.478						0.649		
5	Prospect of staying	–0.100	0.469						0.630		
6	Neighborhood safety	–0.298	0.511							0.445	
6	Safe at home	–0.666	0.542							0.416	
6	Police access	–0.062	0.487							0.429	
6	Protect robbed/attack	–0.214	0.517							0.465	
6	Personal safety	–0.533	0.531							0.326	
7	Do things with others	–0.614	0.494								0.326
7	Time with others	–0.411	0.519								0.257
7	Social interactions	–0.604	0.472								0.346
7	People in general	–0.835	0.385								0.220

In terms of subdomains, items within domains had a high degree of residual association, with an average loading of .406. Consistent with this finding was a significant likelihood ratio test for improvement in fit of the bifactor model over the unidimensional graded response model ($\chi^2_{35} = 2188, \ p < .0001$).

Table 2 displays the observed and expected (in italics) category proportions for each item. In general, there is close agreement between observed and expected response proportions. The root mean square error (RMSE) between observed and expected proportions (over all items and categories) was 0.026, indicating that the model with common category parameters fit these data extremely well. The six category parameters were as follows:

$$-1.395 \quad -.858 \quad -.449 \quad .044 \quad .866 \quad 1.793.$$

Samejima's (1969) model with unique item category parameters produced a significant likelihood ratio test for improvement in fit over the rating scale model ($\chi^2_{169} = 1637, \ p < .0001$), with a decrease in RMSE between observed and expected proportions to 0.010. Factor loadings were almost identical between the two models. Furthermore, there were only minor changes in the estimated item thresholds between the two models, despite the fact that the rating scale model has only one item-specific threshold (and six general thresholds) and Samejima's model has six unique thresholds per item. For example, estimated item thresholds for the first 10 quality-of-life items for both models are presented in Table 3. Table 3 shows that the estimated thresholds are quite similar for the two models. Although the fit of the model is significantly improved when estimating category parameters separately for each item (presumably due to the large number of subjects, items, and categories), the model with common category parameters may be a useful alternative for applications in which the items have the same number of categories.

Finally, a limited-information solution based on polychoric correlations was compared using robust WLS (Muthén, du Toit, & Spisic, 1997), which is available in Mplus software (Muthén & Muthén, 2004) for the full-information approach. The presentation in this study is simply to show that there is alternate estimation method for the bifactor model. A nontechnical review and details of the method are available in Flora and Curran (2004). Because the WLS solution did not converge when a factor loading of the global life satisfaction factor (consisting of only one item) was estimated, it was fixed to its maximum likelihood estimate (0.001). Estimates of the rating scale parameters are obtained from estimates of a mathematically identical model because the category and the intercept parameters cannot be specified directly in Mplus. The obtained parameters are rescaled by dividing by $(1 + a_{1j}^2 + a_{kj}^2)^{1/2}$ to make them comparable to parameters in equation (12). The Mplus program setup and details of reparameterization can be obtained from the author. The first and the second blocks in Table 4 are WLS and the difference between WLS and marginal maximum likelihood (MML) estimates, respectively (MML estimates are in Table 1). Overall, the differences between the limited- and full-information solutions were small. Consistent downward bias in the primary and secondary factor loadings was observed but was more pronounced for the secondary factor, where there is less information.

Discussion

In many practical applications, the bifactor model provides a natural alternative to the traditional conditionally independent unidimensional IRT model. When conditional dependence is likely, as in the case of paragraph comprehension tests, tests in which there are two or more methods of item presentation, or personality or other items that have a two-level structure with an underlying general factor, the item bifactor solution provides an excellent alternative. An

Table 2

Observed and Expected (in Italics) Proportions From the Nine-Dimensional Graded
Bifactor Analysis of Lehman Quality-of-Life Rating Scale Data ($N = 586$)

	1	2	3	4	5	6	7
Life as a whole	.063	.087	.080	.176	.224	.212	.159
	.098	*.084*	*.088*	*.128*	*.232*	*.211*	*.158*
Family	.046	.068	.049	.160	.203	.232	.241
	.078	*.065*	*.069*	*.105*	*.208*	*.224*	*.251*
Amount of family contact	.061	.097	.114	.133	.244	.210	.140
	.104	*.088*	*.090*	*.131*	*.232*	*.206*	*.149*
Family with interaction	.067	.125	.094	.167	.200	.217	.131
	.135	*.092*	*.089*	*.123*	*.211*	*.190*	*.160*
General family stuff	.072	.108	.087	.159	.229	.186	.160
	.134	*.088*	*.084*	*.117*	*.205*	*.192*	*.180*
Total money you get	.138	.155	.137	.128	.235	.143	.063
	.204	*.121*	*.108*	*.137*	*.203*	*.145*	*.081*
Amount pay for basic needs	.077	.121	.106	.145	.276	.195	.080
	.114	*.103*	*.106*	*.149*	*.246*	*.187*	*.096*
Financial well-being	.174	.152	.133	.131	.201	.142	.067
	.240	*.122*	*.104*	*.128*	*.187*	*.136*	*.083*
Money for fun	.147	.171	.148	.109	.208	.135	.082
	.223	*.119*	*.104*	*.129*	*.193*	*.143*	*.090*
Health in general	.048	.063	.051	.113	.392	.215	.118
	.056	*.072*	*.087*	*.140*	*.272*	*.239*	*.133*
Medical care	.043	.039	.055	.135	.258	.311	.160
	.052	*.061*	*.073*	*.119*	*.245*	*.250*	*.199*
How often see doctor	.049	.061	.099	.125	.309	.242	.114
	.070	*.078*	*.089*	*.138*	*.259*	*.228*	*.138*
Talk to therapist	.036	.041	.085	.123	.292	.280	.143
	.055	*.065*	*.078*	*.126*	*.253*	*.247*	*.176*
Physical condition	.034	.072	.084	.119	.261	.283	.147
	.062	*.069*	*.080*	*.127*	*.249*	*.240*	*.173*
Emotional well-being	.065	.087	.104	.157	.273	.195	.119
	.098	*.091*	*.097*	*.141*	*.246*	*.205*	*.122*
Way spend free time	.077	.113	.126	.159	.225	.201	.099
	.126	*.102*	*.102*	*.142*	*.235*	*.185*	*.108*
Amount of free time	.060	.077	.119	.154	.273	.208	.109
	.091	*.090*	*.097*	*.143*	*.252*	*.207*	*.118*
Chance to enjoy time	.053	.082	.087	.130	.218	.241	.189
	.081	*.075*	*.081*	*.122*	*.232*	*.225*	*.186*
Amount of fun	.077	.118	.114	.126	.218	.196	.150
	.130	*.093*	*.091*	*.126*	*.217*	*.192*	*.151*
Amount of relaxation	.077	.080	.108	.131	.259	.225	.119
	.100	*.090*	*.095*	*.137*	*.242*	*.205*	*.131*
Pleasure from TV	.020	.034	.055	.143	.278	.282	.188
	.026	*.046*	*.065*	*.120*	*.276*	*.287*	*.181*
Living arrangements	.073	.070	.085	.131	.271	.210	.159
	.095	*.082*	*.086*	*.127*	*.232*	*.214*	*.165*
Food	.041	.032	.056	.072	.234	.304	.261
	.035	*.045*	*.057*	*.100*	*.227*	*.267*	*.269*

(continued)

Table 2 (continued)

	1	2	3	4	5	6	7
Privacy	.087	.051	.080	.097	.186	.258	.241
	.092	*.069*	*.071*	*.105*	*.202*	*.214*	*.247*
Amount of freedom	.065	.051	.049	.067	.195	.230	.343
	.071	*.054*	*.057*	*.087*	*.179*	*.214*	*.339*
Prospect of staying	.150	.119	.094	.150	.160	.140	.186
	.177	*.099*	*.090*	*.119*	*.196*	*.170*	*.147*
Neighborhood safety	.077	.080	.082	.143	.294	.218	.106
	.106	*.091*	*.094*	*.135*	*.236*	*.202*	*.136*
Safe at home	.055	.043	.061	.111	.280	.280	.171
	.066	*.067*	*.075*	*.117*	*.233*	*.237*	*.205*
Police access	.137	.061	.131	.172	.217	.174	.108
	.134	*.108*	*.107*	*.146*	*.235*	*.176*	*.094*
Protect robbed/attack	.094	.073	.118	.147	.254	.203	.111
	.124	*.097*	*.096*	*.135*	*.229*	*.191*	*.128*
Personal safety	.048	.048	.070	.130	.309	.276	.119
	.066	*.072*	*.083*	*.130*	*.252*	*.235*	*.162*
Do things with others	.031	.032	.067	.142	.341	.254	.133
	.053	*.065*	*.078*	*.127*	*.257*	*.249*	*.171*
Time with others	.036	.063	.080	.167	.317	.247	.089
	.070	*.080*	*.091*	*.141*	*.262*	*.225*	*.130*
Social interactions	.036	.039	.067	.159	.285	.278	.137
	.053	*.065*	*.079*	*.128*	*.259*	*.248*	*.168*
People in general	.019	.027	.032	.128	.302	.321	.171
	.023	*.042*	*.060*	*.114*	*.272*	*.294*	*.195*

attractive by-product of this model is that it requires only the evaluation of a two-dimensional integral, regardless of the number of subtests, paragraphs, or content areas.

In the ordinal response case, the bifactor model provides the advantages previously described for the binary response model and, in addition, provides a very general multidimensional model for graded response data. In mental health measurement, rating scales are typically constructed by sampling items from domains related to a single underlying construct, as in the quality-of-life scale analyzed in the illustration. In these cases, a priori knowledge of which item belongs to which subdomain is available, and the bifactor model is a natural choice. Similarly, in educational measurement problems, tests are often constructed by creating a series of subtests or so-called "testlets" (Wainer & Kiely, 1987) within which items have similar content or focus, and these testlets are then combined to form a test. In this case, item groupings are also known in advance, and the bifactor model applies. Regardless of the number of testlets, the relevant integrals in the full-information maximum marginal likelihood solution always reduce to 2 and can be approximated to any practical degree of accuracy.

Computer Software

A program for estimating the bifactor model for ordinal and dichotomous data is available from the first author. The bifactor model for dichotomous data is available in TESTFACT (Scientific Software International, Lincolnwood, IL).

Table 3

Estimated Category Thresholds for Rating Scale Model (RSM) and Samejima
Model (SM) for the First 10 Quality-of-Life Items

Item		Category					
		1-2	2-3	3-4	4-5	5-6	6-7
Life as a whole	RSM	−1.452	−1.005	−0.673	−0.278	0.361	1.079
	SM	−1.450	−0.994	−0.714	−0.217	0.374	1.072
Family	RSM	−1.578	−1.171	−0.868	−0.508	0.074	0.728
	SM	−1.666	−1.184	−0.950	−0.408	0.110	0.722
Family contact	RSM	−1.435	−0.981	−0.642	−0.241	0.409	1.140
	SM	−1.484	−0.953	−0.575	−0.210	0.409	1.115
Family interaction	RSM	−1.236	−0.831	−0.530	−0.172	0.408	1.059
	SM	−1.441	−0.836	−0.536	−0.085	0.416	1.156
Family stuff	RSM	−1.240	−0.849	−0.558	−0.213	0.347	0.976
	SM	−1.397	−0.876	−0.588	−0.154	0.418	1.023
Total money	RSM	−0.952	−0.521	−0.199	0.182	0.799	1.493
	SM	−1.084	−0.528	−0.134	0.206	0.870	1.552
Basic needs	RSM	−1.395	−0.895	−0.523	−0.082	0.634	1.437
	SM	−1.432	−0.840	−0.486	−0.084	0.643	1.424
Financial well-being	RSM	−0.819	−0.415	−0.114	0.243	0.821	1.471
	SM	−0.913	−0.417	−0.064	0.281	0.869	1.533
Money for fun	RSM	−0.873	−0.467	−0.165	0.194	0.775	1.428
	SM	−1.032	−0.439	−0.046	0.236	0.829	1.430
General health	RSM	−1.876	−1.322	−0.908	−0.418	0.376	1.268
	SM	−1.659	−1.189	−0.945	−0.555	0.465	1.224

Table 4

Weighted Least Squares (WLS) Estimates and Their Difference From Marginal Maximum
Likelihood (MML) Estimates for the First 10 Quality of Life Items

Item	Category						Factor Loading	
	1-2	2-3	3-4	4-5	5-6	6-7	Primary	Method
WLS								
Life as a whole	−1.495	−1.055	−0.716	−0.315	0.342	1.070	0.765	0.001
Family	−1.593	−1.199	−0.896	−0.536	0.053	0.705	0.505	0.642
Family contact	−1.461	−1.014	−0.670	−0.262	0.405	1.144	0.561	0.508
Family interaction	−1.143	−0.790	−0.518	−0.196	0.331	0.915	0.595	0.616
Family stuff	−1.173	−0.825	−0.557	−0.240	0.280	0.855	0.633	0.583
Total money	−0.899	−0.508	−0.207	0.150	0.734	1.381	0.508	0.643
Basic needs	−1.392	−0.910	−0.539	−0.099	0.620	1.418	0.431	0.562
Financial well-being	−0.773	−0.409	−0.128	0.205	0.750	1.354	0.572	0.623
Money for fun	−0.822	−0.452	−0.167	0.171	0.725	1.338	0.550	0.635
General health	−1.780	−1.278	−0.892	−0.435	0.313	1.143	0.506	0.454

(continued)

Table 4 (continued)

Item	Category						Factor Loading	
	1-2	2-3	3-4	4-5	5-6	6-7	Primary	Method
MML-WLS								
Life as a whole	−0.043	−0.050	−0.043	−0.037	−0.019	−0.009	0.071	0.000
Family	−0.015	−0.028	−0.028	−0.028	−0.021	−0.023	0.006	0.076
Family contact	−0.026	−0.033	−0.028	−0.021	−0.004	0.004	0.027	0.065
Family interaction	0.093	0.041	0.012	−0.024	−0.077	−0.144	0.047	0.098
Family stuff	0.067	0.024	0.001	−0.027	−0.067	−0.121	0.036	0.092
Total money	0.053	0.013	−0.008	−0.032	−0.065	−0.112	0.073	0.075
Basic needs	0.003	−0.010	−0.016	−0.017	−0.014	−0.019	0.040	0.085
Financial well-being	0.046	0.006	−0.014	−0.038	−0.071	−0.117	0.069	0.061
Money for fun	0.051	0.015	−0.002	−0.023	−0.050	−0.090	0.059	0.067
General health	0.096	0.044	0.016	−0.017	−0.063	−0.125	0.048	0.184

Appendix
Marginal Maximum Likelihood Estimation

The derivative of the log marginal likelihood to a general item parameter v_j (i.e., a_{jk} and b_j) follows. Let

$$E_{ik}(\theta_1) = \frac{\left[\prod_{h=2}^{s} \int_{\theta_h} L_{ih}(\boldsymbol{\theta}_h^*) g(\theta_h) d\theta_h \right]}{\int_{\theta_k} L_{ik}(\boldsymbol{\theta}_k^*) g(\theta_k) d\theta_k}, \tag{A1}$$

then

$$\frac{\partial \log L}{\partial v_j} = \sum_{i}^{S} \frac{r_i}{P_i} \left(\frac{\partial P_i}{\partial v_j} \right) = \sum_{i=1}^{S} \frac{r_i}{P_i} \int_{\theta_1} \sum_{k=2}^{s} u_{jk} E_{ik}(\theta_1)$$
$$\left\{ \int_{\theta_k} L_{ik}(\boldsymbol{\theta}_k^*) \sum_{t}^{m} \frac{\partial [P_{jt}(\theta_k^*)]^{x_{ijt}}}{\partial v_j} \left(\frac{1}{P_{jt}(\boldsymbol{\theta}_k^*)} \right) g(\theta_k) d\theta_k \right\} g(\theta_1) d\theta_1. \tag{A2}$$

Replacing the integrals with Gauss-Hermite quadrature sums and rearranging terms yields

$$\frac{\partial \log L}{\partial v_j} \cong \sum_{q_1}^{Q} \sum_{k=2}^{s} u_{jk} \sum_{q_k}^{Q} \sum_{t}^{m} \frac{\bar{r}_{jtk}(\mathbf{X}_k^*)}{P_{jt}(\mathbf{X}_k^*)} \frac{\partial P_{jt}(\mathbf{X}_k^*)}{\partial v_j} A(X_{q_k}) A(X_{q_1}), \tag{A3}$$

where

$$\bar{r}_{jtk}(\mathbf{X}_k^*) = \sum_{i=1}^{S} r_i x_{ijt} [E_{ik}(X_{q_1})] L_{ik}(X_{q_1}, X_{q_k})/P_i, \tag{A4}$$

and $\mathbf{X}_k^* = \begin{bmatrix} X_{q_1} \\ X_{q_k} \end{bmatrix}$. $\bar{r}_{jtk}$ represents the expected number of positive responses for item j in category k.
These equations are similar to those in the unrestricted case, except that in the bifactor case, the conditional probability of response pattern $\mathbf{w}_{ik}$ (i.e., responses to items $j = 1, \ldots, n_k$ in subsection

k for response pattern i) is weighted by the factor, $E_{ik}(X_{q_1})$. Furthermore, because each item appears in one subsection only (k), $\bar{r}$ varies with k, in contrast to the unrestricted case.

For the category parameters of the rating scale model, d_k, the likelihood equations are

$$\frac{\partial \log L}{\partial d_g} \cong \sum_{q_1}^{Q} \sum_{k=2}^{s} u_{jk} \sum_{q_k}^{Q} \sum_{j}^{n} \left[\frac{\bar{r}_{jgk}(\mathbf{X}_k^*)}{P_{jg}(\mathbf{X}_k^*)} - \frac{\bar{r}_{j,g+1,k}(\mathbf{X}_k^*)}{P_{j,g+1}(\mathbf{X}_k^*)} \right] \frac{\partial p_{jg}(\mathbf{X}_k^*)}{\partial d_g} A(X_{q_k}) A(X_{q_1}). \tag{A5}$$

From provisional parameter values, each E-step yields $\bar{r}_{jtk}$ and $\bar{N}_k$ (expectations of complete data statistics computed conditionally on incomplete data; see Bock et al., 1988), where

$$\bar{N}_k(\mathbf{X}_k^*) = \sum_{i=1}^{S} r_i [E_{ik}(X_{q_1})] L_{ik}(X_{q_1}, X_{q_k}) / P_i \tag{A6}$$

denotes the effective sample size for subset k at quadrature point (X_{q_1}, X_{q_k}) and $\bar{r}_{jtk}$ the corresponding expected number of positive responses for item j in category k. When weighted by $A(\mathbf{X})$ and summed over quadrature nodes for each subsection, $\bar{N}_k$ yields the total number of respondents, whereas corresponding weighting and summation for $\bar{r}_{jtk}$ yields the total number of respondents rating item j in category t.

The subsequent M-step solves using conventional maximum likelihood ordinal probit analysis, substituting provisional expectations of $\bar{r}_{jtk}$ and $\bar{N}_k$ (see Bock & Jones, 1968). The elements of the information matrix required for the M-step solution of a_{j1}, a_{jk}, and b_j are

$$\mathrm{E}\left(-\frac{\partial^2 \log L}{\partial a_{jk}^2} \right) = \sum_{q1}^{Q} \sum_{qk}^{Q} \bar{N}_k(\mathbf{X}_k^*) \sum_{t}^{m} \frac{1}{P_{jt}(\mathbf{X}_k^*)} \left[\frac{\partial P_{jt}(\mathbf{X}_k^*)}{\partial a_{jk}} \right]^2, \tag{A7}$$

$$\mathrm{E}\left(-\frac{\partial^2 \log L}{\partial c_j^2} \right) = \sum_{q1}^{Q} \sum_{qk}^{Q} \bar{N}_k(\mathbf{X}_k^*) \sum_{t}^{m} \frac{1}{P_{jt}(\mathbf{X}_k^*)} \left[\frac{\partial P_{jt}(\mathbf{X}_k^*)}{\partial c_j} \right]^2, \tag{A8}$$

$$\mathrm{E}\left(-\frac{\partial^2 \log L}{\partial a_{j1} \partial a_{jk}} \right) = \sum_{q1}^{Q} \sum_{qk}^{Q} \bar{N}_k(\mathbf{X}_k^*) \sum_{t}^{m} \frac{1}{P_{jt}(\mathbf{X}_k^*)} \frac{\partial P_{jt}(\mathbf{X}_k^*)}{\partial a_{j1}} \frac{\partial P_{jt}(\mathbf{X}_k^*)}{\partial a_{jk}}, \tag{A9}$$

$$\mathrm{E}\left(-\frac{\partial^2 \log L}{\partial a_{jh} \partial c_j} \right) = \sum_{q1}^{Q} \sum_{qk}^{Q} \bar{N}_k(\mathbf{X}_k^*) \sum_{t}^{m} \frac{1}{P_{jt}(\mathbf{X}_k^*)} \frac{\partial P_{jt}(\mathbf{X}_k^*)}{\partial a_{jh}} \frac{\partial P_{jt}(\mathbf{X}_k^*)}{\partial c_j}. \tag{A10}$$

where

$$\frac{\partial P_{jt}(\mathbf{X}_k^*)}{\partial v} = \phi[z_{jt}(\mathbf{X}_k^*)] \frac{\partial z_{jt}(\mathbf{X}_k^*)}{\partial v} - \phi[z_{j,t-1}(\mathbf{X}_k^*)] \frac{\partial z_{j,t-1}(\mathbf{X}_k^*)}{\partial v}. \tag{A11}$$

Elements of the information matrix required for the M-step solution of the category parameters d_t are

$$\mathrm{E}\left(-\frac{\partial^2 \log L}{\partial d_g^2} \right) = -\sum_{q1}^{Q} \sum_{qk}^{Q} \bar{N}_k(\mathbf{X}_k^*) \sum_{j}^{n} \left[\frac{1}{P_{jg}(\mathbf{X}_k^*)} + \frac{1}{P_{j,g+1}(\mathbf{X}_k^*)} \right] \left[\frac{\partial P_{jg}(\mathbf{X}_k^*)}{\partial d_g} \right]^2, \tag{A12}$$

and

$$\mathrm{E}\left(-\frac{\partial^2 \log L}{\partial d_g \ \partial d_{g-1}}\right) = -\sum_{q1}^{Q} \sum_{qk}^{Q} \bar{N}_k(\mathbf{X}_k^*) \sum_j^n \frac{1}{P_{jg}(\mathbf{X}_k^*)} \frac{\partial P_{jg}(\mathbf{X}_k^*)}{\partial d_g} \frac{\partial P_{j,g-1}(\mathbf{X}_k^*)}{\partial d_{g-1}}. \tag{A13}$$

For $|g - g'| \geq 2$, elements of the information matrix are zero. During the M-step, improved estimates of item and category parameters are obtained separately for the rating scale model but jointly for Samejima's (1969) model. The additional elements of the information matrix required for Samejima's model are of the following general form:

$$\mathrm{E}\left(-\frac{\partial^2 \log L}{\partial d_g \ \partial v_j}\right) = -\sum_{q1}^{Q} \sum_{qk}^{Q} \bar{N}_k(\mathbf{X}_k^*)$$
$$\sum_j^n \left(\frac{1}{P_{jg}(\mathbf{X}_k^*)} \frac{\partial P_{jg}(\mathbf{X}_k^*)}{\partial v_j} - \frac{1}{P_{j,g+1}(\mathbf{X}_k^*)} \frac{\partial P_{j,g+1}(\mathbf{X}_k^*)}{\partial v_j}\right) \frac{\partial p_{jg}(\mathbf{X}_k^*)}{\partial d_g}. \tag{A14}$$

References

American College Testing Service (ACT). (2006). *ACT Assessment Science Test*. Iowa City, IA: Author.

Bartholomew, D. J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society, Series B, 42*, 293-321.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.

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

Bock, R. D., Gibbons, R. D., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*, 261-280.

Bock, R. D., & Jones, L. V. (1968). *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466-491.

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

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

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

Kaziz, E., Miller, D. R., Clark, J., Skinner, K., Lee, A., Rogers, W., et al. (1998). Health-related quality of life in patients served by the Department of Veterans Affairs: Results from the Veterans Health Study. *Archives Internal Medicine, 158*, 626-632.

Lehman, A. F. (1988). A quality of life interview for the chronically mentally ill. *Evaluation and Program Planning, 11*, 51-62.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, Monograph 140.

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

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

Muraki, E. (1983). *Marginal maximum likelihood estimation for three-parameter polychotomous item response models: Application of an EM algorithm*. Unpublished doctoral dissertation, University of Chicago.

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

Muraki, E., & Carlson, J. E. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement, 19*, 73-90.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered, categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115-132.

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

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript.

Muthén, B. O., & Satorra, A. (1995). Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika*, *60*, 489-503.

Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide* (3rd ed.). Los Angeles: Author.

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

Stroud, A. H., & Sechrest, D. (1966). *Gaussian quadrature formulas*. Englewood Cliffs, NJ: Prentice Hall.

Stuart, 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.

Wainer, H., & Kiely, G. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, *24*, 185-201.

Wang, W. C., & Wilson, M. (2005). The Rasch testlet model. *Applied Psychological Measurement*, *29*, 126-149.

Wilson, M., & Adams, R. J. (1995). Rasch models for item bundles. *Psychometrika*, *60*, 181-198.

Acknowledgments

Author's Address

Address correspondence to Robert D. Gibbons, Center for Health Statistics, University of Illinois at Chicago, 1601 W. Taylor, Chicago, IL 60612; e-mail: rdgib@uic.edu.