The Generalized Graded Unfolding Model: A General Parametric Item Response Model for Unfolding Graded Responses

James S. Roberts
John R. Donoghue
James E. Laughlin

Educational Testing Service
Princeton, New Jersey

August 1998
The Generalized Graded Unfolding Model: A General Parametric Item Response Model for Unfolding Graded Responses

James S. Roberts  
Center for Drug and Alcohol Programs  
Medical University of South Carolina

John R. Donoghue  
Educational Testing Service

James E. Laughlin  
University of South Carolina

The work reported in this paper was partially supported by an ETS postdoctoral fellowship (PJ 877-10) to the first author.
The Generalized Graded Unfolding Model

Abstract

Psychologists have long used binary or graded disagree-agree responses to measure attitudes. Such data have traditionally been analyzed with cumulative models, but several researchers have recently argued that unfolding models are generally more appropriate. There have been several parametric item response models proposed to unfold disagree-agree responses. Some of these models allow only for binary responses whereas others permit both binary and graded responses. A new item response model, referred to as the Generalized Graded Unfolding Model (GGUM), is developed in this paper. The GGUM allows for either binary or graded responses and generalizes previous item response models for unfolding in two useful ways. First, it implements a discrimination parameter that varies across items, and thus, items are allowed to discriminate among respondents in different ways. Second, the GGUM allows for distinctively different use of response categories across items. It does this by implementing response category threshold parameters that vary across items. A marginal maximum likelihood algorithm is implemented to estimate GGUM item parameters, whereas person parameters are derived from an expected a posteriori technique. Recovery simulations suggest that accurate item parameter estimates can be obtained with approximately 750 subjects. Additionally, accurate person estimates are derived with approximately 20 6-category items. The applicability of the GGUM to common attitude testing situations is illustrated with real data on student attitudes toward abortion. Index terms: attitude measurement, unfolding model, item response theory, graded unfolding model, generalized graded unfolding model, Thurstone scale, Likert scale.
Several researchers (Andrich, 1996; Roberts, 1995; Roberts, Laughlin & Wedell, 1997; van Schuur & Kiers, 1994) have recently argued that binary or graded disagree-agree responses to attitude statements generally result from an ideal point process (Coombs, 1964) in which an individual endorses an attitude statement to the extent that the sentiment expressed by the statement adequately matches the individual's own opinion. This argument implies that disagree-agree responses are best analyzed with some form of unfolding (i.e., proximity) model that implements a single-peaked response function. (Henceforth, both binary and graded responses will be referred to simply as "disagree-agree responses". The number of response categories should be deduced from the context in which this term is used, unless it is explicitly specified.) Within the context of item response theory, an unfolding model suggests that an individual will agree with a statement to the extent that the individual and the statement are located near each other on an underlying affective continuum - a latent continuum which spans the two poles associated with negative and positive affect.

Several unidimensional item response models are available to unfold disagree-agree responses to attitude statements. Some of these models are appropriate for binary responses whereas others will allow for either binary or graded data. Models for binary data include both parametric (Andrich, 1988; Andrich & Luo, 1993; Desarbo & Hoffman, 1987, Hoijtink, 1990, 1991; Verhelst & Verstralen, 1993) and nonparametric (Cliff, Collins, Zatkin, Gallipeau, & McCormick, 1988; van Schuur, 1984) approaches. Similarly, there are parametric (Andrich, 1996; Roberts & Laughlin, 1996ab) and nonparametric (Cliff et al., 1988; Van Schuur, 1993) models for graded data as well. Although nonparametric models are practical because they make fewer assumptions about the specific form of the item response function, correctly specified parametric models offer
additional measurement advantages. Attitude estimates from correctly specified parametric models are invariant to the actual items used to calibrate the estimates. Additionally, estimates of item locations are independent of the distribution of attitudes in the sample. These two qualities facilitate more complex measurement applications such as item banking and adaptive testing, and thus, are valuable features of parametric models (Hambleton, Swaminathan, & Rogers, 1991; Lord, 1980).

This report will focus on extending parametric item response models for unfolding graded data. Previous models are limited in at least two regards. First, they are based on a rating scale approach in which one assumes that individuals will consistently utilize response categories in an identical manner across all items on an attitude questionnaire. Second, these models assume that the discrimination capabilities of all items are equal. The model presented in this report removes both of these restrictions. Specifically, it allows for differential response category utilization across items, and it permits variable levels of discrimination among items. The model is a generalization of the graded unfolding model (Roberts, 1995; Roberts & Laughlin, 1996ab), and hence, it is referred to as the Generalized Graded Unfolding Model (GGUM).

The Generalized Graded Unfolding Model (GGUM)

The GGUM is developed from a series of four basic premises about the response process. The first premise is that when an individual is asked to express his or her agreement with an attitude statement, then the individual tends to agree with the item to the extent that it is located close to his or her own position on a unidimensional latent attitude continuum. In this context, the degree to which the sentiment of an item reflects the opinion of an individual is given by the proximity of the individual to the item on the attitude continuum. If we let $\delta_i$ denote the position of the $ith$
item on the continuum and let θᵢ denote the location of the jth individual on the continuum, then the individual is more likely to agree with the item to the extent that the distance between θᵢ and δᵢ approaches zero. This is simply a restatement of the fundamental characteristic of an ideal point process (Coombs, 1964).

A second premise of the GGUM is that an individual may respond in a given response category for either of two distinct reasons. As an example, consider an individual with a neutral attitude towards abortion. This individual might strongly disagree with an item that portrays the abortion issue in either a very negative or very positive way. If the item is located far below the individual's position on the attitude continuum (i.e., the item's content is much more negative than the individual's attitude), then we would say that the individual "strongly disagrees from above" the item. In contrast, if the item is located far above the individual's position (i.e., the item's content is much more positive than the individual's attitude), then we would say that the individual "strongly disagrees from below" the item. Hence, there are two possible subjective responses, "strongly disagree from above" and "strongly disagree from below", associated with the single observable response of "strongly disagree". Similarly, the GGUM postulates two subjective responses for each observable response on a rating scale.

The third premise behind the GGUM is that subjective responses to attitude statements follow a cumulative item response model (e.g., Andrich & Luo, 1993). In this paper, we assume that subjective responses follow Muraki's (1992) generalized partial credit model, but other cumulative models could also be used. Muraki's model is used here due to its generality. When applied to subjective responses, the generalized partial credit model is defined as:

\[
Pr[Y_i = y | \theta_j] = \frac{\exp \left( \alpha_i [y (\theta_j - \delta_i) - \sum_{k=0}^{y} \tau_{ik}] \right)}{\sum_{w=0}^{M} \exp \left( \alpha_i [w (\theta_j - \delta_i) - \sum_{k=0}^{w} \tau_{ik}] \right)},
\]

subject to the constraint that:

\[
\sum_{k=0}^{M} \tau_{ik} = 0,
\]

where:

\( Y_i \) = a subjective response to attitude statement \( i \);

\( y = 0, 1, 2, \ldots, M; \ y = 0 \) corresponds to the strongest level of disagreement from below the item whereas \( y = M \) corresponds to the strongest level of disagreement from above the item (see Figure 1);

\( \theta_j \) = the location of individual \( j \) on the attitude continuum,

\( \delta_i \) = the location of attitude statement \( i \) on the attitude continuum,

\( \alpha_i \) = the discrimination of attitude statement \( i \),

\( \tau_{ik} \) = the location of the \( kth \) subjective response category threshold on the attitude continuum relative to the location of the \( ith \) item, and

\( M \) = the number of subjective response categories minus 1.

Note that the value of \( \tau_{in} \) is arbitrarily defined to be zero in equation 1, but it could be set equal to any constant without affecting the resulting probabilities (Muraki, 1992). The model is illustrated in Figure 1 for a hypothetical item with four observable response categories: "strongly disagree", "disagree", "agree", "strongly agree". The abscissa of Figure 1 represents the attitude continuum,
and it is scaled in units of signed distance between an individual's attitude position and the location of the item (i.e., $\theta_j - \delta_i$). The ordinate indexes the probability that an individual's subjective response will fall in one of the 8 possible subjective response categories. [There are 8 subjective response categories and associated probability functions (PFs) due to the fact that an individual may respond in any of the 4 observable response categories because his or her attitudinal position is either above or below the location of the item.] The 7 vertical lines designate the locations where successive subjective response category PFs intersect. These locations are the subjective response category thresholds. In this example, the 7 subjective response category thresholds are successively ordered on the latent continuum. Therefore, these thresholds divide the latent continuum into 8 intervals in which a different subjective response is most likely. The most likely subjective response within each interval is labeled in the figure. Given a particular set of threshold parameters, the dominance of the most likely subjective response within each interval is determined by the discrimination parameter ($\alpha_i$). As this parameter grows more positive, the probability of the most likely response within the interval increasingly dominates the probabilities associated with competing responses.

---

Insert Figure 1 About Here

---

Equation 1 defines an item response model at a subjective response level. However, the model must ultimately be defined in terms of the observable response categories associated with the graded agreement scale. Recall that each observable response category is associated with two possible subjective responses (i.e., one from below the item and one from above the item). Moreover, the two subjective responses corresponding to a given observable response category
are mutually exclusive. Therefore, the probability that an individual will respond using a particular observable category is simply the sum of the probabilities associated with the two corresponding subjective responses:

$$ Pr[Z_i = z \mid \theta_j] = Pr[Y_i = z \mid \theta_j] + Pr[Y_i = (M-z) \mid \theta_j] , $$

(3)

where:

$Z_i =$ an observable response to attitude statement $i$,

$z = 0, 1, 2, \ldots, C$; $z = 0$ corresponds to the strongest level of disagreement and $z = C$ refers to the strongest level of agreement,

$C =$ the number of observable response categories minus 1. Note that $M = 2*C +1$.

The fourth and final premise behind the GGUM is that subjective category thresholds are symmetric about the point $(\theta_j - \delta_i) = 0$, which yields:

$$ \tau_{i(C+1)} = 0 , $$

(4)

and

$$ \tau_{iz} = -\tau_{i(M-z+1)} , \text{ for } z \neq 0 . $$

(5)

At a conceptual level, this premise implies that an individual is just as likely to agree with an item located at either $-h$ units or $+h$ units from the individual's position on the attitude continuum. At an analytical level, this premise leads to the following identity:

$$ \sum_{k=0}^{z} \tau_{ik} = \sum_{k=0}^{M-z} \tau_{ik} . $$

(6)

Incorporating this identity into equation 3 yields the formal definition of the GGUM:
The Generalized Graded Unfolding Model

\[
Pr[Z_i = z | \theta_j] = \frac{\exp \left( \alpha_i [z (\theta_j - \delta_i) - \sum_{k=0}^{z} \tau_{i,k}] \right) + \exp \left( \alpha_i [(M-z)(\theta_j - \delta_i) - \sum_{k=0}^{z} \tau_{i,k}] \right)}{\sum_{w=0}^{C} \left[ \exp \left( \alpha_i [w (\theta_j - \delta_i) - \sum_{k=0}^{w} \tau_{i,k}] \right) + \exp \left( \alpha_i [(M-w)(\theta_j - \delta_i) - \sum_{k=0}^{w} \tau_{i,k}] \right) \right]} \cdot (7)
\]

-------------------------
Insert Figure 2 About Here
-------------------------

The GGUM defines the observable response category PFs associated with the \textit{jth} individual’s objective response to the \textit{i}th item. Figure 2 displays these observable response category PFs for the same hypothetical item referenced in Figure 1. As seen in Figure 2, there is one category PF associated with each observable response available to the individual. Each of these PFs is simply the sum of the two corresponding subjective response category PFs previously shown in Figure 1.

One should note that successive observable response category PFs do not intersect at \(\tau_{11}, \tau_{12}, \ldots, \tau_{1C}\), and therefore, the \(\tau_k\) parameters lose their simple interpretation at the observable score level. Similarly, the \(\alpha_i\) parameters index discrimination at the subjective response level. In contrast, the substantive meaning of both \(\theta_j\) and \(\delta_i\) remains unchanged when moving from a subjective score level to an observable score level. A second point with regard to the \(\tau_k\) parameters is they need not be successively ordered on the latent continuum. These values simply indicate where successive subjective response category PFs intersect, and they have little substantive meaning beyond that. It is necessary, however, that the maximum values associated with successive subjective response category PFs be ordered on the continuum, and this feature is guaranteed by the cumulative model associated with subjective responses. In practice, disordinal thresholds will occur whenever one or more observable response categories is used infrequently.
The Generalized Graded Unfolding Model 10.

by subjects, and our experience suggests that this situation will be encountered often.

-----------------------------------------

Insert Figure 3 About Here

-----------------------------------------

The GGUM is an unfolding model of the response process. This is easily seen by computing the expected value of an observable response for various values of $\theta_j - \delta_i$ using the probability function given in equation 7. Figure 3 portrays the expected value of an objective response for the same hypothetical item with 4 response categories. The categories are coded with the integers 0 to 3 where the codes correspond to the responses of "strongly disagree", "disagree", "agree" and "strongly agree", respectively. As seen in Figure 3, the item elicits greater levels of agreement as the distance between the individual and the item on the attitude continuum decreases.

The Effects of Discrimination and Threshold Parameters on the Expected Value Function

The shape of the expected value function in the GGUM is jointly determined by both the $\alpha_i$ and $\tau_{i,k}$ parameters. Figure 4 portrays the effects of $\alpha_i$ on the expected value function for a 3-category response while holding the $\tau_{i,k}$ values constant. (These values were set to $\tau_{i,1} = -2$ and $\tau_{i,2} = -1$.) Each panel in the figure depicts the function for a different value of $\alpha_i$, ranging from .5 to 30. As $\alpha_i$ increases, the maximum value of the function approaches its upper bound, and the function simultaneously becomes more peaked. When $\alpha_i$ increases without a limit, the expected value function approaches a Guttman-like step function in which the response becomes totally determined by the distance between $\theta_j$ and $\delta_i$.

-----------------------------------------

Insert Figure 4 About Here

-----------------------------------------
The effects of alternative $\tau_{ik}$ values on the shape of the expected value function are illustrated in Figure 5. The function again corresponds to the case of a 3-category response scale, and thus, there are two estimated $\tau_{ik}$ parameters in the GGUM. These parameters have been chosen so that the interthreshold distance is equal across the latent continuum, and this distance is varied from .25 to 1.5 across the four illustrations in Figure 5. The value of $\alpha_i$ has been set equal to 1 across all panels of the figure. As the interthreshold distance increases, the maximum expected value approaches its upper bound, yet the function becomes less steep. This second feature is the opposite of that found when $\alpha_i$ increases.

Insert Figure 5 About Here

---

**Item Parameter Estimation**

Item parameters of the GGUM are estimated using a marginal maximum likelihood approach (Bock & Lieberman, 1970; Bock & Aitkin, 1981). The solution algorithm parallels Muraki's (1992) procedure used in the generalized rating scale model and is based on an expectation-maximization EM strategy. Let $X_s$ be one of $S$ distinct response vectors for a given set of data with $s = 1, 2, ..., S$. Let $x_{si}$ refer to the $ith$ element of $X_s$. Under the assumption of local independence, the conditional probability of observing a particular response vector $s$ given $\theta$ is equal to:

$$Pr[X_s \mid \theta ] = \prod_{i=1}^{I} Pr[Z_i = x_{si} \mid \theta ] . \tag{8}$$

If subjects are sampled from a population with a continuous attitude distribution, denoted as $g(\theta)$, then the marginal probability of observing a particular response pattern, $X_s$, is equal to:
The Generalized Graded Unfolding Model 12.

\[ Pr[X_s] = \int_{-\infty}^{+\infty} Pr[X_s | \theta] g(\theta) \, d\theta. \quad (9) \]

Let \( r_s \) denote the number of subjects with response pattern \( X_s \), and let \( N \) equal the number of subjects in the sample. In this situation, \( r_s \) is multinomially distributed with parameters \( N \) and \( Pr[X_s] \), and the likelihood function is equal to:

\[ L = \frac{N!}{\prod_{s=1}^{S} r_s!} \prod_{s=1}^{S} \left[ Pr[X_s] \right]^{r_s} \quad (10) \]

The log likelihood function is therefore equal to:

\[ \ln(L) = \ln(N!) - \sum_{s=1}^{S} \ln(r_s!) + \sum_{s=1}^{S} r_s \ln(Pr[X_s]) \quad (11) \]

The likelihood equations for \( \alpha_n, \delta_n, \) and \( \tau_{ik} \) are obtained by calculating the first-order partial derivatives of Equation 11 with respect to each parameter and then setting these derivatives equal to 0. The values of \( \alpha_n, \delta_n, \) and \( \tau_{ik} \) which solve these equations are the marginal maximum likelihood estimates.

The general form of the first-order partial derivative of the log likelihood function with respect to a particular item parameter, \( \phi_i \), is given by:
\[ \frac{\partial \ln(L)}{\partial \phi_i} = \sum_{s=1}^{S} \frac{r_s}{Pr[X_s]} \frac{\partial Pr[X_s]}{\partial \phi_i} \]

\[ = \sum_{s=1}^{S} \frac{r_s}{Pr[X_s]} \int \frac{\partial Pr[Z_i = x_{s_i} | \theta]}{\partial \phi_i} \prod_{i'=1}^{I} Pr[Z_{i'} = x_{s_{i'}} | \theta] \, g(\theta) \, d\theta \]

\[ = \sum_{s=1}^{S} \frac{r_s}{Pr[X_s]} \int \frac{\partial Pr[Z_i = x_{s_i} | \theta]}{\partial \phi_i} \left( \prod_{i=1}^{I} Pr[Z_i = x_{s_i} | \theta] \frac{g(\theta)}{Pr[Z_i = x_{s_i} | \theta]} \right) \, d\theta \]

\[ = \sum_{s=1}^{S} \frac{r_s}{Pr[X_s]} \int \frac{\partial Pr[Z_i = x_{s_i} | \theta]}{\partial \phi_i} Pr[X_s | \theta] \frac{g(\theta)}{Pr[Z_i = x_{s_i} | \theta]} \, d\theta . \]

Equation 12 may be approximated using Gauss-Hermite quadrature as follows:

\[ \frac{\partial \ln(L)}{\partial \phi_i} = \sum_{f=1}^{F} \sum_{s=1}^{S} \frac{r_s L_s(V_f) \, A(V_f)}{\tilde{P}_s} \frac{\partial Pr[Z_i = x_{s_i} | V_f]}{\partial \phi_i} \frac{1}{Pr[Z_i = x_{s_i} | V_f]} \]

\[ = \sum_{f=1}^{F} \sum_{z=0}^{C} \sum_{s=1}^{S} \frac{H_{siz} \, r_s L_s(V_f) \, A(V_f)}{\tilde{P}_s} \frac{\partial Pr[Z_i = z | V_f]}{\partial \phi_i} \frac{1}{Pr[Z_i = z | V_f]} \]

\[ = \sum_{f=1}^{F} \sum_{z=0}^{C} \frac{\tilde{r}_{izf}}{Pr[Z_i = z | V_f]} \frac{\partial Pr[Z_i = z | V_f]}{\partial \phi_i} , \]

where:

\[ L_s(V_f) = \prod_{i=1}^{I} Pr[Z_i = x_{s_i} | V_f] , \]

\[ \tilde{P}_s = \sum_{f=1}^{F} L_s(V_f) \, A(V_f) , \]

\[ \tilde{r}_{izf} = \frac{\sum_{s=1}^{S} H_{siz} \, r_s L_s(V_f) \, A(V_f)}{\tilde{P}_s} . \]
and

\( H_{iiz} \) is a dummy variable that is equal to 1 when \( z \) equals \( x_i \) and is equal to 0 otherwise. In

Equation 13, \( V_f \) is a quadrature point (Stroud & Secrest, 1966), and \( A(V_f) \) is the rescaled
standard normal density at \( V_f \). The scale of the \( A(V_f) \) values is such that:

\[
\sum_{f=1}^{F} A(V_f) = 1.
\]  (17)

Additionally, \( L_s(V_f) \) is the conditional probability of response pattern \( X_s \) at quadrature point \( V_f \),

\( \tilde{P}_s \) is the marginal probability of response pattern \( X_s \), and \( \tilde{r}_{izf} \) is the expected frequency of
response \( z \) for item \( i \) at quadrature point \( V_f \). Equation 13 includes a parameter specific
component, \( \partial Pr[Z_i = z | V_f] / \partial \phi_i \), which must be evaluated separately for each parameter in
order to compute the associated first-order partial derivative. The derivation of this component is
given in Appendix A for each item parameter.

An EM algorithm described by Muraki (1992) is used to solve the likelihood equations for \( \alpha_o \),
\( \delta_o \), and \( \tau_{ik} \). In the expectation stage of the algorithm, estimates of the \( \tilde{r}_{izf} \) quantities are
calculated from the observed responses and the provisional item parameter estimates. In the
maximization stage of the algorithm, the \( \tilde{r}_{izf} \) estimates are treated as known constants, and then
the likelihood equations are solved. Because the \( \tilde{r}_{izf} \) estimates are fixed, it is possible solve the
likelihood equations for each item individually. The maximization stage continues until the most
likely item parameter estimates for all items have been computed for a given \( \tilde{r}_{izf} \). The
completion of a single expectation stage followed by a single maximization stage constitutes one
cycle within the EM algorithm. Additional cycles are conducted until the largest change in any
item parameter estimate from one cycle to the next is arbitrarily small (e.g., less than .0005).
The Generalized Graded Unfolding Model.

The maximization stage of the EM algorithm proceeds in two steps. In the first step, the likelihood equations associated with the \( \tau_{ik} \) parameters are solved for each item individually. The solution is computed using Fisher's method of scoring, and thus, the information matrix for the \( \tau_{ik} \) parameters is required for each item. The information matrix for a given item, \( i \), is denoted as:

\[
I_{\tau(i)} = \begin{bmatrix}
\tilde{I}_{\tau_{i1} \tau_{i1}} & \tilde{I}_{\tau_{i1} \tau_{i2}} & \cdots & \tilde{I}_{\tau_{i1} \tau_{ic}} \\
\tilde{I}_{\tau_{i2} \tau_{i1}} & \tilde{I}_{\tau_{i2} \tau_{i2}} & \cdots & \tilde{I}_{\tau_{i2} \tau_{ic}} \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{I}_{\tau_{ic} \tau_{i1}} & \tilde{I}_{\tau_{ic} \tau_{i2}} & \cdots & \tilde{I}_{\tau_{ic} \tau_{ic}}
\end{bmatrix}
\]  

(18)

The elements of the information matrix are derived in Rao (1973) and are equal to:

\[
\tilde{I}_{\tau_{ik} \tau_{ik'}} = \sum_{f=1}^{F} \tilde{N}_{if} \sum_{z=0}^{C} \frac{1}{Pr[Z_i = z | V_f]} \frac{\partial Pr[Z_i = z | V_f]}{\partial \tau_{ik}} \frac{\partial Pr[Z_i = z | V_f]}{\partial \tau_{ik'}}
\]

(19)

where \( \tilde{N}_{if} \) is the expected number of persons at quadrature point \( V_f \) who responded to item \( i \):

\[
\tilde{N}_{if} = \sum_{z=0}^{C} \tilde{f}_{izf}.
\]

(20)

The value of \( \tilde{N}_{if} \) is calculated in the expectation stage of the algorithm and is held constant during the maximization stage.

In the method of scoring, the update function used to calculate \( \tau_{ik} \) parameters on the \( qth \) iteration is given by:
The Generalized Graded Unfolding Model 16

\[
\begin{bmatrix}
\tau_{11} \\
\tau_{12} \\
\vdots \\
\tau_{1C_q}
\end{bmatrix}
=\
\begin{bmatrix}
\tau_{11} \\
\tau_{12} \\
\vdots \\
\tau_{1C_{q-1}}
\end{bmatrix}
+ \left[ \mathcal{I}_{\tau(0)} \right]^{-1}
\begin{bmatrix}
\frac{\partial \ln(L)}{\partial \tau_{11}} \\
\frac{\partial \ln(L)}{\partial \tau_{12}} \\
\vdots \\
\frac{\partial \ln(L)}{\partial \tau_{1C_q}}
\end{bmatrix}
\tag{21}
\]

The $\tau_{ik}$ parameters for a given item are updated in an iterative fashion until there is little change in parameters from one iteration to the next or until some maximum limit of iterations has been reached (e.g., 30 iterations).

In the second step of the maximization stage, the likelihood equations for the $\alpha_i$ and $\delta_i$ parameters are solved for each item individually. The solution is, again, computed using the method of scoring, and the information matrix required in the solution is denoted as:

\[
\mathcal{I}_{\alpha\beta (i)} = \begin{bmatrix}
\tilde{I}_{\alpha_i\alpha_i} & \tilde{I}_{\alpha_i\delta_i} \\
\tilde{I}_{\delta_i\alpha_i} & \tilde{I}_{\delta_i\delta_i}
\end{bmatrix}
\tag{22}
\]

The elements of this matrix are derived in Rao (1973) and are equal to:

\[
\tilde{I}_{\phi_i\phi'_i} = \sum_{f=1}^{F} \sum_{z=0}^{C} \frac{1}{Pr[Z_i = z | V_f]} \frac{\partial Pr[Z_i = z | V_f]}{\partial \phi_i} \frac{\partial Pr[Z_i = z | V_f]}{\partial \phi'_i}
\tag{23}
\]

where $\phi_i$ and $\phi'_i$ denote either $\alpha_i$ or $\delta_i$ in a general manner. The parameters are updated in an iterative fashion, and the update equation for the $qth$ iteration is given by:
The Generalized Graded Unfolding Model 17.

\[
\begin{bmatrix}
\alpha_i \\
\delta_i
\end{bmatrix}_q = \begin{bmatrix}
\alpha_i \\
\delta_i
\end{bmatrix}_{q-1} + \begin{bmatrix}
\frac{\partial \ln(L)}{\partial \alpha_i} \\
\frac{\partial \ln(L)}{\partial \delta_i}
\end{bmatrix} \tilde{I}_{\alpha \delta(i)}^{-1}
\]

(24)

The \( \alpha_i \) and \( \delta_i \) parameters for a given item are updated iteratively until there is little change in parameters from one iteration to the next or until some maximum limit of iterations has been reached.

The two-steps of the maximization stage are performed repeatedly until there is little change in any item parameter estimate from one repetition to the next or until 10 repetitions have been performed. [This two-step maximization procedure is essential when a constant set of subjective category thresholds (\( \tau_k \)) is estimated across all items as is the case for the Graded Unfolding Model (Roberts, 1995; Roberts & Laughlin, 1996ab). In the current model, however, one could easily solve the likelihood equations associated with all the parameters for a given item (i.e., \( \alpha_i \), \( \delta_i \), and \( \tau_{ik} \)) in a single maximization step. Nonetheless, the two-step procedure is maintained here in an effort to promote consistency in the solution algorithm across models.] The conclusion of the maximization stage constitutes the end of a given EM cycle. The stability of parameter estimates is evaluated at the end of each EM cycle, and additional cycles are performed if needed.

Initial Item Parameter Values

The EM algorithm requires a judicious choice of initial item parameter values in order to avoid local maxima. In practice, these “start-up” values are obtained by estimating item parameters from constrained versions of the GGUM. For example, the Graded Unfolding Model can be used to produce \( \delta_i \) and \( \tau_k \) estimates under the assumptions that subjective category thresholds are
equal across items and all $\alpha_i$ are equal to 1. [Start-up values for the Graded Unfolding Model are given in Roberts and Laughlin (1996ab)]. The $\delta_i$ and $\tau_{ik}$ estimates produced with the Graded Unfolding Model can then be used as start-up values in a less restrictive model which allows $\tau_{ik}$ parameters to vary across items. These estimates of $\delta_i$ and $\tau_{ik}$ can then be used as start-up values when estimating parameters of the full GGUM which includes $\alpha_i$. Simulations have indicated that this sequential generation of start-up values provides adequate input to the estimation algorithm.

**Person Parameter Estimation**

The marginal maximum likelihood estimates of item parameters are used in conjunction with the observed responses to derive person parameter estimates. These person parameter estimates constitute the individual “attitude estimates”. In this study, person parameter estimates are obtained using an expected a posteriori (EAP) procedure in which the estimate for the $j$th individual is calculated as:

$$\hat{\theta}_j = \frac{\sum_{f=1}^{F} V_f L_f(V_f) A(V_f)}{\sum_{f=1}^{F} L_f(V_f) A(V_f)}$$

(25)

where $L_f(V_f)$ is the conditional likelihood of observing the $j$th individual’s response vector given that the individual is located at quadrature point $V_f$. The EAP estimate, $\hat{\theta}_j$, is the posterior mean of the $\theta$ distribution for the $j$th individual given the individual’s response vector. It is advantageous because it exists for any response pattern, and its average error in the population is smaller than that for any other estimator (Bock & Mislevy, 1982).
Recovery of Model Parameters

Method

The recovery of GGUM item and person parameters was simulated under 36 alternative conditions that were derived by factorially combining six levels of sample size with six levels of test length. The number of subjects analyzed in these conditions was either 200, 300, 500, 750, 1000 or 2000, and responses to either 5, 10, 15, 20, 25 or 30 attitude items were generated. Items were always located at equally distant positions on the latent continuum and always ranged from -2.0 to 2.0, regardless of the number of items studied. True item discrimination parameters were generated from a uniform distribution which spanned the interval of (.5, 2), and true person parameters were randomly sampled from a normal distribution with \( \mu = 0 \) and \( \sigma = 1.0 \).

The threshold parameters were generated independently for each item. For a given item, the true \( \tau_{iC} \) parameter was generated from a uniform \((-1.4, -0.4)\) distribution. Successive true \( \tau_{ik} \) parameters were then generated with the following recursive equation:

\[
\tau_{i,k-1} = \tau_{i,k} - .25 + e_{i,k-1}, \quad \text{for } k = 2, 3, \ldots, C
\]

(26)

where \( e_{i,k-1} \) denotes a random error term generated from a \( N(0, .04) \) distribution. The \( \tau_{ik} \) parameters derived with this formula were not consistently ordered across the continuum for each item, and they were similar to parameter estimates obtained in preliminary analyses of real data.

The observable response simulated for each person-item combination was on a 6-point scale (e.g., "strongly disagree", "disagree", "slightly disagree", "slightly agree", "agree" and "strongly agree"). Six response probabilities were computed with equation 7 and subsequently used to divide a probability interval (i.e., a closed interval between 0 and 1) into 6 mutually exclusive and
exhaustive segments, where each segment corresponded to a particular observable response category. A random number was then generated from a uniform probability distribution, and the simulated response was that response associated with the probability segment in which the random number fell. After an observable response to each item had been generated for all subjects, the data were used to estimate GUM parameters. The process of generating data and subsequently estimating parameters was replicated 30 times in each condition, and the true values of all parameters remained constant across replications.

Measures of Estimation Accuracy

Three measures of estimation accuracy were investigated in this study. The Root Mean Squared Error (RMSE) was the first of these measures. The RMSE provided an index of the average unsigned discrepancy between a set of true parameters and a corresponding set of estimates. The RMSE was calculated across all the parameters of a given type in any single replication. For example, the RMSE of item location estimates from a particular replication was computed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{I} (\delta_i - \hat{\delta}_i)^2}{I}}, \quad (27)$$

where:

$$\delta_i = \text{the true location of the } ith \text{ item on the attitude continuum},$$

$$\hat{\delta}_i = \text{the estimated location for the } ith \text{ item on the attitude continuum},$$

and

$$I = \text{the number of items on the test}.$$

Analogous quantities were computed for the item discrimination, subjective category threshold
and person parameters in a given replication.

The Pearson correlation between estimated and true parameters was the second measure of accuracy utilized in the investigation. This correlation was computed across a given set of parameters (i.e., $\alpha_i$, $\delta_i$, $\tau_{ik}$, or $\theta_j$) within each replication. Therefore, it provided a simple index of the degree of linearity between estimated and true parameter values.

The third and final measure of accuracy was the average absolute discrepancy between estimated and true expected value functions on the $\theta$ interval of $[-3, 3]$. This measure was calculated by integrating the absolute difference between the expected value function of the $i$th item computed with true parameters and that computed with estimated parameters, and then dividing the result by 6 (i.e., the length of the evaluation interval):

$$D_i = \frac{1}{6} \sum_{z=0}^{3} z \left| \hat{Pr}[Z_i = z \mid \theta] - Pr[Z_i = z \mid \theta] \right| d\theta. \tag{28}$$

Note that $\hat{Pr}[Z_i = z \mid \theta]$ and $Pr[Z_i = z \mid \theta]$ represent the value of Equation 7 when calculated from estimated and true item parameters, respectively. The integral in equation 28 was evaluated numerically for each item using a globally adaptive scheme based on Gauss-Kronrod rules (Piessens, deDoncker-Kapenga, Uberhuber, Kahaner, 1983), and then the resulting $D_i$ values were averaged across all the items within a given replication. The resulting measure (denoted simply as $D$) provided an index of the average similarity between the estimated expected value functions and the true functions for a given set of items. Some researchers have argued that an index like $D$ is the most useful measure of overall estimation accuracy because item parameter estimates can be substantially different from true values yet still yield expected value functions that are quite similar to their corresponding true functions (Hulin, Lissak & Drasgow, 1982; Linn, Levine, Hastings &
Results

Analysis of RMSE and $r$ Measures

Univariate ANOVAs were performed separately for the RMSE and $r$ values associated with a given parameter. The 6 levels of sample size and 6 levels of test length served as the two between-subjects factors in these analyses, and there were 30 replications in each cell of the corresponding factorial design. Both main effects and the two-way interaction were consistently significant ($p < .0001$) regardless of the accuracy measure or parameter under consideration. The proportion of the corrected total sum of squares ($\eta^2$ value) associated with each effect in a given analysis is shown in Table 1. The $\eta^2$ values for item parameter estimates indicated that both the RMSE and $r$ indices were primarily a function of sample size. Test length and its interaction with sample size had relatively less influence on these measures. In contrast, the RMSE and $r$ values associated with $\theta_j$ estimates were overwhelmingly a function of test length.

---------------------------
Insert Table 1 About Here
---------------------------

The average RMSE and average $r$ values obtained with a given sample size are plotted for each item parameter in Figure 6. The average RMSE generally decreased as the sample size increased, regardless of the item parameter in question. However, the magnitude of the average RMSE stabilized by the time the sample size reached 750, and relatively little change was observed with larger samples. With 750 simulees, the average RMSE value for $\delta$ values was equal to 0.104 which represented 8% of the standard deviation of true $\delta$ parameters. The corresponding average RMSE values for $\alpha$ and $\epsilon$ were equal to 0.113 and 0.174, and these values
represented 25% and 38% of the standard deviation associated with each set of true parameters, respectively.

The average correlation coefficients shown in Figure 6 consistently increased for each parameter as the sample size grew larger. However, the relative size of these increases were generally small after the sample size reached 750. With 750 simulees, the average correlation between true and estimated parameters was equal to .998, .971 and .925 for $\delta$, $\alpha$, and $\hat{\tau}$ respectively.

Figure 7 portrays the average RMSE and average correlation between estimated and true $\theta$ parameters as a function of test length. The average RMSE consistently decreased as the test length grew larger, whereas the average correlation consistently increased. Moreover, there was no sample size beyond which these quantities stabilized. However, when the test length was increased to 20 items, the average RMSE fell to .188 which represented 19% of the average standard deviation of true $\theta$ parameters. The average correlation coefficient was equal to .984 at this point, and thus, the estimates were almost completely a linear function of the true values once the test length reached 20 items.

\textit{Analysis of Average Absolute Difference Measures}

A 6 x 6 ANOVA was conducted to assess the main effects and the interaction effect of test length and sample size on the average absolute difference between true and estimated expected
value functions. The main effects and the interaction were all highly significant \((p < .0001)\), but the \(\eta^2\) values for each effect \((\eta^2_T = .02, \eta^2_S = .73, \text{ and } \eta^2_{TS} = .13)\) suggested that sample size was the primary determinant underlying the average absolute difference. The average absolute difference measure is portrayed as a function of sample size in Figure 8. This measure generally decreased as the sample size increased, although decreases began to attenuate once the sample size reached 750 simulees. At that point, the average absolute difference between true and estimated expected value functions was equal to only .082, and thus, the item parameter estimates reproduced the expected value functions quite well.

-----------------------------------
Insert Figure 8 About Here
-----------------------------------

Discussion

The analyses of accuracy measures suggest that reasonable model parameter estimates can be obtained whenever 750 or more examinees respond to approximately 20 or more 6-category items. These data demands are well within the range of many moderately large-scale testing programs. Moreover, if item parameter estimates were available, then the model would also be applicable in many small-scale attitude measurement situations. For example, if GGUM item parameter estimates were available for standardized attitude questionnaires, then attitude estimates for any number of respondents could be calculated using the EAP method. The resulting estimates would be reasonably precise whenever 20 or more uniformly distributed items with 6 or more response categories were used to calculate the estimates. Obviously, this logic depends implicitly on the availability of suitable item parameter estimates, but these estimates
could be obtained for any published attitude questionnaire for which a large norming sample was available.

These results also suggest that \( \alpha_i \) and \( \delta_i \) parameters are estimated more precisely than are \( \tau_{ik} \) parameters. However, all the parameters were estimated reasonably well once the sample size reached 750, and at that point, the average expected value function developed from the estimates was very accurate. Perhaps fewer than 750 examinees might be required if accuracy does not increase linearly between the ranges of 500 and 750 subjects, but this conjecture remains to be studied.

**Standard Errors of Parameter Estimates**

The approximate posterior standard deviation of EAP person parameter estimates has been given by Bock and Mislevy (1982) as:

\[
\sigma_{\hat{\theta}_j} = \sqrt{\frac{\sum_{f=1}^{F} (V_f - \hat{\theta}_j)^2 L_j(V_f) A(V_f)}{\sum_{f=1}^{F} L_j(V_f) A(V_f)}}
\]

where \( V_f \), \( L_j(V_f) \), and \( A(V_f) \) are defined as in Equation 25. These approximations are easily computed at the end of the EM algorithm.

Estimates of item parameter standard errors are derived following the logic implemented in BILOG (Mislevy and Bock, 1990). First, the item parameter information matrix described by Bock and Lieberman (1970) is approximated by considering only the response patterns observed in the sample. This estimate of the information matrix is based on all item parameters and contains \( I \times (C + 2) \times I \times (C + 2) \) elements. The element in the \( pth \) row and the \( qth \) column of the
The Generalized Graded Unfolding Model

The information matrix is given by:

\[
\hat{I}_{pq} = N \sum_{s=1}^{S} \frac{P_{OBS_s}}{\hat{p}_s^2} \left( \begin{array}{cc}
\frac{\partial \hat{p}_s}{\partial \phi_p} & \frac{\partial \hat{p}_s}{\partial \phi_q}
\end{array} \right), \quad \text{for} \quad p = 1, 2, \ldots, I(2 + C), \quad q = 1, 2, \ldots, I(2 + C),
\]

(30)

where

\( P_{OBS_s} \) is the relative frequency of the observed response pattern \( s \),

\( \phi_p \) is an arbitrary parameter \((\alpha_i, \delta_i, \tau_i)\) for one of the \( I \) items,

\( \phi_q \) is an arbitrary parameter \((\alpha_i, \delta_i, \tau_i)\) for one of the \( I \) items, and

\( \hat{p}_s \) is the marginal likelihood of observing response pattern \( s \) as defined in Equation 15.

The matrix defined by Equation 30 can get very large as the number of items grows, and thus, it can be computationally difficult and time consuming to compute and subsequently invert. A more efficient approximation of the information matrix can be derived by considering only those derivatives that correspond to a single item (i.e., the block diagonal elements of the information matrix when the \( p \) rows of the information matrix are ordered by items):

\[
\hat{I}_{\text{adj}(i)} =
\begin{bmatrix}
\hat{I}_{\alpha_i \alpha_i} & \hat{I}_{\alpha_i \delta_i} & \hat{I}_{\alpha_i \tau_{i1}} & \ldots & \hat{I}_{\alpha_i \tau_{ic}} \\
\hat{I}_{\delta_i \alpha_i} & \hat{I}_{\delta_i \delta_i} & \hat{I}_{\delta_i \tau_{i1}} & \ldots & \hat{I}_{\delta_i \tau_{ic}} \\
\hat{I}_{\tau_{i1} \alpha_i} & \hat{I}_{\tau_{i1} \delta_i} & \hat{I}_{\tau_{i1} \tau_{i1}} & \ldots & \hat{I}_{\tau_{i1} \tau_{ic}} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\hat{I}_{\tau_{ic} \alpha_i} & \hat{I}_{\tau_{ic} \delta_i} & \hat{I}_{\tau_{ic} \tau_{i1}} & \ldots & \hat{I}_{\tau_{ic} \tau_{ic}}
\end{bmatrix}
\]

(31)

This approximation is reasonable whenever the number of items is large (greater than 20) due to the fact that the cross derivatives between items will generally be small (Mislevy & Bock, 1990).
The individual elements of $\hat{I}_{a\theta(t)}$ are given by:

$$
\hat{I}_{\phi_i, \omega_i} = N \sum_{s=1}^{S} \frac{P_{\text{obs}}}{\tilde{p}_s^2} \left( \frac{\partial \tilde{p}_s}{\partial \phi_i} \frac{\partial \tilde{p}_s}{\partial \omega_i} \right),
$$

where $\phi_i$ and $\omega_i$ refer generally to distinct pairs of item parameters for item $i$. The general derivative in Equation 32 is equal to:

$$
\frac{\partial \tilde{p}_s}{\partial \phi_i} = \sum_{f=1}^{F} A(V_f) \frac{\partial L_s(V_f)}{\partial \phi_i}
= \sum_{f=1}^{F} A(V_f) \frac{\partial \left( \prod_{i=1}^{I} \frac{1}{\text{Pr}[Z_i = x_{s_i} | V_f]} \right)}{\partial \phi_i}
= \sum_{f=1}^{F} A(V_f) \frac{\partial \text{Pr}[Z_i = x_{s_i} | V_f]}{\partial \phi_i} \left( \prod_{i=1}^{I} \frac{1}{\text{Pr}[Z_i = x_{s_i} | V_f]} \right)
= \sum_{f=1}^{F} L_s(V_f) \frac{\text{Pr}[Z_i = x_{s_i} | V_f]}{A(V_f)} \frac{\partial \text{Pr}[Z_i = x_{s_i} | V_f]}{\partial \phi_i}.
$$

The quantity $\frac{\partial \text{Pr}[Z_i = x_{s_i} | V_f]}{\partial \phi_i}$, referred to in Equation 33, can be calculated for each item parameter of interest using the equations given in Appendix A. With the approximated information matrix for a given set of item parameters computed, one can then derive the approximate standard errors for those parameters as:

$$
\delta_{\phi_i} = \left[ \text{DIAG}(\hat{I}_{a\theta(t)}) \right]^{1/2},
$$
The Item and Test Information Functions

The item information function for the GGUM is equal to:

\[ I_i(\theta_j) = -E \left[ \frac{\partial^2 \ln(L)}{\partial \theta_j^2} \right] \]

\[ = \alpha_i^2 \left[ \sum_{z=0}^C \left( P(Z_i = z) \sigma_{y_i|\theta_j,z}^2 \right) - \sigma_{y_i|\theta_j}^2 \right], \tag{35} \]

and the corresponding test information function is equal to:

\[ I(\theta_j) = \sum_{i=1}^I I_i(\theta_j) \]

\[ = \sum_{i=1}^I \alpha_i^2 \left[ \sum_{z=0}^C \left( P(Z_i = z) \sigma_{y_i|\theta_j,z}^2 \right) - \sigma_{y_i|\theta_j}^2 \right], \tag{36} \]

where

\( \sigma_{y_i|\theta_j,z}^2 \) is the conditional variance of the \( j \)th individual's subjective response to item \( i \) given the individual's observable response to item \( i \), and

\( \sigma_{y_i|\theta_j}^2 \) is the variance of the \( j \)th individual's subjective response to item \( i \).

Details on the derivation of equations 35 and 36 are given in Appendix B.

----------------------------------------------

Insert Figure 9 About Here

----------------------------------------------

Figure 9 illustrates the item information function for a hypothetical item with six response categories. The item is located at \( \delta_i = 0 \) and has an associated \( \alpha_i = 1.0 \). The upper panel of Figure 9 shows how item information varies with corresponding changes in the distance between equally spaced \( \tau_{ik} \) values. The distance between successive \( \tau_{ik} \) values (i.e., the interthreshold distance which is denoted as \( \Psi \)) is held constant within a given curve, but the value of this distance
changes across curves. Specifically, the upper panel portrays the item information functions associated with $\Psi$ values of .2, .4, .6 and 1.0. The information functions are all bimodal and symmetric about the origin. They approach zero whenever $|\theta_j - \delta_i|$ is equal to 0 or is infinitely large itself. Information is highest for smaller values of $\Psi$, and the points (or intervals) on the $\theta_j - \delta_i$ axis at which each maximum occurs becomes more distant from the origin as $\Psi$ increases.

Moreover, the information functions become less peaked as $\Psi$ increases, and thus, the range of $|\theta_j - \delta_i|$ values that yield nearly maximal amounts information gets larger with increasing $\Psi$.

The lower panel of Figure 9 demonstrates how the item information function varies with corresponding changes in $\alpha_i$. The hypothetical item in this figure is again located at $\delta_i = 0$ and has an associated $\Psi = .4$. The curves correspond to alternative $\alpha_i$ values of .5, 1.0, 2.0, and 4.0. As in the upper panel, the item information functions are bimodal, are symmetric about the origin, and approach zero whenever $|\theta_j - \delta_i|$ is equal to 0 or is infinitely large itself. The maximum amount of information is achieved with larger values of $\alpha_i$, and the information functions become more peaked as $\alpha_i$ increases.

When comparing the upper and lower panels of Figure 9, it is apparent that $\alpha_i$ and $\tau_{ik}$ affect the information function in distinctively different ways. The information function becomes larger and more peaked as $\alpha_i$ increases, but it becomes smaller and less peaked as $\Psi$ increases. Thus, maximum measurement precision will be achieved at two symmetric points (or regions) on the latent continuum, and, all other things being equal, items with large discrimination indices and small interthreshold distances will yield the most precision at these points.

**An Example With Real Data**

*Data.* Graded disagree-agree responses to 50 abortion attitude items were obtained from 750
University of South Carolina undergraduates. Students responded to each item using one of six response categories - strongly disagree, disagree, slightly disagree, slightly agree, agree, and strongly agree. Items were presented in a random order for each respondent using a personal computer.

Dimensionality. Davison (1977) has shown that responses from a simple metric unfolding model will exhibit two major principal components, and the component loadings will form a simplex pattern. Simulations of the GGUM have suggested a similar structure with regard to the first two principal components, although the endpoints of the simplex are generally folded inward. Therefore, a principal components analysis was performed to assess which items were least likely to conform to the unidimensionality assumption of the GGUM. Conformability was operationally defined in terms of the item-level communality estimates derived from the first two principal components. An item was discarded if its communality was less than .3. This criterion, although somewhat arbitrary, appeared to be reasonable based on previous simulations. The communality criterion led to the removal of three items from the initial pool.

Selection of Final Scale Items. The 47 items remaining in the item pool were calibrated using the MML algorithm. Item fit was assessed using both infit and outfit chi-square statistics (Wright & Masters, 1982; Linacre & Wright, 1994). Unfortunately, the distributions of these statistics are not known when responses follow the GGUM. Therefore, they were used simply in a heuristic fashion in an effort to identify the most ill-fitting items. No item exhibited misfit that was distinctively worse than other items, and thus, all 47 items remained under consideration for the final scale.

Twenty items were ultimately retained on the final scale. These 20 items were selected on the
basis of their initial calibrations such that all portions of the attitude continuum were reflected in a more or less uniform fashion. The final scale items were recalibrated using the MML algorithm, and EAP person estimates were also derived.

*Item Parameter Estimates.* Table 2 lists the attitude statements from the final scale along with the associated item parameter estimates derived from the GGUM. Statements in the table are listed in order of ascending \( \delta_i \) values. The GGUM ordered the statements in a logical fashion corresponding to negative, neutral, and positive sentiments. Interestingly, there was a distinct quadratic relationship between \( \hat{\delta}_i \) and \( \hat{\tau}_{ik} \) values, for each \( k=1, 2, \ldots, 5 \), such that larger absolute values of \( \hat{\tau}_{ik} \) occurred as \( \hat{\delta}_i \) values became more extreme (i.e., an inverted-U relationship emerged). This relationship suggested that moderate items distinguished among respondents more than extreme items. An examination of the theoretical item characteristic curves derived from the GGUM supported this interpretation. Whether this finding is a general characteristic of attitude items or a particular feature of these data must be verified empirically. The \( \alpha_i \) parameters exhibited a substantial amount of variation and ranged from .6 to 2.0 across items. However, there were no apparent relationships between \( \alpha_i \) parameters and the other item parameters.

Insert Table 2 About Here

*Person Parameter Estimates.* The mean of the \( \hat{\Theta}_j \) distribution was .01 with a standard deviation of .98. The distribution differed significantly from a normal distribution according to the Shapiro-Wilk criterion \( (W=.97, p<.0001) \), and this difference was due to a slightly negative skew (skewness = -.54) and a moderate degree of peakedness (kurtosis = .81). The median \( \hat{\Theta}_j \) value was .13, and this value fell between the statements “My feelings about abortion are very
mixed” and “Abortion should be a woman’s choice, but should never be used as a conventional method of birth control”. Therefore, the typical respondent in this sample possessed a very slight “pro-choice” orientation.

*Global Item Fit.* Fit statistics have not yet been developed for the GGUM. Therefore a descriptive analysis was conducted to gain a better understanding of the global model fit to the data. Specifically, the signed difference between each \((\hat{\theta}_j, \hat{\delta}_i)\) pair was calculated, and then these differences were sorted in ascending order. The sorted differences were divided into 200 roughly homogenous groups of size \(n=75\), and then the mean observed score and mean score expected under the GGUM were calculated within each group. In this way, much of the random variation was averaged out of the observed scores. The relationship between the average observed and expected scores is portrayed in Figure 10 as a function of the corresponding average \(\theta_j - \delta_i\) value corresponding to each of the 200 groups. As shown in the Figure 10, the GGUM arranged both items and individuals on the latent continuum such that more agreement was exhibited when \(|\theta_j - \delta_i|\) became smaller. Additionally, a strong linear relationship was found between mean observed scores and mean expected scores as indicated by a product-moment correlation of .995. (A moderately large correlation, \(r=.82\), existed between observed and expected scores prior to calculating averages.) Thus, the fit of the GGUM to these data appeared reasonable.

---------------------------------------------
Insert Figure 10 About Here
---------------------------------------------

*Differences Among Nested Models.* Constrained versions of the GGUM were developed for the abortion attitude data in which either the \(\alpha_i\) parameters were set equal to 1.0 across items (Model A), the \(\tau_{ik}\) parameters were held constant across items (Model B), or both of these conditions
The Generalized Graded Unfolding Model 33.

were imposed (Model C). Note that all three alternative models are nested versions of the GGUM, and thus, a likelihood ratio statistic can be used to index the incremental model fit (Rao, 1973). This statistic is equal to -2[ln(L*)-ln(L)] where L* and L refer to the marginal likelihood associated with the more constrained model and the less constrained model, respectively. Table 3 summarizes these model comparisons. Models A, B and C all resulted in significantly reduced fit as compared to the GGUM. On a per degree of freedom basis, the reduction in fit resulting from constraints on $\alpha_i$ was roughly similar to that obtained when $\tau_{ik}$ were constrained (e.g., 16.5 versus 18.48). However, when both types of parameters were constrained, the decrease in fit per degrees of freedom was much more pronounced (e.g., 31.0). This suggests that there is some redundancy in the $\alpha_i$ and $\tau_{ik}$ parameters such that misfit due to constraints imposed on one set of parameters can be partially compensated by the removal of constraints on the other set.

Insert Table 3 About Here

Conclusions

The data in the previous example illustrate the fact that attitude items can vary with regard to their discriminability and the ways in which their associated response categories are used. The GGUM generalizes previous IRT models for unfolding disagree-agree responses and provides an opportunity to model both of these phenomena. The example also indicates that the GGUM can fit real data reasonably well, and thus, it can contribute to the measurement arsenal of applied attitude researchers.

The simulations described in this paper suggest that MML estimates of item parameters are
accurate when responses are obtained from approximately 750 subjects, and EAP estimates of individual attitude are quite good when responses to 20 6-category items are obtained. Although these data demands will often exceed the resources of the typical applied researcher, estimates of individual attitudes can be obtained quite easily when published item parameter estimates are available. Therefore, the GGUM has promise in both small-scale and moderately large-scale attitude research programs.

Lastly, the GGUM, being a parametric item response model, holds promise for both item banking and computerized adaptive testing in attitude measurement. However, the item and test information functions that operate in the GGUM are quite different from those found in general cumulative item response models (Donoghue, 1994, Muraki, 1993). Specifically, the GGUM information functions are bimodal and their maximum values are complex functions of the distance between \( \theta_j \) and \( \delta_i \). These features of the model will pose new and interesting challenges for item selection in adaptive testing situations.
References


The Generalized Graded Unfolding Model 39.

Appendix A

The derivatives of the probability function in Equation 7 with respect to parameters \( \alpha_i, \delta_i, \) and \( \tau_{ik} \) are as follows:

\[
\frac{\partial \Pr[Z_i = z \mid V_f]}{\partial \alpha_i} = \frac{[a(z \ t - q) + b ((M-z) \ t - q)] g}{g^2} \left( \sum_{\tilde{w} = 0}^{\tilde{C}} [\tilde{a}(w \ t - \tilde{q}) + \tilde{b} ((M-w) \ t - \tilde{q})] \right) \\
- \frac{(a + b) \left( \sum_{\tilde{w} = 0}^{\tilde{C}} [\tilde{a}(w \ t - \tilde{q}) + \tilde{b} ((M-w) \ t - \tilde{q})] \right)}{g^2} \\
= \Pr[Z_i = z \mid V_f] \frac{a(z \ t - q) + b ((M-z) \ t - q)}{\Pr[Z_i = z \mid V_f] \ g} \left( \sum_{\tilde{w} = 0}^{\tilde{C}} [\tilde{a}(w \ t - \tilde{q}) + \tilde{b} ((M-w) \ t - \tilde{q})] \right) \\
- \frac{(a + b) \left( \sum_{\tilde{w} = 0}^{\tilde{C}} [\tilde{a}(w \ t - \tilde{q}) + \tilde{b} ((M-w) \ t - \tilde{q})] \right)}{g} \\
= \Pr[Z_i = z \mid V_f] \left[ E[Y_i | V_f, Z_i = z] - E(Y_i | V_f) \right] \\
- \Pr[Z_i = z \mid V_f] \left[ E[Y_i | V_f] (t - q) \right] \\
= \Pr[Z_i = z \mid V_f] \left[ t (E[Y_i | V_f, Z_i = z] - E(Y_i | V_f)) + E(\tilde{q} | V_f) - q \right],
\]
\[
\frac{\partial \Pr[Z_i = z | V_f]}{\partial \delta_i} = \frac{\left[a (-\alpha, z) + b (-\alpha, (M - z))\right] g}{g^2} \\
- (a + b) \sum_{w=0}^{C} \frac{\left[\tilde{a} (-\alpha, w) + \tilde{b} (-\alpha, (M - w))\right]}{g^2} \\
= \Pr[Z_i = z | V_f] \frac{\left[a (-\alpha, z) + b (-\alpha, (M - z))\right]}{\Pr[Z_i = z | V_f] g} \\
- (a + b) \sum_{w=0}^{C} \frac{\left[\tilde{a} (-\alpha, w) + \tilde{b} (-\alpha, (M - w))\right]}{g} \\
= \Pr[Z_i = z | V_f] \frac{(-\alpha_i) E[Y_i | V_f, Z_i = z]}{\Pr[Z_i = z | V_f] (\alpha_i) E[Y_i | V_f] - E[Y_i | V_f, Z_i = z]} \\
= \Pr[Z_i = z | V_f] \frac{(-\alpha_i) \left(E[Y_i | V_f] - E[Y_i | V_f, Z_i = z]}{\Pr[Z_i = z | V_f] (\alpha_i) \left(E[Y_i | V_f] - E[Y_i | V_f, Z_i = z]) \\
and
\]

\[
\frac{\partial \Pr[Z_i = z | V_f]}{\partial \tau_{ik}} = \frac{(-\alpha_i U_{zk})(a + b) g}{g^2} \\
- \frac{(a + b) \sum_{w=0}^{C} \left[\tilde{a} (-\alpha, U_{wk}) + \tilde{b} (-\alpha, U_{wk})\right]}{g^2} \\
= \Pr[Z_i = z | V_f] \frac{(-\alpha_i) \left(E[U_{wk} | V_f] - U_{zk}\right)}{\Pr[Z_i = z | V_f] (\alpha_i) \left(E[U_{wk} | V_f] - U_{zk}\right),}
\]

where:

\[
a = \exp \left[\alpha_i \left(z (V_f - \delta_i) - \sum_{k=0}^{C} U_{zk} \tau_{ik}\right)\right],
\]

\[
\tilde{a} = \exp \left[\alpha_i \left(w (V_f - \delta_i) - \sum_{k=0}^{C} U_{wk} \tau_{ik}\right)\right],
\]
The Generalized Graded Unfolding Model 41.

\[ b = \exp \left[ \alpha_i \left( (M - z)(V_f - \delta_i) - \sum_{k=0}^{C} U_{zk} \tau_{i,k} \right) \right], \quad (42) \]

\[ \tilde{b} = \exp \left[ \alpha_i \left( (M - w)(V_f - \delta_i) - \sum_{k=0}^{C} U_{wk} \tau_{i,k} \right) \right], \quad (43) \]

\[ g = \sum_{w=0}^{C} (\tilde{a} + \tilde{b}), \quad (44) \]

\[ q = \sum_{k=0}^{C} U_{zk} \tau_{i,k}, \quad (45) \]

\[ \tilde{q} = \sum_{k=0}^{C} U_{wk} \tau_{i,k}, \quad (46) \]

and

\[ t = V_f - \delta_i. \quad (47) \]

Note that in Equations 39 through 46, \( U_{zk} \) and \( U_{wk} \) are dummy variables that are equal to 1 whenever \( z \leq k \) or \( w \leq k \), respectively, and are otherwise set to 0. Additionally, the term \( E[Y_i \mid V_f, Z_i = z] \) in Equations 37 and 38 represents the conditional expectation of subjective response \( Y_i \) at quadrature point \( V_f \) given an observed response \( z \), whereas \( E[Y_i \mid V_f] \) refers to the unconditional expectation of subjective response \( Y_i \) at quadrature point \( V_f \). The term \( E[\tilde{q} \mid V_f] \) in equation 37 represents the unconditional expectation of \( \tilde{q} \) at quadrature point \( V_f \), and the term \( E[U_{wk} \mid V_f] \) in Equation 39 is the unconditional expectation of variable \( U_{wk} \) at quadrature point \( V_f \).
Appendix B

The probability of observing a particular matrix of responses, $X$, under the GGUM is given by:

$$ Pr[X] = \prod_{j=1}^{N} \prod_{i=1}^{I} Pr[Z_i = x_{ji} | \theta_j], $$

(48)

where:

$Pr[Z_i = x_{ji} | \theta_j]$ is the conditional probability that the $j$th individual will respond to the $i$th item using response $x_{ji}$ given $\theta_j$. The test information function is then defined as:

$$ I(\theta_j) = -E \left[ \frac{\partial^2 \ln(Pr[X])}{\partial \theta_j^2} \right], $$

(49)

where:

$E$ is the expectation operator, and

$$ \ln(Pr[X]) = \sum_{j=1}^{N} \sum_{i=1}^{I} \ln(Pr[Z_i = x_{ji} | \theta_j]). $$

(50)

The first derivative of Equation 50 with respect to $\theta_j$ is equal to:

$$ \frac{\partial \ln(Pr[X])}{\partial \theta_j} = \sum_{i=1}^{I} \frac{\partial Pr[Z_i = x_{ji} | \theta_j]}{\partial \theta_j} \frac{1}{Pr[Z_i = x_{ji} | \theta_j]} \cdot $$

(51)

Let

$$ a = \exp \left[ \alpha_i \left( x_{ji} (\theta_j - \delta_i) - \sum_{k=0}^{V} \tau_{ik} \right) \right], $$

(52)

$$ \bar{a} = \exp \left[ \alpha_i \left( w (\theta_j - \delta_i) - \sum_{k=0}^{w} \tau_{ik} \right) \right], $$

(53)
The Generalized Graded Unfolding Model 43.

\[ b = \exp \left[ \alpha_i \left[ (M - x_{ji}) (\theta_j - \delta_j) - \sum_{k=0}^{x_{ji}} \tau_{ik} \right] \right], \quad (54) \]

\[ \tilde{b} = \exp \left[ \alpha_i \left[ (M - w) (\theta_j - \delta_j) - \sum_{k=0}^{w} \tau_{ik} \right] \right], \quad (55) \]

and

\[ g = \sum_{w=0}^{c} (\tilde{a} + \tilde{b}). \]

Then

\[
\frac{\partial \Pr[Z_i = x_{ji} | \theta_j]}{\partial \theta_j} = \left( \frac{(a' + b') g - (a + b) g'}{g^2} \right)
\]

\[
= \left( \frac{[\alpha_i x_{ji} + \alpha_i (M - x_{ji}) b] g - \sum_{w=0}^{c} [\alpha_i (w) \tilde{a} + \alpha_i (M - w) \tilde{b}] (a + b)}{g^2} \right)
\]

\[= \frac{\alpha_i (x_{ji} a + (M - x_{ji}) b)}{g} \left[ \frac{\alpha_i \sum_{w=0}^{c} [w \tilde{a} + (M - w) \tilde{b}]}{g} \right] \Pr[Z_i = x_{ji} | \theta_j]. \quad (57) \]

The first derivative of Equation 50 with respect to \( \theta_j \) becomes:
\[
\frac{\partial \ln(Pr[X])}{\partial \theta_j} = \sum_{i=1}^{I} \frac{\partial Pr[Z_i = x_{ij} | \theta_j]}{\partial \theta_j} \frac{1}{Pr[Z_i = x_{ij} | \theta_j]}
\]

\[
= \sum_{i=1}^{I} \left[ \alpha_i \left( x_{ij}, a + (M-x_{ij}) b \right) \frac{Pr[Y_i = x_{ij} | \theta_j]}{(g) Pr[Z_i = x_{ij} | \theta_j]} - \frac{\alpha_i \sum_{w=0}^{C} \left[ w \tilde{a} + (M-w) \tilde{b} \right]}{g} \right] 
\]

\[
= \sum_{i=1}^{I} \alpha_i \left[ \frac{x_{ij} Pr[Y_i = x_{ij} | \theta_j]}{(M-x_{ij}) Pr[Y_i = (M-x_{ij}) | \theta_j]} - \sum_{w=0}^{C} \frac{(w Pr[Y_i = w | \theta_j] + (M-w) Pr[Y_i = (M-w) | \theta_j])}{Pr[Z_i = x_{ij} | \theta_j]} \right]
\]

\[
= \sum_{i=1}^{I} \alpha_i \left( E[Y_i | \theta_j, x_{ij}] - E[Y_i | \theta_j] \right),
\]

where:

\[E[Y_i | \theta_j]\] is the expectation of the \(j\)th individual's subjective response to item \(i\), and

\[E[Y_i | \theta_j, x_{ij}]\] is the conditional expectation of the \(j\)th individual's subjective response given that person's observed response. The second derivative of Equation 50 with respect to \(\theta_j\) is then equal to:

\[
\frac{\partial^2 \ln(Pr[X])}{\partial \theta_j^2} = \sum_{i=1}^{I} \alpha_i \left[ \left( x_{ij} a' + (M-x_{ij}) b' \right) (a + b) - \left( x_{ij} a + (M-x_{ij}) b \right) (a' + b') \right] \frac{1}{(a + b)^2}
\]

\[
- \left( \sum_{w=0}^{C} w \tilde{a}' + (M-w) \tilde{b}' \right) g - g' \left( \sum_{w=0}^{C} w \tilde{a} + (M-w) \tilde{b} \right) \right] \frac{1}{g^2}
\]

\[
(59)
\]
\[
\sum_{i=1}^{I} \alpha_i \left[ \frac{\alpha_i x_{ji}^2 (a + \alpha_i (M-x_{ji}))^2 b}{(a + b)} - \left( \frac{x_{ji} a + (M-x_{ji}) b}{(a + b)} \right) \right] \\
- \sum_{w=0}^{C} \left( \frac{\alpha_i w^2 \tilde{a} + \alpha_i (M-w)^2 \tilde{b}}{g} \right) \\
+ \left( \frac{\sum_{w=0}^{C} \left( \alpha_i w \tilde{a} + \alpha_i (M-w) \tilde{b} \right)}{g} \right) \left( \frac{\sum_{w=0}^{C} \left( w \tilde{a} + (M-w) \tilde{b} \right)}{g} \right)
\]

\[
= \sum_{i=1}^{I} \alpha_i \left[ \frac{x_{ji} Pr[Y_i = x_{ji} | \theta_j] + (M-x_{ji})^2 Pr[Y_i = (M-x_{ji}) | \theta_j]}{Pr[Z_i = x_{ji} | \theta_j]} \right] \\
- \alpha_i \left( \frac{x_{ji} Pr[Y_i = x_{ji} | \theta_j] + (M-x_{ji}) Pr[Y_i = (M-x_{ji}) | \theta_j]}{Pr[Z_i = x_{ji} | \theta_j]} \right)^2 \\
- \alpha_i \left( \sum_{w=0}^{C} \left( w^2 Pr[Y_i = w | \theta_j] + (M-w)^2 Pr[Y_i = (M-w) | \theta_j] \right) \right) \\
+ \alpha_i \left( \sum_{w=0}^{C} \left( w Pr[Y_i = w | \theta_j] + (M-w) Pr[Y_i = (M-w) | \theta_j] \right) \right)^2 \\
= \sum_{i=1}^{I} \alpha_i^2 \left[ \left( E[Y_i^2 | \theta_j, x_{ji}] - E[Y_i | \theta_j, x_{ji}] \right)^2 - \left( E[Y_i^2 | \theta_j] - E[Y_i | \theta_j] \right)^2 \right]
\]

\[
= \sum_{i=1}^{I} \alpha_i^2 \left[ \sigma_{Y_i | \theta_j, x_{ji}}^2 - \sigma_{Y_i | \theta_j}^2 \right],
\]

where:

\(\sigma_{Y_i | \theta_j}^2\) is the variance of the \(j\)th individual's subjective response to item \(i\), and

\(\sigma_{Y_i | \theta_j, x_{ji}}^2\) is the conditional variance of the \(j\)th individual's subjective response given that person's observed response. If the expected value of the second derivative is obtained
numerically with respect to \( x_{ji} \), then the test information function can be written as:

\[
I(\theta_j) = -E \left[ \frac{\partial^2 \ln(Pr[ X ])}{\partial \theta_j^2} \right]
\]

\[
= - \sum_{i=1}^{I} \sum_{z=0}^{C} Pr[Z_i = z] \alpha_i^2 \left[ \sigma_{y_i|\theta_j,z}^2 - \sigma_{y_i|\theta_j}^2 \right] - \sum_{i=1}^{I} \alpha_i^2 \left[ \left( \sum_{z=0}^{C} Pr[Z_i = z] \sigma_{y_i|\theta_j,z}^2 \right) - \sigma_{y_i|\theta_j}^2 \right]. \tag{60}
\]

The item information function is simply the term in equation 60 that corresponds to a particular item, \( i \):

\[
I_i(\theta_j) = - \alpha_i^2 \left[ \left( \sum_{z=0}^{C} Pr[Z_i = z] \sigma_{y_i|\theta_j,z}^2 \right) - \sigma_{y_i|\theta_j}^2 \right]. \tag{61}
\]
The Generalized Graded Unfolding Model 47.

Acknowledgments

This research was partially conducted while the first author was a postdoctoral fellow at Educational Testing Service under the direction of the second author. We are indebted to the ETS postdoctoral program, directed by Charles Davis, for its support of this research. We are also grateful to Robert Mislevy for his helpful comments during the many stages this project and to Hank Richardson for proofreading the manuscript.

Author's Address

Send requests for reprints or further information to James S. Roberts, Center for Drug and Alcohol Programs (CDAP), Institute of Psychiatry - 4 North, Medical University of South Carolina, 171 Ashley Avenue, Charleston, SC 29425, U. S. A. Email: roberjam@musc.edu.
Table 1. The Proportion of Variation in Measures of Parameter Estimation Accuracy Accounted for by Test Length ($\eta_{T}^2$), Sample Size ($\eta_{S}^2$) and Their Interaction ($\eta_{TS}^2$).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>RMSE</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\eta_{T}^2$</td>
<td>$\eta_{S}^2$</td>
<td>$\eta_{TS}^2$</td>
</tr>
<tr>
<td>$\delta_i$</td>
<td>.06</td>
<td>.36</td>
<td>.07</td>
</tr>
<tr>
<td>$\alpha_i$</td>
<td>.07</td>
<td>.40</td>
<td>.06</td>
</tr>
<tr>
<td>$\tau_{ik}$</td>
<td>.02</td>
<td>.67</td>
<td>.05</td>
</tr>
<tr>
<td>$\theta_j$</td>
<td>.89</td>
<td>.03</td>
<td>.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\eta_{T}^2$</th>
<th>$\eta_{S}^2$</th>
<th>$\eta_{TS}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_{T}^2$</td>
<td>.02</td>
<td>.28</td>
<td>.05</td>
</tr>
<tr>
<td>$\eta_{S}^2$</td>
<td>.06</td>
<td>.25</td>
<td>.06</td>
</tr>
<tr>
<td>$\eta_{TS}^2$</td>
<td>.01</td>
<td>.61</td>
<td>.17</td>
</tr>
<tr>
<td>$\eta_{T}^2$</td>
<td>.90</td>
<td>.01</td>
<td>.06</td>
</tr>
</tbody>
</table>
### Table 2. GGUM Item Parameter Estimates for 20 Abortion Attitude Statements

<table>
<thead>
<tr>
<th>Statement</th>
<th>$\delta_i$</th>
<th>$\alpha_i$</th>
<th>$\hat{v}_{i1}$</th>
<th>$\hat{v}_{i2}$</th>
<th>$\hat{v}_{i3}$</th>
<th>$\hat{v}_{i4}$</th>
<th>$\hat{v}_{i5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Abortion is unacceptable under any circumstances.</td>
<td>-3.5</td>
<td>1.4</td>
<td>-2.8</td>
<td>-2.9</td>
<td>-2.1</td>
<td>-2.5</td>
<td>-2.2</td>
</tr>
<tr>
<td>2. Abortion is the destruction of one life for the convenience of another.</td>
<td>-3.2</td>
<td>1.1</td>
<td>-3.6</td>
<td>-3.3</td>
<td>-3.3</td>
<td>-2.6</td>
<td>-2.6</td>
</tr>
<tr>
<td>3. Abortion is inhumane.</td>
<td>-2.8</td>
<td>1.6</td>
<td>-2.9</td>
<td>-2.9</td>
<td>-2.6</td>
<td>-2.3</td>
<td>-2.2</td>
</tr>
<tr>
<td>4. Abortion can be described as taking a life unjustly.</td>
<td>-2.5</td>
<td>2.0</td>
<td>-3.0</td>
<td>-2.6</td>
<td>-2.6</td>
<td>-1.9</td>
<td>-1.8</td>
</tr>
<tr>
<td>5. Abortion could destroy the sanctity of motherhood.</td>
<td>-2.2</td>
<td>1.0</td>
<td>-2.1</td>
<td>-2.2</td>
<td>-2.2</td>
<td>-1.1</td>
<td>-1.4</td>
</tr>
<tr>
<td>6. Even if one believes that there may be some exceptions, abortion is</td>
<td>-1.9</td>
<td>1.4</td>
<td>-2.4</td>
<td>-1.9</td>
<td>-2.0</td>
<td>-1.5</td>
<td>-1.2</td>
</tr>
<tr>
<td>7. Abortion should not be made readily available to everyone.</td>
<td>-1.6</td>
<td>0.6</td>
<td>-1.5</td>
<td>-1.7</td>
<td>-2.5</td>
<td>-1.2</td>
<td>-1.8</td>
</tr>
<tr>
<td>8. Abortion is basically immoral except when the woman's physical health</td>
<td>-1.1</td>
<td>1.4</td>
<td>-1.6</td>
<td>-1.3</td>
<td>-1.1</td>
<td>-0.8</td>
<td>-0.0</td>
</tr>
<tr>
<td>9. Abortion should be illegal except in extreme cases involving incest or</td>
<td>-1.0</td>
<td>1.5</td>
<td>-1.1</td>
<td>-1.0</td>
<td>-0.8</td>
<td>-0.9</td>
<td>-0.2</td>
</tr>
<tr>
<td>10. There are some clear situations where abortion should be legal, but it</td>
<td>-0.6</td>
<td>0.9</td>
<td>-1.3</td>
<td>-0.9</td>
<td>-2.0</td>
<td>-0.6</td>
<td>-0.8</td>
</tr>
<tr>
<td>11. I cannot whole-heartedly support either side of the abortion debate.</td>
<td>-0.3</td>
<td>1.5</td>
<td>-0.8</td>
<td>-0.6</td>
<td>-0.7</td>
<td>-0.1</td>
<td>-0.3</td>
</tr>
<tr>
<td>12. My feelings about abortion are very mixed.</td>
<td>-0.2</td>
<td>1.3</td>
<td>-1.1</td>
<td>-0.4</td>
<td>-1.2</td>
<td>0.1</td>
<td>-0.5</td>
</tr>
<tr>
<td>13. Abortion should be a woman's choice, but should never be used simply</td>
<td>0.4</td>
<td>0.7</td>
<td>-1.0</td>
<td>-0.9</td>
<td>-2.8</td>
<td>-1.2</td>
<td>-1.6</td>
</tr>
</tbody>
</table>
The Generalized Graded Unfolding Model 50.

Table 2 (continued).

<table>
<thead>
<tr>
<th>Statement</th>
<th>$\delta_i$</th>
<th>$\hat{\alpha}_i$</th>
<th>$\hat{\tau}_{i1}$</th>
<th>$\hat{\tau}_{i2}$</th>
<th>$\hat{\tau}_{i3}$</th>
<th>$\hat{\tau}_{i4}$</th>
<th>$\hat{\tau}_{i5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>14. Abortion should generally be legal, but should never be used as a conventional method of birth control.</td>
<td>0.6</td>
<td>0.7</td>
<td>-1.1</td>
<td>-0.7</td>
<td>-2.8</td>
<td>-1.2</td>
<td>-1.8</td>
</tr>
<tr>
<td>15. Although abortion on demand seems quite extreme, I generally favor a woman's right to choose.</td>
<td>1.1</td>
<td>1.6</td>
<td>-2.1</td>
<td>-2.1</td>
<td>-1.9</td>
<td>-1.6</td>
<td>-1.1</td>
</tr>
<tr>
<td>16. Regardless of my personal views about abortion, I do believe others should have the legal right to choose for themselves.</td>
<td>1.4</td>
<td>1.3</td>
<td>-2.6</td>
<td>-2.0</td>
<td>-2.6</td>
<td>-2.2</td>
<td>-1.9</td>
</tr>
<tr>
<td>17. Society has no right to limit a woman's access to abortion.</td>
<td>1.8</td>
<td>1.1</td>
<td>-2.5</td>
<td>-2.5</td>
<td>-2.1</td>
<td>-2.3</td>
<td>-1.6</td>
</tr>
<tr>
<td>18. Outlawing abortion violates a woman's civil rights.</td>
<td>2.3</td>
<td>1.4</td>
<td>-3.2</td>
<td>-2.8</td>
<td>-3.3</td>
<td>-2.4</td>
<td>-2.3</td>
</tr>
<tr>
<td>19. A woman should retain the right to choose an abortion based on her own life circumstances.</td>
<td>2.5</td>
<td>1.4</td>
<td>-3.5</td>
<td>-3.0</td>
<td>-3.4</td>
<td>-2.8</td>
<td>-2.4</td>
</tr>
<tr>
<td>20. Abortion should be legal under any circumstances.</td>
<td>2.8</td>
<td>1.3</td>
<td>-2.6</td>
<td>-3.0</td>
<td>-2.5</td>
<td>-2.8</td>
<td>-2.0</td>
</tr>
</tbody>
</table>
Table 3. Likelihood ratio statistics for constrained versions of the GGUM.$^a$

<table>
<thead>
<tr>
<th>Restricted Model</th>
<th>GGUM</th>
<th>Model A</th>
<th>Model B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A</td>
<td>330.56</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>($\alpha_i$ constant across items)</td>
<td>(20)</td>
<td>$p &lt; .0001$</td>
<td>$\text{-}$</td>
</tr>
<tr>
<td>Model B</td>
<td>1756.05</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>($\tau_{ik}$ constant across items)</td>
<td>(95)</td>
<td>$p &lt; .0001$</td>
<td>$\text{-}$</td>
</tr>
<tr>
<td>Model C</td>
<td>3564.73</td>
<td>3234.17</td>
<td>1808.68</td>
</tr>
<tr>
<td>($\alpha_i$ and $\tau_{ik}$ constant across items)</td>
<td>(115)</td>
<td>(95)</td>
<td>(20)</td>
</tr>
<tr>
<td></td>
<td>$p &lt; .0001$</td>
<td>$p &lt; .0001$</td>
<td>$p &lt; .0001$</td>
</tr>
</tbody>
</table>

$^a$ Each entry in the table contains -$2[\ln(L^*) - \ln(L)]$ where $L^*$ is the marginal likelihood for the restricted model and $L$ is the marginal likelihood for the less restricted model. The difference in degrees of freedom associated with the two models is given in parentheses. The $p$-values are derived from a $\chi^2$ distribution with degrees of freedom equal to that for the entry.
The Generalized Graded Unfolding Model 52.

Figure Captions

Figure 1. Subjective response category PFs for a hypothetical 4-category item as a function of $\theta_j - \delta_i$. (Subjective response category thresholds are located at -1.3, -.7, -.3, 0.0, .3, .7, and 1.3. The discrimination parameter is equal to 1.0.)

Figure 2. Observed response category PFs for a hypothetical 4-category item as a function of $\theta_j - \delta_i$.

Figure 3. Expected values of an observable response to a hypothetical 4-category item as a function of $\theta_j - \delta_i$.

Figure 4. Expected values of an observable response to a hypothetical 3-category item as a function of $\theta_j - \delta_i$ and $\alpha_i$. The discrimination parameter varies from .5, 1.0, 1.5, 2.0, 10.0, to 30.0.

Figure 5. Expected values of an observable response to a hypothetical 3-category item as a function of $\theta_j - \delta_i$ and $\tau_{ik}$. The distance between successive subjective category thresholds varies from .25, .5, 1.0, to 1.5.

Figure 6. Average root mean squared error (RMSE) values and average correlations between estimated and true item parameters as a function of the number of subjects utilized in the estimation procedure.

Figure 7. Average root mean squared error (RMSE) values and average correlations between estimated and true person parameters as a function of the number of test items utilized in the estimation procedure.

Figure 8. Average absolute difference between estimated and true expected value functions for each item portrayed as a function of the number of subjects utilized in the estimation procedure.
Figure 9. Theoretical item information functions for the GGUM portrayed as a function of $	heta_j - \delta_i$, $\alpha_i$, and $\tau_{ik}$. In the upper panel, $\alpha_i$ is held constant while the distance between successive $\tau_{ik}$ values (i.e., $\Psi$) is varied. In the lower panel, $\Psi$ is held constant while $\alpha_i$ is varied.

Figure 10. Average observed item responses and average expected item responses as a function of $\theta_j - \delta_i$. 
Figure 1

Probability

\[ \theta - \delta \]
Figure 5

Tau Values: $-0.5, -0.25, 0, 0.25, \text{ and } 0.5$

[Graph 1]

Expected Value

$\theta - \delta$

2

1

0

-3 -2 -1 0 1 2 3

Tau Values: $-1, -0.5, 0, 0.5, \text{ and } 1$

[Graph 2]

Expected Value

$\theta - \delta$

2

1

0

-3 -2 -1 0 1 2 3

Tau Values: $-2, -1, 0, 1, \text{ and } 2$

[Graph 3]

Expected Value

$\theta - \delta$

2

1

0

-3 -2 -1 0 1 2 3

Tau Values: $-3, -1.5, 0, 1.5, \text{ and } 3$

[Graph 4]

Expected Value

$\theta - \delta$
Figure 9

The figure depicts two sets of curves representing information as a function of $\theta - \delta$.

Top graph:
- Curves labeled with $\psi = 0.2$, $\psi = 0.4$, $\psi = 0.6$, and $\psi = 1.0$.

Bottom graph:
- Curves labeled with $\alpha = 0.5$, $\alpha = 1.0$, $\alpha = 2.0$, and $\alpha = 4.0$. The scale on the left side ranges from 0 to 10.
Figure 10

Response

\[ \theta - \delta \]

Response Type  

--- EXPECTED  

■ ■ ■ OBSERVED