# MARGINAL BAYESIAN PARAMETER ESTIMATION IN THE MULTIDIMENSIONAL GENERALIZED GRADED UNFOLDING MODEL

A Dissertation
Presented to
The Academic Faculty

By

Vanessa M. Thompson

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Psychology

Georgia Institute of Technology

May 2014

Copyright ® Vanessa Marie Thompson 2014

Marginal Bayesian Parameter Estimation in the
Multidimensional Generalized Graded Unfolding Model

Approved by:

James S. Roberts, Ph.D., Advisor
School of Psychology
*Georgia Institute of Technology*

Susan E. Embretson, Ph.D.
School of Psychology
*Georgia Institute of Technology*

Lawrence R. James, Ph.D.
School of Psychology
*Georgia Institute of Technology*

Daniel H. Spieler, Ph.D.
School of Psychology
*Georgia Institute of Technology*

Charles K. Parsons, Ph.D.
Scheller College of Business
*Georgia Institute of Technology*

Date Approved: January 27, 2014

## ACKNOWLEDGEMENTS

I would like to express appreciation and gratitude to my advisor, Jim Roberts, for his guidance, patience, and commitment to teaching. In addition, my research and teaching skills have benefitted from the knowledge of members of the Quantitative Psychology faculty and my committee members: Christopher Hertzog, Susan Embretson, Larry James, Dan Spieler, and Chuck Parsons. Finally, I am grateful for the support and friendship of Hi Shin Shim as we traversed graduate school together.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## SUMMARY

The Multidimensional Generalized Graded Unfolding Model (MGGUM) is a proximity-based, noncompensatory item response theory (IRT) model. It has applications in the context of attitude, personality, and preference measurement. Initial development of the MGGUM used fully Bayesian Markov Chain Monte Carlo (MCMC) parameter estimation (Roberts, Jun, Thompson, & Shim, 2009a; Roberts & Shim, 2010). Research has shown several challenges can arise while estimating MGGUM parameters using this method. For instance, the meaning of dimensions can switch during the MCMC estimation process. In addition, difficulties in obtaining informative starting values may lead to increased identification of local maxima. Furthermore, researchers must contend with lengthy MCMC computer processing time. Previous research has shown alternative estimation methods perform just as well as, if not better than, MCMC for the unidimensional Generalized Graded Unfolding Model (GGUM; Roberts & Thompson, 2011). Specifically, marginal maximum *a posteriori* (MMAP) item parameter estimation paired with expected *a posteriori* (EAP) person parameter estimation is a viable alternative to MCMC in the GGUM. The present work implements MMAP/EAP parameter estimation with the multidimensional model using rectangular quadrature. Additionally, item location initial values are derived from detrended correspondence analysis (DCA) based on previous work implementing correspondence analysis with the GGUM (Polak, 2011). A parameter recovery study is used to demonstrate the accuracy of two-dimensional MGGUM MMAP/EAP parameter estimates and a comparative analysis of MMAP/EAP and MCMC demonstrates equal

accuracy, yet much improved efficiency of the former method. Analysis of real attitude measurement data provides an illustrative application of the model.

# CHAPTER 1

# OVERVIEW

Parameter estimation in multidimensional item response theory (MIRT) is a complex process. MIRT models allow item parameters to vary across items and dimensions. Therefore, in a MIRT model with only two item parameters allowed to vary per dimension (e.g. location and discrimination), it is necessary to estimate *D*-dimensions of latent traits and 2**D* item parameters. One such highly parameterized multidimensional model is the recently developed multidimensional extension of the Generalized Graded Unfolding Model (GGUM; Roberts, Donoghue, & Laughlin, 2000). The GGUM measures attitudes using graded levels of agreement responses to statements. The multidimensional extension of the GGUM is known as the Multidimensional Generalized Graded Unfolding Model (MGGUM; Roberts, Jun, Thompson, & Shim, 2009a; Roberts & Shim, 2010). Within this model, person parameter estimates are obtained for each dimension. In addition, item location and discrimination parameters may vary across items and dimensions. Furthermore, other item parameters, known as subjective response category thresholds, can also vary across items, but are constant across dimensions.

Previous research has shown reasonable MGGUM parameter estimates can be obtained using Markov Chain Monte Carlo (MCMC) techniques (Roberts et al., 2009a; Roberts & Shim, 2010). The primary drawback of this process is that it is vastly time-consuming. Using a fast personal computer running WinBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2007), this estimation method can take days (e.g. 14 days for 20,000

iterations) to converge on a solution for a single data set comprised of 2000 subjects, 20 items, and 6 response categories. The purpose of this research is to ascertain the advantages of an alternative approach to parameter estimation in the MGGUM. This work estimates item parameters using a marginal maximum *a posteriori* (MMAP; Mislevy, 1986) approach. Person parameters are estimated with an expected *a posteriori* (EAP; Bock & Mislevy, 1982) technique. Comparative parameter estimation research in the unidimensional GGUM has demonstrated the superiority of MMAP/EAP. It was shown to produce parameter estimates that are just as accurate, if not more so, than those derived from MCMC (Roberts & Thompson, 2011). In addition, MMAP/EAP took seconds as opposed to hours to reach a solution for a typical set of GGUM item responses. As shown later in this document, the implementation of MMAP/EAP to estimate MGGUM parameters in this study is a more efficient and equally accurate method.

Marginal item parameter estimation in multidimensional models like the MGGUM requires integration over latent space. Integration of person parameters ($\theta_1$, …, $\theta_D$) out of the likelihood function can be accomplished using the Expectation-Maximization algorithm (EM; Dempster, Laird, & Rubin, 1977). Within the EM algorithm, rectangular quadrature is commonly used to approximate integration over a unidimensional latent continuum. After this integration has been accomplished, the item parameter estimates that maximize the marginal posterior distribution are found. Because integration of person parameters requires known item parameter values (i.e. the expectation component), the process of integration followed by item parameter

estimation via maximization continues iteratively until there is little change in the estimates.

The iterative estimation process is likely to reach a solution faster with more informative starting values of item parameters. Previous MGGUM research has suggested the need for more informative item location starting values to increase estimation accuracy (Roberts & Shim, 2010). In this study, initial values for item location parameters are obtained using detrended correspondence analysis (DCA; Hill & Gauch, 1980). DCA is a variant of traditional correspondence analysis (CA; Greenacre, 2007) to counteract estimation artifacts, which shall be addressed further in Chapter 5. Past research has shown CA was able to produce GGUM item location estimates comparable to those obtained using GGUM estimation software (Polak, 2011).

The forthcoming pages in Part I provide a historical foundation for multidimensional measurement, parameter estimation, and the MGGUM. Part II generalizes the MMAP/EAP parameter estimation approach to the MGGUM, discusses the implementation of DCA with the MGGUM, details a simulation design and presents subsequent results investigating the application of the new approach, presents a comparative analysis of MMAP/EAP and MCMC MGGUM parameter estimates, and describes an application of the new approach to real attitude measurement data.

# PART I: INTRODUCTION

# CHAPTER 2

# BACKGROUND

## 2.1 Classical Unfolding

### 2.1.1 Coombs Unfolding Model

The concept of unfolding has its foundation in psychological scaling literature. Coombs (1950; 1960) was the first to explicitly propose an unfolding model based on preference rankings between stimuli. His model yields single-peaked preference functions. This conceptualization is based on the idea that individuals and stimuli lie on a joint continuum or *J* scale. Figure 1 depicts the location of an individual, *X*, and the location of four stimuli, *A* through *D*, on such a scale.

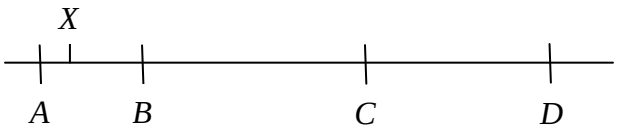



*Figure 1.* Joint distribution of stimuli and an individual.

Individuals and stimuli have fixed positions along this continuum. Using observed preference rankings, Coombs suggested a model to find the locations of individuals and items along the continuum. Unfolding involves the notion of folding the *J* scale at the individual's location such that all preference rankings increase in magnitude. Coombs proposed that as the distance of an individual from an item decreases on a continuum, the probability of endorsement or preference for the item increases. The continuum here represents a type of preference dimension underlying the stimuli in question.

### 2.1.2 Multidimensional Unfolding

One of the limitations of Coombs' (1950; 1960) unfolding model is that it was restricted to a single latent continuum. Bennett and Hays (1960) extended unfolding models of preference rankings to multiple dimensions, or continua. In a multidimensional context, preferences are based on multiple characteristics/attributes of a single stimulus. According to this model, individuals have an ideal point on each dimension. Their reported overall preference ranking of a stimulus is a function of how closely the stimulus location coincides with ideal points across dimensions. Preference functions are now single-peaked surfaces in multidimensional space. Still distance-based, this implies the probability of endorsement or preference increases to the extent an individual is close to the location of the stimulus in multidimensional space (Busing, 2010).

The extension of unidimensional unfolding to multiple latent dimensions is theoretically well-grounded. However, applications of such models lead to cumbersome mathematics reducing the feasibility of their implementation. As a result, Bennett and Hays' approach to multidimensional unfolding was blended with the field of multidimensional scaling (Shepard, 1962a, 1962b; Kruskal & Wish, 1978). Multidimensional scaling provides a method of locating points in multidimensional space based on similarity or dissimilarity judgments (e.g. rankings or ratings). The similarity or dissimilarity provides an indication of distances between points. When combined with Bennett and Hays' work, this approach models preference ranking in a *D*-dimensional joint space which locates both persons and stimuli.

A concern when working with multidimensional models is that researchers need to determine the appropriate number of dimensions underlying a given model. Bennett and Hays (1960) initially proposed three methods of identifying the minimum number of dimensions present: mutual boundary, cardinality, and groups. However, the most common method utilized to assess dimensionality in the multidimensional scaling approach involves analyzing model fit via Kruskal's concept of stress (1964). Stress is a fit index for *D*-dimensional models. It is essentially a loss function where lower levels of stress indicate greater levels of model fit.

Stress can be measured in a variety of ways. Raw stress is directly dependent upon the size of the *D*-dimensional design. Within multidimensional unfolding it is the sum of squared deviations of distances ($d_{ji}$) and distance estimates ($\hat{d}_{ji}$) between the coordinate locations of the *j*th individual and *i*th item in *D*-dimensional space. Kruskal's (1964) initial concept of stress is known as Stress-1. It is raw stress divided by the sum of squared distance estimates (Kruskal & Carroll, 1969). This value is calculated for each individual and then averaged across the sample:

$$S = S_1 = \frac{1}{J}\sum_{j=1}^{J}\sqrt{\frac{\sum_{i=1}^{I}\left(d_{ji} - \hat{d}_{ji}\right)^2}{\sum_{i=1}^{I}\hat{d}_{ji}^2}} \qquad (1)$$

where *J* is the total number of subjects and *I* is the total number of items.

Kruskal and Carroll's (1969) Stress-2 uses an alternative constant within the denominator that is proportional to the variance of distance estimates. Again, this value is computed for each individual and averaged across the sample:

$$S_2 = \frac{1}{J}\sum_{j=1}^{J}\sqrt{\frac{\sum_{i=1}^{I}\left(d_{ji}-\hat{d}_{ji}\right)^2}{\sum_{i=1}^{I}\left(\hat{d}_{ji}-\hat{d}_{\cdot\cdot}\right)^2}} \tag{2}$$

where $\hat{d}_{\cdot\cdot}$ is the grand mean of all distance estimates.

S-Stress is a measure of stress that comes in two forms: S-Stress-1 and S-Stress-2. These two forms differ from the former quantities in that squared versus raw distances are used in the calculation of stress, but are still averages of the sample values:

$$SS_1 = \frac{1}{J}\sum_{j=1}^{J}\sqrt{\frac{\sum_{i=1}^{I}\left(d_{ji}^2-\hat{d}_{ji}^2\right)^2}{\sum_{i=1}^{I}\left(\hat{d}_{ji}^2\right)^2}} \tag{3}$$

and

$$SS_2 = \frac{1}{J}\sum_{j=1}^{J}\sqrt{\frac{\sum_{i=1}^{I}\left(d_{ji}^2-\hat{d}_{ji}^2\right)^2}{\sum_{i=1}^{I}\left(\hat{d}_{ji}^2-\hat{d}_{\cdot\cdot}^2\right)^2}}\,. \tag{4}$$

One potential problem with either formulation of S-Stress is that it can lead to situations where extreme distances are exacerbated and small distances are trivialized (Takane, Young, & De Leeuw, 1977; Busing, Groenen, & Heiser, 2005). However, minimizing any of these forms of stress is thought to improve model fit.

### 2.1.3 Degeneracy in Multidimensional Unfolding

Minimizing stress should improve model fit, but it will not always lead to a meaningful solution. This can be a common occurrence with multidimensional unfolding models. An optimal solution for a multidimensional unfolding 1) minimizes stress, 2) is interpretable, 3) is parsimonious, and 4) is reliable (Shepard, 1974). Degenerate solutions occur when stress is minimized, but points are tightly clustered in multidimensional space. These solutions are not interpretable because there is not enough differentiation between points despite almost, if not, perfect model fit (Busing, 2010). It has been argued that degeneracies will occur in almost all multidimensional unfolding situations. Consequently, researchers have attempted to penalize or correct the calculation of stress in multidimensional unfolding to avoid degenerate solutions (e.g. Busing et al., 2005). Unfortunately, these adjustments appear to only be successful some of the time. Degenerate solutions can still result despite adjustments to stress calculations (Busing, 2010; Busing et al., 2005). Other psychometric approaches, such as item response theory (IRT), may be more promising. A method like IRT may allow researchers to select an unfolding model capable of estimating latent traits and features of stimuli with less risk of degenerate solutions due to additional information provided by the probability function (Roberts, Shim, Jun, Thompson, & McIntyre, 2009b).

## 2.2 Thurstone and Likert Attitude Measurement

Thus far, the discussion of unfolding models has centered on preference rankings. Unfolding is also relevant to direct ratings of a single stimulus. In order to study unfolding in such situations, it is necessary to delve deeper into historical attitude measurement literature. While Coombs (1950; 1960) proposed the first formal unfolding

model, notions of unfolding are present in earlier work by Thurstone (1927; 1928). Although Thurstone did not officially propose a model for the locations of individuals on an attitude continuum, he presented a rationale for determining the locations of individuals after scaling the locations of questionnaire items on that same continuum. Moreover, his rationale is consistent with the notion of an ideal-point response process. Likert (1932), on the other hand, introduced a method to measure individuals without previously scaling questionnaire items. His approach is consistent with the notion of a dominance-based response process. Thus, the ideal-point and dominance-based response processes have their roots in classical attitude measurement, as do unfolding and cumulative measurement models. The details of these two perspectives are discussed further below.

### 2.2.1 Thurstone Attitude Measurement

In order to understand Thurstone's (1927; 1928) approach, it is necessary to understand his law of comparative judgment. This is the basis for what is commonly referred to as Thurstone scaling. The law of comparative judgment involves the notion of statements possessing varying degrees of *affective value* along a continuum. The likelihood of statement *A* being judged as having a greater (or lesser) affective value than statement *B* is the result of both statements' affective values as seen in Figure 2.

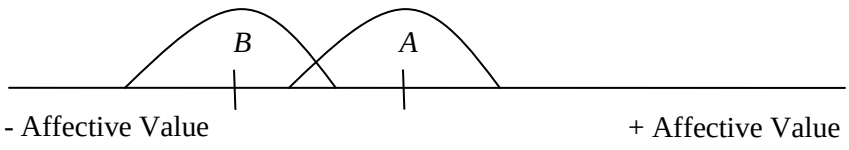



*Figure 2.* Distribution of affective value ratings for statements.

Each statement has a distribution of affective values indicating its variability across persons. Moreover, the scale value is the mode of a given affective value distribution. The affective value of a statement is presumably not related to the particular attitude of the judge.

The process of Thurstone scaling begins with the creation of statements covering all possible degrees of affective value towards a particular stimulus. Statements are scaled onto a continuum using affective values assigned by judges. The methods of successive intervals and equal-appearing intervals are commonly used in the scaling process. Final statements are selected so there are roughly an equal number of statements representing all portions of the continuum. Agreement ratings to the final statements (i.e. a second source of data) are used to identify the locations of persons along the continuum. Thurstone scaling is consistent with the notion of a proximity-based unfolding model: an individual is more likely to endorse or agree with statements whose scale values are closer to his/her location on the continuum. Therefore, person locations are derived by calculating the average or median scale value of items to which an individual agrees.

### 2.2.2 Likert Attitude Measurement

In contrast to the Thurstone paradigm, Likert (1932) proposed an alternative method of rating-based measurement. Statements are written clearly in favor or not in favor of a stimulus. Only statements with unmistakably positive or negative affective values towards a particular stimulus are included. Potentially neutral or moderate statements are not present. In contrast to Thurstone, the Likert procedure does not scale

statements and select them based on affective values. This makes Likert's method much more efficient than Thurstone's. Likert's procedure requires that individuals provide agreement ratings for each statement. After obtaining these ratings, researchers can use such methods as item-total correlations, Cronbach's alpha, or principal components analysis to select the final statements. The final statements are presented to individuals with a graded agreement scale. Reverse scoring of responses to statements with negative affective values is necessary, after which a total item sum score is constructed for each person. This score serves as a proxy for the person's location on the continuum. For instance, an individual who expresses some level of agreement with a positive statement will have a more positive attitude than an individual who expresses disagreement. Thus, the Likert procedure is consistent with a dominance-based response process in which greater overall agreement is an indication of a more positive attitude and greater overall disagreement suggests a more negative attitude, taking into account reverse-scored items. However, the affective value of each statement is ignored when constructing the total item score.

### 2.2.3 Classical Models from an Item Response Theory Perspective

Researchers have argued that the Thurstone, as opposed to Likert, procedure is more consistent with unfolding IRT models (Andrich, 1996; Roberts, Laughlin, & Wedell, 1999; Drasgow, Chernyshenko, & Stark, 2010). The two methods generally yield highly correlated measures of individuals with moderate attitudes or preferences. However, Likert's method has been shown to incorrectly estimate the attitudes or preferences of individuals at the extreme ends of the continuum when responses follow from a proximity-based response process (Roberts et al., 1999). Whether at the extreme

positive or negative end of the continuum, Likert's procedure suggests those individuals had more moderate attitudes. The Thurstone method, on the other hand, is able to adequately estimate attitudes or preferences across the entire range of the latent continuum. While the Likert procedure may be a more efficient method, the inability to realistically measure all individuals is a substantial limitation in the eyes of many researchers. This is not to say the Likert procedure should be entirely discarded. Using this method, it is possible to efficiently obtain accurate parameter estimates for responses that follow from a dominance-based response process. Only when the data follow a proximity-based response process should researchers be wary (Roberts et al., 1999). In light of this, the Thurstone paradigm is recommended as a viable option in the context of unfolding measurement models.

### 2.3 Unfolding in Item Response Theory

Unfolding IRT seeks to model the responses of individuals to items on a test/questionnaire and estimate their latent trait level. Items may possess different characteristics, such as varying locations (e.g. difficulty or valence) or the ability to discriminate between respondents (Rasch, 1960; Lord & Novick, 1968). In traditional measurement applications, most IRT models are cumulative in nature. This implies higher levels of a latent trait should lead to increased probabilities of obtaining higher item and resulting test scores. On the other hand, unfolding IRT models are proximity-based like the Thurstone measurement technique. In these models higher item scores have a greater probability of occurring when an individual is located close to an item on the latent continuum. As individuals are located farther away from an item, lower item scores are more likely to occur. Graphical representations of such models yield single-

peaked expected value functions along the latent continuum like that seen in Figure 5a (see page 24).

### 2.3.1 Applications of Unfolding

Within the context of IRT, studies have shown unfolding models are capable of utilizing binary or polytomous scales (e.g. Andrich & Luo, 1993; Andrich, 1996; Roberts & Laughlin, 1996; Roberts et al., 2000). One area in which unfolding IRT models are frequently applied is attitude research. Studies have investigated attitudes towards such issues as national pride (Javaris & Ripley, 2007), abortion (Roberts et al., 2000), capital punishment (Andrich, 1989), and work satisfaction (Carter & Dalal, 2010). Unfolding IRT models have also been implemented in personality research (e.g. Chernyshenko, Stark, Drasgow, & Roberts, 2007; Weekers & Meijer, 2008; Drasgow et al., 2010). For instance, studies have found that perceptions of personality traits like conscientiousness are measured well with an unfolding IRT model (Stark, Chernyshenko, Drasgow, & Williams, 2006; Carter, Lake, & Zickar, 2010). Furthermore, unfolding IRT models have also been utilized in research on individual change. A study on smoking cessation assessed individuals' propensity to change to determine the likelihood of success in smoking cessation (Noel, 1999). Proximity-based, ideal-point models in these areas provide just as good model fit, if not better, than dominance models.

### 2.3.2 Unfolding Item Response Theory Models

#### 2.3.2.1 Squared Simple Logistic Model (SSLM)

Using Thurstone's (1927; 1928) approach to attitude measurement, one of the first unidimensional unfolding IRT models was presented by Andrich (1988). Andrich

proposed a parametric unfolding IRT model based on the premise of Rasch's (1960) simple logistic model, but utilizing a different distance metric. The dichotomous IRT model known as the Squared Simple Logistic Model (SSLM) takes the form:

$$P\left(Z_i = z \middle| \theta_j,\ \delta_i\right) = \frac{\exp\left[-z\left(\theta_j - \delta_i\right)^2\right]}{1 + \exp\left[-\left(\theta_j - \delta_i\right)^2\right]} \tag{5}$$

where

$Z_i$ is a response to the *i*th item,

$\theta_j$ is the location of the *j*th individual on the latent continuum,

$\delta_i$ is the location of the *i*th item on the latent continuum, and

$z$ indicates an observable response of 1 (agreement) or 0 (disagreement).

It has been shown model parameters (person and item locations) can be recovered in a simulation study consisting of 200 persons and 20 items (Andrich, 1988). Using joint maximum likelihood estimation (JML; see Section 2.4.1) the obtained parameter estimates were of adequate accuracy levels as measured by root mean square deviation (RMSD) of estimates to true values despite some statistical inconsistencies. Some researchers find the inability to use graded levels of agreement as a limitation of this model since responses are treated in a binary fashion. A second limitation of this model is the response probability ceiling of 0.5 in situations where a person and item are located at the exact same point on the latent continuum ($\theta_j$ - $\delta_i$ = 0). Unfortunately, the only option to overcome this restriction is to select another model.

2.3.2.2 Parallelogram Analysis (PARELLA) Model

An alternative dichotomous unfolding IRT model suitable for assessing attitudes or preferences was derived by Hoitjink (1990). With foundations in Coombs' (1964) work, the Parallelogram Analysis (PARELLA) model is no longer restricted by the SSLM probability ceiling of 0.5 when the distance between a person and an item is zero. The probability of agreement in the PARELLA model takes the form:

$$P\left(Z_i = z \middle| \theta_j,\ \delta_i,\ \gamma\right) = \frac{\left(\left[\left(\theta_j - \delta_i\right)^2\right]^{-\gamma}\right)^{1-z}}{1+\left[\left(\theta_j - \delta_i\right)^2\right]^{-\gamma}} \tag{6}$$

where $\gamma$ is a power parameter indicating response interference and the remaining parameters are as defined in Equation 5.

Larger values of the power parameter ($\gamma$ > 10) indicate responses occur in a deterministic fashion. Smaller values ($\gamma$ < 10) suggest a probabilistic-based response mechanism. Again, model parameters were estimated and adequately recovered comparing estimates to true values in simulation studies (Hoitjink, 1990). Using marginal maximum likelihood estimation (MML; see Section 2.4.2), greater accuracy of parameter estimates were obtained as sample size increased from 100 to 900 subjects. In addition, as the value of the power parameter increased better estimates were obtained.

2.3.2.3 Hyperbolic Cosine Model (HCM)

The SSLM and PARELLA model are restricted to dichotomous disagree or agree responses. The Hyperbolic Cosine Model (HCM; Andrich & Luo, 1993; Andrich, 1995) provides an alternative approach which considers subjectively why an individual

disagrees or agrees with a particular item. An individual who agrees lies relatively close to the given item on a latent continuum. However, disagreement can occur for either of two reasons – an individual is located so far above or below the item that it no longer represents them well. Using this logic, the HCM consists of three ordered subjective response categories even though there are only two possible observable response options. The three subjective response categories are disagreement from below the item, agreement, and disagreement from above the item. Observed disagree responses are coded as 0, while observed agree responses are coded as 1. Using the symmetry of the hyperbolic cosine (cosh) function, the probability of a given observed response takes the form:

$$P\left(Z_i = 1 \middle| \theta_j, \delta_i, \vartheta_i\right) = \frac{\exp\left(\vartheta_i\right)}{\exp\left(\vartheta_i\right) + 2\cosh\left(\theta_j - \delta_i\right)} \tag{7}$$

$$P\left(Z_i = 0 \middle| \theta_j, \delta_i, \vartheta_i\right) = \frac{2\cosh\left(\theta_j - \delta_i\right)}{\exp\left(\vartheta_i\right) + 2\cosh\left(\theta_j - \delta_i\right)} \tag{8}$$

where

$\vartheta_i = (\tau_{i2} - \tau_{i1})/2$ is the unit parameter for the *i*th item and

$\tau_{ik}$ is the *k*th threshold between levels of subjective agreement for the *i*th item.

The HCM implements thresholds, $\tau_{ik}$, which are symmetric about the point $\theta_j$ - $\delta_i$ for each item. Thresholds occur where successive subjective response probability curves cross each other, and thus, there are two thresholds in the binary case as seen in Figure 3.

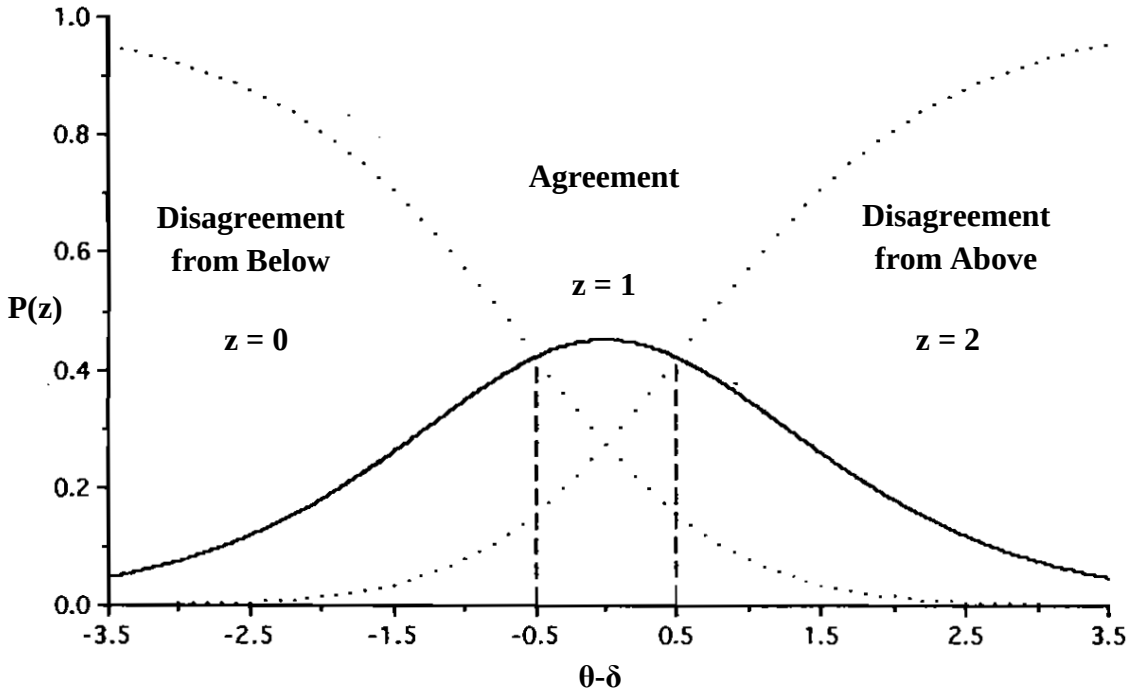



*Figure 3.* Subjective response probability curves for a binary item. From "A Hyperbolic Cosine Latent Trait Model for Unfolding Dichotomous Single-Stimulus Responses," by D. Andrich and G. Luo, 1993, *Applied Psychological Measurement*, *17*, p. 256.

When thresholds are ordered on the latent continuum, then the subjective response category (*z*) with the greatest probability of occurring between thresholds is considered to be the dominant response category. Individuals located within this interval along the latent continuum have a higher probability of utilizing the dominant response category. Subjective responses follow a cumulative model in that as the level of $\theta$ increases, an individual is more likely to use a higher order subjective response category. Andrich and Luo (1993) summed the subjective response category curves associated with either type of disagreement to obtain the probability of observed disagreement. The probability associated with the single subjective agree response category was, therefore, the observed probability of an agree response.

2.3.2.4 General Hyperbolic Cosine Model (GHCM)

An extension of the HCM, the General Hyperbolic Cosine Model (GHCM), was developed by Andrich (1996). The GHCM incorporates graded response categories representing levels of agreement from strongly disagree to strongly agree. In this polytomous model, individuals theoretically have subjective levels of agreement from below or above an item given their location on the latent continuum in relation to the item. Extreme agreement has only one subjective response probability curve, but all remaining observed response options possess two subjective response probability curves, as in the HCM. For each of these levels of disagreement, symmetric thresholds, $\tau_{ik}$, still exist, but have increased in number as there are more possible response options. The units (i.e. $\tau_{i(k+1)} - \tau_{ik}$) also need not be equal across categories or items allowing for greater model flexibility. The GHCM takes the form:

$$P\left(Z_i = z \middle| \theta_j,\ \delta_i,\ z < C\right) = \frac{\left[\exp \kappa_{ik}\right] 2\cosh\left[\left(C - z\right)\left(\theta_j - \delta_i\right)\right]}{\sum_{z=0}^{C-1}\left(\left[\exp \kappa_{ik}\right] 2\cosh\left[\left(C - z\right)\left(\theta_j - \delta_i\right)\right]\right)} \tag{9}$$

$$P\left(Z_i = C \middle| \theta_j,\ \delta_i\right) = \frac{\exp \kappa_{ik}}{\sum_{z=0}^{C-1}\left(\left[\exp \kappa_{ik}\right] 2\cosh\left[\left(C - z\right)\left(\theta_j - \delta_i\right)\right]\right)} \tag{10}$$

where

$z$ indicates the level of agreement ranging from 0 (strongest disagreement) to $C$ (strongest agreement),

$\kappa_{ik} = -\sum_{k=1}^{z} \tau_{ik}$ is the summation of negated subjective response category thresholds, $\tau_{ik}$, for the $i$th item, and

$C$ is the number of observable response categories minus 1.

As seen in Figure 4, subjective response probability curves for each of the observable response options are symmetric about the item location ($\delta_i$) and hence the point $\theta_j$ - $\delta_i$.

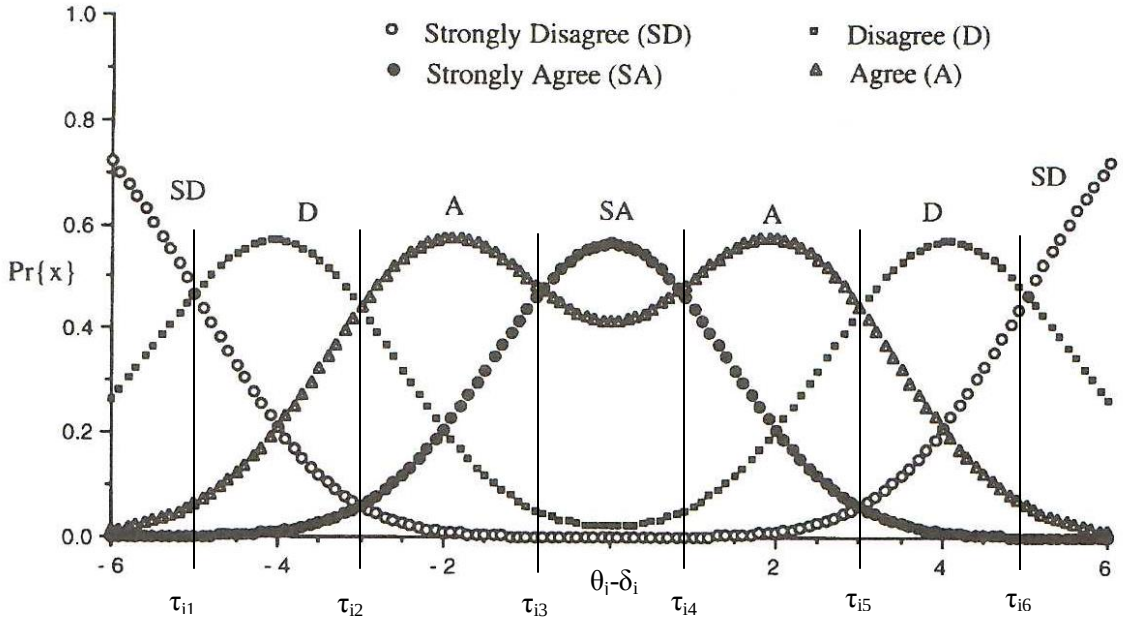



*Figure 4.* Subjective response category probability curves for a four category item. From "A Hyperbolic Cosine Latent Trait Model for Unfolding Polytomous Resposnes: Reconciling Thurstone and Likert Methodologies," by D. Andrich, 1996, *Journal of Mathematical and Statistical Psychology, 49*, p. 357.

The subjective response curve depicting extreme agreement (strongly agree) is symmetric with a single mode. Extreme disagreement possesses two symmetric subjective response curves that are unimodal. This curve approaches a probability of 1 at the extreme ends of the latent continuum and a probability of 0 when a person is located closest to the item. Curves expressing some other level of disagreement (e.g. agree, disagree) are symmetric and bimodal. Like the HCM and PARELLA model, the GHCM is not restricted by the

probability ceiling found in the SSLM. This freedom and the ability to incorporate more response options made the GHCM a useful addition to the field of unfolding IRT models.

2.3.2.5 Graded Unfolding Model (GUM)

Similar to the HCM and GHCM, the Graded Unfolding Model (GUM; Roberts & Laughlin, 1996) posits graded subjective response categories follow a cumulative model. Unlike the GHCM, the GUM presumes there are two subjective response categories associated with each observable response option – including the response indicating the highest level of agreement. These, in turn, are combined to model observed responses that represent incremental levels of agreement. The GUM takes the form:

$$P\left(Z_i = z \middle| \theta_j\right) = \frac{\exp\left[ z\left(\theta_j - \delta_i\right) - \sum_{k=0}^{z} \tau_k \right] + \exp\left[ \left(M - z\right)\left(\theta_j - \delta_i\right) - \sum_{k=0}^{z} \tau_k \right]}{\sum_{w=0}^{C} \left\{ \exp\left[ w\left(\theta_j - \delta_i\right) - \sum_{k=0}^{w} \tau_k \right] + \exp\left[ \left(M - w\right)\left(\theta_j - \delta_i\right) - \sum_{k=0}^{w} \tau_k \right] \right\}} \qquad (11)$$

where $M = 2*C + 1$, $\tau_0 = 0$, and all other terms are as previously defined.

As shown in Equation 11, the probability of a particular observed response option is the result of the summation of two subjective response probability curves. These subjective response probability curves follow Andrich's (1978) rating scale model. Probability curves for observed responses remain symmetric about the item location and the point $\theta_j$ - $\delta_i$. The expected value of a response increases to the extent a person is located close to an item on the latent continuum. In this formulation of the GUM, the subjective response category thresholds are constant across items.

On the other hand, it is also possible to conceptualize the GUM using a partial credit model (PCM; Masters, 1982) formulation. In this approach subjective response

category thresholds ($\tau_{ik}$) are allowed to vary across items with $\tau_{i1} = 0$ so that the GUM becomes (Roberts & Laughlin, 1996):

$$P\left(Z_i = z|\theta_j\right) = \frac{\exp\left[z\left(\theta_j - \delta_i\right) - \sum_{k=0}^{z}\tau_{ik}\right] + \exp\left[\left(M - z\right)\left(\theta_j - \delta_i\right) - \sum_{k=0}^{z}\tau_{ik}\right]}{\sum_{w=0}^{C}\left\{\exp\left[w\left(\theta_j - \delta_i\right) - \sum_{k=0}^{w}\tau_{ik}\right] + \exp\left[\left(M - w\right)\left(\theta_j - \delta_i\right) - \sum_{k=0}^{w}\tau_{ik}\right]\right\}}. \quad (12)$$

In both formulations, item discrimination values are constrained to equal one across items. While the SSLM, PARELLA model, and HCM are limited to dichotomous responses, the GHCM and GUM allow for graded responses. The GHCM and the GUM both possess subjective response curves symmetric about the point $\theta_j$ - $\delta_i$ on the latent continuum. However, the difference between the GHCM and the GUM is the inclusion of two subjective response categories for the strongest level of observed agreement in the latter model. Consequently, the GUM can be represented with a single parametric equation, whereas the GHCM requires a piecewise function.

2.3.2.6 Generalized Graded Unfolding Model (GGUM)

A broader conceptualization of the GUM is found within the Generalized Graded Unfolding Model (GGUM; Roberts et al., 2000). This model is a unidimensional, polytomous IRT model with single-peaked, nonmonotonic response functions. It is the foundation for the focal model in this research, which will be discussed in further detail in Chapter 3. The primary difference between the GUM and GGUM is the incorporation of item discrimination ($\alpha_i$) parameters that are allowed to vary across items. For a particular item, given an individual's latent trait ($\theta_j$), the GGUM depicts the probability of an observable response as:

$$P\left(Z_i = z|\theta_j\right) = \frac{\exp\left\{\alpha_i\left[z\left(\theta_j - \delta_i\right) - \sum_{k=0}^{z}\tau_{ik}\right]\right\} + \exp\left\{\alpha_i\left[\left(M-z\right)\left(\theta_j - \delta_i\right) - \sum_{k=0}^{z}\tau_{ik}\right]\right\}}{\sum_{w=0}^{C}\left(\exp\left\{\alpha_i\left[w\left(\theta_j - \delta_i\right) - \sum_{k=0}^{w}\tau_{ik}\right]\right\} + \exp\left\{\alpha_i\left[\left(M-w\right)\left(\theta_j - \delta_i\right) - \sum_{k=0}^{w}\tau_{ik}\right]\right\}\right)} \quad (13)$$

where $\alpha_i$ is the discrimination of the *i*th item (free to vary across items) and the remaining parameters are defined as in Equations 11 and 12.

Figure 5a depicts the expected value function of a hypothetical GGUM item located at $\delta_i = 0$ with $\alpha_i = 1$ and $\tau_{i0} = 0$, $\tau_{i1} = -1.3$, $\tau_{i2} = -0.7$, and $\tau_{i3} = -0.3$. As is the case with the GUM, GGUM expected value functions are symmetric about item locations ($\delta_i$), and hence the point $\theta_j - \delta_i$. Moreover, the expected value of the GGUM increases as the absolute value of the distance between $\theta_j - \delta_i$ decreases. Figures 5b and 5c demonstrate the impact of changes in certain item parameters on the expected value function. As item discrimination ($\alpha_i$) values increase, the curve becomes steeper approaching an upper bound of the expected value function, as seen in Figure 5b moving from Panel I to Panel II. Increases in interthreshold distances again result in the expected value function approaching its upper bound. However, these increases also produce curves with more gradual slopes, which contrast the effects of increases in discrimination values, as evident in Figure 5c moving from Panel I to Panel II.

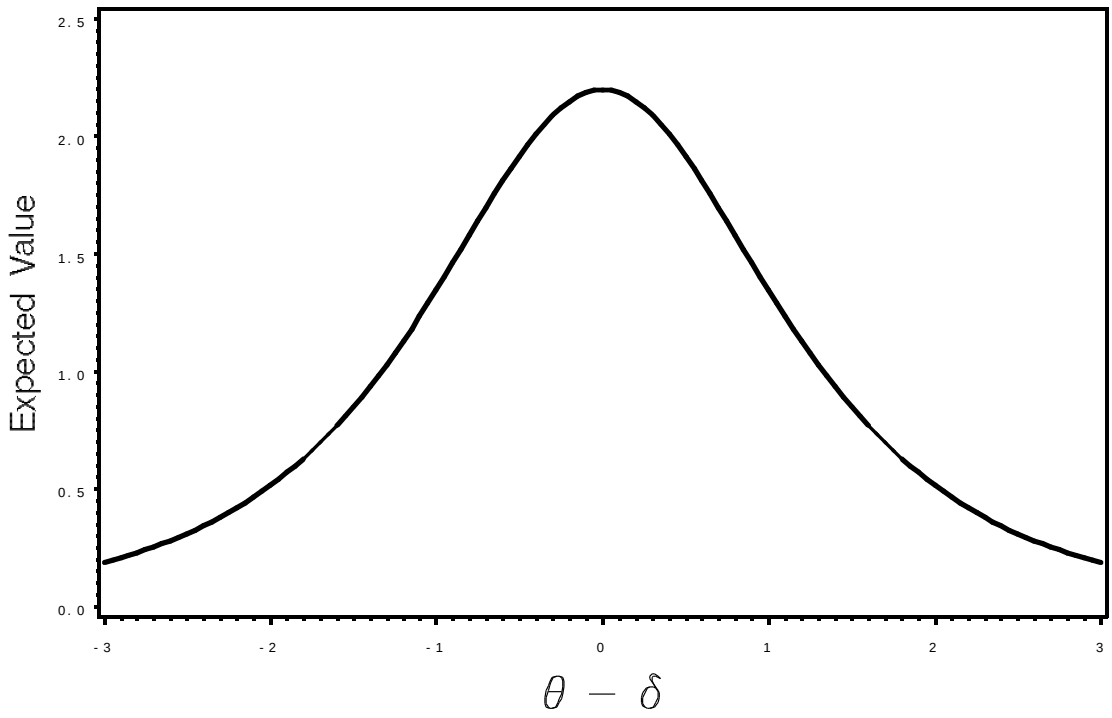
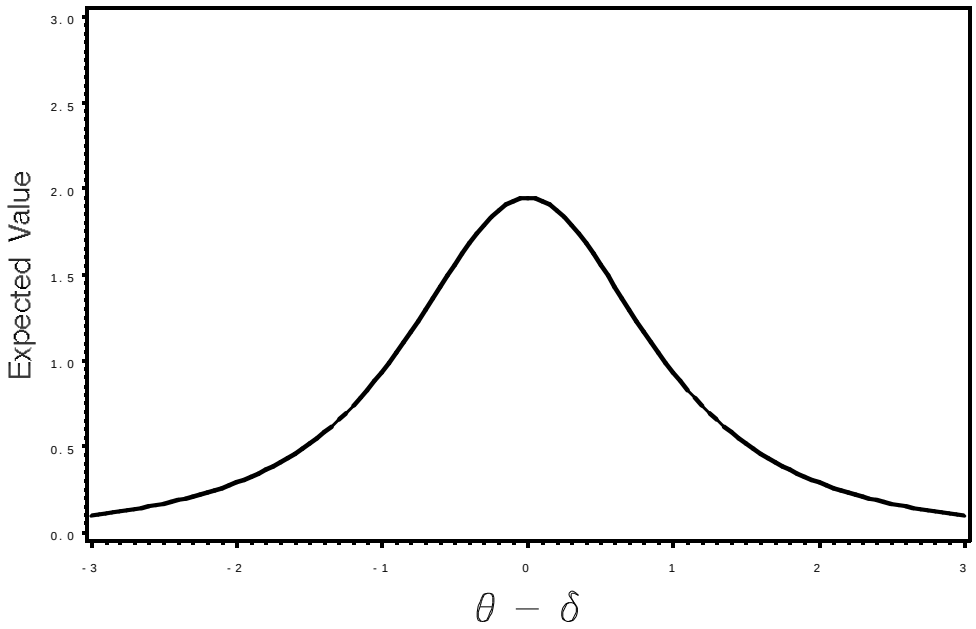



*Figure 5a.* The expected value of a hypothetical four response category unidimensional GGUM item.

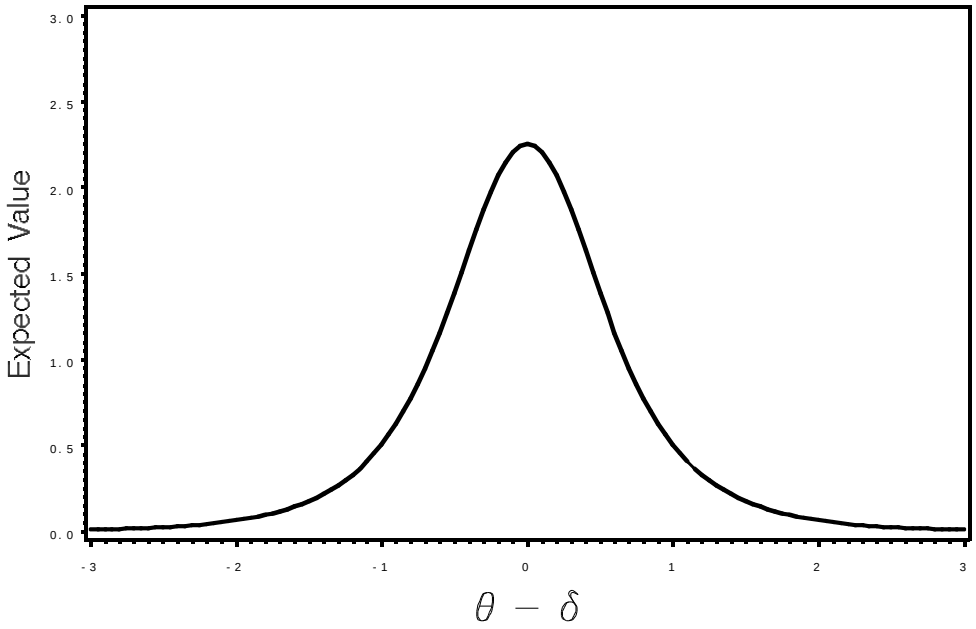
Panel I




$\alpha_i = 1$

$\tau_{i0} = 0,\ \tau_{i1} = -0.6,\ \tau_{i2} = -0.4,\ \tau_{i2} = -0.2$

Panel II

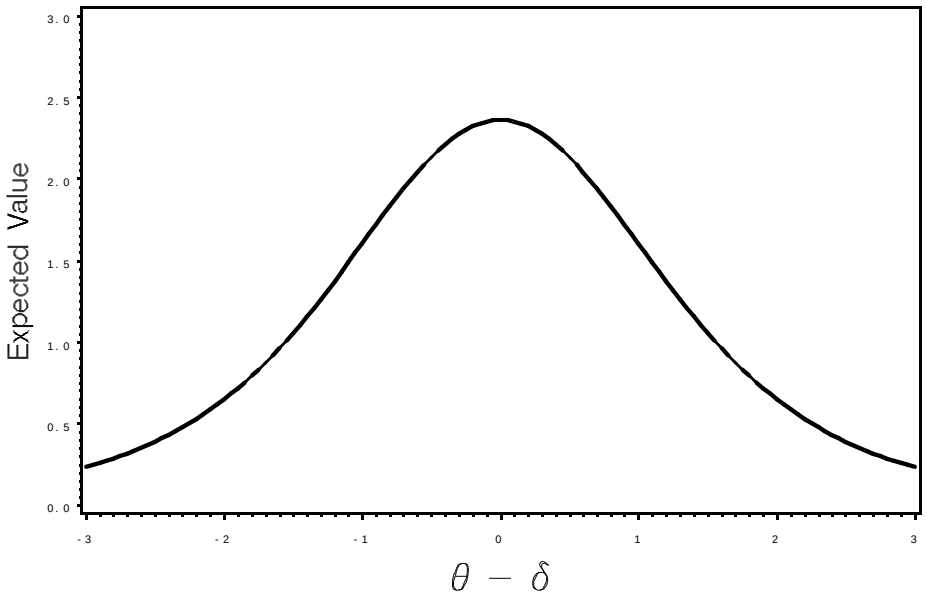



$\alpha_i = 2$

$\tau_{i0} = 0,\ \tau_{i1} = -0.6,\ \tau_{i2} = -0.4,\ \tau_{i2} = -0.2$

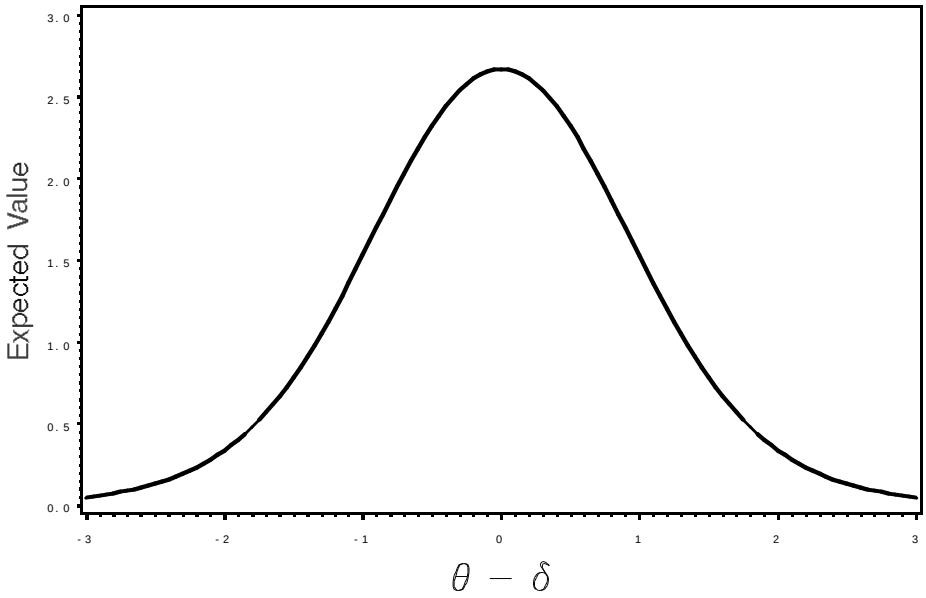
*Figure 5b.* The expected value of a hypothetical four response category unidimensional GGUM item as discrimination values change.

Panel I




$\alpha_i = 1$
$\tau_{i0} = 0$, $\tau_{i1} = -1.5$, $\tau_{i2} = -1$, $\tau_{i3} = -0.5$

Panel II




$\alpha_i = 2$
$\tau_{i0} = 0$, $\tau_{i1} = -1.5$, $\tau_{i2} = -1$, $\tau_{i3} = -0.5$

*Figure 5c.* The expected value of a hypothetical four response category unidimensional GGUM item as interthreshold distances and discrimination values change.

An indication of precision of measurement for the GGUM can be obtained through item and test information. Test information is the result of summing item information for all items on a questionnaire. The inverse of test information provides the error variance for latent ability estimates. Information from a single unidimensional GGUM item is obtained by:

$$I_i\left(\theta_j\right) = -\mathrm{E}\left[\frac{\partial^2 \ln\left(L_i\right)}{\partial \theta_j^2}\right] \tag{14}$$

where, using Equation 13,

$$L_i = \prod_{j=1}^{J} P\left(Z_i = x_{ji} \middle| \theta_j\right) \tag{15}$$

such that

$x_{ij}$ is an observable response to the *i*th item for the *j*th individual, while

$P\ (\mathrm{Z}_i = x_{ij} \mid \theta_j)$ indicates the model probability of a particular response to the *i*th item.

Figure 6 illustrates the item information function for a hypothetical GGUM item with four response categories while setting $\delta_i = 0$, $\alpha_i = 1$, $\tau_{i0} = 0$, $\tau_{i1} = -1.3$, $\tau_{i2} = -0.7$, and $\tau_{i3} = -0.3$. As seen in the figure, information is symmetric about the item's location.

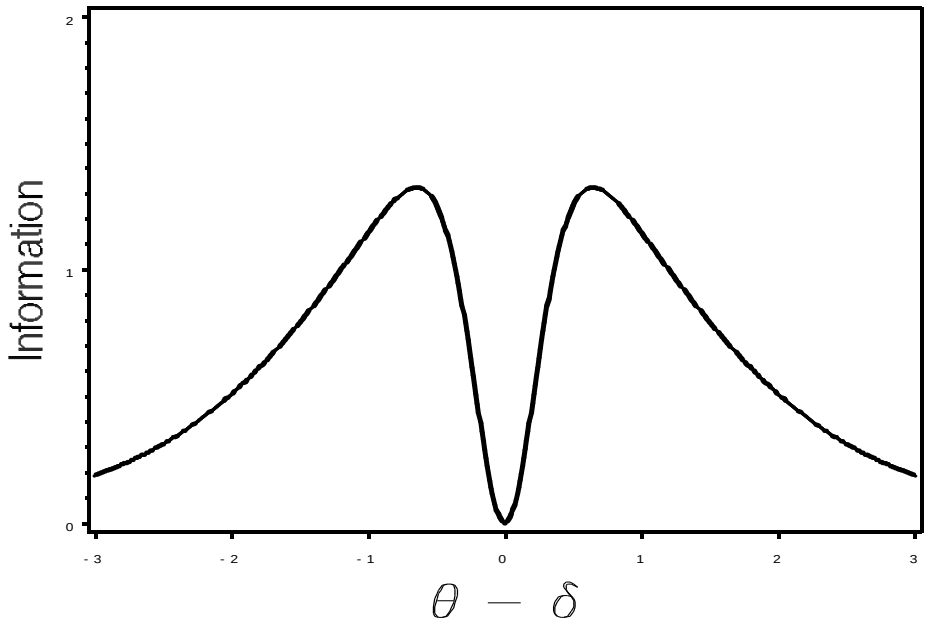



*Figure 6.* Item information for a hypothetical four category unidimensional GGUM item.

As the distance between a person and item approaches infinity or zero, the amount of information obtained approaches zero. Strong observed agreement suggests a person is located near or at the item, but they could in reality be located slightly above or below the item. Consequently, the response conveys no directional information regarding the individual's location relative to the item. Strong observed disagreement results from an individual being located very far above or below an item. Again, the exact location of the individual relative to the item is unknown. Thus, the amount of information obtained approaches zero in those situations as well. Therefore, maximum information can be obtained when an individual is located far enough away such that they are on the cusp between disagreement and agreement.

## 2.4 Parameter Estimation in the Univariate GGUM Family of Models

### 2.4.1 Joint Maximum Likelihood (JML)

In the GUM, parameter estimation was accomplished with joint maximum likelihood (JML) estimation (Roberts & Laughlin, 1996). This method involves jointly estimating both person ($\theta$) and item ($\delta$, $\tau$) parameters. Maximizing the log of the likelihood function in iterative steps results in model parameter estimates. Researchers may choose to iterate using the Newton-Raphson algorithm (N-R; Lord, 1986; Baker, 1987) or Fisher scoring (Rao, 1973), which uses the information matrix as opposed to the Hessian matrix. Regardless of the iterative method selected, the first step involves developing initial values for all model parameters. Then, person parameters are treated as known, fixed values in order to solve for item parameters. Next, estimated item parameters are fixed in order to solve for person parameters. In each iteration constraints are imposed to provide an (arbitrary) origin and scale. The log likelihood function for the GUM takes the form:

$$\ln(L)=\sum_{j=1}^{J}\sum_{i=1}^{I}\left\{\begin{array}{l}\left(-\sum_{k=0}^{x_{ji}}\tau_k\right)+\ln\left\{\exp\left[x_{ji}\left(\theta_j-\delta_i\right)\right]+\exp\left[\left(M-x_{ji}\right)\left(\theta_j-\delta_i\right)\right]\right\}\\ -\ln\left\{\sum_{w=0}^{C}\left[\exp\left(-\sum_{k=0}^{w}\tau_k\right)\right]\left\{\exp\left[w\left(\theta_j-\delta_i\right)\right]+\exp\left[\left(M-w\right)\left(\theta_j-\delta_i\right)\right]\right\}\right\}\end{array}\right\} \quad (16)$$

where all terms are as defined in Equation 11. Maximizing Equation 16 with respect to each parameter results in GUM estimates.

Statistical concerns emerge in the JML approach as estimates may not be consistent unless samples are large enough and tests are long enough (e.g. Haberman, 1977). However, as more persons and items are added there are more parameters to

estimate. With more parameters to estimate the estimation process becomes more complex (Bock & Aitkin, 1981; Harwell & Baker, 1991). Despite increased complexity, GUM person and item parameter estimates increased in accuracy, as measured by RMSD, when the number of items increased to at least 15 and the number of subjects increased to at least 100 (Roberts & Laughlin, 1996).

One issue of concern is that JML is problematic for individuals consistently utilizing extreme disagreement response categories. Person parameter estimates ($\hat{\theta}_j$) for these individuals are infinite, as are item location estimates ($\hat{\delta}_i$) for those items to which all individuals express extreme disagreement. As such, those individuals and items must be discarded from analysis with the GUM when using JML. Another concern is the identification of local maxima. In search of the global maximum of the log likelihood, local maxima may be identified, which can substantially slow down the estimation process. This is especially true when solving for person parameter estimates ($\hat{\theta}_j$). Local maxima necessitate implementation of slower estimation techniques such as grid search checks of surrounding parameter values along the latent continuum. Despite these issues, JML was able to adequately recover both person ($\theta_j$) and item parameters ($\delta_i$, $\tau_{ik}$) of the GUM (Roberts & Laughlin, 1996).

### 2.4.2 Marginal Maximum Likelihood (MML)

With the advent of the GGUM, an alternative estimation method was implemented to circumvent some of the operational drawbacks of JML (e.g. statistical inconsistency, estimation in the presence of extreme scores, local maxima, etc.). Marginal maximum likelihood (MML) is a method of item parameter estimation that

incorporates a prior distribution for person parameters ($\theta_j$). Person parameters ($\theta_j$) are then integrated out of the likelihood function to solve for item parameter ($\hat{\delta}_i$, $\hat{\alpha}_i$, $\hat{\tau}_{ik}$) estimates. Unlike JML, person parameters ($\hat{\theta}_j$) are only obtained after conducting a separate estimation process such as expected *a posteriori* (EAP) estimation (see Section 2.4.5).

The general marginal likelihood function for an item estimated with MML takes the form:

$$L_i = \prod_{j=1}^{J} \left[ \int \left( P\left(Z_i = x_{ji} \mid \theta_j\right) g\left(\theta_j \mid \mu, \sigma^2\right) \right) d\theta_j \right] \tag{17}$$

where $g(\theta_j|\mu,\sigma^2)$ is an arbitrary prior distribution for $\theta_j$ with population mean $\mu$ and variance $\sigma^2$. Maximizing the log of the general marginal likelihood function with respect to each item parameter will again yield item parameter estimates, as in JML. However, MML is capable of producing more accurate estimates with smaller samples (fewer subjects/items) where JML would produce biased estimates (Harwell, Baker, & Zwarts, 1988; Lord, 1986).

Using the Expectation-Maximization algorithm (EM; Dempster et al., 1977), MML parameter estimates are obtained by first estimating the number of individuals expected to be at particular points along the latent continuum. These points, known as quadrature points, are artificial divisions of the latent continuum that allow one to approximate integrals over a continuous space. The most common form of quadrature used in MML is rectangular quadrature in which equally spaced midpoints of rectangles are used as the quadrature points (Bock & Aitkin, 1981; Mislevy, 1986).

The expectation step involves calculating the expected number of individuals utilizing a specific response category at a particular quadrature point. The expected values, or pseudocounts, are then treated as fixed quantities and used to estimate item parameters during the maximization step in order to find the global maximum of the log marginal likelihood. Iterations between the expectation and maximization steps continue until estimates change minimally between quadrature points. However, the very first iteration is conducted using initial item parameter values identified by the researcher (Bock & Aitkin, 1981; Harwell, Baker & Zwarts, 1988). As is the case with JML, a researcher may choose to implement either the N-R algorithm or Fisher scoring in the maximization step.

As seen in Roberts et al. (2000), the conditional probability of a distinct response vector is:

$$P\left(\mathbf{X}_s \middle| \theta\right) = \prod_{i=1}^{I} P\left(Z_i = x_{si} \middle| \theta\right) \tag{18}$$

where

$S$ is the possible number of distinct response vectors,

$\mathbf{X}_s$ is a distinct response vector, and

$x_{si}$ is the response to the $i$th item in response vector $\mathbf{X}_s$.

Thus, the marginal probability of response vector $\mathbf{X}_s$ is:

$$P\left(\mathbf{X}_s\right) = \int P\left(\mathbf{X}_s \middle| \theta\right) g\left(\theta\right) d\theta. \tag{19}$$

Using the marginal probability, the likelihood function for a set of item response data is:

$$L = \frac{J!}{\prod_{s=1}^{S} r_s!} \prod_{s=1}^{S} \left[ P\left(\mathbf{X}_s\right)^{r_s} \right]. \tag{20}$$

Taking the log of Equation 20 will yield the log marginal likelihood function under the GGUM:

$$\ln\left(L\right) = \ln\left(J!\right) - \sum_{s=1}^{S} \ln\left(r_s!\right) + \sum_{s=1}^{S} r_s \ln\left[P\left(\mathbf{X}_s\right)\right] \tag{21}$$

where $r_s$ is the number of individuals with response vector $\mathbf{X}_s$. Solving for the maximum of Equation 21 with respect to an arbitrary item parameter ($\pi_i$) produces an estimate of that parameter, which is approximated in quadrature form as:

$$\frac{\partial \ln\left(L\right)}{\partial \pi_i} = \sum_{q=1}^{Q} \sum_{z=0}^{C} \frac{\bar{r}_{izq}}{P\left(Z_i = z \middle| A_q\right)} \frac{\partial P\left(Z_i = z \middle| A_q\right)}{\partial \pi_i} \tag{22}$$

where

$$\bar{r}_{izq} = \sum_{s=1}^{S} \frac{H_{siz} r_s L_s\left(A_q\right) W\left(A_q\right)}{\tilde{P}_s} \tag{23}$$

is the expected number of individuals using a particular response category for item *i* at the *q*th quadrature point, $A_q$, having quadrature weight $W(A_q)$ with $H_{siz}$ taking on a value of 1 when *z* equals $x_{si}$ and 0 otherwise. The probability of a using a particular response category at the *q*th quadrature point, $A_q$, is:

$$P\left(Z_i = z \middle| A_q\right). \tag{24}$$

The derivative of Equation 24 with respect to an arbitrary item parameter ($\pi_i$) is then computed as:

$$\frac{\partial P\left(Z_i = z \middle| A_q\right)}{\partial \pi_i}. \tag{25}$$

Finally, the likelihood of a particular response pattern at the $q$th quadrature point, $A_q$, is:

$$L_s\left(A_q\right) = \prod_{i=1}^{I} P\left(Z_i = x_{si} \middle| A_q\right) \tag{26}$$

and $\tilde{P}_s$ is the marginal probability of a particular response pattern in quadrature form:

$$\tilde{P}_s = \sum_{q=1}^{Q} L_s\left(A_q\right) W\left(A_q\right). \tag{27}$$

Utilizing Fisher scoring within the EM algorithm, Roberts, Donoghue, and Laughlin (1998) were able to recover item parameter estimates ($\hat{\delta}_i$, $\hat{\alpha}_i$, $\hat{\tau}_{ik}$) in a simulation study using initial values from a constrained model. Accurate item parameter estimates, as measured by RMSD, were obtained with as few as 750 subjects, 20 items, and 6 response categories. Interestingly, the inclusion of additional subjects did not substantially improve accuracy of item parameter estimates. In addition, Roberts and colleagues (2000) applied this model to simulated and real data. Using MML they found good estimates of true values as measured by RMSD. Data demand studies have shown accurate MML item parameter estimates are generally obtained when there are at least 1250 subjects, 15-20 items, and 4 response categories, again using RMSD (Cui, Roberts, & Bao, 2004). With fewer response categories, MML parameter estimates for extreme

item ($\hat{\delta}_i$, $\hat{\alpha}_i$, $\hat{\tau}_{ik}$) and person ($\hat{\theta}_j$) parameters may drift (Roberts & Thompson, 2011). With respect to non-extreme parameters, fewer subjects are needed to achieve accurate parameter estimates when more response categories are involved. With 6 response categories and 750-1000 subjects, accurate item parameter estimates can be still be obtained (Cui et al., 2004). Similar results indicate little to no change in accuracy levels as the number of subjects increases beyond 1000 (Roberts & Thompson, 2011). These studies demonstrate the existence of variations in data demands when using MML with the GGUM as a function of sample size, test length, and the number of response categories. Therefore, researchers should be cognizant of the effects of their study design on parameter accuracy when implementing MML.

### 2.4.3 Markov Chain Monte Carlo (MCMC)

JML and MML are not the only estimation methods that have been implemented within the GGUM. De la Torre, Stark, and Chernyshenko (2006) performed a comparative study of MML and Markov Chain Monte Carlo (MCMC) estimation in the GGUM. MCMC is a fully Bayesian technique involving the specification of prior distributions for all parameters, initial values for all parameters, and then jointly estimating all parameters. Point estimates of posterior distributions are used as estimates in this sampling-based method (Béguin & Glas, 2001), but one of the luxuries of the method is the ability to look at the form of the posterior distribution of each model parameter.

De la Torre and colleagues (2006) determined MCMC is capable of producing more accurate item parameter estimates than MML with smaller RMSD of estimates to

true values based on a GGUM simulation study. Their study used Metropolis-Hastings (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) within Gibbs sampling (Geman & Geman, 1984). Gibbs sampling generates *n* samples from a joint probability distribution of person and item parameters with known conditional distributions in order to obtain the joint posterior distribution. Beginning with initial values for each parameter, individual samples are obtained that draw each parameter from its associated full conditional distribution:

$$P\left(\theta|\mathbf{X},\delta,\alpha,\tau_1,\tau_2,\ldots,\tau_C\right) \propto L\left(\mathbf{X}\right)P\left(\theta\right), \tag{28}$$

$$P\left(\delta|\mathbf{X},\theta,\alpha,\tau_1,\tau_2,\ldots,\tau_C\right) \propto L\left(\mathbf{X}\right)P\left(\delta\right), \tag{29}$$

$$P\left(\alpha|\mathbf{X},\theta,\delta,\tau_1,\tau_2,\ldots,\tau_C\right) \propto L\left(\mathbf{X}\right)P\left(\alpha\right), \tag{30}$$

$$P\left(\tau_1|\mathbf{X},\theta,\delta,\alpha,\tau_2,\ldots,\tau_C\right) \propto L\left(\mathbf{X}\right)P\left(\tau_1\right), \tag{31}$$

$$\vdots$$

$$P\left(\tau_C|\mathbf{X},\theta,\delta,\alpha,\tau_1,\tau_2,\ldots,\tau_{C-1}\right) \propto L\left(\mathbf{X}\right)P\left(\tau_C\right) \tag{32}$$

where **X** is a raw data matrix. A distribution is formed from the samples, which, in theory, converges to a stationary joint posterior distribution. Parameter estimates are obtained from the associated mean values of the joint posterior distribution.

If conditional distributions are not known, Metropolis-Hastings can be implemented within Gibbs sampling. Metropolis-Hastings involves taking random samples from a probability distribution where direct sampling is not possible (Patz &

Junker, 1999). In each iteration, *t*, a sample value (i.e. $\theta^*$, $\delta^*$, $\alpha^*$, $\tau_1{}^*$, …, $\tau_C{}^*$) is drawn from a proposal distribution. The sample is accepted and each parameter is set equal to the value from the proposal distribution with probability:

$$p\left(\theta^{(t-1)},\theta^*\right)=\min\left\{1,\frac{P\left(\theta^*\left|\mathbf{X},\delta^{(t-1)},\alpha^{(t-1)},\tau_1^{(t-1)},\tau_2^{(t-1)},\ldots,\tau_C^{(t-1)}\right.\right)}{P\left(\theta^{(t-1)}\left|\mathbf{X},\delta^{(t-1)},\alpha^{(t-1)},\tau_1^{(t-1)},\tau_2^{(t-1)},\ldots,\tau_C^{(t-1)}\right.\right)}\right\}, \tag{33}$$

$$p\left(\delta^{(t-1)},\delta^*\right)=\min\left\{1,\frac{P\left(\delta^*\left|\mathbf{X},\theta^t,\alpha^{(t-1)},\tau_1^{(t-1)},\tau_2^{(t-1)},\ldots,\tau_C^{(t-1)}\right.\right)}{P\left(\delta^{(t-1)}\left|\mathbf{X},\theta^t,\alpha^{(t-1)},\tau_1^{(t-1)},\tau_2^{(t-1)},\ldots,\tau_C^{(t-1)}\right.\right)}\right\}, \tag{34}$$

$$p\left(\alpha^{(t-1)},\alpha^*\right)=\min\left\{1,\frac{P\left(\alpha^*\left|\mathbf{X},\theta^t,\delta^t,\tau_1^{(t-1)},\tau_2^{(t-1)},\ldots,\tau_C^{(t-1)}\right.\right)}{P\left(\alpha^{(t-1)}\left|\mathbf{X},\theta^t,\delta^t,\tau_1^{(t-1)},\tau_2^{(t-1)},\ldots,\tau_C^{(t-1)}\right.\right)}\right\}, \tag{35}$$

$$p\left(\tau_1^{(t-1)},\tau_1^*\right)=\min\left\{1,\frac{P\left(\tau_1^*\left|\mathbf{X},\theta^t,\delta^t,\alpha^t,\tau_2^{(t-1)},\ldots,\tau_C^{(t-1)}\right.\right)}{P\left(\tau_1^{(t-1)}\left|\mathbf{X},\theta^t,\delta^t,\alpha^t,\tau_2^{(t-1)},\ldots,\tau_C^{(t-1)}\right.\right)}\right\}, \tag{36}$$

$$\vdots$$

$$p\left(\tau_C^{(t-1)},\tau_C^*\right)=\min\left\{1,\frac{P\left(\tau_{k-1}^*\left|\mathbf{X},\theta^t,\delta^t,\alpha^t,\tau_1^t,\ldots,\tau_{C-1}^t\right.\right)}{P\left(\tau_{k-1}^{(t-1)}\left|\mathbf{X},\theta^t,\delta^t,\alpha^t,\tau_1^t,\ldots,\tau_{C-1}^t\right.\right)}\right\}. \tag{37}$$

Consequently the new parameter value is either accepted with the corresponding probability given above or the value remains unchanged from the previous draw. De la Torre et al. (2006) obtained parameter estimates by performing a single Metropolis-Hastings parameter update within each of the Gibbs sampling iterations.

MCMC has been shown to produce more accurate GGUM estimates than MML with respect to estimates of extreme items measured via RMSD of estimates to true

values (de la Torre et al., 2006; Roberts & Thompson, 2011). This was especially true for items with few response categories (e.g. binary items). In these instances, MCMC had substantially lower RMSD of parameter estimates than MML. While these findings may induce researchers to implement MCMC in their own work, it is important to note the substantial processing time of MCMC. De la Torre et al. (2006) found convergence using MCMC for a particular data set took just over an hour while an MML solution was reached in minutes. Other researchers have found MCMC can take even longer (e.g. hours to days) to converge upon a solution as the number of subjects, items, and or response categories increase (e.g. Roberts & Thompson, 2011). The duration of estimation is, no doubt, affected by the type of computer program used to implement the MCMC algorithm. De la Torre and colleagues (2006) demonstrated the Ox-based program (Doornik, 2003) can produce GGUM parameter estimates ($\hat{\theta}_j$, $\hat{\delta}_i$, $\hat{\alpha}_i$, $\hat{\tau}_{ik}$) much faster than the WinBUGS program (Spiegelhalter et al., 2007) for general MCMC parameter estimation. Computer programs optimized for GGUM parameter estimation may shorten the wait for convergence upon a solution.

### 2.4.4 Marginal Maximum A Posteriori (MMAP)

If time is not an issue, researchers may be inclined to use MCMC with the GGUM as opposed to MML, given the evidence presented thus far. Should time constraints be an issue, the efficiency of MML suggests it may be a viable choice, especially when the number of response categories is greater than four. However, recent research has shown the performance of an alternative marginal method of item parameter estimation, marginal maximum *a posteriori* (MMAP; Mislevy, 1986), supersedes that of MCMC and MML in the GGUM (Roberts & Thompson, 2011). While MML is able to adequately

estimate GGUM item parameters, it is not without its shortcomings. In general, MML has been known to result in less accurate estimates in smaller samples and with fewer items than MMAP (Mislevy, 1986; Cui et al., 2004; Gao & Chen, 2005). In addition, research has shown use of MML leads to inconsistent estimates of extreme items compared with MCMC, which is exacerbated with binary items (de la Torre et al., 2006; Roberts & Thompson, 2011).

Bayesians would thus counter the argument to use MML in the GGUM with the notion that the inclusion of prior distributions for item ($\delta_i$, $\alpha_i$, $\tau_{ik}$) and person ($\theta_j$) parameters provides additional sources of information during the MCMC estimation process. This can lead to more accurate parameter estimates. Prior distributions play larger roles in the estimation process when the data are less informative, or rather, more extreme. The prior distributions restrict estimates from drifting to extreme values by imposing a distributional assumption as a guide for the estimation process. When the data are more informative, there is less sensitivity to any imposed prior distributions (Harwell et al., 1988; Mislevy, 1986).

MMAP is an estimation method that incorporates prior distributions for all parameters. However, unlike MCMC, the person parameter ($\theta_j$) is integrated out of the likelihood function during the estimation process (Mislevy, 1986). Within IRT, there will generally be more person parameters than item parameters regardless of the model. Including person parameters in the estimation process substantially increases the number of necessary calculations. Integration of these terms out of the likelihood function simplifies the item parameter estimation process. This method of marginal Bayesian item parameter estimation is, like MML, an iterative procedure requiring initial values only for

item parameters in which the EM algorithm can be implemented to maximize the posterior likelihood, which is the product of prior distributions and a likelihood function.

As demonstrated in Roberts and Thompson (2011), in the context of the GGUM, the posterior likelihood can be found up to a proportionality constant via:

$$\begin{aligned} L &\propto \prod_{j=1}^{J} \int \left( \prod_{i=1}^{I} \left( P\left(Z_i = z \mid \theta_j\right) g\left(\theta_j \mid \mu, \sigma^2\right) b\left(\delta_i\right) a\left(\alpha_i\right) \prod_{k=1}^{C} t\left(\tau_{ik}\right) \right) \right) d\theta_j \\ &\propto \prod_{j=1}^{J} \left\{ \int \prod_{i=1}^{I} P\left(Z_i = z \mid \theta_j\right) g\left(\theta_j \mid \mu, \sigma^2\right) d\theta_j \right\} \prod_{i=1}^{I} \left( b\left(\delta_i\right) a\left(\alpha_i\right) \prod_{k=1}^{C} t\left(\tau_{ik}\right) \right) \end{aligned} \tag{38}$$

where

> $P\ (Z_i = z \mid \theta_j)$ is the probability of the *j*th individual's response to the *i*th item as given in Equation 13,
>
> $b\ (\delta_i)$ is a prior distribution for the *i*th item location,
>
> $a\ (\alpha_i)$ is a prior distribution for the *i*th item discrimination, and
>
> $t\ (\tau_{ik})$ is a prior distribution for the *k*th subjective response category for the *i*th item.

Taking the logarithm of Equation 38 yields:

$$\begin{aligned} \ln\left(L\right) &\propto \ln\left( \left\{ \prod_{j=1}^{J} P(\mathbf{X}_j) \right\} \left\{ \prod_{i=1}^{I} \left( b\left(\delta_i\right) a\left(\alpha_j\right) \prod_{k=1}^{C} t\left(\tau_{ik}\right) \right) \right\} \right) \\ &= \left\{ \sum_{j=1}^{J} \ln P(\mathbf{X}_j) \right\} + \left\{ \sum_{i=1}^{I} \left( \ln\left[ b\left(\delta_i\right) \right] + \ln\left[ a\left(\alpha_i\right) \right] + \sum_{k=1}^{C} \ln\left[ t\left(\tau_{ik}\right) \right] \right) \right\} \end{aligned} \tag{39}$$

where the marginal probability of response vector **X** for the *j*th individual is denoted by $P(\mathbf{X}_j)$. Using Fisher scoring, this function is maximized with respect to each item parameter ($\delta_i$, $\alpha_i$, $\tau_{ik}$). Following Roberts and Thompson (2011), MMAP estimates are obtained in quadrature form by finding the root (posterior maximum) of:

$$\frac{\partial \ln(L)}{\partial \delta_i} = \left( \sum_{q=1}^{Q} \sum_{z=0}^{C} \frac{\bar{r}_{izq}}{P\left(Z_i = z \middle| A_q\right)} \frac{\partial P\left(Z_i = z \middle| A_q\right)}{\partial \delta_i} \right) + \frac{\partial \ln\left[b(\delta_i)\right]}{\partial \delta_i}, \quad (40)$$

$$\frac{\partial \ln(L)}{\partial \alpha_i} = \left( \sum_{q=1}^{Q} \sum_{z=0}^{C} \frac{\bar{r}_{izq}}{P\left(Z_i = z \middle| A_q\right)} \frac{\partial P\left(Z_i = z \middle| A_q\right)}{\partial \alpha_i} \right) + \frac{\partial \ln\left[a(\alpha_i)\right]}{\partial \alpha_i}, \quad (41)$$

and

$$\frac{\partial \ln(L)}{\partial \tau_{ik}} = \left( \sum_{q=1}^{Q} \sum_{z=0}^{C} \frac{\bar{r}_{izq}}{P\left(Z_i = z \middle| A_q\right)} \frac{\partial P\left(Z_i = z \middle| A_q\right)}{\partial \tau_{ik}} \right) + \frac{\partial \ln\left[t(\tau_{ik})\right]}{\partial \tau_{ik}}. \quad (42)$$

These equations require the derivative, with respect to a given item parameter, of the probability function evaluated at a particular quadrature point. The final terms in Equations 40 through 42 are the derivatives of the log of the parameter's prior distribution. Derivations of these terms and the derivatives of the GGUM probability function at the *q*th quadrature point with respect to a parameter ($\delta_i$, $\alpha_i$, $\tau_{ik}$) are available in the appendix of Roberts and Thompson (2011).

Research has shown there are many benefits to utilizing MMAP. There tends to be less drifting of item location ($\hat{\delta}_i$) and discrimination ($\hat{\alpha}_i$) estimates for extreme items using MMAP compared to MML (Lim & Drasgow, 1990). The prior distributions appear to rein in those items and regress estimates back towards the mean of the prior distribution. In larger samples there is less of an opportunity for this to occur given the increased amount of information in the data used in the estimation process. In addition, studies have shown MMAP is able to produce more accurate item parameter estimates than MML in smaller samples with fewer items (Gao & Chen, 2005). While research has

shown marginal techniques may have excessive standard errors for extreme items, this appears to be more common in MML than MMAP (Roberts & Thompson, 2011).

Roberts and Thompson (2011) performed a comparative study of MML, MMAP, and MCMC within the GGUM. When the number of response categories, items, and subjects were large, there was little differentiation between the methods. However, MMAP resulted in greater accuracy of item parameter estimates ($\hat{\delta}_i$, $\hat{\alpha}_i$, $\hat{\tau}_{ik}$) compared to MML and MCMC with fewer response categories, items, and subjects, as measured via RMSD. MMAP was also able to produce accurate estimates of extreme items, thus performing better than MML in such instances. With respect to processing time, MMAP's efficiency was similar to MML, but took substantially less time than MCMC. Whereas MCMC could take hours or days to converge on a solution for a single replication, MMAP took no more than a few seconds to a few minutes to reach a solution. Thus, MMAP appears to be a more appropriate estimation method for the GGUM than MCMC or MML, given enhanced or comparable accuracy and decreased processing time. Recommended data demands with MMAP in the GGUM are similar to those of MML based on the work by Roberts and Thompson (2011).

### 2.4.5 Expected A Posteriori (EAP) Estimates of θ

While MMAP is recommended for item parameter estimation ($\hat{\delta}_i$, $\hat{\alpha}_i$, $\hat{\tau}_{ik}$) in the GGUM, it does not produce person parameter estimates ($\hat{\theta}_j$). A method of person parameter estimation routinely paired with marginal item parameter estimation is expected *a posteriori* (EAP) estimation (Bock & Mislevy, 1982). EAP is non-iterative

process that requires the specification of a prior distribution for $\theta_j$. A posterior distribution is obtained via:

$$P\left(\theta\middle|\mathbf{X}_j\right)=\frac{L\left(\mathbf{X}_j|\theta\right)g\left(\theta_j\middle|\mu,\sigma^2\right)}{P\left(\mathbf{X}_j\right)} \tag{43}$$

where $L(\mathbf{X}_j|\theta)$ is the likelihood of response vector **X** for the *j*th individual. The mean of the posterior distribution is the EAP estimate for the person parameter ($\hat{\theta}_j$). Unlike estimates in JML, an EAP estimate can be calculated for any response pattern. In addition, the average population error of an EAP estimate is lower, and thus more accurate, than maximum a posteriori and maximum likelihood person parameter estimates when the population matches the prior distribution (Bock & Mislevy, 1982).

Item parameter estimation ($\hat{\delta}_i$, $\hat{\alpha}_i$, $\hat{\tau}_{ik}$) research with the GGUM has utilized EAP to estimate person parameters ($\hat{\theta}_j$). The GGUM EAP estimates can be obtained using quadrature via:

$$\hat{\theta}_j \approx \frac{\sum_{q=1}^{Q} A_q L\left(\mathbf{X}_j\middle|A_q\right)W\left(A_q\right)}{\sum_{q=1}^{Q} L\left(\mathbf{X}_j\middle|A_q\right)W\left(A_q\right)} \tag{44}$$

where $L(\mathbf{X}_j|A_q)$ is the conditional likelihood of the *j*th individual's response vector given that the individual is located at the *q*th quadrature point, $A_q$. Research has shown GGUM EAP person parameter ($\theta_j$) accuracy, measured via RMSD, improves as test length increases from 10 to 30 items (Roberts, Donoghue, & Laughlin, 2002; Roberts &

Thompson, 2011). In addition, person parameter estimates ($\hat{\theta}_j$) had lower RMSD as the number of response categories increased beyond a binary condition. Interestingly, there were no observed differences in EAP person parameter estimate ($\hat{\theta}_j$) accuracy based on MML, MMAP, or MCMC item parameter estimates. Thus, EAP estimation can be implemented in the GGUM with any of these item parameter estimation methods (Roberts & Thompson, 2011).

### 2.4.6 Initial Values

Item parameter estimation using MMAP (or MML) does not require the specification of initial values for person parameters ($\theta_j$). However, the process of marginal item parameter estimation does require starting points for the iterative process. Initial values for GGUM item parameters provide a "best guess" of where the item lies on the latent continuum ($\delta_i$), how discriminating the item is ($\alpha_i$), and the locations of subjective response category thresholds ($\tau_{ik}$). GGUM initial values are derived from GUM item parameter estimates (Roberts et al., 2000; Roberts, Fang, Cui, & Wang, 2006; Roberts & Thompson, 2011). GUM estimates provide informative initial values for the GGUM given that one model is a generalized version of the other. Uninformative initial values can lengthen the time it takes to obtain item parameter estimates. More iterations may be required to move along the latent continuum while searching for the optimal solution located at the global maximum of the corresponding log marginal likelihood function. However, progressing through more of the latent continuum increases the chance of locating a local maximum. More informative initial values should avoid local maxima while speeding up the estimation process (Roberts & Laughlin, 1996). While some researchers have argued the selection of initial values should not impact parameter

estimates (Bock, 1991), others suggest evidence exists to the contrary (Nader, Tran, & Formann, 2011). Therefore, using informed initial values will likely not hinder, but rather may benefit researchers.

# CHAPTER 3

## MULTIDIMENSIONAL GENERALIZED GRADED UNFOLDING MODEL

Within the context of unfolding, the most widely used IRT models are unidimensional in nature – involving one latent continuum. While this simplifies the model and resulting conclusions, it is possible an item deemed unidimensional actually assesses more than one dimension. Failing to include a dimension in a model can lead to incorrect results due to the violation of local independence. Generally speaking, preferences are multifaceted constructs. For example, graded responses indicating coffee preference could be a function of the amount of cream present, the amount of sugar present, the strength of the coffee, etc. Thus, a researcher using a multidimensional model to assess such preferences can enhance the validity of their study. In addition, using a single questionnaire comprised of multidimensional items analyzed with a multidimensional model can save time over using multiple questionnaires each assessing a single dimension. Thus, multidimensional models, as found in multidimensional unfolding IRT, are more appropriate in these situations.

Multidimensional IRT (MIRT) models allow for the assessment of multiple latent traits regardless of whether some or all items load onto each dimension. Unfolding MIRT models remain proximity-based and are noncompensatory. In a noncompensatory model the probability of endorsing an item increases when an individual is located close to an item on all dimensions (Reckase, 2009). As the number of dimensions increase in a multidimensional model, there are corresponding increases in the number of parameters

to be estimated. Increases in model complexity result in a more complicated parameter estimation process.

The Multidimensional Generalized Graded Unfolding Model (MGGUM; Roberts et al., 2009a; Roberts & Shim, 2010) is a recent multidimensional extension of the GGUM. The MGGUM probability function takes the form:

$$P\left(Z_i = z|\underline{\theta}_j\right) = \frac{\exp\left[\left(z\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right)-\sum_{k=0}^{z}\psi_{ik}\right]+\exp\left[\left((M-z)\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right)-\sum_{k=0}^{z}\psi_{ik}\right]}{\sum_{w=0}^{C}\left(\exp\left[\left(w\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right)-\sum_{k=0}^{w}\psi_{ik}\right]+\exp\left[\left((M-w)\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right)-\sum_{k=0}^{w}\psi_{ik}\right]\right)} \quad (45)$$

where

$\underline{\theta}_j = \left(\theta_{j1}\ldots\theta_{jD}\right)$ is a vector of $D$ location coordinates for the $j$th individual in the latent space,

$\theta_{jd}$ is the location of the $j$th individual on the $d$th dimension,

$\delta_{id}$ is the location of the $i$th item on the $d$th dimension,

$\alpha_{id}$ is the discrimination of the $i$th item on the $d$th dimension,

and

$$\psi_{ik} = \sum_{d=1}^{D}\alpha_{id}\tau_{ik}. \quad (46)$$

Note that $\psi_{ik}$ is a dimensionless quantity which varies by item and response category. Similar to the GGUM, an individual will receive higher item scores to the degree the individual is located close to an item in multidimensional space. The MGGUM is an unfolding model yielding single-peaked expected value surfaces that are symmetric about the item's location in the latent space, as seen in Figure 7a with two dimensions. This is

a two-dimensional extension of the unidimensional GGUM expected value curve found in Figure 5a. The contour plot of the expected value function found in Figure 7b depicts concentric circles symmetric about the item's location in latent space. This hypothetical MGGUM item is located at $\delta_{i1} = 0$, $\delta_{i2} = 0$ with equal discriminations of $\alpha_{i1} = 1$, $\alpha_{i2} = 1$ and subjective response category thresholds of $\tau_{i0} = 0$, $\tau_{i1} = -1.3$, $\tau_{i2} = -0.7$, and $\tau_{i3} = -0.3$.

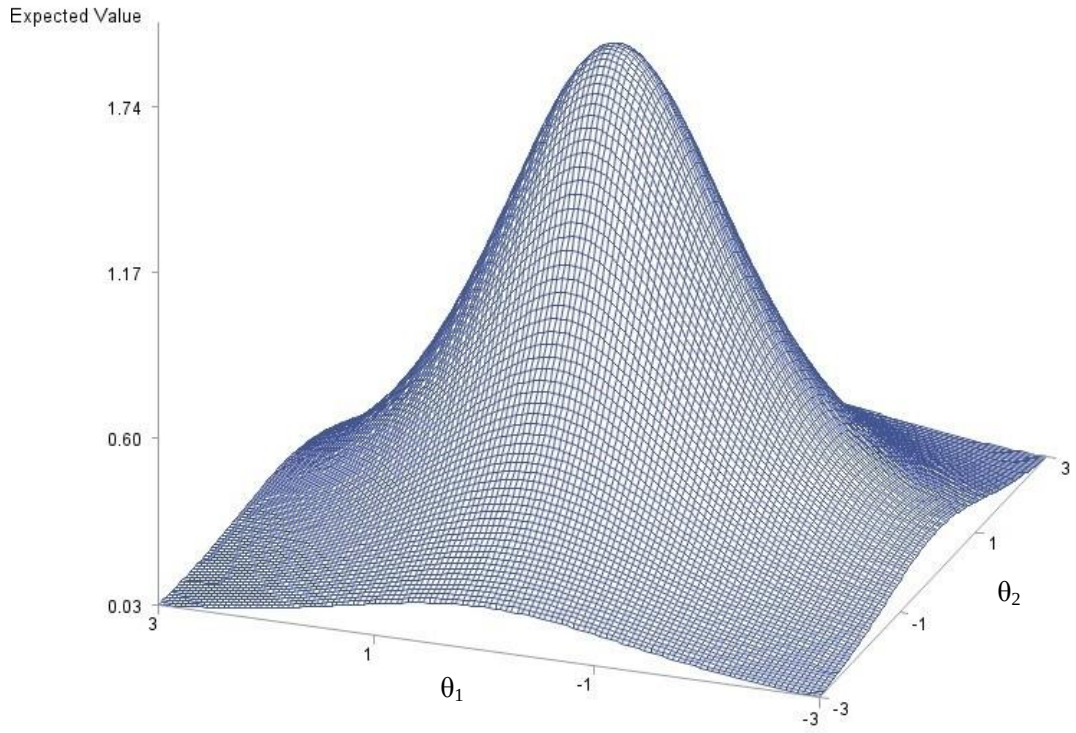



*Figure 7a.* Expected value surface for the MGGUM.

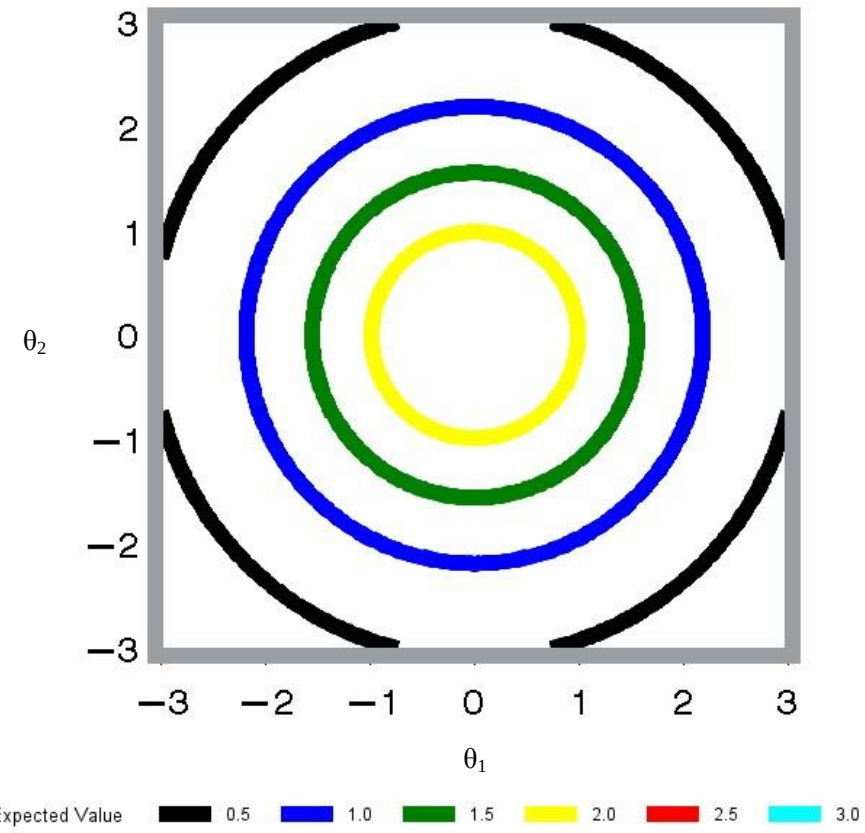



*Figure 7b.* Contour plot of the expected value surface for the MGGUM.

Changes in the discrimination values and interthreshold distances affect the MGGUM expected value surface in a manner similar to the unidimensional GGUM. As seen in Figures 5b and 5c, increases in unidimensional GGUM discrimination values result in steeper expected value curves, while increasing the interthreshold distance results in more diffuse expected value curves. Increases in the MGGUM dimensional discrimination values result in steeper expected value surfaces. In the presence of equally weighted dimensional discrimination values, equal increases of these values increase the slopes on both dimensions. Increases leading to unequal dimensional discrimination values result in a steeper slope for the dimension with the larger discrimination value

compared to other dimensions. Changes in interthreshold distances continue to produce more diffuse expected value surfaces as in the GGUM. These relationships are depicted in the contour plots found in Figure 8 for a four category MGGUM item located at $\delta_{i1} = 0$ and $\delta_{i2} = 0$. Moving from Panel I to Panel II, increases in the discrimination value for a single dimension result in concentric elliptical contour plots versus concentric circular plots. This implies there is greater differentiation between individuals along the dimension with the higher discrimination value in terms of their likelihood of utilizing a particular subjective response category. Increases in interthreshold distances, moving from Panel I to Panel III, result in greater distances between the concentric circles. Moving from Panel I to Panel IV, increases in both the discrimination value for a single dimension and interthreshold distances result in concentric ellipses with greater distances in between ellipses.

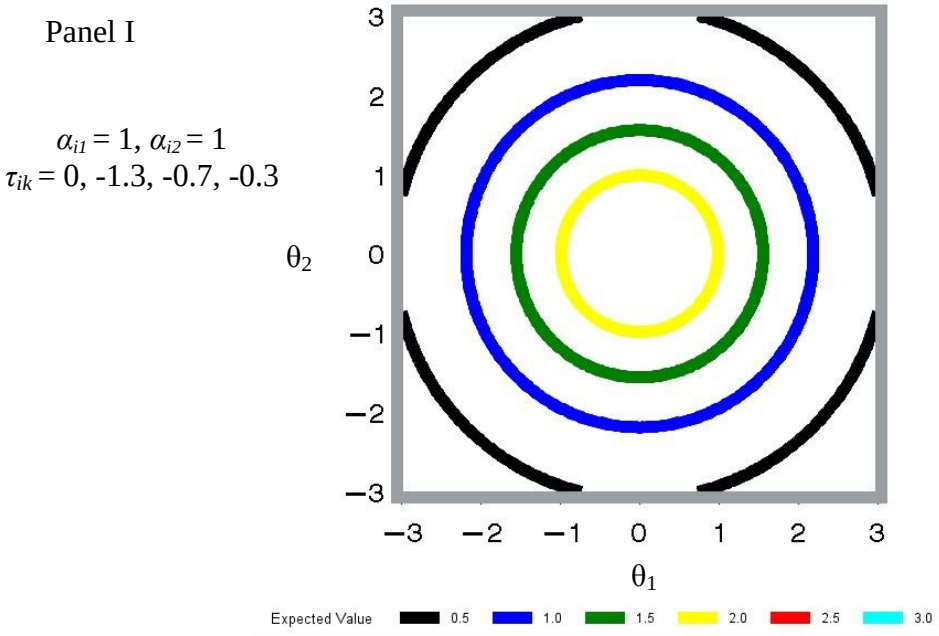
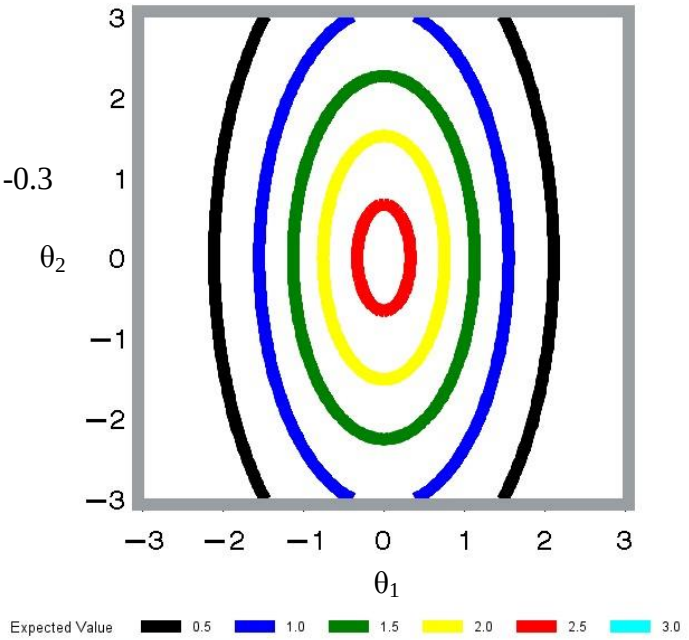



*Figure 8.* Contour plots of a hypothetical four category MGGUM item.

Panel II

$\alpha_{i1} = 2$, $\alpha_{i2} = 1$
$\tau_{ik} = 0$, -1.3, -0.7, -0.3




Panel III

$\alpha_{i1} = 1$, $\alpha_{i2} = 1$
$\tau_{ik} = 0$, -2.6, -1.4, -0.6

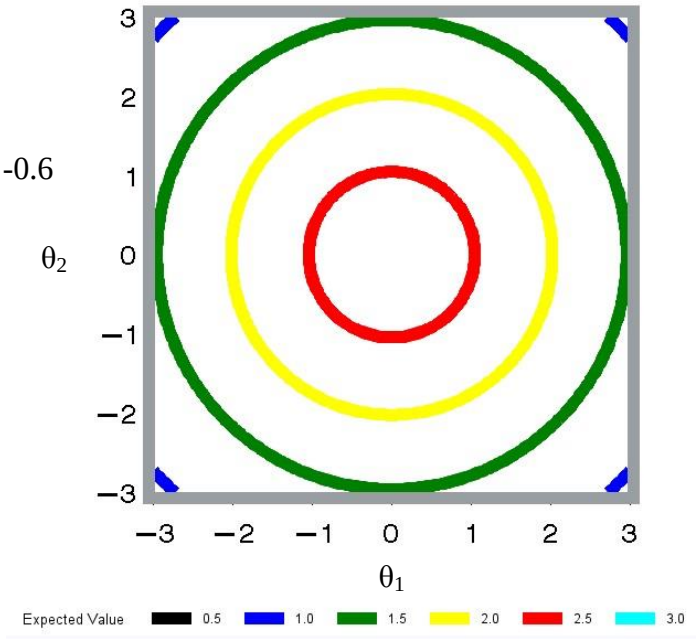



*Figure 8.* Continued.

Panel IV

$\alpha_{i1} = 2$, $\alpha_{i2} = 1$
$\tau_{ik} = 0$, -2.6, -1.4, -0.6

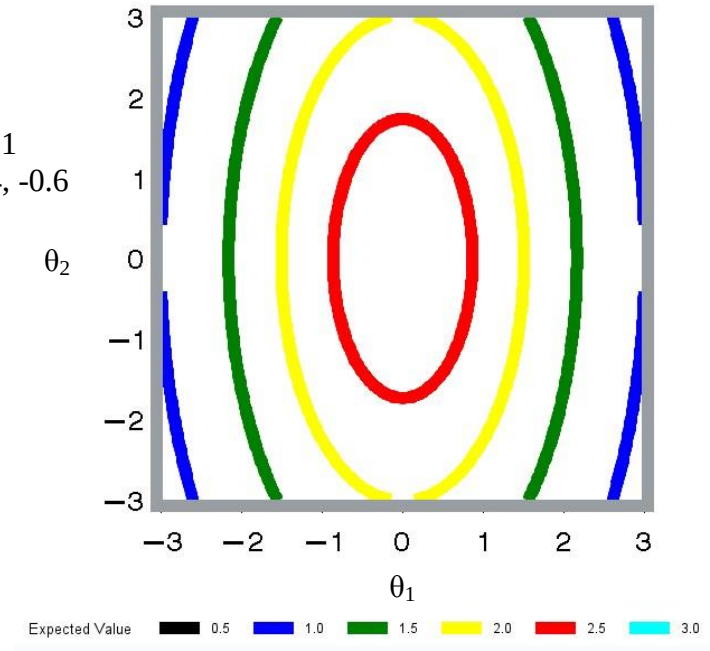



*Figure 8.* Continued.

Item information within the MGGUM can be calculated using Ackerman's (1994) matrix method. As the MGGUM is a noncompensatory model, there is no one particular vector of interest and thus, Ackerman's procedure is more appropriate than the directional vector approach of Reckase and McKinley (1991). In a two-dimensional case Ackerman's procedure defines information as:

$$I_i\left(\underline{\theta}\right) = -\mathrm{E}\begin{pmatrix} \dfrac{\partial^2 \ln\left(L_i\right)}{\partial\theta_1^2} & \dfrac{\partial^2 \ln\left(L_i\right)}{\partial\theta_1\,\partial\theta_2} \\ \dfrac{\partial^2 \ln\left(L_i\right)}{\partial\theta_1\,\partial\theta_2} & \dfrac{\partial^2 \ln\left(L_i\right)}{\partial\theta_2^2} \end{pmatrix} \tag{47}$$

where, using Equation 45,

$$L_i = \prod_{j=1}^{J} P\left(Z_i = x_{ji} \middle| \underline{\theta}_j\right). \tag{48}$$

Item information that is unique to each dimension in a two-dimensional model is determined by the elements on the main diagonal of the matrix, whereas joint information is determined by the off diagonal elements in the matrix. Thus, item information becomes:

$$I_i(\theta_1) = -\mathrm{E}\left(\frac{\partial^2 \ln(L_i)}{\partial \theta_1^2}\right) = -\mathrm{E}\left\{\left[\frac{\partial^2 P\left(Z_i = z|\underline{\theta}_j\right)}{\partial \theta_1^2} * \frac{1}{P\left(Z_i = z|\underline{\theta}_j\right)}\right] - \left[\left(\frac{\partial P\left(Z_i = z|\underline{\theta}_j\right)}{\partial \theta_1}\right)^2 * \frac{1}{\left(P\left(Z_i = z|\underline{\theta}_j\right)\right)^2}\right]\right\}, \quad (49)$$

$$I_i(\theta_2) = -\mathrm{E}\left(\frac{\partial^2 \ln(L_i)}{\partial \theta_2^2}\right) = -\mathrm{E}\left\{\left[\frac{\partial^2 P\left(Z_i = z|\underline{\theta}_j\right)}{\partial \theta_2^2} * \frac{1}{P\left(Z_i = z|\underline{\theta}_j\right)}\right] - \left[\left(\frac{\partial P\left(Z_i = z|\underline{\theta}_j\right)}{\partial \theta_2}\right)^2 * \frac{1}{\left(P\left(Z_i = z|\underline{\theta}_j\right)\right)^2}\right]\right\}, \quad (50)$$

and

$$\begin{aligned} I_i(\theta_1\theta_2) &= -\mathrm{E}\left(\frac{\partial^2 \ln(L_i)}{\partial \theta_1 \, \partial \theta_2}\right) \\ &= -\mathrm{E}\left\{\left[\frac{\partial^2 P\left(Z_i = z|\underline{\theta}_j\right)}{\partial \theta_1 \, \partial \theta_2} * \frac{1}{P\left(Z_i = z|\underline{\theta}_j\right)}\right] - \left[\frac{\partial P\left(Z_i = z|\underline{\theta}_j\right)}{\partial \theta_1} * \frac{\partial P\left(Z_i = z|\underline{\theta}_j\right)}{\partial \theta_2} * \frac{1}{\left[P\left(Z_i = z|\underline{\theta}_j\right)\right]^2}\right]\right\}. \end{aligned} \quad (51)$$

The full derivation of MGGUM item information for two dimensions is available in Appendix A.

Item information surfaces take on different forms depending upon the nature of the item structure. Simple structure occurs when an item loads onto only one dimension. Figures 9a and 9b depict item information for a hypothetical four response category MGGUM item with simple structure located at $\delta_{i1} = 0$, $\delta_{i2} = 0$ with unequal discriminations $\alpha_{i1} = 1$, $\alpha_{i2} = 0$ and subjective response category thresholds set at $\tau_{i0} = 0$, $\tau_{i1} = -1.3$, $\tau_{i2} = -0.7$, and $\tau_{i3} = -0.3$. As seen in Figure 9a, MGGUM item information will

resemble a unidimensional GGUM item information curve along the primary dimension with constant values along all other dimensions. This occurs because the only non-zero value in the information matrix is associated with the dimension on which the item loads, which is the negative expected value of the second derivative of the log likelihood function. All other elements in the information matrix will have a value of zero in the case of simple structure. The contour plot in Figure 9b depicts how information remains constant along the non-measured dimension for simple structure items. Information with respect to the measured dimension changes for different levels of latent ability $\theta_1$, while it remains constant across $\theta_2$.

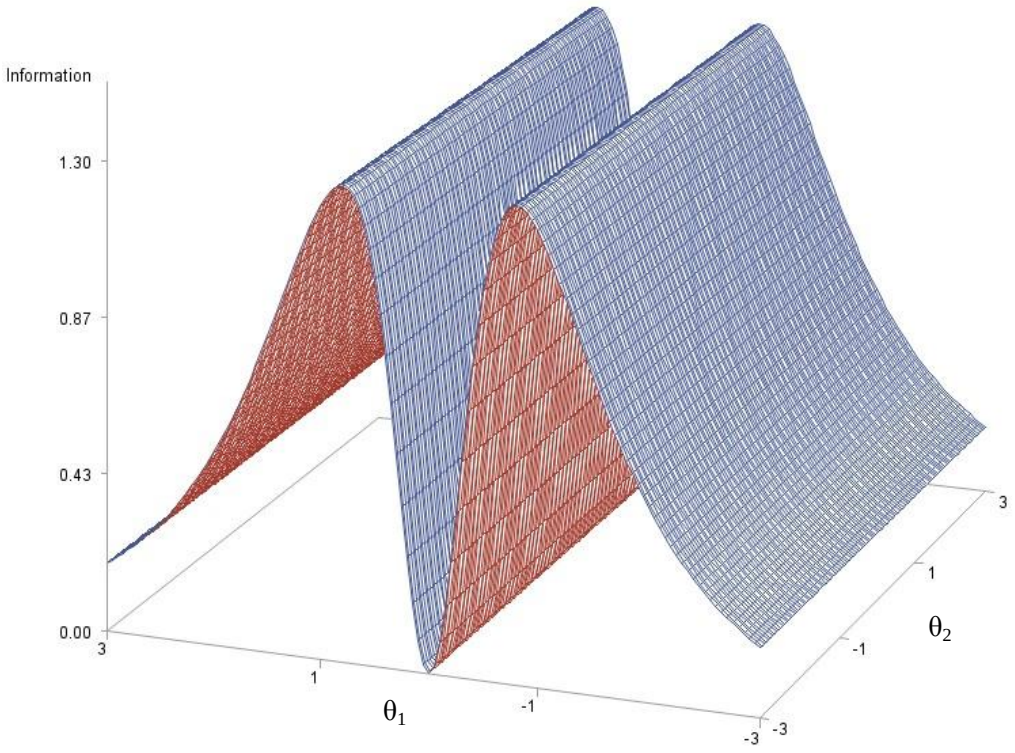



*Figure 9a.* Information surface for two-dimensional simple structure MGGUM item.

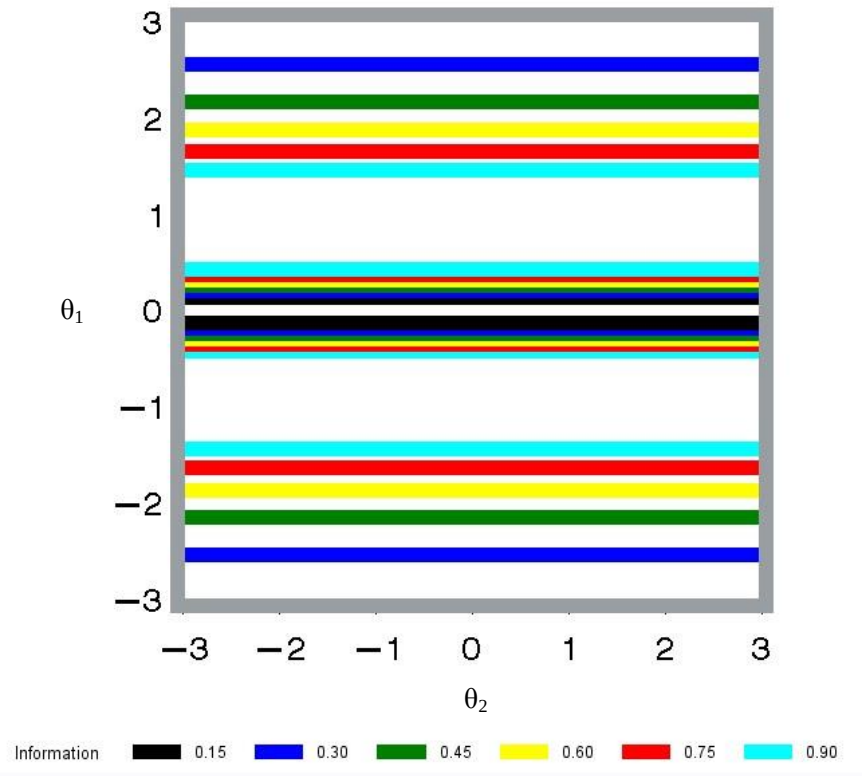



*Figure 9b.* Contour plot of item information surface for two-dimensional simple structure MGGUM item.

Complex structure occurs when an item loads on multiple dimensions. Here, the item information surface will remain symmetric about the item's location in latent space. However, it will take on a different graphical form due to multiple distinct non-zero elements in the information matrix. Figure 10a depicts the information surfaces with respect to $\theta_1$, $\theta_2$, and both dimensions simultaneously for a hypothetical four response category MGGUM item. This item is located at $\delta_{i1} = 0$, $\delta_{i2} = 0$ with equal discriminations $\alpha_{i1} = 1$, $\alpha_{i2} = 1$ and subjective response category thresholds of $\tau_{i0} = 0$, $\tau_{i1} = -1.3$, $\tau_{i2} = -0.7$, and $\tau_{i3} = -0.3$. Each panel represents a unique element from the 2 x 2 information matrix. The information surfaces with respect to $\theta_1$ and $\theta_2$ are bimodal, while the information

surface with respect to both dimensions simultaneously is quad-modal. The associated contour plots found in Figure 10b depict symmetry about the location of the item in latent space and again each panel corresponds to a unique element of the information matrix. Information approaches zero at the point in latent space where $|\theta_{jd} - \delta_{id}| = 0$. In addition, information also approaches zero where $|\theta_{jd} - \delta_{id}| = \pm\infty$. Maximum information is obtained at coordinate locations in the latent space when an individual expresses moderate levels of agreement (Thompson & Roberts, 2010).

Panel I

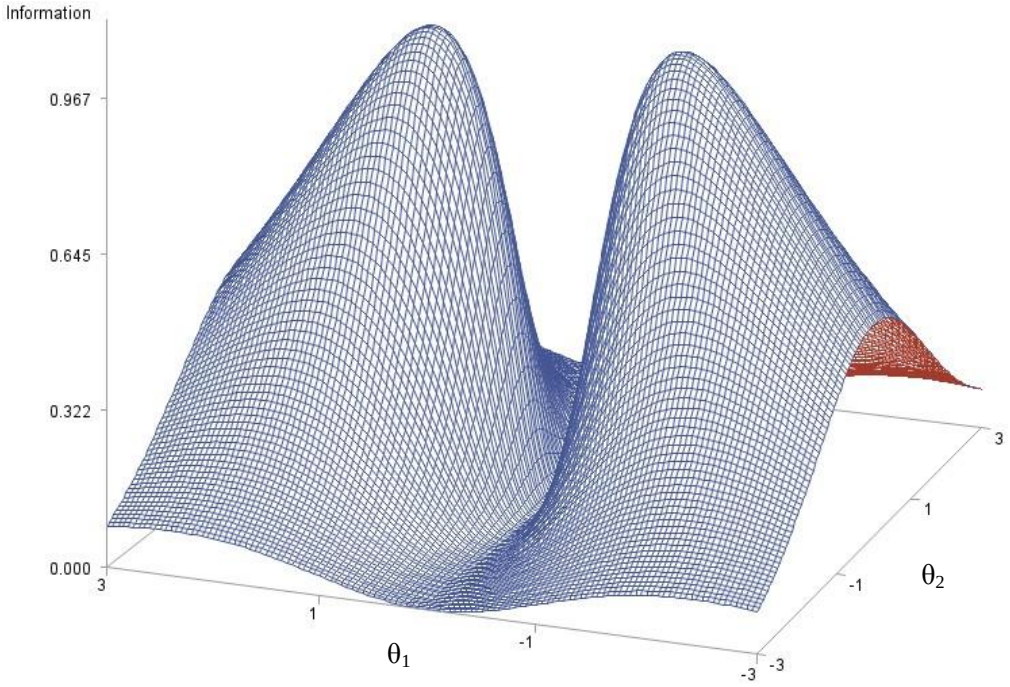
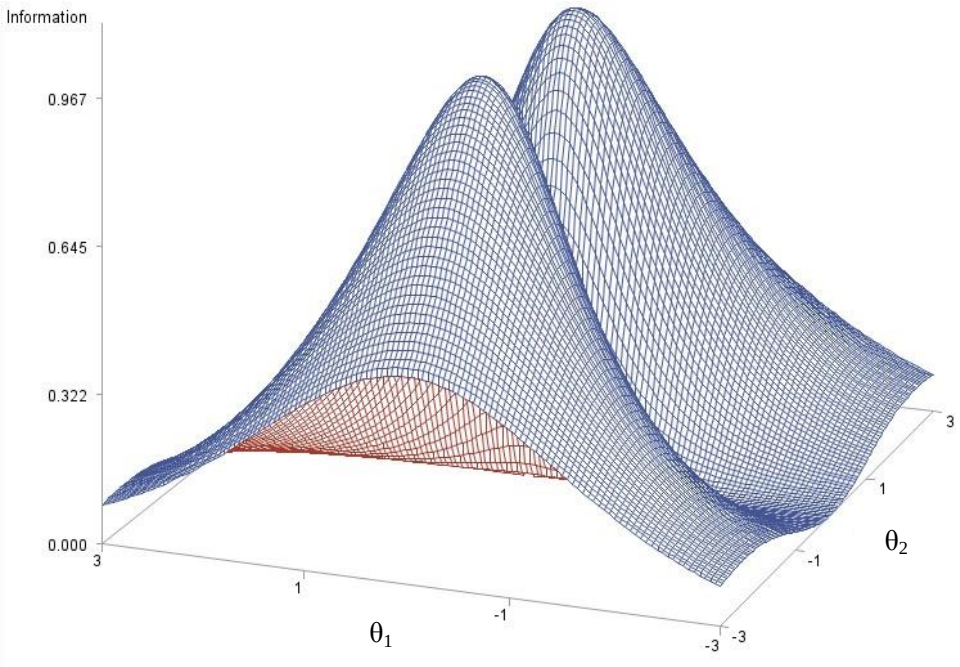



*Figure 10a.* Information surface for two-dimensional complex structure MGGUM item.

Panel II




Panel III

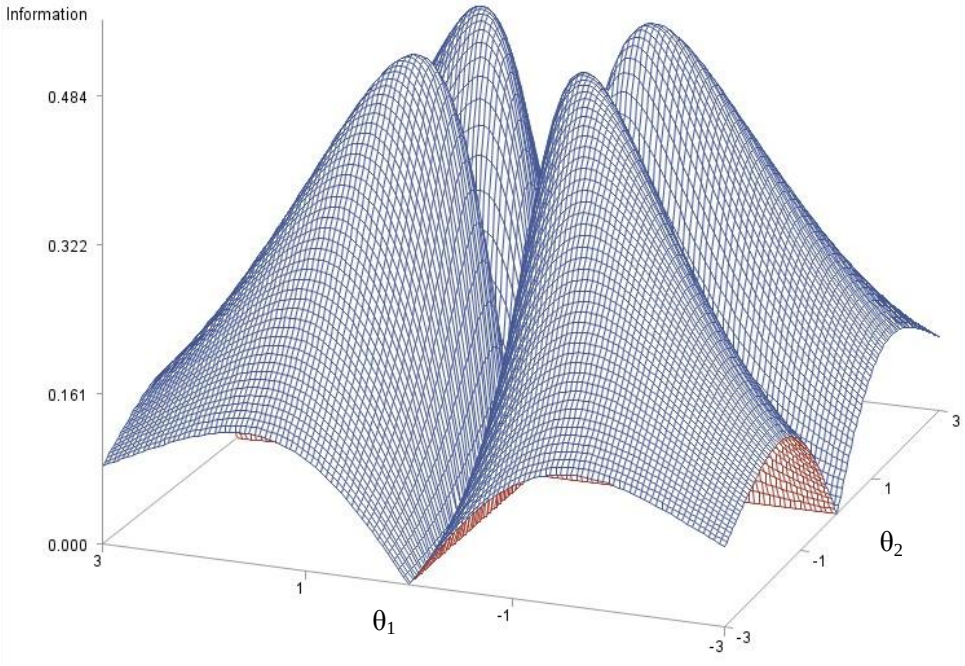
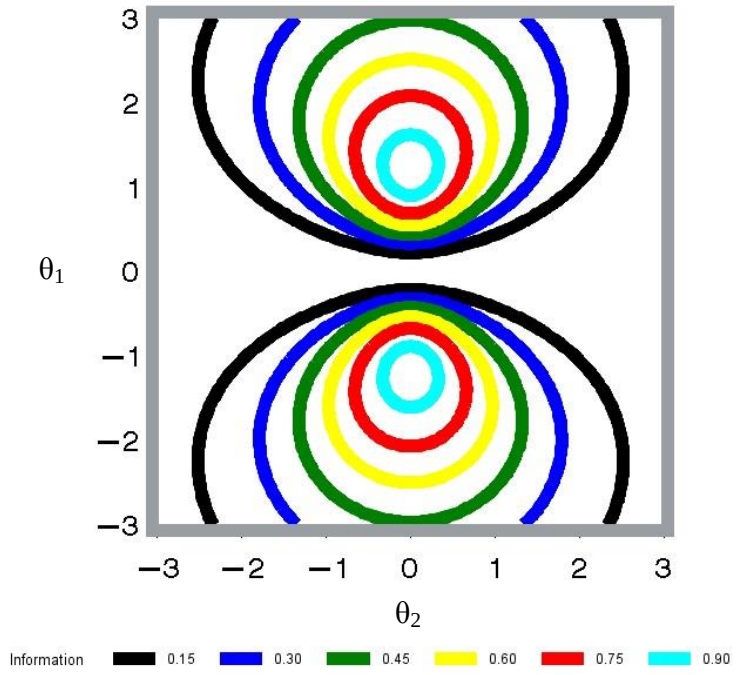



*Figure 10a.* Continued.

Panel I

$\theta_1$ Information




Panel II

$\theta_2$ Information

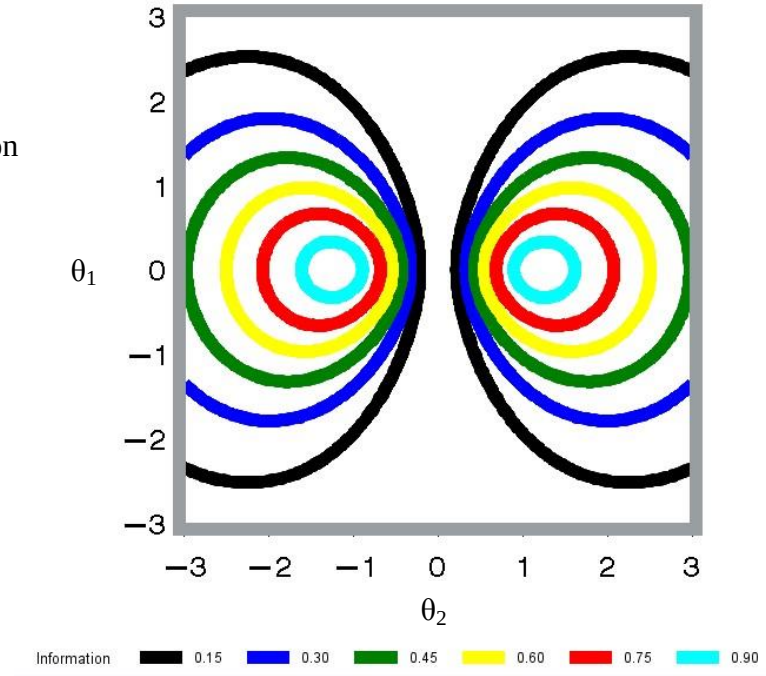



*Figure 10b.* Contour plot of item information surface for two-dimensional complex structure MGGUM item.

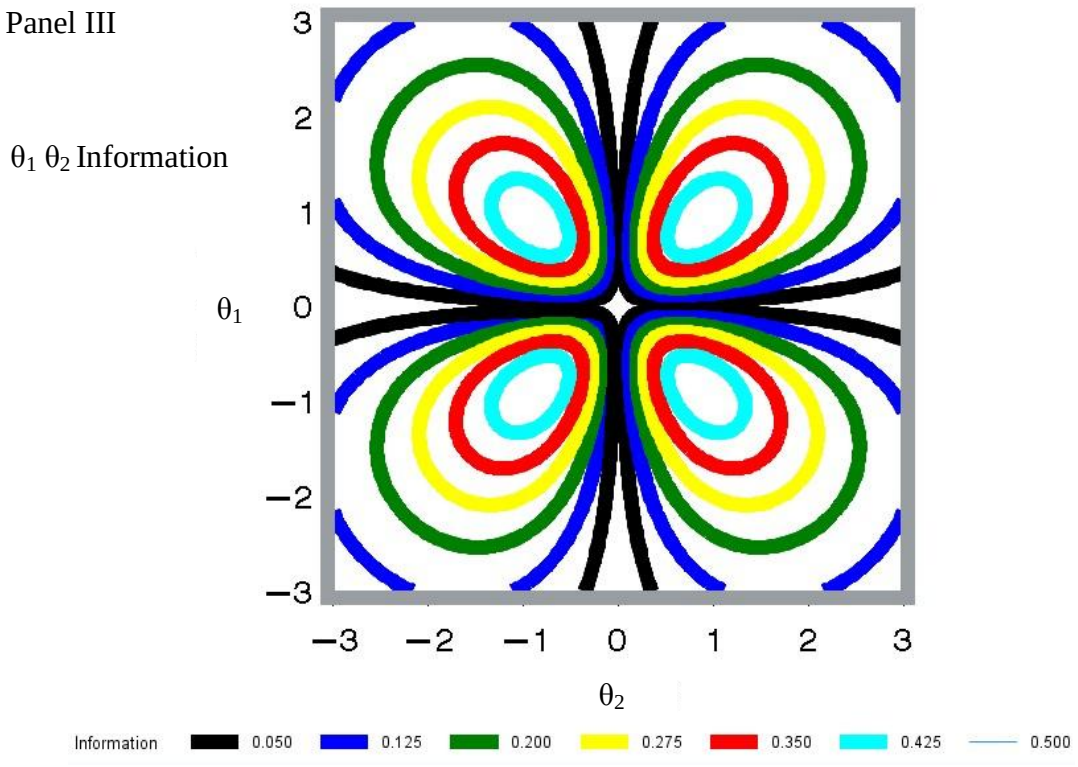



*Figure 10b.* Continued.

It is argued that within the context of the MGGUM, there is only one optimal orientation of the dimensions, which will be identified during the estimation process (Roberts & Shim, 2010). The idea of optimal orientation bears similarities to Carroll and Chang's (1970) INDSCAL model in multidimensional scaling. It may be the case that other, or even all, noncompensatory MIRT models have a unique optimal orientation given the structure of the models. The discrimination matrix needs to be a diagonal matrix for each item. Rotation, however, will result in a non-diagonal matrix. Therefore, the meaning of a rotated solution conflicts with the basic definition of the MGGUM (Roberts et al., 2009a; Roberts & Shim, 2010). Thus, the existence of the singular

optimal orientation of dimensions suggests there is a unique orientation which may have a substantive psychological meaning (Carroll & Chang, 1970).

### 3.1 Estimation in the Multidimensional Generalized Graded Unfolding Model

At present, only fully Bayesian parameter estimation techniques implemented via MCMC have been utilized with the MGGUM (Roberts et al., 2009a; Roberts & Shim, 2010). In previous MCMC studies estimates of item locations ($\hat{\delta}_{id}$) and person parameters ($\hat{\theta}_1,\ldots,\hat{\theta}_D$) were obtained using Metropolis-Hastings sampling (Hastings, 1970; Metropolis et al., 1953). Slice sampling was used to estimate discrimination ($\hat{\alpha}_{id}$) and subjective response category threshold ($\hat{\tau}_{ik}$) parameters (Neal, 2003).

Research has shown MCMC is able to adequately recover MGGUM item parameters ($\delta_{id}$, $\alpha_{id}$, $\tau_{ik}$) and EAP person parameters ($\theta_{jd}$) using RMSD of estimates to true values as an indicator of accuracy (Roberts et al., 2009a; Roberts & Shim, 2010). Similar to the performance of MCMC with the unidimensional GGUM (Roberts & Thompson, 2011), estimation accuracy for MGGUM item location ($\delta_{id}$) and subjective response category threshold parameters ($\tau_{ik}$) is enhanced as sample size increases (Roberts et al., 2009a; Roberts & Shim, 2010). However, dimensional structure of items within the MGGUM appears to affect accuracy of parameter estimation as well. Complex structure enhances accuracy of discrimination ($\alpha_{id}$) and subjective response category thresholds ($\tau_{ik}$), but has the opposite effect on person ($\theta_{jd}$) and item location ($\delta_{id}$) parameters. Estimation accuracy these parameters decreases in complex structure conditions, but improves with simple structure. As noted by Roberts et al. (2009a), the reason item structure affects item location estimates in this manner is not entirely clear. It may be

related to the calculation of RMSD. RMSD in complex structure has been calculated using item location estimates for all dimensions. However, in the case of simple structure, RMSD is calculated using only item location estimates for the measured dimension.

With respect to person parameters ($\theta_{jd}$), certain conditions have led to increased accuracy of estimates in the presence of complex structure items. Longer tests with complex structure items lead to greater accuracy (Roberts et al., 2009a). Although, with longer tests there will be more item parameters to estimate, which will likely be estimated with less accuracy (larger RMSD). Despite differences in estimation accuracy for parameters in certain dimensional structure conditions, all estimates were of an acceptable level of accuracy (Roberts et al., 2009a; Roberts & Shim, 2010). It is possible that similar findings regarding the effect of item structure on item location and person parameter estimates may occur with other estimation methods.

In the unidimensional GGUM, MCMC/EAP estimation is an extremely time-consuming process taking hours to converge upon a solution given the volume and complexity of the data (number of subjects, items, and response categories; Roberts & Thompson, 2011). In the multidimensional model, the computational efficiency of MCMC/EAP estimation is substantially worse (Roberts et al., 2009a; Roberts & Shim, 2010). As model complexity increases, more time is required to estimate parameters. Days of estimation in a unidimensional model become weeks in a multidimensional model. There is some appeal to implementing MCMC/EAP as it avoids calculation of potentially cumbersome derivatives and is able to conduct joint estimation of all model parameters (Béguin & Glas, 2001; Yao & Schwartz, 2006). However, unless researchers

have ample time to conduct a study; it is not the most realistic option. Waiting weeks for parameter estimates only to find out a model does not appear to fit the data is a luxury few can afford. A marginal approach is likely to reach a solution much more efficiently.

Efficiency may also be impacted by the selection of initial values, whether using a marginal or fully Bayesian estimation method. Previous research implementing MCMC/EAP with a two-dimensional MGGUM identified initial values using unidimensional GGUM estimates (Roberts et al., 2009a; Roberts & Shim, 2010). Initial values for the first dimension came from unidimensional estimates regardless of item structure, while values for the second dimension were set to zero. Initial subjective response category thresholds were then computed based on item locations and the number of response categories following the technique of Roberts and Thompson (2011).

Using the initial value strategy detailed above there is relatively no information guiding a starting point to estimate the second dimension. In the case of simple structure items this may seem like a non-issue. However, if some items assess one dimension and others assess a different dimension, this strategy implies that some of the items are being estimated from an item location and discrimination starting point of zero on their measured dimension. In the case of complex structure, where items load onto all dimensions, there is at least some initial information for one dimension, but all items begin the estimation process with item locations and discriminations starting at zero on the remaining dimensions. Therefore, a different initial value strategy which provides information relative to all dimensions may be more appropriate and increase efficiency. Detrended correspondence analysis (DCA; Hill & Gauch, 1980), a method used primarily in ecological research, is described in Section 5.4 as such an alternative.

In addition to the issue of efficiency, previous research with MCMC/EAP in the MGGUM has raised other concerns. Roberts et al. (2009a) and Roberts and Shim (2010) found that, on occasion, the meaning of dimensions could switch during a Markov chain. This is not a novel finding as Bolt and Lall (2003) observed something similar in their work with the Multicomponent Latent Trait Model (Whitely, 1980) using MCMC. In addition, MCMC/EAP in the MGGUM incorrectly estimated the sign on the non-measured dimension for a few individuals with extreme scores on that non-measured dimension (Roberts et al., 2009a). Utilizing a different form of parameter estimation like MMAP/EAP may eliminate these problems. For example, the misestimation of extreme person parameters may be related to the fact that MCMC jointly estimates parameters. Thus, it requires initial values for all types of parameters. Simplifying the process and estimating one set of parameters at a time may improve accuracy of estimates. In addition, EAP estimates based on marginal item estimates, like MMAP, are computed after obtaining item parameter estimates. It is possible EAP estimation accuracy may improve when calculated using potentially more accurate item parameters.

The reversal of dimensional meaning may also be related to the estimation method or perhaps the initial values of item parameters. It is possible the identities of dimensions may still switch using MMAP/EAP. However, integration of person parameters out of the solution may help stabilize the meaning of dimensions. Regardless, the MMAP/EAP procedure should be relatively quick, and if dimensional switching still occurs, then its impact on the solution time should be relatively less than that seen with MCMC/EAP. MCMC/EAP takes time to converge upon a posterior distribution after processing a number of burn-in iterations. If dimensions switch within the MCMC/EAP

procedure additional burn-ins may be required, which would increase the duration of an already computationally burdensome method.

# CHAPTER 4

## AIMS OF THE PRESENT STUDY

As previously discussed, research has shown MMAP item and EAP person parameter estimation is equal or superior to MML/EAP and fully Bayesian MCMC/EAP approaches implemented in the GGUM with respect accuracy and computational efficiency (Roberts & Thompson, 2011). The primary argument of the present research is that similar findings result when MMAP/EAP are applied to the multidimensional extension of the GGUM.

Fully Bayesian techniques are presently implemented in the MGGUM via MCMC which provides EAP estimates of all model parameters. The extensive computing time (i.e. days) required to use this method is not realistic for large-scale applications. MMAP/EAP offers an alternative estimation method that takes seconds/minutes to converge for the unidimensional GGUM. Given the increase in model complexity, the use of MMAP/EAP in the MGGUM will be somewhat slower than in the GGUM. However, MGGUM parameter estimation with MMAP/EAP should take considerably less time than with MCMC/EAP, at least for situations with limited dimensionality. Therefore, this study implements MMAP item parameter estimation in a two-dimensional MGGUM. The characteristics of this new approach are assessed using simulation techniques and the method is applied to real data to further illustrate its utility. In light of the previous work with MCMC/EAP in the MGGUM, a direct comparative analysis of MMAP/EAP and MCMC/EAP is included as well to address the issues of accuracy and efficiency.

As researchers have suggested the selection of initial values may affect estimation accuracy and efficiency (Nader et al., 2011), this study utilizes an alternative method of selecting initial values. Based on previous unidimensional work (Polak, 2011), detrended correspondence analysis (DCA; Hill & Gauch, 1980; see Section 5.4) is used to identify initial values for item and person location estimates. Initial subjective response category thresholds are then computed based on the item locations estimated with DCA. These initial values should be more informative, and hence lead to greater accuracy in a shorter amount of time when estimating MGGUM parameters.

# PART II: PROCEDURE

# CHAPTER 5

# PARAMETER ESTIMATION

## 5.1 Item Parameter Estimation

Implementing MMAP item parameter estimation in the MGGUM is a more complex process than with the unidimensional GGUM. Integrating over the latent space within the EM algorithm becomes more computationally burdensome with each additional dimension in the model. Typically 20-30 rectangular quadrature points are recommended in GGUM data demand studies (Roberts et al., 2002). If this rationale is applied to the MGGUM, then each additional dimension added to the model would exponentially increase the number of quadrature points. Thus, a simple two-dimensional MGGUM will require 900 quadrature points (i.e. 30 x 30). In addition, as the number of quadrature points increase, there will be an associated increase in duration of the estimation process, and the computation time could become substantial. In two-dimensional models, MMAP will likely take less time than MCMC. However, computational speed may still be an issue with MMAP in higher dimensional models. Research has suggested alternative estimation methods like adaptive quadrature may improve efficiency in such cases (Schilling & Bock, 2005). This is beyond the scope of the current proposal, but is a direction for future research.

The process of MMAP item parameter estimation in the MGGUM begins with deriving the likelihood function. Within the MGGUM, the conditional probability of a response vector $\mathbf{X}_j$ given the *j*th individual's latent trait levels is:

$$P\left(\mathbf{X}_j \left| \underline{\theta}_j \right.\right) = \prod_{i=1}^{I} P\left(Z_i = x_{ji} \left| \underline{\theta}_j \right.\right). \tag{52}$$

The marginal probability of a response vector becomes:

$$P\left(\mathbf{X}_j\right) = \int \ldots \int P\left(\mathbf{X}_j \left| \theta_{j1} \ldots \theta_{jD} \right.\right) g_1\left(\theta_{j1}\right) \ldots g_D\left(\theta_{jD}\right) d\theta \ldots \, d\theta_D \tag{53}$$

where $g_1(\theta_{j1})$ ... $g_D(\theta_{jD})$ are prior population distributions for each of the $D$ dimensions. Higher order integration is necessary with each additional dimensional in the MGGUM. Building upon this, the marginal posterior likelihood for the MGGUM becomes:

$$L \propto \prod_{j=1}^{J} \int \ldots \int \left( \prod_{i=1}^{I} \begin{pmatrix} P\left(Z_i = x_{ji} \mid \theta_{j1} \ldots \theta_{jD}\right) g_1\left(\theta_{j1}\right) \ldots g_D\left(\theta_{jD}\right) \\ b_1\left(\delta_{i1}\right) \ldots b_D\left(\delta_{iD}\right) a_1\left(\alpha_{i1}\right) \ldots a_D\left(\alpha_D\right) \prod_{k=1}^{C} t\left(\tau_{ik}\right) \end{pmatrix} \right) d\theta_{j1} \ldots \, d\theta_{jD} \tag{54}$$

where

$x_{ji}$ is the observed response for the $j$th individual to the $i$th item,

$b_d$ ($\delta_{id}$) is a prior distribution for the $i$th item location on the $d$th dimension,

$a_d$ ($\alpha_{id}$) is a prior distribution for the $i$th item discrimination on the $d$th dimension, and

$t$ ($\tau_{ik}$) is a prior distribution for the $k$th subjective response category for the $i$th item.

Taking the logarithm of the marginal posterior likelihood in Equation 54 results in:

$$\begin{aligned} \ln\left(L\right) &\propto \ln\left( \left\{ \prod_{j=1}^{J} P(\mathbf{X}_j) \right\} \left\{ \prod_{i=1}^{I} \left( b_1\left(\delta_{i1}\right) \ldots b_D\left(\delta_{iD}\right) a_1\left(\alpha_{i1}\right) \ldots a_D\left(\alpha_D\right) \prod_{k=1}^{C} t\left(\tau_{ik}\right) \right) \right\} \right) \\ &= \left\{ \sum_{j=1}^{J} \ln P(\mathbf{X}_j) \right\} + \left\{ \sum_{i=1}^{I} \left( \sum_{d=1}^{D} \left[ \ln\left(b_d\left(\delta_{id}\right)\right) + \ln\left(a_d\left(\alpha_{id}\right)\right) \right] + \sum_{k=1}^{C} \ln\left(t\left(\tau_{ik}\right)\right) \right) \right\}. \end{aligned} \tag{55}$$

MGGUM item parameters can be obtained by finding the respective roots of first-order partial derivatives of Equation 55. However, these estimates are approximated in quadrature form (full derivations are available in Appendix B):

$$\frac{\partial \ln(L)}{\partial \delta_{id}} = \left( \sum_{q_1}^{Q_1} \cdots \sum_{q_d}^{Q_D} \left[ \sum_{z=0}^{C} \frac{\bar{r}_{izq_1 \ldots q_d}}{P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)} \frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \delta_{id}} \right] \right) + \frac{\partial \ln\left[b(\delta_{id})\right]}{\partial \delta_{id}}, \quad (56)$$

$$\frac{\partial \ln(L)}{\partial \alpha_{id}} = \left( \sum_{q_1}^{Q_1} \cdots \sum_{q_d}^{Q_D} \left[ \sum_{z=0}^{C} \frac{\bar{r}_{izq_1 \ldots q_d}}{P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)} \frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \alpha_{id}} \right] \right) + \frac{\partial \ln\left[a(\alpha_{id})\right]}{\partial \alpha_{id}}, \quad (57)$$

and

$$\frac{\partial \ln(L)}{\partial \tau_{ik}} = \left( \sum_{q_1}^{Q_1} \cdots \sum_{q_d}^{Q_D} \left[ \sum_{z=0}^{C} \frac{\bar{r}_{izq_1 \ldots q_d}}{P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)} \frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \tau_{ik}} \right] \right) + \frac{\partial \ln\left[t(\tau_{ik})\right]}{\partial \tau_{ik}} \quad (58)$$

where $A_{q_1} \ldots A_{q_d}$ are quadrature coordinates (i.e. a quadrature point in *d*-dimensional space), and

$$P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right) \quad (59)$$

is the probability of response *z* evaluated at a quadrature point. Equations 56 through 58 also involve the derivatives of Equation 59 with respect to a particular item parameter ($\delta_{id}$, $\alpha_{id}$, $\tau_{ik}$), which are available in Appendix C.

The expected number of individuals using a particular response category, *z*, for item *i* at a quadrature point is:

$$\bar{r}_{izq_1 \ldots q_d} = \sum_{j=1}^{J} \frac{H_{jiz} L_j \left( A_{q_1} \ldots A_{q_d} \right) W\left( A_{q_1} \right) \ldots W\left( A_{q_d} \right)}{\tilde{P}_j} \tag{60}$$

where $H_{jiz}$ has a value of 1 when $z$ equals $x_{ji}$ and 0 otherwise. The weight at a quadrature point in latent space is $W\left( A_{q_d} \right)$ and

$$L_j \left( A_{q_1} \ldots A_{q_d} \right) = \prod_{i=1}^{I} P\left( Z_i = x_{ji} \middle| A_{q_1} \ldots A_{q_d} \right) \tag{61}$$

is the conditional probability of a particular response pattern at a quadrature point in latent space. The marginal probability of a particular response pattern is:

$$\tilde{P}_j = \sum_{q_1}^{Q_1} \ldots \sum_{q_d}^{Q_D} L_j \left( A_{q_1} \ldots A_{q_d} \right) W\left( A_{q_1} \right) \ldots W\left( A_{q_d} \right). \tag{62}$$

MMAP item parameter estimation with rectangular quadrature necessitates the specification of quadrature points along each dimension, forming a grid in the two dimensional case. The EM algorithm is then implemented by evaluating item parameter estimates at each quadrature point along one dimension, holding the coordinate values for all other dimensions constant. Next, the researcher evaluates estimates at quadrature points along the second dimension holding the coordinate values of all other dimensions constant. This process is continued until all dimensions have been evaluated. Final item parameter estimates are obtained when the global maximum of the log marginal likelihood function is identified across all dimensions.

Following the approach of Roberts et al. (2000) and Roberts and Thompson (2011), Fisher scoring is used in the EM cycles. Fisher scoring requires calculating

information with respect to a particular item parameter. In a *D*-dimensional MGGUM, information for subjective response category thresholds estimates ( $\hat{\tau}_{ik}$ ) is equal to:

$$\tilde{\mathbf{I}}_{\boldsymbol{\tau}_{\mathbf{i}}} = \begin{pmatrix} \tilde{I}_{\tau_{i1}\tau_{i1}} & \cdots & \tilde{I}_{\tau_{i1}\tau_{iC}} \\ \vdots & \ddots & \vdots \\ \tilde{I}_{\tau_{iC}\tau_{i1}} & \cdots & \tilde{I}_{\tau_{iC}\tau_{iC}} \end{pmatrix}. \tag{63}$$

Subjective response category thresholds can then be estimated iteratively using the parameter gradient along with Equation 63 as follows:

$$\begin{pmatrix} \tau_{i1} \\ \tau_{i2} \\ \vdots \\ \tau_{iC} \end{pmatrix}_{J} = \begin{pmatrix} \tau_{i1} \\ \tau_{i2} \\ \vdots \\ \tau_{iC} \end{pmatrix}_{J-1} + \left(\tilde{\mathbf{I}}_{\tau_{ik}}\right)^{-1} \begin{pmatrix} \dfrac{\partial \ln L}{\partial \tau_{i1}} \\ \dfrac{\partial \ln L}{\partial \tau_{i2}} \\ \vdots \\ \dfrac{\partial \ln L}{\partial \tau_{iC}} \end{pmatrix}. \tag{64}$$

When estimating item location ( $\hat{\delta}_{id}$ ) and discrimination ( $\hat{\alpha}_{id}$ ) parameters that may vary across dimensions, information with respect to these parameters is defined as:

$$\tilde{\mathbf{I}}_{\alpha_{id}\delta_{id}} = \begin{pmatrix} \tilde{I}_{\alpha_{i1}\alpha_{i1}} & \cdots & \tilde{I}_{\alpha_{i1}\alpha_{id}} & & \tilde{I}_{\alpha_{i1}\delta_{i1}} & \cdots & \tilde{I}_{\alpha_{i1}\delta_{id}} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \tilde{I}_{\alpha_{id}\alpha_{i1}} & \cdots & \tilde{I}_{\alpha_{id}\alpha_{id}} & & \tilde{I}_{\alpha_{id}\delta_{i1}} & \cdots & \tilde{I}_{\alpha_{id}\delta_{id}} \\ & \vdots & & \ddots & & \vdots & \\ \tilde{I}_{\alpha_{i1}\delta_{i1}} & \cdots & \tilde{I}_{\alpha_{i1}\delta_{id}} & & \tilde{I}_{\delta_{i1}\delta_{i1}} & \cdots & \tilde{I}_{\delta_{i1}\delta_{id}} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \tilde{I}_{\alpha_{id}\delta_{i1}} & \cdots & \tilde{I}_{\alpha_{id}\delta_{id}} & & \tilde{I}_{\delta_{id}\delta_{i1}} & \cdots & \tilde{I}_{\delta_{id}\delta_{id}} \end{pmatrix}. \tag{65}$$

In a two-dimensional model this becomes:

$$\tilde{\mathbf{I}}_i = \begin{pmatrix} \tilde{I}_{\alpha_{i1}\alpha_{i1}} & \tilde{I}_{\alpha_{i1}\alpha_{i2}} & \tilde{I}_{\alpha_{i1}\delta_{i1}} & \tilde{I}_{\alpha_{i2}\delta_{i1}} \\ \tilde{I}_{\alpha_{i1}\alpha_{i2}} & \tilde{I}_{\alpha_{i2}\alpha_{i2}} & \tilde{I}_{\alpha_{i1}\delta_{i2}} & \tilde{I}_{\alpha_{i2}\delta_{i2}} \\ \tilde{I}_{\alpha_{i1}\delta_{i1}} & \tilde{I}_{\alpha_{i2}\delta_{i1}} & \tilde{I}_{\delta_{i1}\delta_{i1}} & \tilde{I}_{\delta_{i1}\delta_{i2}} \\ \tilde{I}_{\alpha_{i1}\delta_{i2}} & \tilde{I}_{\alpha_{i2}\delta_{i2}} & \tilde{I}_{\delta_{i1}\delta_{i2}} & \tilde{I}_{\delta_{i2}\delta_{i2}} \end{pmatrix}. \tag{66}$$

Information is then used to iteratively update parameter estimates:

$$\begin{pmatrix} \alpha_{i1} \\ \vdots \\ \alpha_{id} \\ \delta_{i1} \\ \vdots \\ \delta_{id} \end{pmatrix}_J = \begin{pmatrix} \alpha_{i1} \\ \vdots \\ \alpha_{id} \\ \delta_{i1} \\ \vdots \\ \delta_{id} \end{pmatrix}_{J-1} + \left(\tilde{\mathbf{I}}_i\right)^{-1} \begin{pmatrix} \frac{\partial \ln L}{\partial \alpha_{i1}} \\ \vdots \\ \frac{\partial \ln L}{\partial \alpha_{id}} \\ \frac{\partial \ln L}{\partial \delta_{i1}} \\ \vdots \\ \frac{\partial \ln L}{\partial \delta_{id}} \end{pmatrix}, \tag{67}$$

which in two dimensions is:

$$\begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \delta_{i1} \\ \delta_{i2} \end{pmatrix}_J = \begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \delta_{i1} \\ \delta_{i2} \end{pmatrix}_{J-1} + \left(\tilde{\mathbf{I}}_i\right)^{-1} \begin{pmatrix} \frac{\partial \ln L}{\partial \alpha_{i1}} \\ \frac{\partial \ln L}{\partial \alpha_{i2}} \\ \frac{\partial \ln L}{\partial \delta_{i1}} \\ \frac{\partial \ln L}{\partial \delta_{i2}} \end{pmatrix}. \tag{68}$$

Preliminary work determined that a dimension by dimension approach was required to estimate item location ($\hat{\delta}_{id}$) and discrimination ($\hat{\alpha}_{id}$) parameters. In the case of items with simple structure, the information matrix associated with Equation 66 contains both zero and non-zero values. Non-zero values exist in the rows and columns

for the measured dimension parameters, while zero values pervade the rows and columns for the non-measured dimension. Therefore, inversion of this matrix, as seen in Equation 67, is not possible. There are entire rows and columns containing values of zero for the non-measured dimension. For consistency, the same method was applied to complex structure items because item structure is not always known before estimating item parameters. Therefore, using a multidimensional likelihood, parameter estimates from each dimension are obtained using dimension-specific information:

$$\tilde{\mathbf{I}}_{id} = \begin{pmatrix} \tilde{I}_{\alpha_{id}\alpha_{id}} & \tilde{I}_{\alpha_{id}\delta_{id}} \\ \tilde{I}_{\alpha_{id}\delta_{id}} & \tilde{I}_{\delta_{id}\delta_{id}} \end{pmatrix} \tag{69}$$

in the updating of:

$$\begin{pmatrix} \alpha_{id} \\ \delta_{id} \end{pmatrix}_{J} = \begin{pmatrix} \alpha_{id} \\ \delta_{id} \end{pmatrix}_{J-1} + \tilde{\mathbf{I}}_{id}^{-1} \begin{pmatrix} \frac{\partial \ln L}{\partial \alpha_{id}} \\ \frac{\partial \ln L}{\partial \delta_{id}} \end{pmatrix}. \tag{70}$$

The iterative updating of parameters occurs until there is little to no change in values from one iteration to the next. Derivations of elements in the information matrices denoted in Equations 63 and 69 can be found in Appendix D.

### 5.2 Person Parameter Estimation

EAP estimates for *D* dimensions of latent traits are based on the expectation of an individual's latent traits given the person's response vector. The conditional likelihood of the response vector, $L(\mathbf{X}_j|\theta_{j1}\ldots\theta_{jD})$, is multiplied by prior distributions for each

dimension, *d*. Taking into account the observed response pattern, $X_j$, integration over the latent space leads to the EAP estimate for a particular dimension:

$$\mathrm{E}\left(\theta_{jd}\left|\mathbf{X}_j\right.\right)=\frac{\int\ldots\int\theta_{jd}L\left(\mathbf{X}_j\left|\theta_{j1}\ldots\theta_{jD}\right.\right)g_1\left(\theta_{j1}\right)\ldots g_D\left(\theta_{jD}\right)d\theta_{j1}\ldots d\theta_{jD}}{\int\ldots\int L\left(\mathbf{X}_j\left|\theta_{j1}\ldots\theta_{jD}\right.\right)g_1\left(\theta_{j1}\right)\ldots g_D\left(\theta_{jD}\right)d\theta_{j1}\ldots d\theta_{jD}}. \qquad (71)$$

Equation 71 can be approximated with quadrature as:

$$\hat{\theta}_{jd}\approx\frac{\sum_{q_1}^{Q_1}\ldots\sum_{q_d}^{Q_D}A_{q_d}L_j\left(A_{q_1}\ldots A_{q_d}\right)W\left(A_{q_1}\right)\ldots W\left(A_{q_d}\right)}{\sum_{q_1}^{Q_1}\ldots\sum_{q_d}^{Q_D}L_j\left(A_{q_1}\ldots A_{q_d}\right)W\left(A_{q_1}\right)\ldots W\left(A_{q_d}\right)}. \qquad (72)$$

Thus, with a set of fixed rectangular quadrature coordinates, EAP produces MGGUM person parameter estimates ($\hat{\theta}_{jd}$) for the *j*th individual on the *d*th dimension using Equation 72.

### 5.3 MCMC / EAP Parameter Estimation

Using the procedures of Roberts et al. (2009a) and Roberts & Shim (2010), MCMC/EAP estimation in the MGGUM is also implemented in this study for a subset of the replications in the simulation design to provide a direct comparison of MMAP/EAP to MCMC/EAP. Metropolis-Hastings sampling is used to estimate item location ($\hat{\delta}_{id}$) and person ($\hat{\theta}_1,\ldots,\hat{\theta}_D$) parameters while slice sampling is used to estimate discrimination ($\hat{\alpha}_{id}$) and subjective response category threshold ($\hat{\tau}_{ik}$) parameters. No rotational constraints are required in light of the previous discussion regarding identifying the optimal orientation of dimensions in the estimation process.

### 5.4 Detrended Correspondence Analysis (DCA) for Initial Values

Research suggests the selection of initial values for item parameters may impact the accuracy of resulting estimates (Nader et al., 2011). For instance, it has been suggested estimation of item locations using initial values which are not in the vicinity of true item locations may lead to increased identification of local maxima (Roberts & Laughlin, 1996). More informative initial values should avoid or at least identify fewer local maxima. In addition, initial values closer to the true item parameters should decrease the time necessary to locate the global maximum. Therefore, it is desirable to obtain informative initial values for MGGUM item parameters.

Previous research implementing correspondence analysis (CA; Greenacre, 2007) with data that are consistent with the unidimensional GGUM (Polak, 2011) shows promise for an alternative initial value strategy for the MGGUM. CA, also known as reciprocal averaging (RA), is widely used in the field of ecology for categorical data and has only more recently been extended to other domains. Based on a table of frequency counts, **F**, CA uses singular value decomposition to scale or obtain dimensional structure (coordinate locations) of variables from an $n$ x $m$ contingency table, **P**, with observations $p_{nm}$. Each identified CA axis/dimension is orthogonal to other axes/dimensions (ter Braak & Prentice, 1988). Similar to principal components analysis, the first dimension will account for the greatest variance, the second dimension will account for the maximum variance that is orthogonal to the first dimension, and so on. Generally speaking, the dimensions accounting for the greatest variance and are subjectively interpretable are retained in the final solution (ter Braak, 1995).

Within the GGUM, **F** is a *J* x *I* matrix of observed responses ($x_{ji}$). This matrix is then transformed into a correspondence matrix, **P**, also known as a contingency table,

$$p_{ji} = x_{ji} / x_{++} \tag{73}$$

where $x_{++}$ is the sum over respective rows and columns. As noted by Greenacre (2007), the general algorithm is computed by first obtaining a matrix of standardized residuals, **S**, using the correspondence matrix:

$$\mathbf{S} = \mathbf{D}_r^{-\frac{1}{2}} \left( \mathbf{P} - \mathbf{r}\mathbf{c}^{\mathrm{T}} \right) \mathbf{D}_c^{-\frac{1}{2}} \tag{74}$$

where

**r** = **P1** is a vector of row marginal probabilities,

**c** = $\mathbf{P}^{\mathrm{T}}\mathbf{1}$ is a vector of column marginal probabilities,

$\mathbf{D}_r$ is the diagonal matrix of row marginal probabilities, and

$\mathbf{D}_c$ is the diagonal matrix of column marginal probabilities.

The **S** matrix now consists of standardized deviations of locations in space based on a metric referred to as chi-square distances:

$$s_{ji} = \frac{p_{ji} - p_{+i} p_{j+}}{\sqrt{p_{+i} p_{j+}}}. \tag{75}$$

Next, the singular value decomposition of **S** is obtained through an iterative process, available in Appendix E,

$$\mathbf{S} = \mathbf{U}\mathbf{D}_a\mathbf{V}^{\mathrm{T}} \tag{76}$$

where

**U** contains $u_{j1}, \ldots, u_{j[min(j-1, i-1)]}$,

**V** contains $v_{i1}, \ldots, v_{i[min(j-1, i-1)]}$,

$\mathbf{U}^{T}\mathbf{U} = \mathbf{V}^{T}\mathbf{V} = \mathbf{I}$, and

$\mathbf{D}_{a}$ is the min(j-1, i-1) *x* min(j-1, i-1) square diagonal matrix of singular values, *a*, in descending order.

This singular value decomposition of a rectangular matrix is akin to eigenvector-eigenvalue decomposition of a square symmetric matrix.

After completion of this singular value decomposition, the standardized coordinates are computed separately for rows, **Φ**, and columns, **Γ**,

$$\mathbf{\Phi} = \mathbf{D}_{r}^{-\frac{1}{2}}\mathbf{U} \tag{77}$$

$$\mathbf{\Gamma} = \mathbf{D}_{c}^{-\frac{1}{2}}\mathbf{V}. \tag{78}$$

The final row, **R**, and column, **C**, coordinates are calculated using the standardized coordinates from Equations 77 and 78, respectively,

$$\mathbf{R} = \mathbf{D}_{r}^{-\frac{1}{2}}\mathbf{U}\mathbf{D}_{a} = \mathbf{\Phi}\mathbf{D}_{a} \tag{79}$$

$$\mathbf{C} = \mathbf{D}_{c}^{-\frac{1}{2}}\mathbf{V}\mathbf{D}_{a} = \mathbf{D}_{a}\mathbf{\Gamma}. \tag{80}$$

One drawback of traditional CA is that an artifact known as the *arch effect* may be present. This effect is identified when the estimated locations lie in an arch or parabolic shape along higher order dimensions (Peet, Knox, Case, & Allen, 1988).

Arches can occur anywhere beyond the first dimension regardless of whether these dimensions are interpretable (Jongman, ter Braak, & van Tongeren, 1995). With a unidimensional solution, an arch may be present along the second dimension. As seen in Figure 11 with a one-dimensional interpretable solution, item and person locations lie along an arch on the second dimension. However, in the presence of a two-dimensional interpretable solution an arch may be present on the second dimension, as well as higher order dimensions. In such situations the parabolic coordinate locations on higher order dimensions are artifacts of the estimation process.

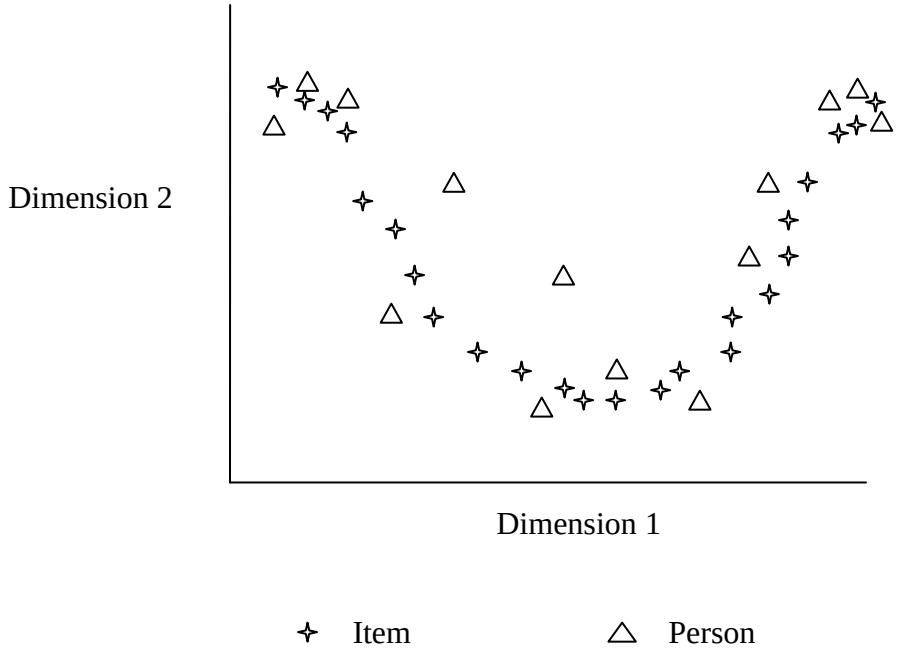



*Figure 11.* CA location estimates for two dimensions.

In the case of known unidimensional data, any additional dimensions identified by CA are typically ignored. Therefore, the arch effect is not an issue. However, with exploratory analyses, it becomes necessary to determine exactly how many dimensions

should be retained. There are no definitive standards for retaining dimensions in the literature. The common practice is to subjectively decide the proper number of dimensions based on interpretability. Using this method, researchers will likely have to contend with arch effects.

Another area of concern with CA is known as the *edge effect* (Greenacre, 2007). This occurs when there is decreased variation in the outer ranges of the latent space for higher order dimensions, whether interpretable or not. This artificial compression of points along the extremes of a dimension can also be seen in Figure 11 for two dimensions. The items and people located at the extreme ends the first dimension have decreased variation in their locations on the second dimension. Again, with known unidimensional data, this effect is a non-issue. However, higher order dimensional solutions will need to contend with edge effects because additional dimensions may be real dimensions versus artifacts of the estimation process. Artificial compression along the extremes of an interpretable dimension will lead to less accurate location estimates. A solution to combat both the arch and edge effects in such situations will be discussed shortly.

Using CA with row principal normalization, Polak (2011) conducted a parameter recovery study in the context of the unidimensional GGUM. CA item location estimates of generated GGUM items were extremely close to actual item locations. The high correspondence rate between estimates and true locations suggests CA is a sufficient method to obtain GGUM item location estimates. While CA performed well with the GGUM, it should be noted that a second dimension was observed containing an arch effect. As unidimensional data was used in the study, the second dimension was

discarded. It should be noted there are other benefits to utilizing CA with the GGUM. CA does not require the specification of initial values or prior distributions in the estimation process. In addition, CA is able to simultaneously estimate the location of persons and items, whereas marginal estimation methods require separate estimation of persons and items (Polak, 2011; Roberts & Thompson, 2011).

Given the performance of CA within the unidimensional GGUM in estimating item locations, it seems to reason that CA could be implemented within the MGGUM to provide informative initial values for item locations. However, with multidimensional data the possibility of obtaining arch and edge effects must be addressed. As an arch effect was observed in unidimensional data, it is quite likely one will also be observed in multidimensional data. With respect to the edge effect, this may be a concern when there are few extreme MGGUM items such that there is decreased variability at the ends of a dimension. The present study hopes to address and/or avoid these issues by implementing a variant of CA developed to alleviate, or at least diminish, these effects regardless of model dimensionality.

Detrended correspondence analysis (DCA; Hill & Gauch, 1980) is a modification of CA, still based on chi-square distances, thought to be a more appropriate for analyzing multidimensional data (Peet et al., 1988). After conducting CA to identify the number of dimensions and produce initial location estimates in multidimensional space, an additional step is added, namely the process of detrending. This detrending is able to reduce the effects of both the edge and arch effects. After detrending in a multidimensional solution, the first dimension will still account for the greatest amount of variance with additional dimensions accounting for decreasing amounts of variance.

Detrending is carried out such that locations on the second dimension are detrended with respect to the first dimension, locations on the third dimension are detrended with respect to the first and second dimensions, and so on with all higher order dimensions detrended with respect to lower order dimensions (Jongman et al., 1995).

There are two common ways to detrend, but both methods will flatten an existing arch effect. An arch on the second dimension can be thought of as the result of folding along the first dimension. In order to reduce this arch, one detrending method is to segment the first dimension and normalize locations on the second dimension within each of the segments. The edge effect is then reduced by re-segmenting and re-normalizing multiple times using different starting points along a dimension (Hill, 1994). The number of segments used is arbitrary and set by the researcher.

The second detrending method is to detrend by higher order polynomials. A second extracted dimension is identified so that it is orthogonal to the first dimension. Thinking of an arch as the result of folding the first dimension, the presence of an arch suggests the second dimension is quadratically related to the first dimension. Therefore, detrending by polynomials requires an extracted dimension be orthogonal to a lower order dimension and orthogonal to the square, cube, etc., of that lower order dimension (Jongman et al, 1995). It is argued this is a 'more stable' method of detrending since detrending by segments can still lead to artificial compression of variance, otherwise known as the edge effect (Jongman et al, 1995, p.108; Minchin, 1987). Upon completion of the detrending process, regardless of the detrending method, the resulting location estimates should resemble those found in Figure 12 for two interpretable dimensions with reductions in the arch and edge effects compared to Figure 11.

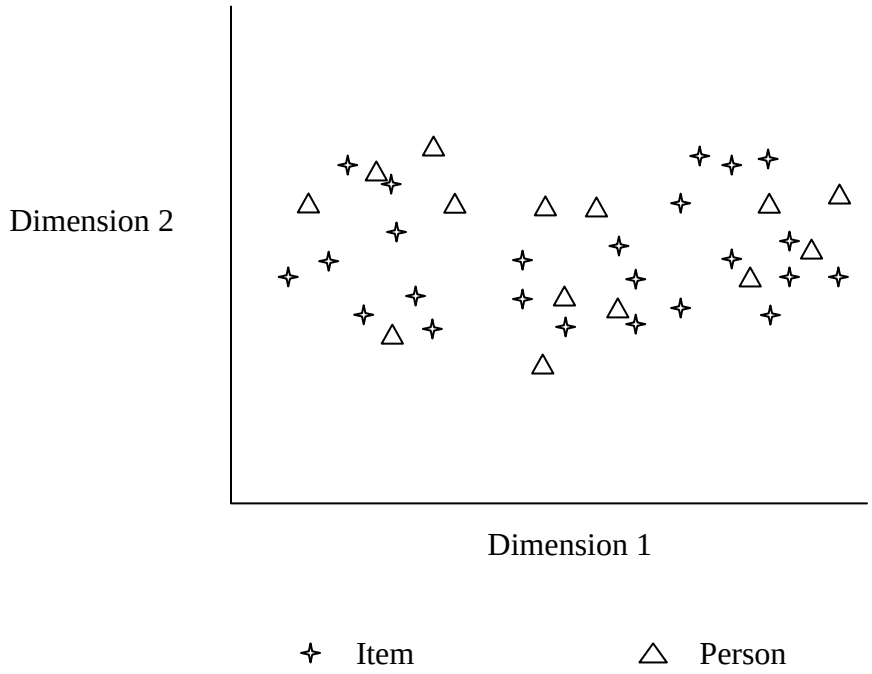



*Figure 12.* DCA location estimates for two dimensions.

Implementing DCA within the MGGUM for item location initial values involves first obtaining CA estimates of item locations for each dimension based on the aforementioned computations in Equation 80, as well Appendix E. Item locations are estimated using:

$$\hat{\delta}_{id} = \frac{a_d^{\kappa} v_{id}}{\sqrt{p_{+j}}} \tag{81}$$

where $\kappa$ takes on a value of 0 for row principal normalization, 1 for column principal normalization, or ½ for symmetric normalization. This study utilizes row principal normalization such that locations are normalized within a row or person (Polak, 2011).

Detrending by polynomials is then conducted to obtain estimates for higher order dimensions by extracting each additional dimension such that it is orthogonal not only to lower order dimensions, but any polynomial function of lower order dimensions. Coordinate locations on the second dimension are identified such that they are orthogonal to locations on the first dimension, $\hat{\delta}_{i2} \perp \hat{\delta}_{i1}$, and orthogonal to the square of locations on the first dimension, $\hat{\delta}_{i2} \perp \left(\hat{\delta}_{i1}\right)^2$. Locations on the third dimension are identified such that they are orthogonal to locations on the first dimension, $\hat{\delta}_{i3} \perp \hat{\delta}_{i1}$, orthogonal to the square of locations on the first dimension, $\hat{\delta}_{i3} \perp \left(\hat{\delta}_{i1}\right)^2$, and orthogonal to the cube of locations on the first dimension, $\hat{\delta}_{i3} \perp \left(\hat{\delta}_{i1}\right)^3$. In an analogous fashion, locations on the third dimension will also be orthogonal to locations on the second dimension to the first, second, and third power. The extraction of further higher order dimensions continues in the same manner (Jongman et al., 1995). This is implemented in the iterative singular value decomposition process found in Appendix E. Thus, in a two-dimensional MGGUM, this implies DCA item locations on the first dimension will be equal to CA item locations on the first dimension. Only locations on higher order dimensions will differ between the two methods given the detrended dimensional extraction.

In light of the past success identifying GGUM item locations with CA and the ability to counteract the arch and edge artifacts using DCA, the present study implements DCA with detrending by polynomials to identify initial values for MGGUM item locations ($\delta_{id}$). While DCA is utilized to identify initial values for item location

parameters, item discrimination and subjective response category threshold parameters also require initial values for MMAP estimation, which is discussed further in Chapter 6.

# CHAPTER 6

# SIMULATION DESIGN

## 6.1 Estimation Programs

Following the design of previous research with the MGGUM (Roberts et al., 2009a; Roberts & Shim, 2010) and GGUM (Roberts & Thompson, 2011), a parameter recovery study and analysis of real attitude measurement data was conducted. Multidimensional MMAP item parameter estimates ($\hat{\delta}_{id}$, $\hat{\alpha}_{id}$, $\hat{\tau}_{ik}$) were obtained with rectangular quadrature using Fisher scoring. EAP person parameter estimates ($\hat{\theta}_{jd}$) were then calculated using final MMAP item parameter estimates. The MMAP/EAP estimation was performed with a modified version of the GGUM2004 software (Roberts et al., 2006). The FORTRAN modifications were performed by the author to estimate MGGUM parameters using MMAP/EAP. MCMC estimation of parameters was accomplished using OpenBUGS (Lunn, Spiegelhalter, Thomas, & Best, 2009), an open-source version of WinBUGS (Spiegelhalter et al., 2007). All work was conducted on computers running a 64-bit operating system with either Intel® Core™ 2 Quad processors or Intel® Core™ i7 processors and 8 GB of memory.

## 6.2 Factorial Design

This study utilized a two-dimensional MGGUM. The performance of MMAP/EAP was examined while varying test length, sample size, number of response categories, and dimensional structure. Three test lengths (10, 20, and 30 items), six sample sizes (500 to 2000 subjects in increments of 250), three response category conditions (2, 4, and 6 response categories), and two dimensional structure conditions

(simple and complex structure) were included. Following recommendations with the unidimensional GGUM (Roberts et al., 2002), there were 30 replications in each cell. Therefore, the MMAP/EAP parameter recovery study utilized a 3 *x* 7 *x* 3 *x* 2 (test length *x* sample size *x* response category *x* dimensional structure) design for a total of 126 cells with 30 replications each; resulting in 3,780 total replications.

Efficiency and accuracy of MMAP/EAP and MCMC/EAP was examined in a subset of the aforementioned design. Previous GGUM research indicated a difference in accuracy of parameter estimates between MMAP/EAP and MCMC/EAP manifested while varying the number of response categories (Roberts & Thompson, 2011). The estimates were most discrepant with two response categories, but little difference was observed with six response categories. No differences were found between the estimation methods while varying sample size or test length. Thus, MGGUM estimates in the current study were compared for two and six response categories, while holding the number of subjects (2000) and items (20) constant. Divergent accuracy results were thought to occur more often when comparing estimates from complex structure items, so item structure was held constant as well. Finally, in light of the time required to implement MCMC/EAP with the MGGUM (Roberts et al., 2009a; Roberts & Shim, 2010) only five replications were examined in each of these two cells of interest.

### 6.3 True Parameters

True person parameters ($\theta_{jd}$) were sampled from a multivariate normal distribution $\underline{\Theta}_j \sim \text{MVN}(\underline{0}, I)$. True values of multidimensional item parameters ($\delta_{id}$, $\alpha_{id}$) were developed from past MMAP estimates of GGUM parameters (Roberts, Lin, & Laughlin, 2001; Roberts & Thompson, 2011). MMAP item parameter locations of

generated GGUM data were used to classify items into five equally-spaced intervals. In the case of simple structure, stratified random sampling with replacement resulted in the selection of an equal number of items from each interval. Half of the items were assigned GGUM estimates as truth on the first dimension with the second dimension zeroed out. The remaining items reversed the assignment of GGUM estimates on dimensions. This resulted in an equal number of items spanning each dimension. For complex structure items, stratified random sampling with replacement was conducted twice, once for each dimension, resulting in an equal number of items from each interval. The samples for each dimension were randomly paired and the associated GGUM estimates were assigned to an item. In accordance with previous research, all item discrimination ($\alpha_{id}$) true values were rescaled by $1/\sqrt{2}$ so that similar information from different latent dimensions was available as dimensional structure varied (Roberts et al., 2009a). Subjective response category thresholds, on the other hand, were held constant across dimensions for an item. Therefore, true $\tau_{ik}$ parameters came from the average of resampled unidimensional GGUM $\tau_{ik}$ parameters.

### 6.4 Item Response Generation

True person and item parameters were used as input in the calculation of the probability of utilizing a specific response category using Equation 45. These probabilities then divided a [0, 1] interval into corresponding segments. After generating a random uniform number, the location of the number within the interval (within a probability range of a particular response category) indicated the generated response for that item. A contingency table check of the generated item responses verified all response categories were utilized across all items within a replication.

### 6.5 Prior Distributions

Person ($\theta_{jd}$), item location ($\delta_{id}$), and subjective response category threshold ($\tau_{ik}$) parameters were estimated in MMAP/EAP and MCMC using the following prior distributions: $\theta_{jd} \sim$ MVN(**0**, ***I***), $\delta_{id} \sim N(0, 4)$, and $\tau_{ik} \sim$ Lognormal($\mu_{ik}$, 1), where $\mu_{ik}$ is a linear function dependent upon the number of response categories and item location (difficulty) extremity. This function led to prior values for successive subjective response category thresholds that generally fell closer together as the number of response categories increased. It also shifted the prior values away from the item location ($\delta_{id}$) as the number of categories decreased.

Prior distributions for item discrimination ($\alpha_{id}$) parameters varied by estimation method. Pilot testing of a *lognormal*(0, 2) distribution with MMAP/EAP resulted in problematic item discrimination estimates ($\hat{\alpha}_{id}$) for simple structure items. Item discrimination estimates ($\hat{\alpha}_{id}$) for the non-measured dimension were at times noticeably different from zero (e.g. $\hat{\alpha}_{id} = 0.37$). Therefore, following the work of Béguin and Glas (2001), a normal prior, $N(0,1)$, was used for item discrimination ($\alpha_{id}$) estimation in MMAP/EAP. This prior had no noticeable effect on the MMAP/EAP estimates of complex structure items. Therefore, it was used for both complex and simple structure items. Pilot testing implementing the same normal prior in MCMC, where only complex structure items are considered, resulted in discrepant item discrimination estimates ($\alpha_{id}$). Item discrimination estimates ($\hat{\alpha}_{id}$) for measured dimensions were at times noticeably close to zero (e.g. $\hat{\alpha}_{id} = 0.08$). Thus, MCMC item discrimination ($\alpha_{id}$) estimation utilized a lognormal(0, 2) prior distribution used in earlier studies (Roberts & Shim, 2010).

## 6.6 Initial Values

The computer program CANOCO (ter Braak & Smilauer, 2002) was used to perform DCA on a generated item response matrix for each replication. This analysis resulted in item location ($\delta_{id}$) and person location ($\theta_{jd}$) coordinates in the same two-dimensional space. Initial subjective response category thresholds ($\tau_{ik}$) for both methods were computed using the initial item location ($\delta_{id}$) coordinates following the technique suggested in Roberts and Laughlin (1996). These thresholds need not be equidistant, but are constant across dimensions with $\tau_{i0} = 0$ for all items. Given that subjective response category thresholds are impacted by the extremity of an item, it is first necessary to identify the origin or location, *O*, with respect to the interthreshold distance. This is accomplished by approximating:

$$\tau_{i1} = \ln[O] = \ln\left[.961 + .707\sqrt{\delta_{i1}^2 + \delta_{i2}^2}\right]. \quad (82)$$

Next, an interthreshold distance quantity, $\Delta$, based on the number of response categories, is incorporated into the approximation formula. This yields initial estimates for successive item thresholds as *k* increases:

$$\tau_{ik} = \ln[O + \Delta] = \ln\left[\left(.961 + .707\sqrt{\delta_{i1}^2 + \delta_{i2}^2}\right) + \left([k-2]\left[.825 - .732C + .096C^2\right]\right)\right]. \quad (83)$$

If the initial value of $\tau_{ik}$ was less than zero, it was set to the arbitrary small value of ln(0.1).

Initial values for item discrimination ($\alpha_{id}$) parameters were set to 0.25 for all dimensions regardless of item structure and estimation method.

## 6.7 Quadrature

Numerical integration was performed in both the MMAP and EAP algorithms using rectangular quadrature in which there were 30 quadrature points spanning from -4.0 to 4.0 on each dimension. This gave rise to a 900 point grid on the two-dimensional latent plane. Quadrature points and the associated densities were static throughout both the MMAP and EAP procedures.

# CHAPTER 7

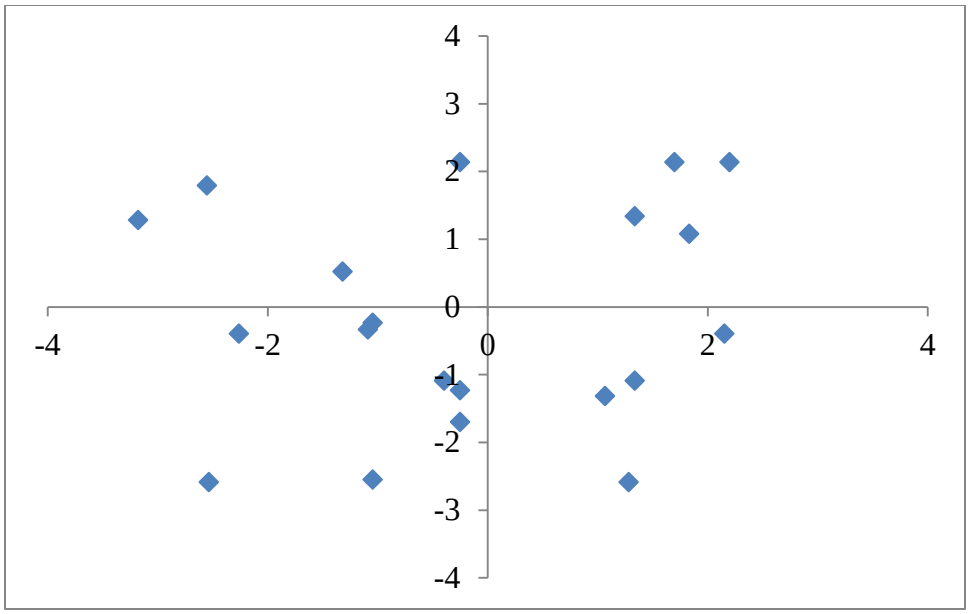
# RESULTS

## 7.1 DCA Initial Values

Prior to implementing the resultant DCA coordinates as initial values, the correlations between true multidimensional coordinate values and the DCA coordinate estimates were examined. Strong correlations were observed for items (average absolute $r$ = 0.85) and persons (average absolute $r$ = 0.90). On occasion, the DCA process reversed the poles of a dimension such that positive estimates corresponded with negative true generated values and vice versa, but this is irrelevant to the MGGUM likelihood function.

As an example, Figures 13 and 14 present one replication of DCA coordinates for items and persons, respectively, with 2000 subjects, 20 items, and 6 response categories in the complex structure condition. Panel I of each figure depicts the true parameter values, while Panel II depicts the DCA coordinate estimates.

Panel I




Panel II

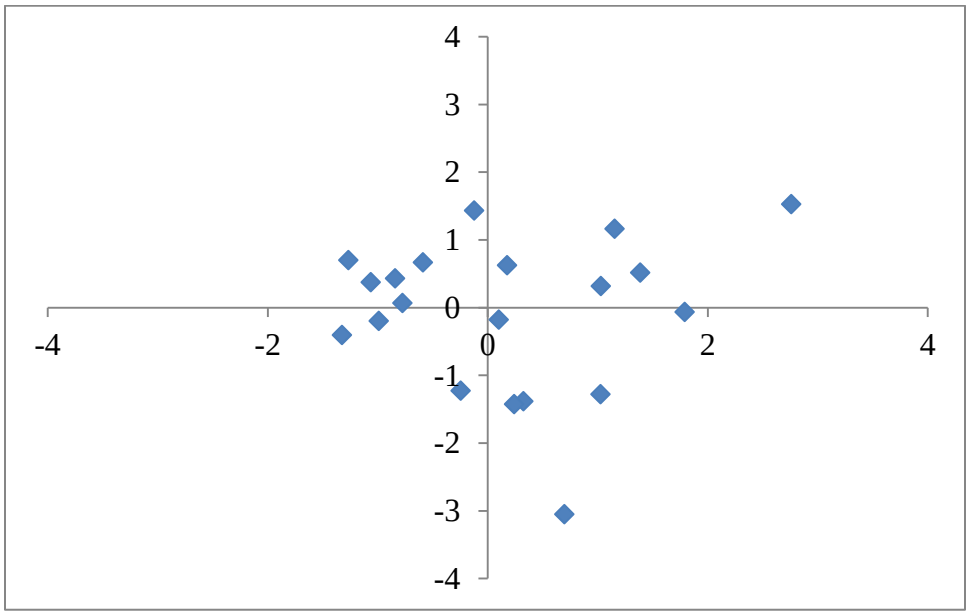
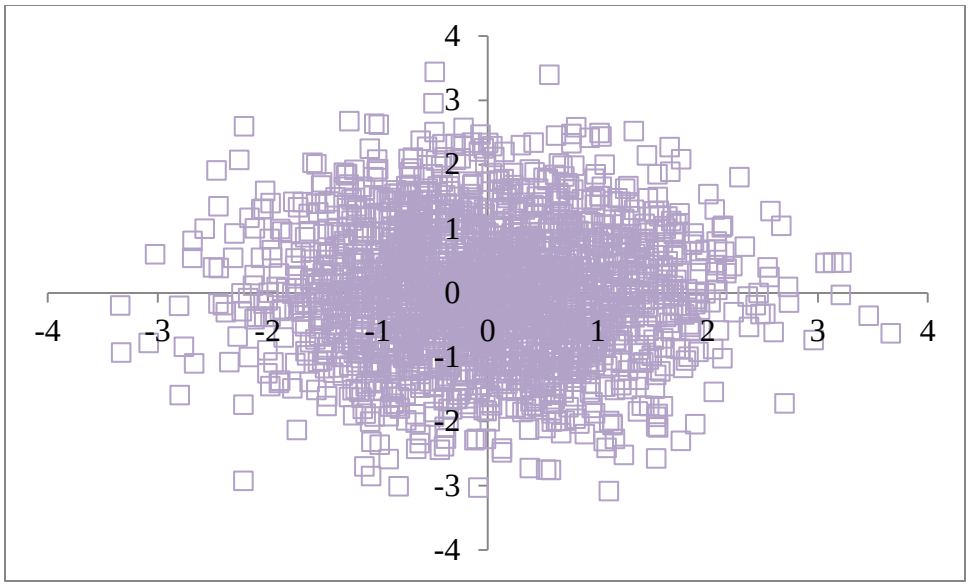



*Figure 13.* True (Panel I) and estimated DCA (Panel II) item locations.

Panel I




Panel II

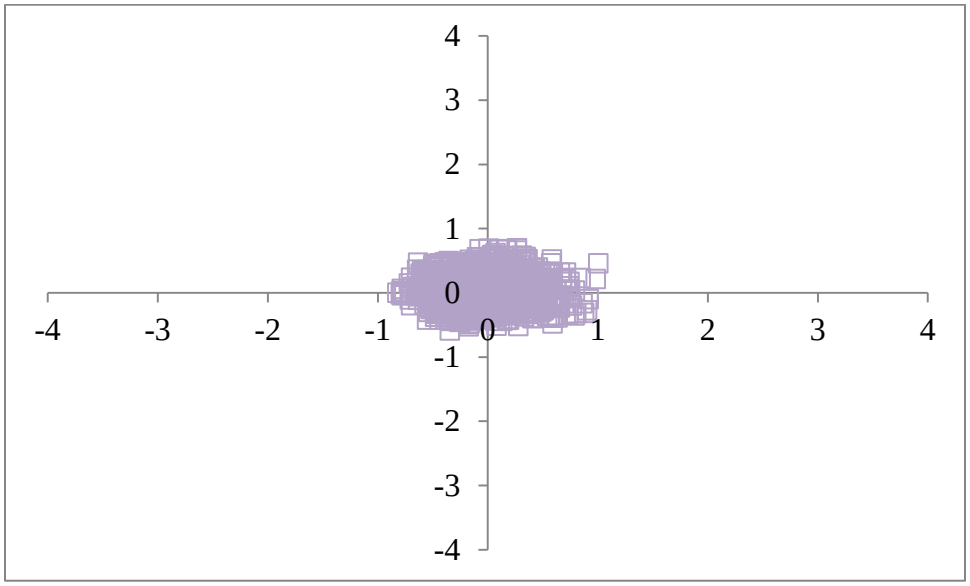



*Figure 14.* True (Panel I) and estimated DCA (Panel II) person locations.

In light of compression around the origin for the DCA estimates visible in Figures 13 and 14, coordinate locations of items and persons were rescaled to match their corresponding prior distributions, $\delta_{id} \sim N(0, 4)$ and $\theta_{jd} \sim \text{MVN}(\mathbf{0}, \boldsymbol{I})$. Figures 15 and 16

present the rescaled DCA coordinates of items and persons, respectively, for the same replication of 2000 subjects, 20 items, and 6 response categories in the complex structure condition.

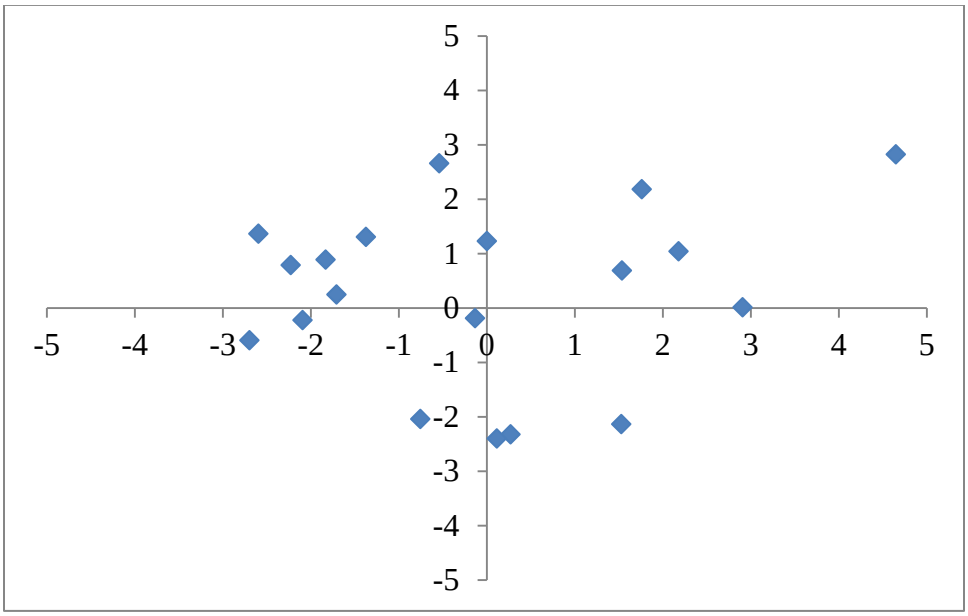



*Figure 15.* Rescaled DCA item location coordinates.

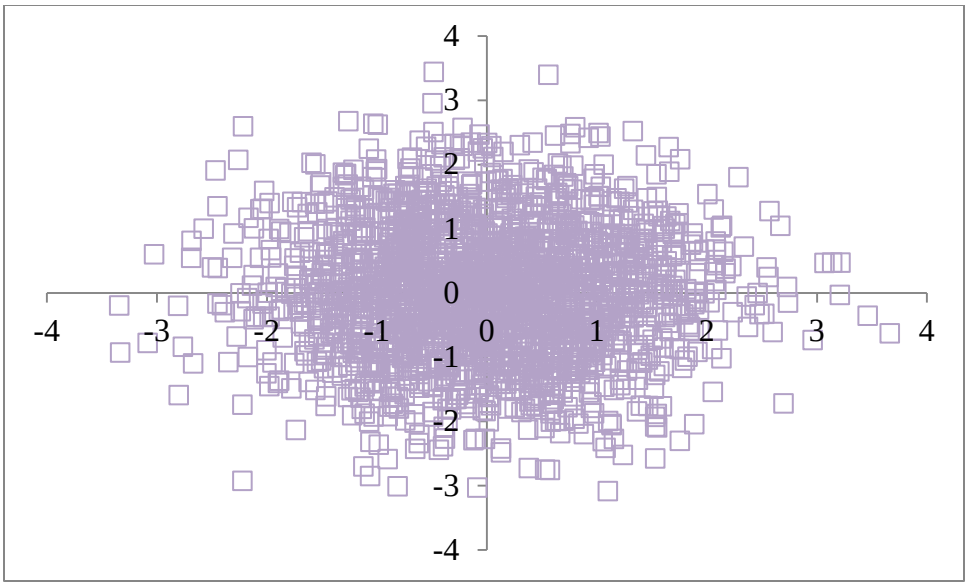



*Figure 16.* Rescaled DCA person location coordinates.

The rescaled item location ($\delta_{id}$) coordinates were then used as initial values in MMAP and MCMC estimation. Initial values for person parameters were also required for MCMC, thus the rescaled DCA person location coordinates were utilized.

### 7.2 Parameter Recovery

Comparisons of true and estimated item and person parameters were examined using the average RMSD across parameters in a given replication. The RMSD is a function of the similarity of means, variances, and covariance between true and estimated values. In this simulation study, the RMSDs for estimates of item locations ($\hat{\delta}_{id}$), item discriminations ($\hat{\alpha}_{id}$), and person parameters ($\hat{\theta}_{jd}$) were calculated as:

$$RMSD = \sqrt{\frac{\sum_{g=1}^{G}\sum_{d=1}^{2}\left(V_{gd}\right)\left(\hat{\lambda}_{gd} - \lambda_{gd}\right)^2}{2G}}. \quad (84)$$

Here

$g$ is a specific item or person,

$G$ is the total number of items or persons,

$d$ is the specific dimension,

$\lambda_{gd}$ is the true parameter value for $\delta_{id}$, $\alpha_{id}$, or $\theta_{jd}$, respectively, and

$\hat{\lambda}_{gd}$ is the estimated parameter value for $\delta_{id}$, $\alpha_{id}$, or $\theta_{jd}$.

$V_{gd}$ was set to one when estimating the RMSDs for $\hat{\alpha}_{id}$ and $\hat{\theta}_{jd}$. When estimating $\hat{\delta}_{id}$ in the simple structure condition, $V_{gd}$ was set to one for the measured dimension and zero for the non-measured dimension. For the complex structure condition $V_{gd}$ was set to one for both dimensions. Following Roberts et al. (2009a) and Roberts et al. (2002), mean square deviations for subjective response category thresholds were pooled across $C$ response categories when calculating the RMSD:

$$RMSD_{\tau_{ik}} = \sqrt{\frac{\sum_{i=1}^{I}\sum_{k=1}^{C}\left(\hat{\tau}_{ik} - \tau_{ik}\right)^2}{IC}} \tag{85}$$

where

$\tau_{ik}$ is the true parameter value and

$\hat{\tau}_{ik}$ is the estimated parameter value.

Four between-subjects factorial analyses of variance were conducted using the RMSD for a given parameter as the dependent measure in order to evaluate the accuracy of MMAP/EAP estimates. Each ANOVA examined the main effect of test length, sample size, number of response categories, and item structure along with all possible

interactions. Results were deemed interpretable when the effect size ($\eta^2$) was greater than or equal to 5% and observed statistical significance was below the Bonferroni corrected statistical significance level of $p \leq .05/4$. The Type I error rate was divided by four because the same ANOVA model was used to assess the accuracy of each of the four types of MGGUM parameters.

Separate split-plot analyses were conducted to compare the accuracy of MMAP/EAP and MCMC/EAP parameter estimates. Each of these analyses examined the main effects of estimation method and number of response categories (limited to either 2 or 6 categories) and their interaction. Estimation method was treated as a within-replications factor whereas the number of response categories was a between-replications effect. Following Roberts and Thompson (2011), the effect size within a family ($\eta_w^2$) was calculated for a given effect based on the sum of squares associated with either the between-replications or within-replications part of the design. Again, results were considered worthy of interpretation when $\eta_w^2 > .05$ and $p \leq .05/4$.

### 7.2.1 MMAP / EAP Parameters

Preliminary testing implementing a normal prior distribution to estimate MGGUM item discriminations ($\hat{\alpha}_{id}$) allowed the resulting estimates to have positive or negative values. This was most evident when estimating parameters in the simple structure condition. Following the guidelines in previous research using normal prior distributions, negative item discrimination estimates ($\hat{\alpha}_{id}$) were treated as missing data and ignored in subsequent analyses of RMSDs (Béguin & Glas, 2001; Finkelman, Hooker, & Wang, 2010).

In addition, use of a normal prior distribution for item discriminations required the implementation of a new check on the feasibility of estimating a given item location. Specifically, if an estimated discrimination parameter had a value too close to zero during a maximization loop (i.e. within the interval [-0.075, +0.075]), then the associated item location estimate ($\hat{\delta}_{id}$) for that dimension was set to zero, regardless of the item's true latent structure. The logic behind this constraint has foundations in Sympson's (1978) work wherein there are an infinite number of possible item locations ($\delta_i$) for a given item if the corresponding item discrimination ($\alpha_i$) is truly zero. When item discrimination values ($\alpha_{id}$) are near zero, the likelihood function is essentially flat across the associated dimension. Therefore, attempting to locate the maximum of such a function is extremely difficult. Dimensional item location estimates ($\hat{\delta}_{id}$) were thus constrained to have a value of zero when the associated dimension's item discrimination estimates ($\hat{\alpha}_{id}$) were too close to zero. Should item discrimination estimates ($\hat{\alpha}_{id}$) move outside the interval cutoff prior to meeting the iterative convergence criterion, then item locations ($\hat{\delta}_{id}$) were again freely estimated.

This artificial constraint on item location estimates ($\hat{\delta}_{id}$) due to small item discrimination estimates ($\hat{\alpha}_{id}$) was most prominent in the simple structure condition. In this condition, 85.47% of all item discrimination estimates ($\hat{\alpha}_{id}$) were outside the constraint interval, with 14.53% within it. For the measured dimension, where the true discrimination values ($\alpha_{id}$) were non-zero, 99.77% of the estimates were outside the constraint interval. Only 0.23% of the item discrimination estimates ($\hat{\alpha}_{id}$) for the

measured dimension were within the constraint interval, thus forcing artificial item locations ($\hat{\delta}_{id}$) of zero. For the non-measured dimension, where true discrimination values ($\alpha_{id}$) were zero, 71.16% of the item discrimination estimates ($\hat{\alpha}_{id}$) were outside the constraint interval, while 28.84% were within it. There was only one instance in the complex structure condition (one dimension of one item) where a small item discrimination estimate ($\hat{\alpha}_{id}$) led to a constraint on an item location estimate ($\hat{\delta}_{id}$).

Under the assumption that researchers may not always have a priori knowledge of the true structure of an item, all non-negative item discrimination estimates ($\hat{\alpha}_{id}$) outside the cutoff interval are included in the forthcoming analyses regardless of whether they are associated with true measured or non-measured dimensions. In the complex structure condition, all estimates of item locations ($\hat{\delta}_{id}$) are examined, except for the single dimension item location estimate ($\hat{\delta}_{id}$) that was constrained to be zero due the small corresponding item discrimination estimate ($\hat{\alpha}_{id}$). However, in the simple structure condition, item location estimates ($\hat{\delta}_{id}$) are examined only for the measured dimensions when corresponding true discrimination values ($\alpha_{id}$) are non-zero and the associated item discrimination estimates ($\hat{\alpha}_{id}$) are outside the constraint interval.

With respect to subjective response category thresholds ($\tau_{ik}$), these MGGUM parameters do not vary by dimension. However, the challenging or sometimes impossible task of identifying an item's location ($\delta_{id}$) when the discrimination value ($\alpha_{id}$) is very small also leads to problems with properly locating the corresponding subjective response category thresholds ($\tau_{ik}$). As such, only information from the measured

dimensions is helpful in locating these estimates. Thus, in the present study all subjective response category threshold estimates ($\hat{\tau}_{ik}$) are examined in the complex structure condition. Yet, in the simple structure condition subjective response category threshold estimates ($\hat{\tau}_{ik}$) are only included in the recovery analyses when the item location estimates ($\hat{\delta}_{id}$) for the measured dimension were not subject to the artificial constraint. However, as a reminder, the locations were only constrained 0.23% of the time for the measured dimensions, so almost all of the subjective response category threshold estimates ($\hat{\tau}_{ik}$) were analyzed using this strategy.

Person parameter estimates ($\hat{\theta}_{jd}$) were assumed to be substantially unaffected by the constraint because they are computed based on all estimated item parameters for all items on a simulated questionnaire. As no replication resulted in every item being subject to the constraint, there were still informative items available to compute person parameter estimates ($\hat{\theta}_{jd}$). Thus, all person parameter estimates ($\hat{\theta}_{jd}$) are included in the recovery analysis regardless of item structure.

7.2.1.1 Convergence

Convergence for MMAP item parameter solutions was operationally defined as parameter values changing less than 0.0005 from one iteration to the next. A maximum of 1000 expectation (outer) cycles, 30 maximization (inner) cycles, 50 Fisher scoring iterations for subjective response category thresholds, and 50 Fisher scoring iterations for item locations and item discriminations were allowed. All replications converged using these criteria (expectation cycle mean = 117, s.d. = 125), however some replications

converged faster than others (expectation cycle minimum = 20, expectation cycle maximum = 974). Item structure, complex versus simple, impacted convergence with complex structure requiring more cycles (expectation cycle mean = 143, s.d. = 117) than simple structure (expectation cycle mean = 92, s.d. = 133). Table 1 presents the average number of expectation cycles required for convergence across the different conditions of the factorial design. Convergence was fairly consistently achieved with a similar average number of expectation cycles across the different sample size conditions in the respective item structure conditions. However, as the number of response categories increased, the average number of expectation cycles required decreased in both item structure conditions. While a general decrease in the average number of expectation cycles was observed in the complex structure condition as test length increased, the simple structure condition required more expectation cycles, on average, as test length increased.

Related to the average number of expectation cycles needed to achieve convergence is the duration of computer time required. Also presented in Table 1 is the average duration, in minutes, required for parameter estimation across replications. Overall, each solution took, on average, 53 minutes (mean = 53.10, s.d. = 88.42). Across the structure conditions as the number of subjects increased the duration required decreased, but substantially more so with simple structure. As test length increased there was an associated increase in duration, but again this increase was more pronounced for simple structure. Increases in the number of response categories also increased the time required for simple structure, but there was a slight decrease in duration for complex structure. In general, complex structure solutions (mean = 13.83, s.d. = 28.84) took substantially less time to converge relative to simple structure solutions (mean = 92.37,

s.d. = 148.00). The discrepancy between estimation in the complex structure condition taking less time than the simple structure condition, yet requiring more expectation cycles may be related to fewer cycles being required within the maximization loop. Conversely, estimation in the simple structure condition may have required more cycles within the maximization loop. However, this is just one possible explanation, as only the number of expectation cycles were tracked during the estimation process in this study. Given the shorter duration, these results indicate MMAP/EAP parameter estimation in the MGGUM is more efficient with complex structure items compared to parameter estimation with simple structure items.

*Table 1.* Convergence of MMAP/EAP replications by item structure condition.

| *Factorial Condition* | *Average Expectation Cycles* | | *Average Duration* (minutes) | |
|---|---|---|---|---|
| | Complex | Simple | Complex | Simple |
| Sample Size | | | | |
| 500 | 158 | 90 | 33.98 | 154.37 |
| 750 | 140 | 108 | 17.37 | 103.68 |
| 1000 | 144 | 108 | 11.08 | 97.19 |
| 1250 | 141 | 78 | 9.34 | 82.02 |
| 1500 | 132 | 81 | 9.80 | 77.86 |
| 1750 | 140 | 78 | 8.69 | 64.83 |
| 2000 | 144 | 100 | 9.02 | 66.69 |
| Test Length | | | | |
| 10 | 170 | 63 | 6.24 | 19.51 |
| 20 | 123 | 98 | 12.67 | 75.80 |
| 30 | 134 | 143 | 22.60 | 181.82 |
| Response Categories | | | | |
| 2 | 203 | 118 | 17.21 | 54.29 |
| 4 | 122 | 98 | 10.87 | 88.05 |
| 6 | 104 | 72 | 13.36 | 134.79 |

7.2.1.2 Parameter Recovery

Interpretable effects were identified after analyzing the average RMSD using the aforementioned ANOVA models. However, prior to those analyses it was necessary to match the true and estimated dimensions, as well as the proper signs corresponding to a particular end of the latent continuum. Magnitude and direction of correlations were used to identify the proper dimensional assignment and the poles of the latent continuum. Of the total 3,780 replications, there were 1,698 instances of dimensions needing to be switched. Of the total 7,560 possible instances of sign flipping within a dimension, 5,230 dimensions within replications required a sign adjustment.

In addition, it was necessary to rescale the MMAP/EAP estimates to the metric of the true parameters prior to computing the RMSDs because these estimates were overconstrained by the fixed hyperparameters used with the item prior distributions. True person parameter values were regressed on the EAP estimates by dimension within each replication. The resulting intercepts and slopes were used to set the origin and unit of MMAP/EAP estimates to that of the true parameters. Person parameter ($\hat{\theta}_{jd}$) and item location ($\hat{\delta}_{id}$) estimates were multiplied by the slope and shifted along the dimension according to the value of the intercept. Subjective response category threshold estimates ($\hat{\tau}_{ik}$) were multiplied by the slope, while item discrimination estimates ($\hat{\alpha}_{id}$) were multiplied by the inverse of the slope.

The average RMSDs of all parameter estimates across all conditions were equal to 0.241, 0.166, 0.297, and 0.335 for MMAP item ($\hat{\delta}_{id}$, $\hat{\alpha}_{id}$, $\hat{\tau}_{ik}$) and EAP person parameter

($\hat{\theta}_{jd}$) estimates, respectively. Table 2 presents the average RMSDs in the factorial design conditions, while statistical results from the analyses of variance are presented in Table 3.

*Table 2.* Average RMSD of parameter estimates by condition.

| *Factorial Condition* | $\hat{\delta}_{id}$ | $\hat{\alpha}_{id}$ | $\hat{\tau}_{ik}$ | $\hat{\theta}_{jd}$ |
|---|---|---|---|---|
| Sample Size | | | | |
| 500 | 0.330 | 0.232 | 0.388 | 0.344 |
| 750 | 0.281 | 0.205 | 0.341 | 0.341 |
| 1000 | 0.253 | 0.175 | 0.313 | 0.336 |
| 1250 | 0.223 | 0.152 | 0.277 | 0.332 |
| 1500 | 0.216 | 0.144 | 0.271 | 0.331 |
| 1750 | 0.199 | 0.131 | 0.249 | 0.330 |
| 2000 | 0.187 | 0.124 | 0.242 | 0.329 |
| Test Length | | | | |
| 10 | 0.289 | 0.204 | 0.319 | 0.427 |
| 20 | 0.226 | 0.154 | 0.291 | 0.315 |
| 30 | 0.209 | 0.140 | 0.281 | 0.262 |
| Response Categories | | | | |
| 2 | 0.313 | 0.266 | 0.304 | 0.422 |
| 4 | 0.225 | 0.141 | 0.287 | 0.304 |
| 6 | 0.187 | 0.091 | 0.301 | 0.277 |

*Table 2.* Continued.

| *Factorial Condition* | $\hat{\delta}_{id}$ | $\hat{\alpha}_{id}$ | $\hat{\tau}_{ik}$ | $\hat{\theta}_{jd}$ |
|---|---|---|---|---|
| Item Structure | | | | |
| Complex | 0.316 | 0.167 | 0.204 | 0.357 |
| Simple | 0.166 | 0.165 | 0.390 | 0.312 |

*Table 3.* $\eta^2$ values for analysis of variance effects.

| *Factorial Condition* | $\hat{\delta}_{id}$ | $\hat{\alpha}_{id}$ | $\hat{\tau}_{ik}$ | $\hat{\theta}_{jd}$ |
|---|---|---|---|---|
| Sample Size | **9.78%** | **8.85%** | **12.30%** | 0.27% |
| Test Length | **5.23%** | 4.87% | 1.30% | **46.55%** |
| Response Categories | **12.58%** | **35.13%** | 0.28% | **38.82%** |
| Sample Size *x* Test Length | 0.19% | 0.48% | 0.13% | 0.02% |
| Sample Size *x* Response Categories | 0.92% | 2.86% | 0.13% | 0.24% |
| Test Length *x* Response Categories | 0.67% | 2.50% | 0.18% | 0.82% |
| Sample Size *x* Test Length *x* Response Categories | 0.16% | 0.56% | 0.42% | 0.05% |
| Item Structure | **25.14%** | 0.01% | **43.75%** | 4.90% |
| Item Structure *x* Sample Size | 1.49% | 0.92% | 0.08% | 0.14% |
| Item Structure *x* Test Length | 0.24% | 0.75% | 0.05% | 0.18% |
| Item Structure *x* Response Categories | **6.08%** | **6.46%** | **6.81%** | 2.03% |
| Item Structure *x* Sample Size *x* Test Length | 0.21% | 0.28% | 0.21% | 0.03% |
| Item Structure *x* Sample Size *x* Response Categories | 0.58% | 2.03% | 1.22% | 0.22% |
| Item Structure *x* Test Length *x* Response Categories | 1.39% | 1.96% | 0.94% | 0.03% |
| Item Structure *x* Sample Size *x* Test Length *x* Response Categories | 0.28% | 0.55% | 0.37% | 0.04% |

Note: Values in bold were statistically significant effects at the $p < 0.0125$ level with $\eta^2$ larger than 5%, and thus, are deemed interpretable.

A main effect of the number of subjects was interpretable for all item parameter estimates ($\hat{\delta}_{id}$, $\hat{\alpha}_{id}$, $\hat{\tau}_{ik}$). The same effect was also observed in previous GGUM and MGGUM research (Roberts & Thompson, 2011; Roberts et al., 2009a; Roberts & Shim, 2010). As seen in Figures 17, 18, and 19, there is a decrease in average RMSD of parameter estimates for item locations ($F(6,3654) = 169.86$, $p < 0.0001$), item discriminations ($F(6,3654) = 169.64$, $p < 0.0001$), and subjective response category thresholds ($F(6,3654) = 390.50$, $p < 0.0001$) as the number of subjects increase, from which an increase in estimation accuracy can be inferred. As sample size increased from 500 to 2000 simulees, item location ($\hat{\delta}_{id}$) average RMSDs decreased from 0.330 to 0.187, item discrimination ($\hat{\alpha}_{id}$) average RMSDs decreased from 0.232 to 0.124, and subjective response category threshold ($\hat{\tau}_{ik}$) average RMSDs decreased from 0.388 to 0.242. Across these parameters there appears to be less of a decrease in average RMSD values when increasing sample sizes beyond 1250 subjects. Thus, the greatest benefit in accuracy of parameter estimation is likely achieved when increasing sample sizes up to 1250 subjects.

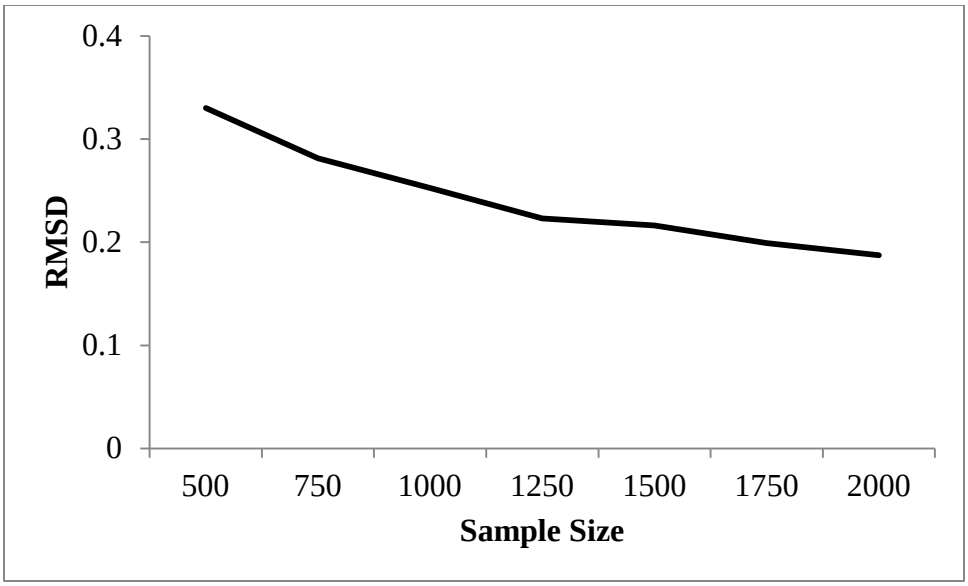



*Figure 17.* Average RMSD of item location estimates across subject conditions.

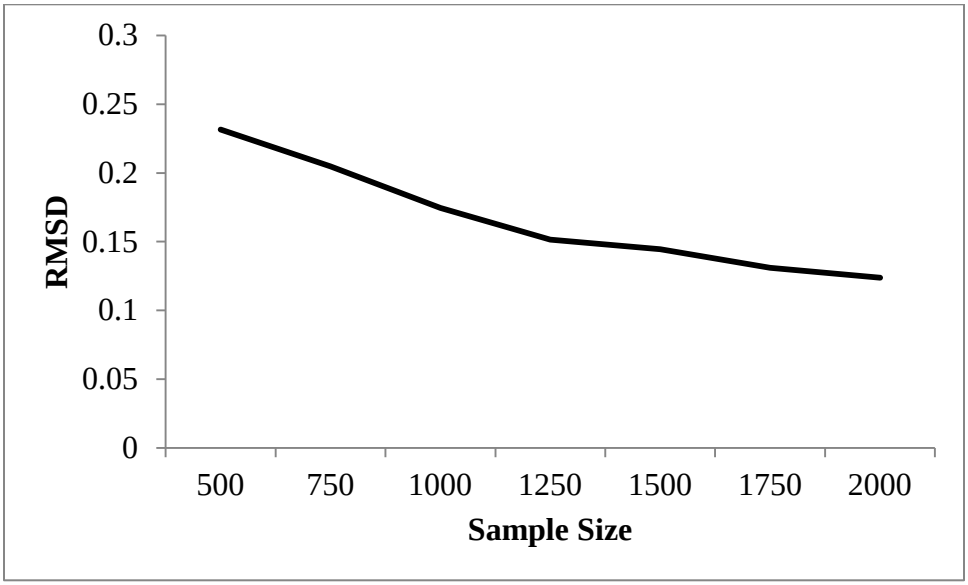



*Figure 18.* Average RMSD of item discrimination estimates across subject conditions.

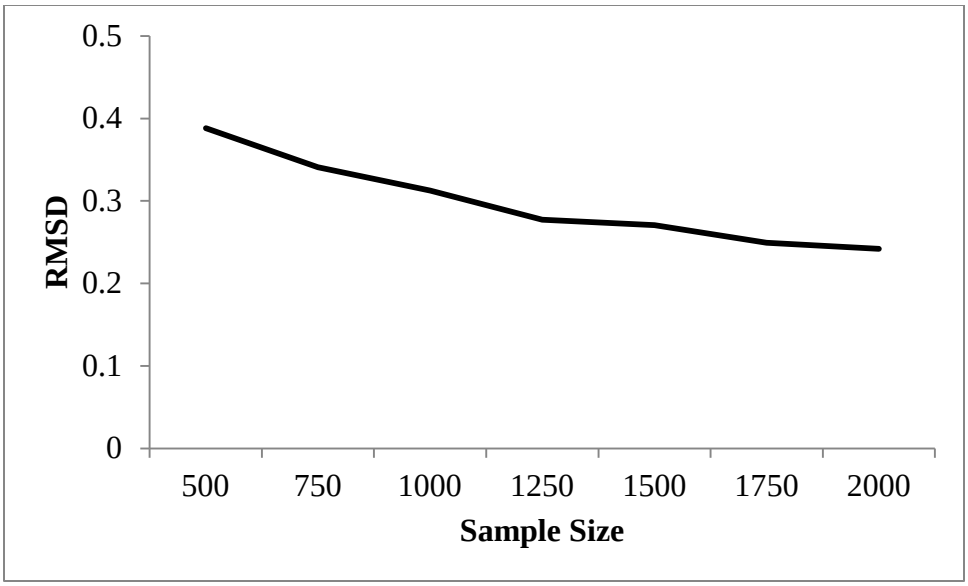



*Figure 19.* Average RMSD of subjective response category threshold estimates across subject conditions.

Contrary to previous GGUM and MGGUM findings, the present research identified an interpretable main effect of test length on estimates of item locations ($F(2,3654) = 272.73$, $p < 0.0001$). Increases in test length led to smaller RMSD, as seen in Figure 20. There is a greater decrease in RMSD, hence a greater increase in accuracy, when moving from 10 to 20 items compared to moving from 20 to 30 items. One explanation for this interpretable effect, albeit a small one, may be related to obtaining better quadrature estimates with longer tests.

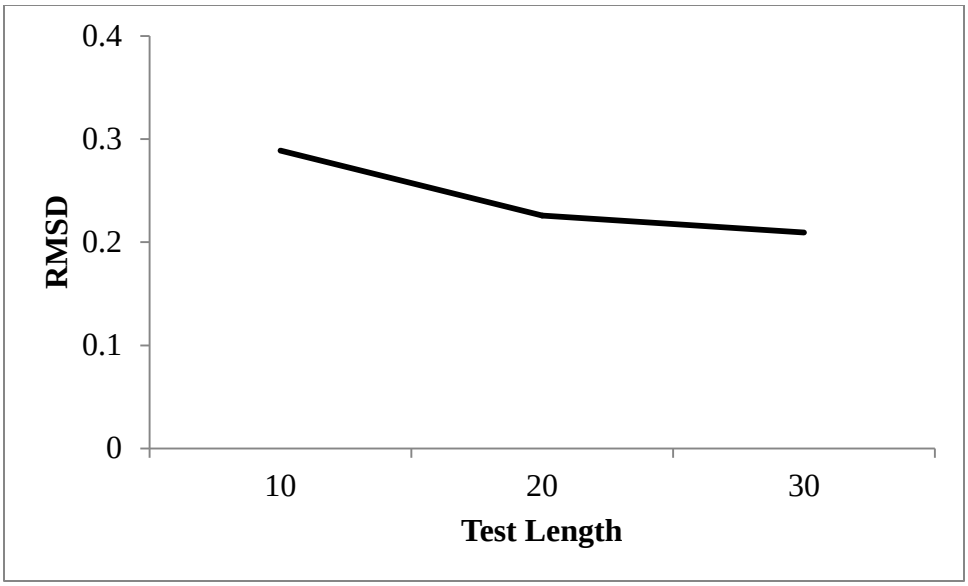



*Figure 20.* Average RMSD of item location estimates across test length conditions.

With respect to the number of response categories and item parameter accuracy, a main effect was observed only for item location ($F(2,3654) = 655.98$, $p < 0.0001$) and item discrimination ($F(2,3654) = 2019.85$, $p < 0.0001$) estimates. In previous MGGUM research (Roberts et al., 2009a; Roberts & Shim, 2010), the number of response categories was held constant. Thus, it is only possible to compare these results to findings from unidimensional work. While GGUM research found main effects for all item parameters (Roberts & Thompson, 2011), the current study failed to identify a main effect of the number of response categories for subjective response category threshold estimates ($\hat{\tau}_{ik}$). Item structure is thought to be responsible for the lack of this main effect. This is discussed further in the forthcoming pages.

As seen in Figures 21 and 22, average RMSD of item location ($\hat{\delta}_{id}$) and item discrimination ($\hat{\alpha}_{id}$) parameters decreased as the number of response categories

increased; leading to enhanced accuracy of parameter estimates. Item location ($\hat{\delta}_{id}$) average RMSDs decreased from 0.313 to 0.187 and item discrimination ($\hat{\alpha}_{id}$) average RMSDs decreased from 0.266 to 0.091. As such, it appears as though estimation accuracy of both of these parameters benefits from increasing the number of response categories.

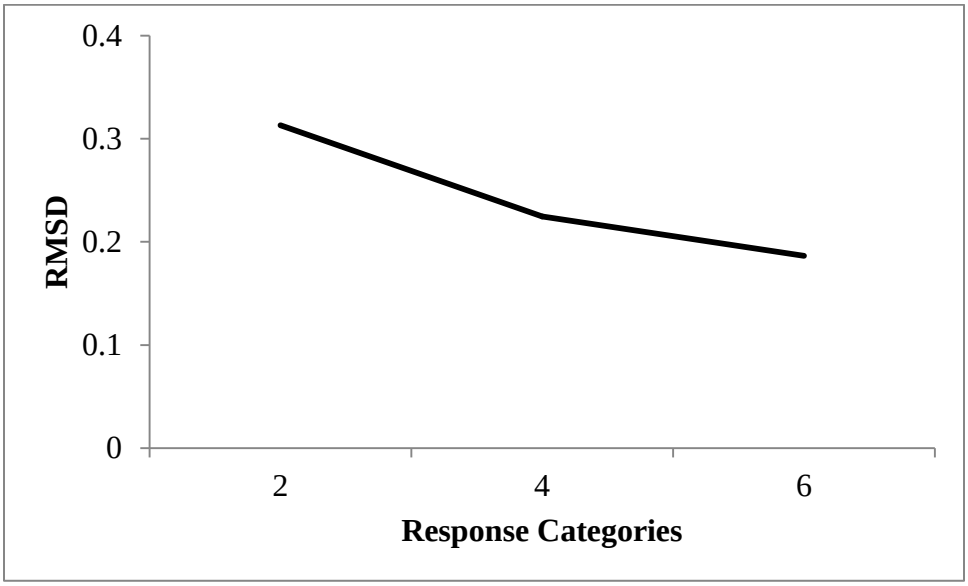



*Figure 21.* Average RMSD of item location estimates across response category conditions.

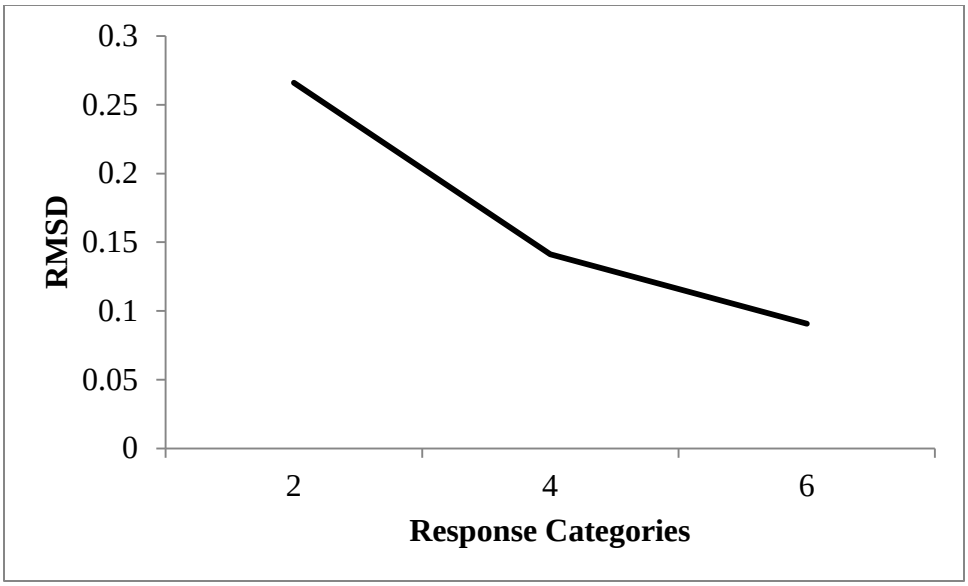



*Figure 22.* Average RMSD of item discrimination estimates across response category conditions.

A main effect of the number of response categories was also interpretable for MGGUM person parameter estimates ($F(2,3654) = 12542.10$, $p < 0.0001$). As seen in Figure 23, the average RMSDs for person parameters decreased from 0.422 to 0.277 as the number of response categories increased, but this was most notable between the two and four response category conditions. This effect mimics previous GGUM results (Roberts & Thompson, 2011), but no comparison can be made to MGGUM research given that response categories were not varied (Roberts et al., 2009a; Roberts & Shim 2010). In addition, a main effect of test length was observed for person parameter estimates ($F(2,3654) = 15041.90$, $p < 0.0001$), as seen in Figure 24. Increases in test length produced more accurate person parameter estimates ($\hat{\theta}_{jd}$), as average RMSDs decreased from 0.427 to 0.262. This main effect replicates findings from the GGUM (Roberts & Thompson, 2011) and MGGUM (Roberts et al., 2009a; Roberts & Shim,

2010) wherein having responses to more items provides more information to better estimate person parameters ($\hat{\theta}_{jd}$).

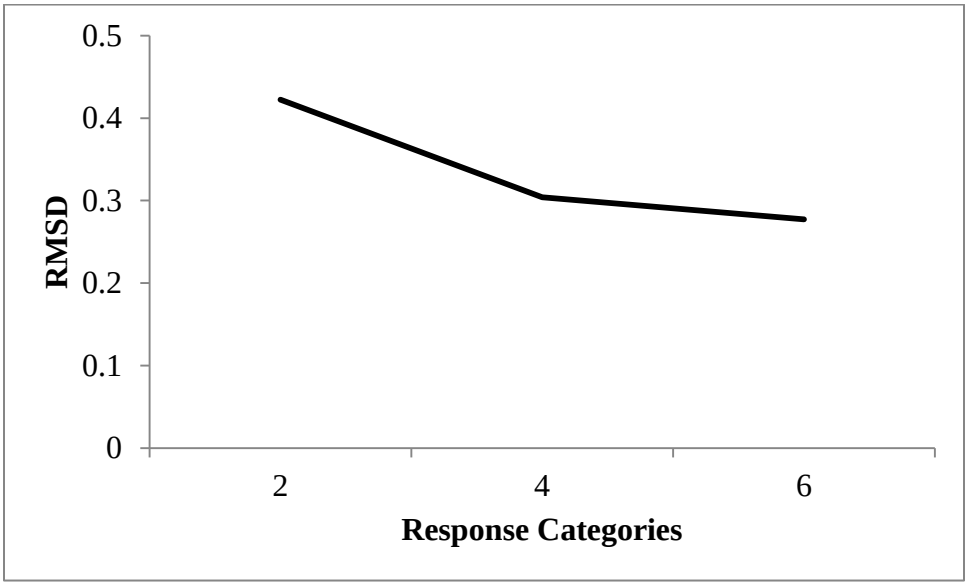



*Figure 23.* Average RMSD of person parameter estimates across response category conditions.

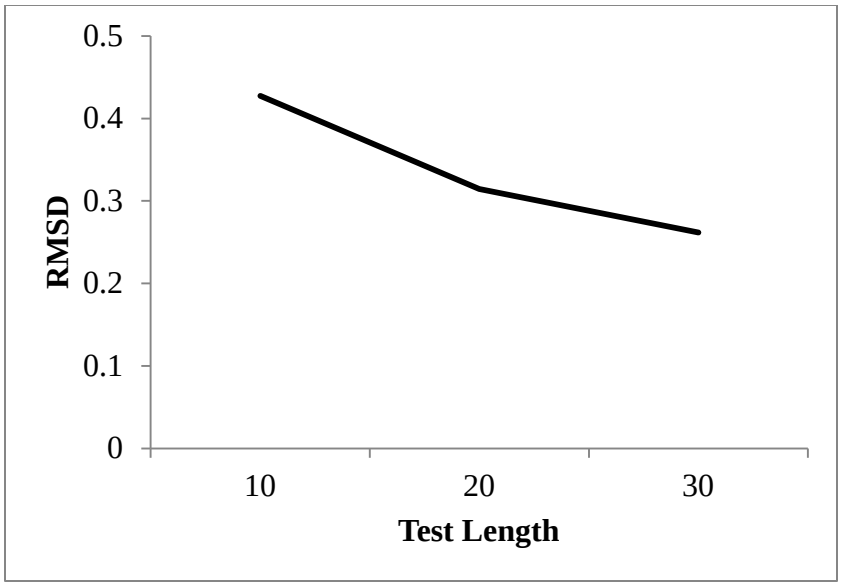



*Figure 24.* Average RMSD of person parameter estimates across test length conditions.

The most substantial decreases in average RMSDs which enhance estimation accuracy of item ($\hat{\delta}_{id}$, $\hat{\alpha}_{id}$) and person ($\hat{\theta}_{jd}$) parameters appear to occur when moving beyond binary response categories. Furthermore, increasing test length from 10 to 20 items appears to provide more improvement in accuracy of person parameter ($\hat{\theta}_{jd}$) and item location ($\hat{\delta}_{id}$) estimates compared to moving from 20 to 30 items. Therefore, given the results presented thus far, MMAP/EAP should provide reasonably accurate MGGUM parameter estimates with a minimum of 1250 subjects, 20 items, and 4 response categories.

Up to this point the results discussed have only examined effects of varying the number of subjects, test length, and the number of response categories. However, also of interest in this study is the impact of item structure on MGGUM parameter estimates. As seen in Table 4, there were no interpretable main or interaction effects involving item

structure for person parameter estimates ($\hat{\theta}_{jd}$). Previous MGGUM research using MCMC/EAP found the opposite to be true; identifying a main effect of item structure and an interaction effect of item structure with test length (Roberts et al., 2009a; Roberts & Shim 2010). The current study used MMAP/EAP, incorporated greater variation in the factorial design, and increased the number of replications within a cell. The use of a different estimation method and/or the limited sample of the earlier work may have contributed to these discrepant results. It should be noted, however, that the average RMSDs for person parameter estimates ($\hat{\theta}_{jd}$) appear more similar to those obtained in previous research for the complex structure condition as opposed to the simple structure condition.

Several interpretable main and interactions effects involving item structure were observed for item parameter estimates ($\hat{\delta}_{id}$, $\hat{\alpha}_{id}$, $\hat{\tau}_{ik}$). As a reminder, these results are based on all positive item discrimination estimates ($\hat{\alpha}_{id}$) and those item location estimates ($\hat{\delta}_{id}$) not subject to the artificial constraint for the measured dimension(s), where true item discrimination values ($\alpha_{id}$) were non-zero. However, this led to the inclusion of 71.16% of item discrimination estimates ($\hat{\alpha}_{id}$) on the non-measured dimensions of items in the simple structure condition. As researchers typically have little to no a priori knowledge of appropriate item structure, this is arguably a valid approach. In addition, subjective responses category threshold estimates ($\hat{\tau}_{ik}$) were included for all complex structure items, but only included for simple structure items when an item location estimate ($\hat{\delta}_{id}$) for a measured dimension was not subject to the constraint. Given

that only 0.23% of the measured dimensions' item location estimates ($\hat{\delta}_{id}$) were constrained in the simple structure condition, the representativeness of these results is not likely impacted. However, the majority (72%) of item discrimination estimates ($\hat{\alpha}_{id}$) for non-measured dimensions were larger than the cutoff and thus included in analyses. Thus, accuracy of estimates in the simple structure condition may be impacted by this selectivity.

The present results identified interpretable main effects of item structure for item location ($F(1,3654) = 2620.52$, $p < 0.0001$) and subjective response category threshold ($F(1,3654) = 5020.62$, $p < 0.0001$) estimates. As seen in Figure 25, the average RMSD of item location estimates ($\hat{\delta}_{id}$) was larger in the complex structure condition, 0.316, compared to that obtained in the simple structure condition, 0.166. This suggests enhanced accuracy within simple structure. A potential rationale is that the increased dimensional complexity makes identifying an item's location more challenging. These findings also coincide with previous MGGUM estimation work (Roberts et al., 2009a; Roberts & Shim, 2010). Figure 26, on the other hand, identified the reverse effect for estimates of subjective response category thresholds ($\hat{\tau}_{ik}$). Again, this effect replicated that found in previous MGGUM research (Roberts et al., 2009a; Roberts & Shim 2010). The average RMSD of subjective response category threshold estimates ($\hat{\tau}_{ik}$) in the complex structure condition, 0.204, was significantly smaller than in the simple structure condition, 0.390. Regardless of item structure, subjective response category thresholds ($\tau_{ik}$) remain constant across dimensions. Therefore, it is conceivable the additional information provided by the second measured dimension enhances parameter estimation

accuracy of complex structure items compared to simple structure items, where information from only one measured dimension is considered.

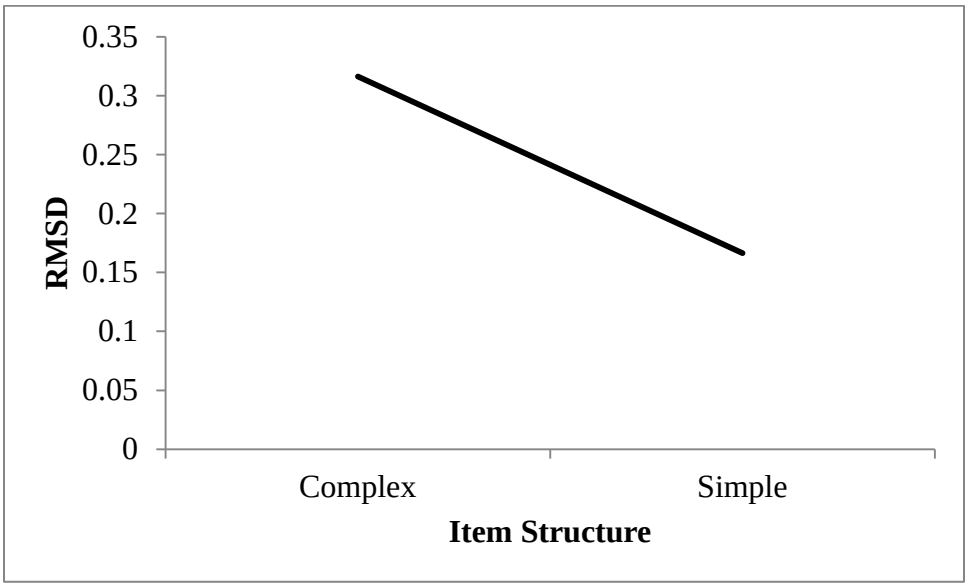



*Figure 25.* Average RMSD of item location estimates by item structure condition.

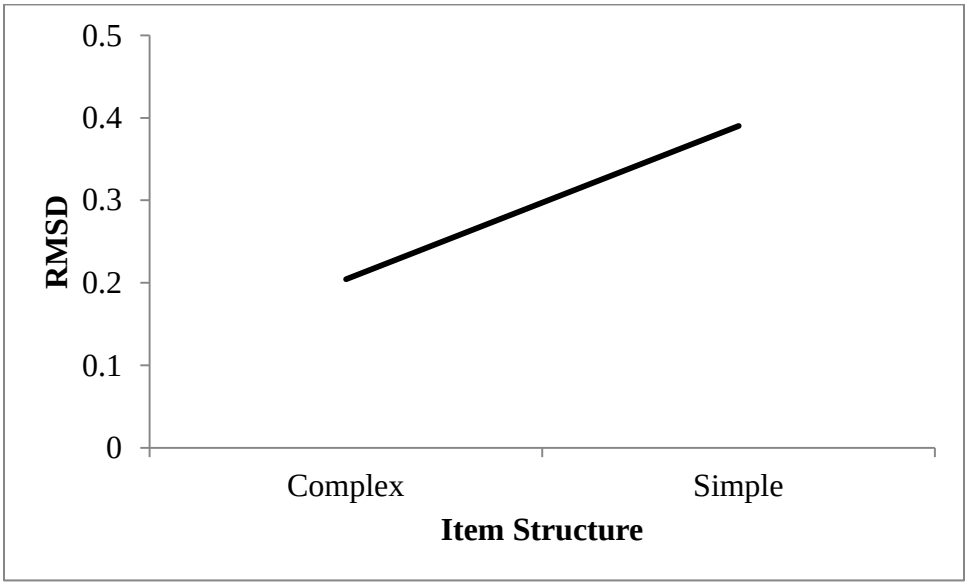



*Figure 26.* Average RMSD of subjective response category threshold estimates by item structure condition.

The lack of a main effect of item structure for item discrimination estimates ($\hat{\alpha}_{id}$), as identified in earlier MGGUM work (Roberts et al., 2009a; Roberts & Shim 2010), may be due to differences in the factorial designs of the studies. The present study varied the number of response categories which led to interpretable two-way interaction effects of item structure and the number of response categories for all item parameter estimate (item location: $F(2,3654) = 317.18$, $p < 0.0001$; item discrimination: $F(2,3654) = 37145$, $p < 0.0001$; subjective response category threshold: $F(2,3654) = 390.50$, $p < 0.0001$). These interpretable interaction effects varied by parameter type, as seen in Figures 27, 28, and 29.

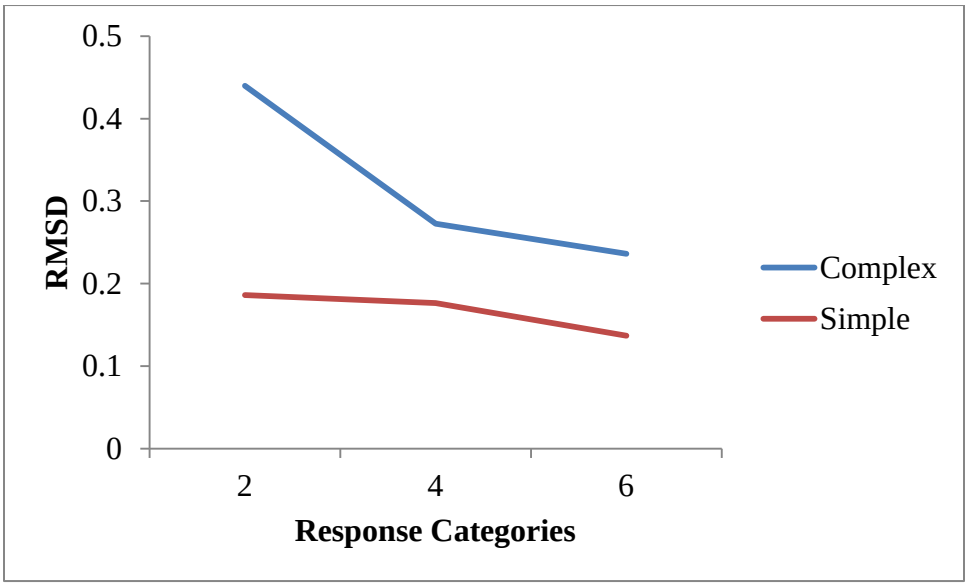



*Figure 27.* Average RMSD of item location estimates for item structure by response category condition.

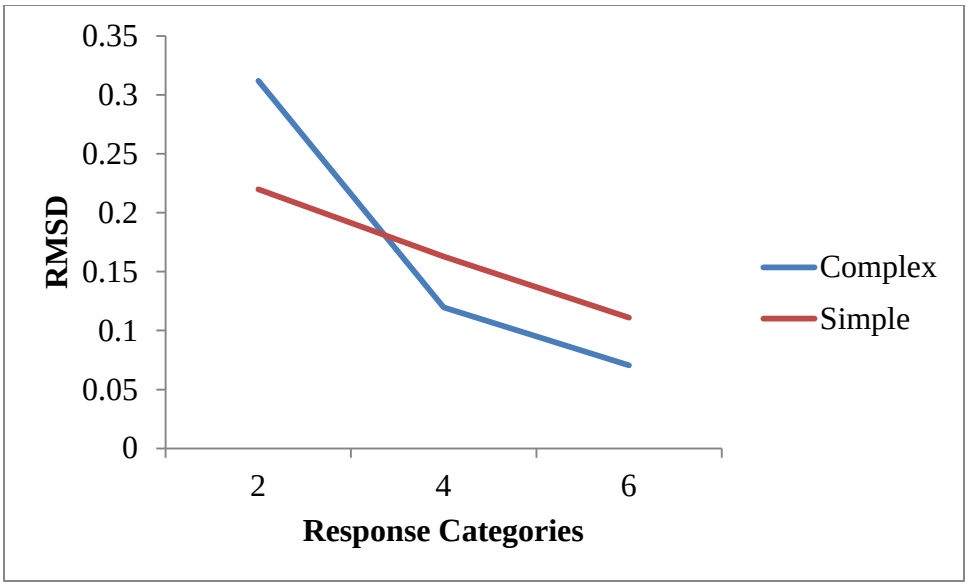



*Figure 28.* Average RMSD of item discrimination estimates for item structure by response category condition.

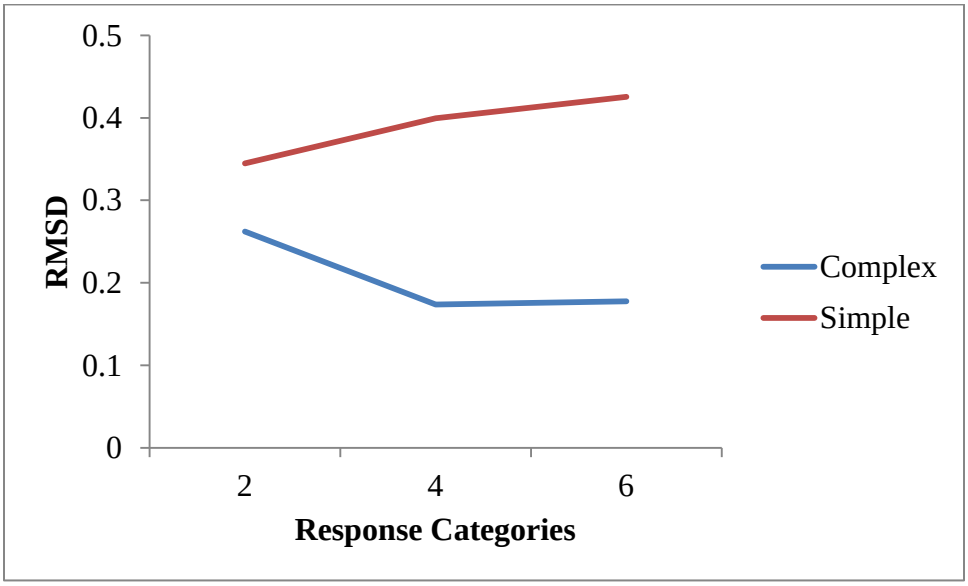



*Figure 29.* Average RMSD of subjective response category threshold estimates for item structure by response category condition.

The interaction effect present in Figure 27 identifies differential changes in accuracy of item location parameter estimates ($\hat{\delta}_{id}$) according to the structure of an item while varying the number of response categories. There are small increases in accuracy of these estimates in the simple structure condition when the number of response categories increases, as evident by the decreasing RMSDs. Accuracy in the complex structure condition increases substantially more when moving from two to four response categories compared to when moving from four to six response categories. It would seem that the relative lack of information provided by binary response categories for a multidimensional item make locating the item much more difficult than when there are more response categories to choose from.

However, a slightly different effect was observed in Figure 29 for subjective response category threshold estimates ($\hat{\tau}_{ik}$). While there is still an overall decrease in average RMSDs of these estimates in the complex structure condition as the number of response categories increase, the same cannot be said for average RMSDs of subjective response category threshold estimates ($\hat{\tau}_{ik}$) in the simple structure condition. Increases in the number of response categories lead to increases in average RMSD for estimates of subjective response category thresholds ($\hat{\tau}_{ik}$) in the simple structure condition. Though at first, these results may seem unexpected, it should be noted that previous GGUM research observed a similar trend with increasing average RMSDs of subjective response category threshold estimates ($\hat{\tau}_{ik}$) as the number of response categories increased (Roberts & Thompson, 2011). A likely explanation is, again, that these parameters do not vary across dimensions and the additional information provided by the second

measured dimension while increasing the number of response categories may be enhancing accuracy.

As is mentioned above, it is not possible to compare these results directly to previous MGGUM research, as earlier work did not investigate effects while varying the number of response categories. However, performance comparisons in the six response category condition are possible given that it was implemented in previous MCMC simulation studies (Roberts et al., 2009a; Roberts & Shim 2010). When considering the accuracy of item location ($\hat{\delta}_{id}$) and subjective response category threshold ($\hat{\tau}_{ik}$) estimates in this response category condition, it should be noted that the present results are akin to previous MCMC findings for the MGGUM (Roberts et al., 2009a; Roberts & Shim 2010).

The relationship observed in Figure 28 for item discrimination estimates ($\hat{\alpha}_{id}$) also indicates estimation accuracy varies depending on both item structure and the number of response categories. In both the complex and simple structure item conditions the average RMSDs of item discrimination estimates ($\hat{\alpha}_{id}$) decrease as the number of response categories increase. There is a consistent decrease in average RMSDs for estimates in the simple structure condition as the number of response categories increase. However, moving from two to four response categories provides a greater decrease in average RMSD for the complex structure condition than does moving from four to six response categories. Overall, item discrimination estimates ($\hat{\alpha}_{id}$) in the simple structure condition are more accurate than in the complex structure condition when there are only two response categories. When there are more response categories, the reverse is true

with item discrimination estimates ($\hat{\alpha}_{id}$) being more accurate in the complex structure condition. This likely explains the lack of the previously identified main effect of item structure on accuracy of item discrimination estimates ($\hat{\alpha}_{id}$) given that only the six response category condition was examined in that earlier work (Roberts et al., 2009a; Roberts & Shim 2010). The interaction effect found here is such that the simple main effect of item structure reverses as the number of response categories increases. This leads to different conclusions about the effect of item structure on parameter estimation accuracy depending on the number of response categories under consideration. If only the six category condition is considered, then the current results are consistent with those found previously.

These results suggest that if using complex structure items, a minimum of four response categories should be sufficient for accurate parameter estimates. If using simple structure items, there appears to be a tradeoff in estimation accuracy as the number of response categories vary. As the number of response categories increase, there is greater accuracy of item location and discrimination estimates. However, there is an accompanying decrease in accuracy of subjective response category thresholds.

7.2.1.3 Standard Errors

Average standard errors (ASEs) of the parameter estimates suggest greater variation in estimation of person parameter estimates ($\hat{\theta}_{jd}$ s.e. = 0.403) compared to item parameters estimates ($\hat{\delta}_{id}$ s.e. = 0.137, $\hat{\alpha}_{id}$ s.e. = 0.067, $\hat{\tau}_{ik}$ s.e. = 0.190). The ASEs by condition in the factorial design are presented in Table 4.

*Table 4.* Average standard error of parameter estimates by condition.

| *Factorial Condition* | $\hat{\delta}_{id}$ | $\hat{\alpha}_{id}$ | $\hat{\tau}_{ik}$ | $\hat{\theta}_{jd}$ |
|---|---|---|---|---|
| Sample Size | | | | |
| 500 | 0.212 | 0.100 | 0.265 | 0.399 |
| 750 | 0.163 | 0.081 | 0.226 | 0.400 |
| 1000 | 0.135 | 0.069 | 0.196 | 0.399 |
| 1250 | 0.127 | 0.061 | 0.178 | 0.403 |
| 1500 | 0.113 | 0.056 | 0.167 | 0.403 |
| 1750 | 0.109 | 0.052 | 0.157 | 0.405 |
| 2000 | 0.100 | 0.049 | 0.145 | 0.406 |
| Test Length | | | | |
| 10 | 0.163 | 0.078 | 0.202 | 0.513 |
| 20 | 0.136 | 0.067 | 0.189 | 0.380 |
| 30 | 0.128 | 0.062 | 0.186 | 0.316 |
| Response Categories | | | | |
| 2 | 0.177 | 0.088 | 0.113 | 0.496 |
| 4 | 0.137 | 0.070 | 0.169 | 0.373 |
| 6 | 0.099 | 0.043 | 0.281 | 0.340 |

*Table 4.* Continued.

| *Factorial Condition* | $\hat{\delta}_{id}$ | $\hat{\alpha}_{id}$ | $\hat{\tau}_{ik}$ | $\hat{\theta}_{jd}$ |
|---|---|---|---|---|
| Item Structure | | | | |
| Complex | 0.113 | 0.080 | 0.151 | 0.421 |
| Simple | 0.161 | 0.056 | 0.230 | 0.386 |

ASEs of item parameter estimates ($\hat{\delta}_{id}$, $\hat{\alpha}_{id}$, $\hat{\tau}_{ik}$) decreased as sample size increased, but there was little change in the ASEs of person parameter estimates ($\hat{\theta}_{jd}$). Increasing test length led to smaller ASEs for all parameter estimates, but this was most noticeable for person parameter estimates ($\hat{\theta}_{jd}$). Increasing the number of response categories also led to decreases in ASEs for person parameter ($\hat{\theta}_{jd}$) estimates as well as for item location ($\hat{\delta}_{id}$) and item discrimination ($\hat{\alpha}_{id}$) estimates. The ASEs for subjective response category threshold estimates ($\hat{\tau}_{ik}$), on the other hand, increased as the number of response categories increased. One explanation for these results is that with more response categories, there are more thresholds to estimate and hence, greater variability of the estimates may occur (Shaftel, Nash, & Gillmor, 2012). Finally, the differences between ASEs of parameter estimates in the complex structure condition compared to those in the simple structure condition were negligible for estimated item location ($\hat{\delta}_{id}$), item discrimination ($\hat{\alpha}_{id}$), and person ($\hat{\theta}_{jd}$) parameters. The differences in the ASEs of estimated subjective response category thresholds ($\hat{\tau}_{ik}$) were slightly more pronounced

with ASEs in the complex structure condition being lower than those in the simple structure condition. It is likely the additional information available in the second measured dimension for the complex structure condition accounted for this result.

### 7.2.2 MMAP/EAP and MCMC/EAP Comparison

In order to assess differences in MGGUM estimation accuracy and efficiency between the MMAP/EAP procedure and the previously applied MCMC/EAP procedure (Roberts et al., 2009a; Roberts & Shim, 2010), a reduced-factorial design was analyzed varying only the number of response categories (2 or 6 categories). The number of subjects (2000), items (20), and item structure (complex) were held constant.

#### 7.2.2.1 Convergence

The MMAP/EAP estimates for this comparative analysis were taken from the first five replications of the relevant cells in the full-factorial simulation design. Item responses from the same replications were also analyzed with the MCMC/EAP technique using the OpenBUGS freeware program (Lunn, Spiegelhalter, Thomas, & Best, 2009) with estimates obtained after 19,000 burn-in iterations. Trace and quantile plots for item parameters were examined after 9,000 and 19,000 burn-in iterations to determine if a chain had converged. After 9,000 burn-in iterations, some posterior distributions on a few replications had not yet stabilized. Those distributions appeared stationary after 19,000 burn-in iterations, thus MCMC/EAP estimates were derived from the 1,000 samples following the burn-in phase.

As anticipated, there was a very noticeable difference in duration of the estimation processes. MMAP/EAP was a much faster procedure and took on average

approximately 7 minutes to converge (mean = 6.60, s.d. = 1.26) compared to MCMC/EAP's average duration of 5,029 minutes (mean = 5,029.73, s.d. = 2,781.13), which is almost 84 hours or close to 3 ½ days. The duration of MMAP/EAP was fairly consistent in both the two and six response category conditions. However, as seen in Table 5, there was a vast difference in the duration of MCMC/EAP depending on the number of response categories. Utilizing only two categories took substantially less time. If fewer MCMC/EAP burn-in iterations had been used, there would likely still be a marked difference in duration compared to MMAP/EAP. Regardless, these results indicate MMAP/EAP parameter estimation is an exceptionally more efficient method of MGGUM parameter estimation than MCMC/EAP.

*Table 5.* Duration of MMAP/EAP and MCMC/EAP replications.

| *Replication* | *Average Duration* (minutes) | |
|---|---|---|
| | 2 Categories | 6 Categories |
| MMAP / EAP | | |
| 1 | 7.31 | 7.28 |
| 2 | 5.12 | 6.64 |
| 3 | 8.11 | 5.85 |
| 4 | 8.86 | 5.30 |
| 5 | 5.83 | 5.70 |
| MCMC / EAP | | |
| 1 | 2,385.63 | 7,730.17 |
| 2 | 2,402.73 | 7,738.57 |
| 3 | 2,401.48 | 7,801.00 |
| 4 | 2,388.82 | 7,542.32 |
| 5 | 2,383.98 | 7,522.60 |

7.2.2.2 Parameter Recovery

Having established the superior efficiency of MMAP/EAP to MCMC/EAP, potential differential accuracy was examined by analyzing average RMSDs of the

resulting parameter estimates. Prior to analyzing the results it was necessary to ensure the true and estimated dimensions were properly matched, as were the signs corresponding to a particular end of the dimension. Again, correlations of true ($\theta_{jd}$) and estimated person parameters ($\hat{\theta}_{jd}$) by dimension indicated any necessary adjustments based on magnitude and direction. In addition, before formal comparative analyses could begin the scales of the MMAP/EAP and MCMC/EAP estimates were equated. This was accomplished just as in the full MMAP/EAP factorial design by regressing true person parameter values on the associated estimates separately for each dimension in a given replication. Again, person parameter ($\hat{\theta}_{jd}$) and item location ($\hat{\delta}_{id}$) estimates were multiplied by the slope and shifted according to the value of the intercept from the corresponding regression, while subjective response category thresholds ($\hat{\tau}_{ik}$) were multiplied by the slope and item discrimination estimates ($\hat{\alpha}_{id}$) were multiplied by the inverse of the slope. Rescaling each set of estimates ensured the same metrics as the true parameters were utilized. Average RMSDs of all rescaled parameter estimates are presented in Table 6.

*Table 6.* Average RMSD of MMAP/EAP and MCMC/EAP parameter estimates by condition.

| *Factorial Condition* | $\hat{\delta}_{id}$ | $\hat{\alpha}_{id}$ | $\hat{\tau}_{ik}$ | $\hat{\theta}_{jd}$ |
|---|---|---|---|---|
| MMAP / EAP | 0.204 | 0.109 | 0.137 | 0.339 |
| 2 Categories | 0.258 | 0.173 | 0.155 | 0.414 |
| 6 Categories | 0.149 | 0.045 | 0.120 | 0.263 |
| MCMC / EAP | 0.279 | 0.123 | 0.149 | 0.340 |
| 2 Categories | 0.343 | 0.191 | 0.155 | 0.415 |
| 6 Categories | 0.216 | 0.055 | 0.143 | 0.265 |
| Overall | 0.241 | 0.116 | 0.143 | 0.339 |

The results describing the effects of varying the number of response categories and estimation method on the item ($\hat{\delta}_{id}$, $\hat{\alpha}_{id}$, $\hat{\tau}_{ik}$) and person parameter ($\hat{\theta}_{jd}$) estimates can be found in Table 7. Effect sizes were computed within a family using the appropriate errors terms in order to obtain a better sense of variability in the data attributed to a particular effect. The effect size for the between-replications effect, varying the number of response categories, was computed separately from the effect sizes for the within-replication effects, estimation method and the interaction of estimation method and varying the number of response categories. While the effect sizes of several tested conditions were greater than the 5% cutoff, some failed to meet the significance criteria ($p \leq .05/4$) and are not identified as interpretable. The lack of statistical

significance is likely a sign of low power to detect modest effects. In the future, incorporation of additional replications may increase the power to detect such effects, and thus, resolve the issue.

*Table 7*. $\eta_w^2$ values for MMAP/EAP and MCMC/EAP analysis of variance effects.

| *Factorial Condition* | $\hat{\delta}_{id}$ | $\hat{\alpha}_{id}$ | $\hat{\tau}_{ik}$ | $\hat{\theta}_{jd}$ |
|---|---|---|---|---|
| Response Categories | **73.75%** | **92.75%** | 15.54% | **98.05%** |
| Estimation Method | 49.97% | 31.04% | 3.80% | 8.53% |
| Response Categories *x* Estimation Method | 0.73% | 2.33% | 4.06% | 0.42% |

Note: Values in bold were statistically significant effects at the $p < 0.0125$ level with effect sizes larger than 5%, and thus, are deemed interpretable.

As shown in Table 8, the magnitudes of the Type III sums of squares used in the analyses were clearly different for the within- versus between-replication effects in this design. The Type III sum of squares associated with the between-replication component of the design was substantially larger than those in the within-replication component. Consequently, the moderately-sized values of $\eta_w^2$ associated with the main effect of estimation method in the analyses of item location ($\hat{\delta}_{id}$) and item discrimination ($\hat{\alpha}_{id}$) RMSDs represent substantial proportions of an extremely small within-replication share of the total sum of squares. In this light, it is not surprising that these effects were not statistically significant. In short, despite the small number of replications examined, the response category effect was noticeably the most prominent effect.

*Table 8*. Type III sum of squares values for the MMAP/EAP and MCMC/EAP analysis effects.

| *Factorial Condition* | $\hat{\delta}_{id}$ | $\hat{\alpha}_{id}$ | $\hat{\tau}_{ik}$ | $\hat{\theta}_{jd}$ |
|---|---|---|---|---|
| Response Categories | 0.06949 | 0.08675 | 0.00281 | 0.11394 |
| Estimation Method | 0.02847 | 0.00098 | 0.00064 | 9.211 E-6 |
| Response Categories *x* Estimation Method | 0.00042 | 0.00007 | 0.00068 | 4.528 E-7 |

Just as in the full-factorial MMAP/EAP parameter recovery study, a main effect of the number of response categories was interpretable for all estimated parameters with the exception of the subjective response category threshold estimates ( $\hat{\tau}_{ik}$ ). Moving from two to six response categories resulted in decreasing the average RMSDs, and, in turn, increasing accuracy of item location ($F(1,9) = 22.47$, $p < 0.0001$), item discrimination ($F(1,9) = 102.38$, $p < 0.0001$), and person parameter ($F(1,9) = 401.55$, $p < 0.0001$) estimates, as seen in Figure 30.

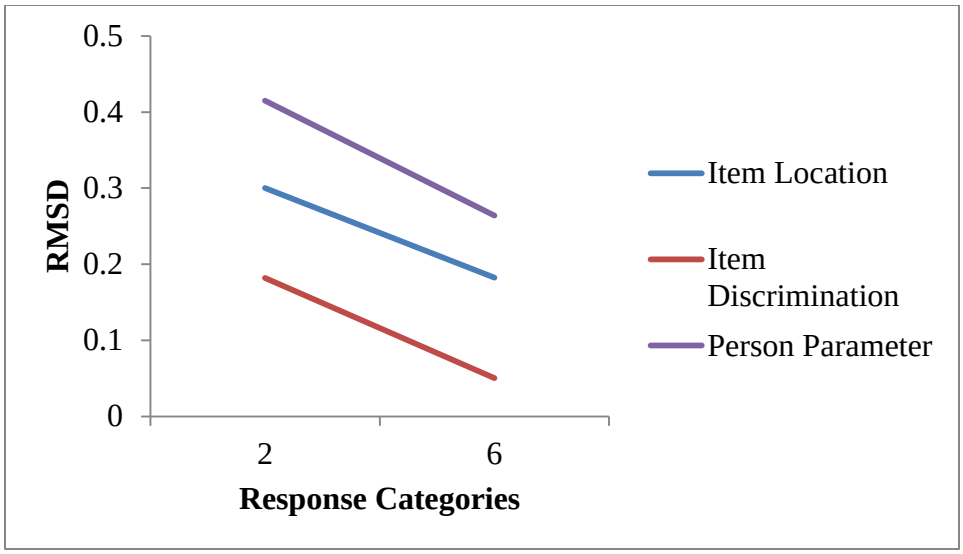



*Figure 30.* Combined MMAP/EAP and MCMC/EAP average RMSD of item and person parameter estimates varying the number of response categories.

Interestingly there were no interpretable main or interaction effects of estimation method for any of the parameters. Previous GGUM research identified an interpretable interaction of estimation method and varying the number of response categories for item location ($\hat{\delta}_{id}$) and subjective response category threshold ($\hat{\tau}_{ik}$) estimates (Roberts & Thompson, 2011). Those results suggested MCMC/EAP estimates were less accurate than MMAP/EAP with fewer response categories, but such differences faded as the number of response categories increased. One marked difference between the previous research and the current study is that the present design only examined accuracy differences of MMAP/EAP and MCMC/EAP parameter estimates in the complex structure condition. Again, perhaps the additional information provided by the second dimension influenced these results. Regardless, from the current study it can be concluded the MMAP/EAP and MCMC/EAP procedures produce very similar MGGUM parameter estimates.

7.2.2.3 Standard Errors

The ASEs of parameter estimates, presented in Table 9, were rescaled to the metric of true parameters using the same method as previously described. However, it should be noted that ASEs are computed differently within each estimation method. The MMAP/EAP ASEs are asymptotic and analytically calculated from the information matrix associated with the estimates. MCMC/EAP ASEs, on the other hand, are calculated as the standard deviations of parameter estimates across draws from the posterior distribution.

*Table 9.* Average standard error of MMAP/EAP and MCMC/EAP parameter estimates by condition.

| *Factorial Condition* | $\hat{\delta}_{id}$ | $\hat{\alpha}_{id}$ | $\hat{\tau}_{ik}$ | $\hat{\theta}_{jd}$ |
|---|---|---|---|---|
| MMAP / EAP | 0.079 | 0.053 | 0.110 | 0.422 |
| 2 Categories | 0.106 | 0.070 | 0.078 | 0.515 |
| 6 Categories | 0.052 | 0.035 | 0.145 | 0.329 |
| MCMC / EAP | 0.192 | 0.121 | 0.101 | 0.424 |
| 2 Categories | 0.236 | 0.183 | 0.102 | 0.517 |
| 6 Categories | 0.148 | 0.058 | 0.101 | 0.330 |
| Overall | 0.136 | 0.087 | 0.106 | 0.423 |

For the most part, varying the number of response categories led to changes in ASEs of parameters within each estimation method. The ASEs for MMAP/EAP and MCMC/EAP person parameter estimates ($\hat{\theta}_{jd}$) were fairly similar across methods and decreased as the number of response categories increased. A similar decrease was observed with respect to the ASEs of item location ($\hat{\delta}_{id}$) and item discrimination ($\hat{\alpha}_{id}$) estimates, again within each estimation method. Conversely, an increase in the ASEs of subjective response category threshold estimates ($\hat{\tau}_{ik}$) was observed, but only with MMAP. The MCMC ASEs of subjective response category threshold estimates ($\hat{\tau}_{ik}$) appear to have remained fairly stable despite changes in the number of response

categories. This observation is likely due to the computational differences in ASEs between the estimation methods.

These results deviate somewhat from findings in the unidimensional GGUM with respect to the ASEs of estimated item parameters ($\hat{\delta}_{id}$, $\hat{\alpha}_{id}$, $\hat{\tau}_{ik}$; Roberts & Thompson, 2011). Previously, GGUM research found similar ASEs of MMAP and MCMC for item location ($\hat{\delta}_{id}$) and subjective response category threshold ($\hat{\tau}_{ik}$) estimates regardless of the number of response categories. However, smaller ASEs of MMAP item discrimination estimates ($\hat{\alpha}_{id}$) were observed in the presence of fewer response categories. The multidimensional nature of the current data may be a contributing factor in these discrepant findings. However, this is only speculation as no MGGUM studies have previously examined the effects of varying the number of response categories (Roberts et al., 2009a; Roberts & Shim, 2010). Future work is recommended to further explore changes in ASEs within and across estimation methods.

# CHAPTER 8

# REAL DATA APPLICATION

Having established the ability to estimate MGGUM parameters with MMAP/EAP, real data from a questionnaire assessing attitudes towards abortion were analyzed. Attitude responses to 19 statements were obtained from 1,562 university students and all responses were obtained using a six-point graded agreement scale. The statements are presented in Appendix F.

Previous studies estimating MGGUM parameters with MCMC/EAP suggested the presence of two dimensions in this data, with the first dimension having a stronger presence than the second (Roberts et al., 2009a; Roberts & Shim, 2010). Thus, two-dimensional MMAP item ($\hat{\delta}_{id}$, $\hat{\alpha}_{id}$, $\hat{\tau}_{ik}$) and EAP person ($\hat{\theta}_{jd}$) parameters were estimated using the computer program implemented in the parameter recovery study. Identical prior distributions as implemented in the parameter recovery study were used for all parameters. The same initial values for item discrimination ($\alpha_{id}$) parameters, 0.25, were used as well. Initial values for item locations ($\delta_{id}$) were obtained from DCA using the CANOCO software (ter Braak & Smilauer, 2002) and rescaled as in the parameter recovery study. Subjective response category threshold ($\tau_{ik}$) initial values were then computed following the technique described in Roberts and Laughlin (1996) using the rescaled initial item locations ($\delta_{id}$).

Utilizing the same convergence criterion as in the recovery study, a solution was reached after 214 expectation (outer) cycles, taking 99.11 minutes. The item parameter estimates ($\hat{\delta}_{id}$, $\hat{\alpha}_{id}$, $\hat{\tau}_{ik}$) are presented in Table 10.

*Table 10*. MGGUM item parameter estimates of abortion attitude data.

| *Item* | $\hat{\delta}_{i1}$ | $\hat{\delta}_{i2}$ | $\hat{\alpha}_{i1}$ | $\hat{\alpha}_{i2}$ | $\hat{\tau}_{i2}$ | $\hat{\tau}_{i3}$ | $\hat{\tau}_{i4}$ | $\hat{\tau}_{i5}$ | $\hat{\tau}_{i6}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2.44 | -3.21 | 1.57 | 0.55 | -1.63 | -1.78 | -1.31 | -1.85 | -1.26 |
| 2 | 2.58 | -1.91 | 1.82 | 0.95 | -2.23 | -2.04 | -1.99 | -1.67 | -1.57 |
| 3 | 2.76 | -1.16 | 2.57 | 1.75 | -2.06 | -1.88 | -1.81 | -1.59 | -1.42 |
| 4 | 2.47 | -1.82 | 3.13 | 1.63 | -2.24 | -1.95 | -1.82 | -1.58 | -1.40 |
| 5 | 2.14 | -0.74 | 1.32 | 1.00 | -1.50 | -1.35 | -1.44 | -1.00 | -0.87 |
| 6 | 1.72 | -0.38 | 1.81 | 1.25 | -1.54 | -1.18 | -1.19 | -1.00 | -0.83 |
| 7 | 1.88 | 0.58 | 0.67 | 0.50 | -1.39 | -1.31 | -1.53 | -1.27 | -1.07 |
| 8 | 1.10 | -0.08 | 1.52 | 0.73 | -1.28 | -0.91 | -0.88 | -0.68 | -0.14 |
| 9 | 0.98 | 0.41 | 1.27 | 0.59 | -1.02 | -0.76 | -0.70 | -0.72 | -0.17 |
| 10 | 0.06 | 0.21 | 2.73 | 1.90 | -0.77 | -0.51 | -0.58 | -0.29 | -0.31 |

*Table 10*. Continued.

| *Item* | $\hat{\delta}_{i1}$ | $\hat{\delta}_{i2}$ | $\hat{\alpha}_{i1}$ | $\hat{\alpha}_{i2}$ | $\hat{\tau}_{i2}$ | $\hat{\tau}_{i3}$ | $\hat{\tau}_{i4}$ | $\hat{\tau}_{i5}$ | $\hat{\tau}_{i6}$ |
|---|---|---|---|---|---|---|---|---|---|
| 11 | 0.01 | 0.18 | 2.58 | 1.93 | -0.83 | -0.54 | -0.66 | -0.30 | -0.31 |
| 12 | -0.67 | -0.15 | 1.23 | 0.83 | -1.16 | -0.79 | -1.32 | -0.88 | -0.72 |
| 13 | -0.82 | -0.17 | 1.21 | 0.70 | -1.05 | -0.81 | -1.32 | -0.88 | -0.85 |
| 14 | -1.09 | -3.67 | 3.11 | 0.81 | -1.89 | -1.63 | -1.56 | -1.31 | -1.04 |
| 15 | -1.41 | -1.44 | 1.93 | 1.06 | -1.86 | -1.51 | -1.61 | -1.47 | -1.29 |
| 16 | -1.87 | -2.29 | 2.11 | 1.03 | -2.05 | -1.81 | -1.71 | -1.59 | -1.32 |
| 17 | -1.91 | -2.09 | 2.31 | 1.01 | -2.16 | -1.85 | -1.95 | -1.52 | -1.39 |
| 18 | -1.86 | -1.83 | 2.47 | 1.13 | -2.10 | -1.77 | -1.87 | -1.54 | -1.30 |
| 19 | -2.11 | -4.08 | 1.69 | 0.54 | -1.75 | -1.99 | -1.73 | -1.83 | -1.30 |

A graphical representation of the estimates can be found in Figure 31. This plot displays the dimension location estimates ($\hat{\delta}_{id}$) for each item with 'spokes' extending from a point indicating the dimensional discrimination estimates ($\hat{\alpha}_{id}$). The estimated person parameters ($\hat{\theta}_{jd}$) are presented in Figure 32, while person parameter ($\hat{\theta}_{jd}$) and item location ($\hat{\delta}_{id}$) estimates are presented jointly in Figure 33. The estimated item locations ($\hat{\delta}_{id}$), as seen in Figures 31 and 33, appear to have greater variability along the first dimension relative to the second. However, estimates of person parameters ($\hat{\theta}_{jd}$) appear to approximate a two-dimensional multivariate normal distribution.

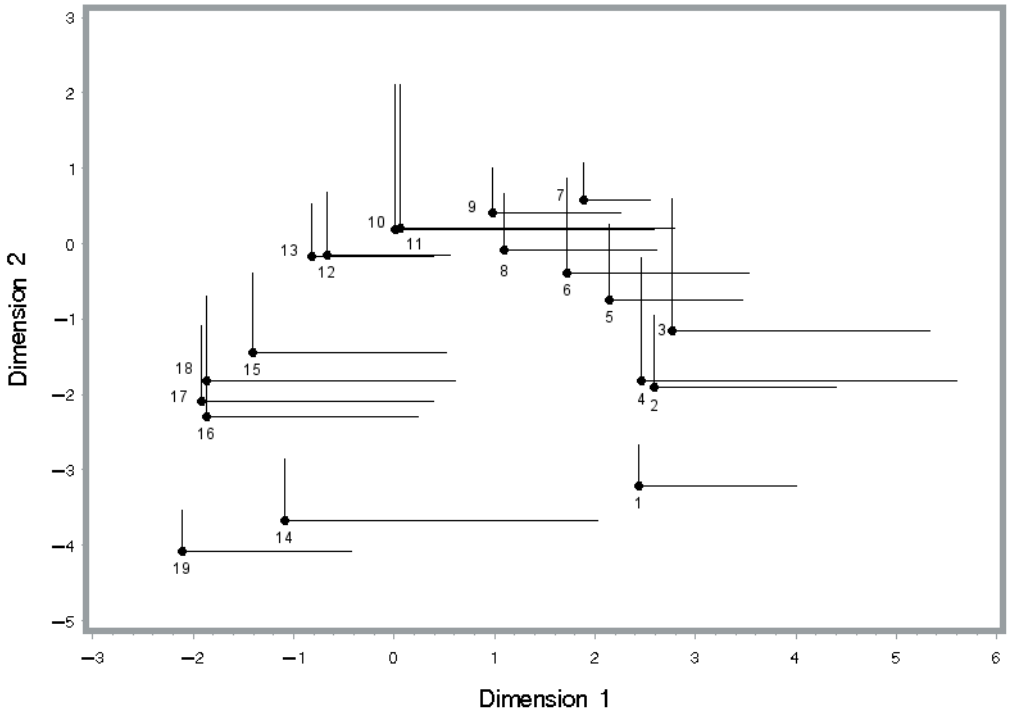



*Figure 31.* Estimated two-dimensional item locations with discrimination 'spokes.'

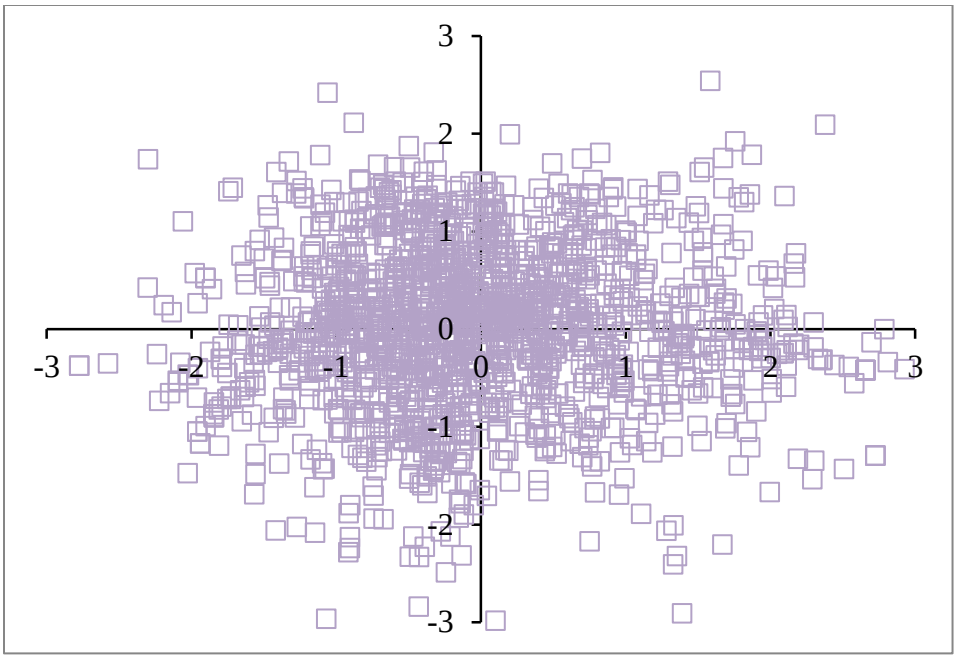



*Figure 32.* Estimated two-dimensional person parameters.

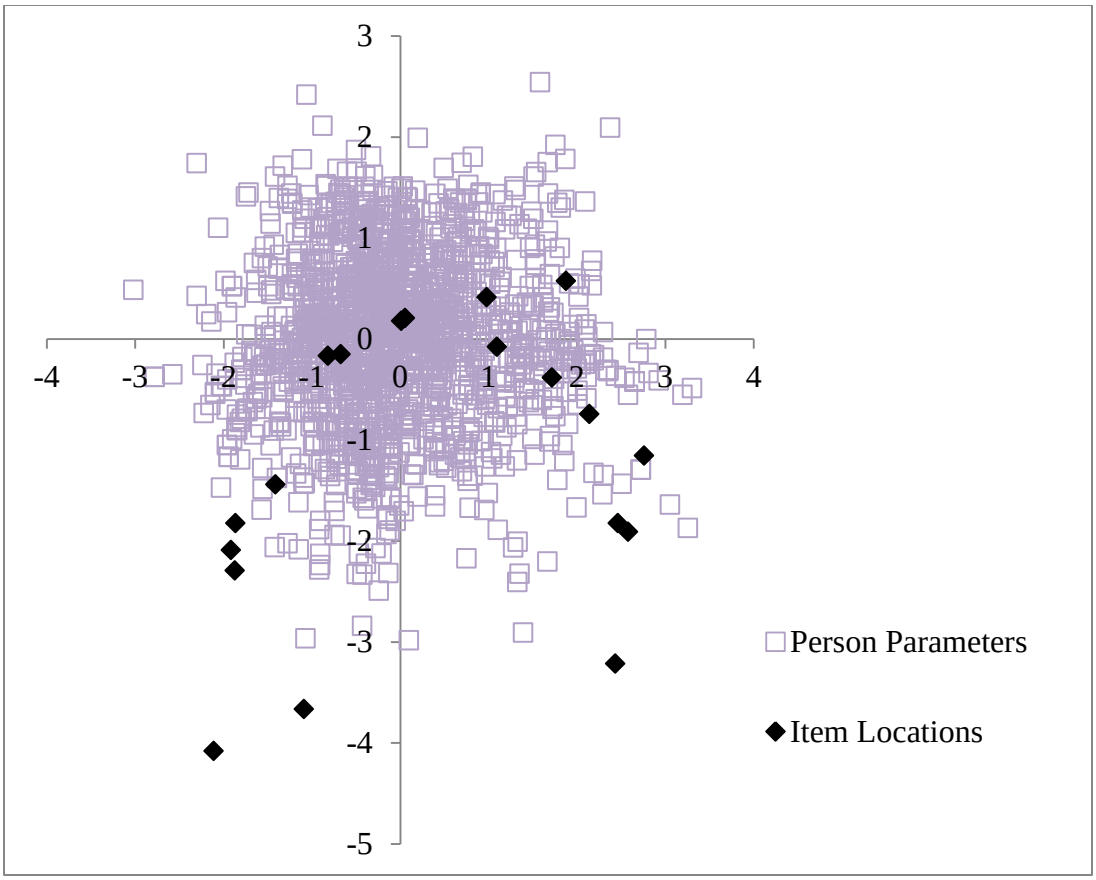



*Figure 33.* Estimated two-dimensional item locations and person parameters.

The normality of the person parameter estimates ($\hat{\theta}_{jd}$) is supported given the fit of the Mahalanobis distance by chi-square scatterplot presented in Figure 34 (Burdenski, 2000). However, the greater variation observed in the Mahalanobis distance by chi-square scatterplot of item location estimates ($\hat{\delta}_{id}$) seen in Figure 35 supports the notion that both dimensions are not equally assessed using these items; an idea that is also clear from the spokes in Figure 31. Item discrimination estimates ($\hat{\alpha}_{id}$) are proportionally larger for the first dimension; however the lack of small item discrimination estimates ($\hat{\alpha}_{id}$) on the

second dimension supports the presence of two dimensions in the data. As such, it appears as though the first dimension differentiates individual to a greater extent.

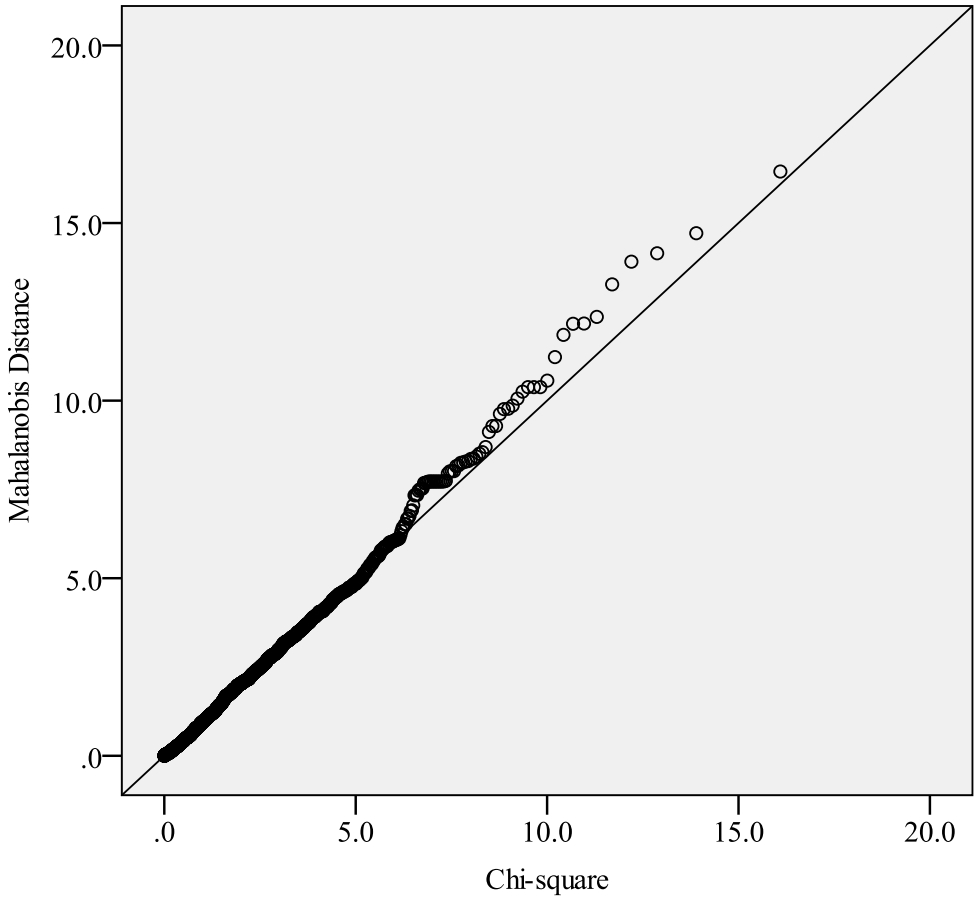



*Figure 34.* Mahalanobis distance by chi-square scatterplot of person parameter estimates.

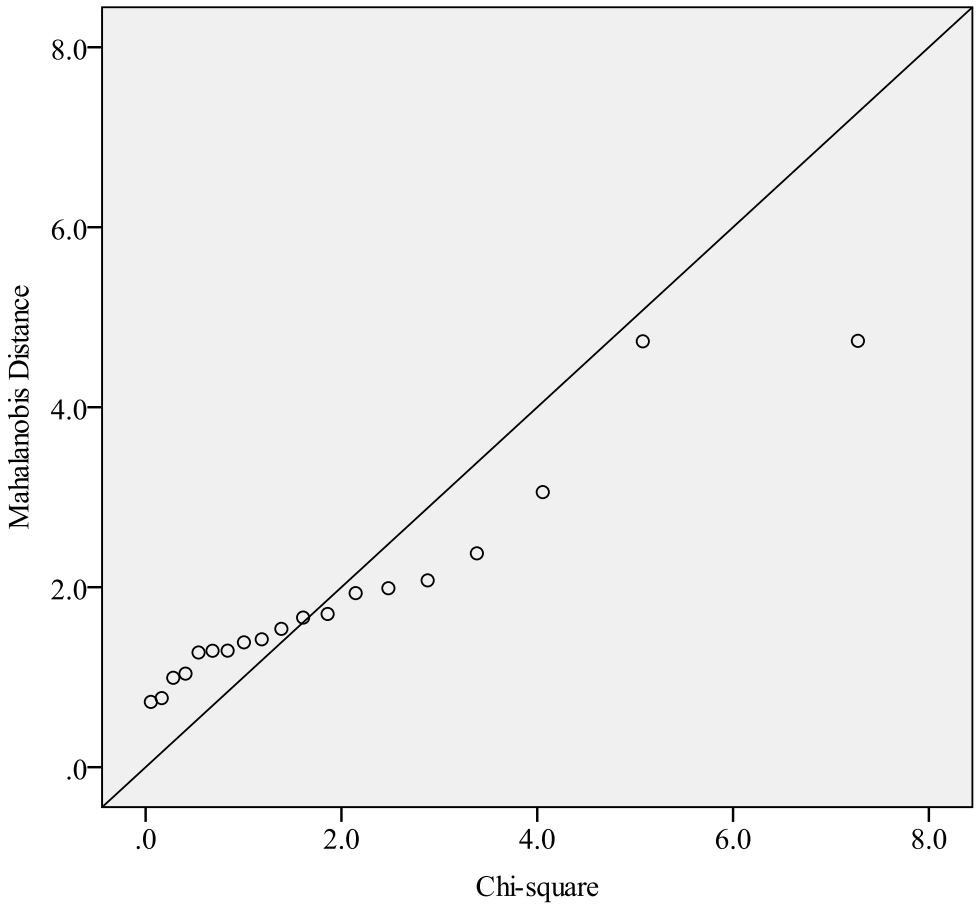



*Figure 35.* Mahalanobis distance by chi-square scatterplot of item location estimates.

Relative fit indices were lower for a multidimensional versus a unidimensional solution (two dimensions: AIC = 56774.10; BIC = 58947.07; CAIC = 59289.07; one dimension: AIC = 58323.34; BIC = 60013.43; CAIC = 60279.43). In keeping with previous research (Roberts et al., 2009a; Roberts & Shim, 2010), the first dimension in the two-dimensional solution accounts for approximately 89% of the variance of GGUM item location estimates ($\hat{\delta}_{id}$), while the second dimension accounts for approximately 7.5%. In addition, approximately 50% of the variation in GGUM item discrimination estimates ($\hat{\alpha}_{id}$) is accounted for by the first dimension, while the second dimension

accounts for approximately 39%. Therefore, the first dimension of the two-dimensional solution corresponds with the single dimension obtained in a unidimensional solution.

Interpreting dimensions is best accomplished by examining the content of items with moderate to high discrimination estimates ($\hat{\alpha}_{id}$) and their associated estimated item locations ($\hat{\delta}_{id}$). Thus, given the estimates presented in Table 10, the first dimension appears to indicate a pro-life to pro-choice continuum, which was well accounted for by the unidimensional GGUM continuum in previous research (Roberts, Donoghue, & Laughlin, 2002; Roberts & Thompson, 2011). An interpretation of the second dimension is less clear and open for discussion. While there are many items with moderate to high discrimination estimates ($\hat{\alpha}_{id}$), item location estimates ($\hat{\delta}_{id}$) do not span the entire range of the continuum. Therefore, considering the content of items with the most discrepant locations along the second dimension, this dimension might involve assessing judicial qualities given that items with high discrimination estimates ($\hat{\alpha}_{id}$) tend to use the words "right(s)" or "legal."

Given the uncertainty of an interpretation for the second dimension, some might question whether it truly exists or if it is merely an artifact of the estimation process. It could be argued that the pattern of estimated item locations ($\hat{\delta}_{id}$) suggests the presence of an arch-like effect, or rather improper fit of a multidimensional model to unidimensional data, despite the detrending of initial values using DCA. In addition, the non-trivial item discrimination estimates ($\hat{\alpha}_{id}$) could be seen by some to support the presence of a larger arch rather than a true second dimension. Arch effects are discussed primarily within the context of correspondence analysis (Greenacre, 2007), however there is little to no

mention of them when using other techniques – particularly never within the context of multidimensional unfolding.

In the simulation portion of this study, parameter estimates were obtained for data generated to fit a two-dimensional model. There were no instances of any arch effects amongst any of those results. Thus, it seems to reason that data which perfectly fits a two-dimensional MGGUM should also not result in an arch effect. On the other hand, it may be that arch effects appear when there is some degree of model misfit to the data. In this case, possibly overfitting unidimensional data may lead to an observable arch effect on higher order dimensions. As such, a unidimensional model may be a more appropriate fit for this data and more interpretable as well. Additional support of overfitting the data comes from the time required to obtain estimates for this data set. Given the average duration to estimate parameters of complex structure items observed in the simulation study, the estimation process for the real data was comparatively long. In the presence of an arch effect the lengthy estimation time could be attributed to a lack of information from a meaningful second dimension. Finally, the uncertain meaning of the second dimension also generates suspicion about its authenticity beyond a mere arch effect. Further applications of this method with real data may provide greater insight into whether arch effects truly exist and what precipitates their presence.

# CHAPTER 9

# DISCUSSION

The primary purpose of this research was to examine the performance of a marginal parameter estimation method, namely MMAP/EAP, within the proximity-based, non-compensatory MGGUM through a simulation study and analysis of real data. Efficiency and accuracy were the focal points of the simulation study, while MMAP/EAP was applied to real data as an indication of the appropriateness of this method in practice. All in all, MMAP/EAP was found to efficiently recover reasonably accurate parameter estimates of a simulated two-dimensional MGGUM. Similar trends to those observed in unidimensional and multidimensional GGUM research with respect to accuracy of estimates when varying the number of subjects, items, and response categories were identified (Roberts & Thompson, 2011; Roberts et al. 2009a, Roberts & Shim, 2010). In addition, MMAP/EAP was able to produce two-dimensional estimates of a real data set measuring attitudes towards abortion.

## 9.1 Suitability of MMAP / EAP with the MGGUM

Despite the increased popularity of implementing MCMC/EAP in research today, using an alternative method such as MMAP/EAP can, at times, benefit researchers. There were no observed differences in estimation accuracy between the methods in the present study, despite implementing different prior distributions for item discrimination parameters ($\alpha_{id}$) and the additional specification of initial values for person parameters ($\theta_{jd}$) with MCMC/EAP. The only observable benefit came with respect to computational efficiency, but this benefit was substantial.

Depending on the complexity of the design (i.e. the number of subjects, test length, number of response categories, and latent structure of the items) MMAP/EAP estimates of two-dimensional MGGUM parameters were obtained in anywhere from approximately 6 to 180 minutes, with an average, across all conditions, of just under an hour. To provide a direct comparison, parameters from a subset of the full simulation design were also estimated with MCMC/EAP in light of previous research requiring days to achieve convergence with this method (Roberts et al., 2009a; Roberts & Shim, 2010). Of particular interest is the substantial discrepancy in computing time that was observed in this study when implementing MMAP/EAP and MCMC/EAP. Estimating the same parameters and using the same subset of data, MMAP/EAP took a little over six and a half minutes to reach a solution, on average, whereas MCMC/EAP took almost three and a half days. Such disparities can hardly be ignored. However, attention should also be drawn to the fact that parameter estimation in this study was limited to just two-dimensions. The numerical integration process required in MMAP is likely to become computationally challenging and inefficient as model dimensionality increases. In such cases, MCMC, although more time-consuming, may be a viable option as it does not require numerical integration. Researchers have also suggested implementing techniques such as adaptive quadrature (Schilling & Bock, 2005), Metropolis-Hastings Robbins-Monro (MHRM; Cai, 2010), or reversible jump MCMC (Green & Hastie, 2009) to estimate parameters of multidimensional models.

Within the context of a two-dimensional MGGUM, the efficiency of MMAP/EAP estimation is impacted by the latent structure of items. It appears as though the identifiability of item parameters may be responsible for discrepancies in efficiency.

Attempting to identify item parameters on non-measured dimensions substantially slows down the estimation process relative to instances where all items measure both dimensions. It remains to be seen how efficient MMAP/EAP estimation would be in situations where there is a mix of complex and simple structure items. This is something worthy of future exploration.

The latent structure of items also affects the accuracy of two-dimensional MGGUM parameter estimates. Items measuring multiple dimensions are more challenging to locate in multidimensional space ($\delta_{id}$). Conversely, when an item only measures one dimension its location is more accurately identified. Estimating the location of multidimensional (i.e. complex structure) items can be likened to an item's location being pulled simultaneously in different directions. If one assumes that information about an item's location is constant, then those items pulled only in one direction with one measured dimension essentially have more information available to guide the estimation of that single coordinate compared to sharing what information resources are available to estimate multiple coordinates.

Despite more accurate estimation of item locations ($\delta_{id}$) for unidimensional (i.e. simple structure) items, subjective response category threshold estimates ( $\hat{\tau}_{ik}$ ) were more accurate with multidimensional (i.e. complex structure) items. The most likely explanation for this finding is that MGGUM subjective response category thresholds ($\tau_{ik}$) are constant across dimensions, which may lead to more stable and accurate estimates when an item measures multiple dimensions. In this case, all of the information contained in an item with regard to these parameters can be focused on a single set of coordinates rather than a different set of coordinates for each dimension.

Moving beyond simulations, in exploratory analyses researchers may have little to no a priori knowledge of item structure, thus there could be uncertainty regarding accuracy of the resulting estimates. It should be noted than an exploratory MGGUM is not an identifiable model when there are one or more items that fail to discriminate at least to some extent on all estimated dimensions. In such instances, it may be beneficial to attempt to simultaneously fit multiple correlated simple structure dimensions using a confirmatory MIRT approach to the model.

Analysis of the MMAP/EAP estimates of the real data in this study suggest a two-dimensional model is more appropriate than a unidimensional one, as indicated by relative fit indices (e.g. AIC, BIC, CAIC). Though given the 'weakness' of the second dimension, an argument could be made that the second dimension is, in fact, not a real dimension, but rather an artifact similar to the arch effect often observed with correspondence analysis. MMAP/EAP estimation of MGGUM parameters in the stimulation study did not identify the presence of any arch effects; however those data were generated to be two-dimensional. It is quite possible that the real data in this study are unidimensional. If that is the case, then the data were overfit by the model. The effects of overfitting a model would be best explored through further simulation research. For instance, attempting to fit a two-dimensional model to simulated unidimensional data, with and without some degree of model misfit may shed light onto whether arch effects are the result of incorporating an unwarranted dimension in the model.

At present, there are no model-specific fit indices to formally confirm the multidimensional structure of data with an MGGUM. Generalizing the $S\text{-}X^2$ indices implemented in the unidimensional GGUM (Roberts, 2008) amongst other fit indices

(e.g., Maydeu-Olivares & Garcia-Forero, 2010; Maydeu-Olivares, Cai, & Hernandez, 2011) are suggested avenues of future research. An even better approach would be to develop a model-free assessment of dimensionality (e.g., DIMTEST: Stout, 1987; DETECT: Zhang & Stout, 1999), although this seems more difficult due to the absence of an observable response scoring function that is suitable for proximity-based responses.

### 9.2 Implementation of DCA

An ancillary goal of the present research was to implement an alternative method of identifying initial values to be used in the estimation processes. In theory, informative initial values may increase the probability of properly identifying the global maximum of the log marginal likelihood instead of local maxima, and may locate it quicker as well (Roberts & Laughlin, 1996). The previously implemented method of using unidimensional GGUM estimates as initial values in the MGGUM estimation process (Roberts et al., 2009a; Roberts & Shim, 2010) imposes a unidimensional structure on data that is knowingly multidimensional, specifically in parameter recovery simulation studies. In such instances, DCA is arguably a more appropriate method given that it attempts to identify a multidimensional structure with multidimensional data. The implementation of DCA in both the parameter recovery study and real data analysis was a simple, straightforward process and is recommended for use over the previous approach.

No direct manipulations designed to ascertain the benefits of DCA over cruder estimates that have been previously used in MGGUM estimation studies were incorporated into this study. Nonetheless, there does appear to be at least one advantage of DCA initial values when comparing results across studies. In former MCMC

estimation studies, the signs of person parameter estimates ($\hat{\theta}_{jd}$) were occasionally misestimated for respondents with the most extreme opinions. In these cases, there was little information about the valance of an attitude because the respondents in question generally disagreed with all of the statements that measured a given dimension. Indeed, Roberts and Shim (2010) attempted to alleviate this problem by estimating the sign for extreme estimates midway through the MCMC chain using a maximum likelihood procedure. Their efforts noticeably reduced the frequency in which sign misestimation occurred, but it did not eliminate it. In contrast, use of DCA estimates for initial values eliminated this problem in the corresponding MCMC replications studied here. Although a more rigorous test of this finding remains for future research, the current results suggest that DCA is a preferable method to develop initial estimates of person parameters ($\hat{\theta}_{jd}$), if needed by a given estimation technique.

## 9.3 Conclusion

Despite the aforementioned limitations, this research has identified notable and novel findings. Particularly, it is the first study to investigate a more computationally efficient parameter estimation algorithm for the recently developed MGGUM. MMAP/EAP performed as expected thereby reinforcing its position as a viable estimation method for the MGGUM and with other MIRT models as well. Its accuracy is comparable to the previously implemented MCMC/EAP, but it is considerably more efficient. In addition, this study is the first to implement DCA to identify multidimensional initial values for a multidimensional noncompensatory IRT model.

This research has the ability to foster the application of this innovative model in more diverse areas involving preference ratings, where more than one latent dimension is generally assumed to operate. Moreover, it makes such applications computationally feasible for practitioners by providing the foundational background and derivation.

.

## APPENDIX A

Item information is derived from Ackerman's (1994) matrix formulation

$$I_i\left(\underline{\theta}\right) = -\mathrm{E}\begin{pmatrix} \dfrac{\partial^2 \ln\left(L_i\right)}{\partial\theta_1^2} & \dfrac{\partial^2 \ln\left(L_i\right)}{\partial\theta_1\,\partial\theta_2} \\ \dfrac{\partial^2 \ln\left(L_i\right)}{\partial\theta_1\,\partial\theta_2} & \dfrac{\partial^2 \ln\left(L_i\right)}{\partial\theta_2^2} \end{pmatrix} \tag{A.1}$$

where, using Equation 45,

$$L_i = \prod_{j=1}^{J} P\left(Z_i = X_{ji} \left| \underline{\theta}_j \right.\right). \tag{A.2}$$

The log of Equation A.2 is

$$\ln\left(L_i\right) = \sum_{j=1}^{J} \ln\left[ P\left(Z_i = z \left| \underline{\theta}_j \right.\right)\right]. \tag{A.3}$$

The first derivative of Equation A.3 with respect to one dimension ($\theta_d$) is

$$\frac{\partial \ln\left[ P\left(Z_i = z \left| \underline{\theta}_j \right.\right)\right]}{\partial\theta_d} = \sum_{i=1}^{I}\left( \frac{\partial P\left(Z_i = z \left| \underline{\theta}_j \right.\right)}{\partial\theta_d} * \frac{1}{P\left(Z_i = z \left| \underline{\theta}_j \right.\right)} \right) \tag{A.4}$$

where the first derivative of the MGGUM probability function (Equation 45) with respect to one dimension ($\theta_d$) is

$$\frac{\partial P\left(Z_i = z \left| \underline{\theta}_j \right.\right)}{\partial\theta_d} = \frac{G\left(A'_\theta + B'_\theta\right) - G'_\theta(A+B)}{G_\theta^2} \tag{A.5}$$

setting

$$A = \exp\left[\left(z\sqrt{\sum_{d=1}^{D}\alpha_{id}^{2}(\theta_{jd}-\delta_{id})^{2}}\right)-\sum_{k=0}^{z}\psi_{ik}\right], \tag{A.6}$$

$$A'_{\theta} = \frac{\partial A}{\partial \theta_{d}} = \frac{z\left(\alpha_{id}\right)^{2} A\left(\theta_{jd}-\delta_{id}\right)}{\sqrt{\sum_{d=1}^{D}\alpha_{id}^{2}(\theta_{jd}-\delta_{id})^{2}}}, \tag{A.7}$$

$$B = \exp\left[\left((M-z)\sqrt{\sum_{d=1}^{D}\alpha_{id}^{2}(\theta_{jd}-\delta_{id})^{2}}\right)-\sum_{k=0}^{z}\psi_{ik}\right], \tag{A.8}$$

$$B'_{\theta} = \frac{\partial B}{\partial \theta_{d}} = \frac{(M-z)\left(\alpha_{id}\right)^{2} B\left(\theta_{jd}-\delta_{id}\right)}{\sqrt{\sum_{d=1}^{D}\alpha_{id}^{2}(\theta_{jd}-\delta_{id})^{2}}}, \tag{A.9}$$

$$G = \sum_{w=0}^{C}\left(\begin{array}{l}\exp\left[\left(w\sqrt{\sum_{d=1}^{D}\alpha_{id}^{2}(\theta_{jd}-\delta_{id})^{2}}\right)-\sum_{k=0}^{w}\psi_{ik}\right] \\ +\exp\left[\left((M-w)\sqrt{\sum_{d=1}^{D}\alpha_{id}^{2}(\theta_{jd}-\delta_{id})^{2}}\right)-\sum_{k=0}^{w}\psi_{ik}\right]\end{array}\right), \tag{A.10}$$

and

$$G'_{\theta} = \frac{\partial G'}{\partial \theta_{d}} = \sum_{w=0}^{C}\left(A'+B'\right). \tag{A.11}$$

Taking the second derivative of Equation A.4 yields

$$\frac{\partial^2 \ln\left[P\left(Z_i = z\middle|\underline{\theta}_j\right)\right]}{\partial\theta_d^2} = \sum_{i=1}^{I}\left\{\begin{array}{l}\left[\dfrac{\partial^2 P\left(Z_i = z\middle|\underline{\theta}_j\right)}{\partial\theta_d^2} * \dfrac{1}{P\left(Z_i = z\middle|\underline{\theta}_j\right)}\right] \\ -\left[\left(\dfrac{\partial P\left(Z_i = z\middle|\underline{\theta}_j\right)}{\partial\theta_d}\right)^2 * \dfrac{1}{\left(P\left(Z_i = z\middle|\underline{\theta}_j\right)\right)^2}\right]\end{array}\right\} \quad \text{(A.12)}$$

where

$$\begin{aligned}\frac{\partial^2 P\left(Z_i = z\middle|\underline{\theta}_j\right)}{\partial\theta_d^2} &= \frac{A''_\theta + B''_\theta}{G} - \frac{G''_\theta\left(A+B\right)}{G^2} \\ &\quad - \frac{2G'_\theta\left(A'_\theta + B'_\theta\right)}{G^2} + \frac{2\left(G'_\theta\right)^2\left(A+B\right)}{G^3}\end{aligned} \quad \text{(A.13)}$$

$$\begin{aligned}A''_\theta &= \frac{\partial^2 A}{\partial\theta_d^2} \\ &= \frac{\left\{z\left(\alpha_{id}\right)^2\left(\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right)\left(A + A'_\theta\left[\theta_{jd}-\delta_{id}\right]\right)\right\}}{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2} \\ &\quad - \frac{\left\{z\left(\alpha_{id}\right)^4 A(\theta_{jd}-\delta_{id})^2\right\}}{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2},\end{aligned} \quad \text{(A.14)}$$

$$\begin{aligned}B''_\theta &= \frac{\partial^2 B}{\partial\theta_d^2} \\ &= \frac{\left\{\left(M-z\right)\left(\alpha_{id}\right)^2\left(\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right)\left(B + B'_\theta\left[\theta_{jd}-\delta_{id}\right]\right)\right\}}{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2} \\ &\quad - \frac{\left\{\left(M-z\right)\left(\alpha_{id}\right)^4 B(\theta_{jd}-\delta_{id})^2\right\}}{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2},\end{aligned} \quad \text{(A.15)}$$

and

$$G'' = \frac{\partial^2 G}{\partial \theta_d} = \sum_{w=0}^{C} \left( A''_{\theta} + B''_{\theta} \right). \tag{A.16}$$

Therefore, item information with respect to one dimension is calculated as

$$\begin{aligned} I_i\left(\theta_d\right) = & -\mathrm{E}\left( \frac{\partial^2 \ln\left(L_i\right)}{\partial \theta_d^2} \right) \\ & -\mathrm{E}\left\{ \begin{array}{l} \left[ \frac{\partial^2 P\left(Z_i = z \middle| \underline{\theta}_j\right)}{\partial \theta_d^2} * \frac{1}{P\left(Z_i = z \middle| \underline{\theta}_j\right)} \right] \\ -\left[ \left( \frac{\partial P\left(Z_i = z \middle| \underline{\theta}_j\right)}{\partial \theta_d} \right)^2 * \frac{1}{\left(P\left(Z_i = z \middle| \underline{\theta}_j\right)\right)^2} \right] \end{array} \right\}. \end{aligned} \tag{A.17}$$

Joint item information is based on the mixed derivative of Equation A.4 with respect to both dimensions

$$\frac{\partial^2 \ln\left[ P\left(Z_i = z \middle| \underline{\theta}_j\right) \right]}{\partial \theta_1 \partial \theta_2} = \sum_{i=1}^{I} \left[ \begin{array}{l} \left( \frac{\partial P\left(Z_i = z \middle| \underline{\theta}_j\right)}{\partial \theta_1 \partial \theta_2} * \frac{1}{P\left(Z_i = z \middle| \underline{\theta}_j\right)} \right) \\ -\left( \frac{\partial P\left(Z_i = z \middle| \underline{\theta}_j\right)}{\partial \theta_1} * \frac{\partial P\left(Z_i = z \middle| \underline{\theta}_j\right)}{\partial \theta_2} * \frac{1}{\left(P\left(Z_i = z \middle| \underline{\theta}_j\right)\right)^2} \right) \end{array} \right] \tag{A.18}$$

where

$$\frac{\partial P\left(Z_i = z \middle| \underline{\theta}_j\right)}{\partial\theta_1\partial\theta_2} = \frac{mA''_\theta + mB''_\theta}{G} - \frac{mG''_\theta\left(A+B\right)}{G^2} - \frac{\frac{\partial G}{\partial\theta_1}\left(\frac{\partial A}{\partial\theta_1}+\frac{\partial B}{\partial\theta_1}\right)}{G^2} - \frac{\frac{\partial G}{\partial\theta_2}\left(\frac{\partial A}{\partial\theta_2}+\frac{\partial B}{\partial\theta_2}\right)}{G^2} + \frac{2\frac{\partial G}{\partial\theta_1}\frac{\partial G}{\partial\theta_2}\left(A+B\right)}{G^3} \tag{A.19}$$

setting

$$mA''_\theta = \frac{\partial^2 A}{\partial\theta_1\partial\theta_2} = \frac{\left[z\left(\alpha_{i1}\right)^2\frac{\partial A}{\partial\theta_2}\left(\theta_{j1}-\delta_{i1}\right)\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right]}{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2} - \frac{\left[\frac{z\left(\alpha_{i1}\right)^2\left(\alpha_{i2}\right)^2 A\left(\theta_{j1}-\delta_{i1}\right)\left(\theta_{j2}-\delta_{i2}\right)}{\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}}\right]}{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}, \tag{A.20}$$

$$mB''_{\theta} = \frac{\partial^2 B}{\partial\theta_1\partial\theta_2}$$
$$= \frac{\left[(M-z)(\alpha_{i1})^2 \frac{\partial B}{\partial\theta_2}(\theta_{j1}-\delta_{i1})\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right]}{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}$$
$$- \frac{\left[\frac{(M-z)(\alpha_{i1})^2(\alpha_{i2})^2 B(\theta_{j1}-\delta_{i1})(\theta_{j2}-\delta_{i2})}{\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}}\right]}{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}, \tag{A.21}$$

and

$$mG''_{\theta} = \frac{\partial^2 G}{\partial\theta_1\partial\theta_1} = \sum_{w=0}^{C}\left(mA''_{\theta} + mB''_{\theta}\right). \tag{A.22}$$

Thus, joint information is obtained via

$$I_i(\underline{\theta}) = -\mathrm{E}\left(\frac{\partial^2 \ln(L_i)}{\partial\theta_1\partial\theta_2}\right)$$
$$= -\mathrm{E}\left\{\left[\frac{\partial^2 P(Z_i = z|\underline{\theta}_j)}{\partial\theta_1\partial\theta_2} * \frac{1}{P(Z_i = z|\underline{\theta}_j)}\right] - \left[\frac{\partial P(Z_i = z|\underline{\theta}_j)}{\partial\theta_1} * \frac{\partial P(Z_i = z|\underline{\theta}_j)}{\partial\theta_2} * \frac{1}{\left[P(Z_i = z|\underline{\theta}_j)\right]^2}\right]\right\}. \tag{A.23}$$

## APPENDIX B

The marginal probability of response vector $\mathbf{X}_j$ is

$$P\left(\mathbf{X}_j\right)=\int\ldots\int P\left(\mathbf{X}_j\left|\theta_{j1}\ldots\theta_{jD}\right.\right)g_1\left(\theta_{j1}\right)\ldots g_D\left(\theta_{jD}\right)d\theta_1\ldots d\theta_D \tag{B.1}$$

where $g_1(\theta_{j1})$ ... $g_D(\theta_{jD})$ are population prior distributions for each of the $D$ dimensions. The likelihood of Equation B.1 is:

$$L=\frac{J!}{\prod_{j=1}^{J}r_j!}\prod_{j=1}^{J}\left[P\left(\mathbf{X}_j\right)^{r_j}\right] \tag{B.2}$$

where $r_j$ is the number of individuals with response vector $\mathbf{X}_j$. Taking the logarithm of the likelihood function in Equation B.2 yields

$$\ln\left(L\right)=\ln\left(J!\right)-\sum_{j=1}^{J}\ln\left(r_J!\right)+\sum_{j=1}^{J}r_j\ln\left[P\left(\mathbf{X}_j\right)\right]. \tag{B.3}$$

The derivative of Equation B.3 with respect to an item location for the $i$th item on the $d$th dimension $\delta_{id}$ becomes

$$\begin{aligned}\frac{\partial\ln\left(L\right)}{\partial\delta_{id}}&=\sum_{j=1}^{J}\frac{r_j}{P\left(\mathbf{X}_j\right)}\frac{\partial P\left(\mathbf{X}_j\right)}{\partial\delta_{id}}\\&=\sum_{j=1}^{J}\frac{r_j}{P\left(\mathbf{X}_j\right)}\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty}\left[\begin{array}{l}\dfrac{\partial P\left(Z_i=x_{ji}\left|\underline{\theta}_j\right.\right)}{\partial\delta_{id}}\\*\prod_{i=1}^{I}P\left(Z_i=x_{ji}\left|\underline{\theta}_j\right.\right)\dfrac{\Gamma\left(\theta\right)\Delta\left(\delta\right)\Upsilon\left(\alpha\right)}{P\left(Z_i=x_{ji}\left|\underline{\theta}_j\right.\right)}\\*\prod_{k=1}^{C}t\left(\tau_{ik}\right)\end{array}\right]d\theta_{j1}\ldots d\theta_{jD}\end{aligned} \tag{B.4}$$

where $\Gamma(\theta) = g_1(\theta_{j1}) \ldots g_D(\theta_{jD})$, $\Delta(\delta) = b_1(\delta_{i1}) \ldots b_D(\delta_{iD})$, and

$\Upsilon(\alpha) = a_1(\alpha_{i1}) \ldots a_D(\alpha_D)$.

Simplifying Equation B.4 leads to

$$\frac{\partial \ln(L)}{\partial \delta_{id}} = \sum_{j=1}^{J} \frac{r_j}{P(\mathbf{X}_j)} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left[ \begin{array}{l} \dfrac{\partial P(Z_i = x_{ji} | \underline{\theta}_j)}{\partial \delta_{id}} \\ * P(\mathbf{X}_j | \underline{\theta}_j) \\ * \dfrac{\Gamma(\theta) \Delta(\delta) \Upsilon(\alpha)}{P(Z_i = x_{ji} | \underline{\theta}_j)} \\ * \prod_{k=1}^{C} t(\tau_{ik}) \end{array} \right] d\theta_{j1} \ldots d\theta_{jD}. \qquad \text{(B.5)}$$

Equation B.5 can be approximated in quadrature form as

$$\frac{\partial \ln(L)}{\partial \delta_{id}} = \sum_{q_1}^{Q_1} \cdots \sum_{q_d}^{Q_D} \left[ \sum_{j=0}^{J} \left( \begin{array}{l} \dfrac{r_j L_j\left(A_{q_1} \ldots A_{q_d}\right) W\left(A_{q_1} \ldots A_{q_d}\right)}{\tilde{P}_j} \\ * \dfrac{\partial P\left(Z_i = x_{ji} \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \delta_{id}} \\ * \dfrac{1}{P\left(Z_i = x_{ji} \middle| A_{q_1} \ldots A_{q_d}\right)} \end{array} \right) \right] + \frac{\partial \ln\left[b\left(\delta_{id}\right)\right]}{\partial \delta_{id}}$$

$$= \left( \sum_{q_1}^{Q_1} \cdots \sum_{q_d}^{Q_D} \left[ \sum_{z=0}^{C} \sum_{j=1}^{J} \left( \begin{array}{l} \dfrac{r_j H_{jiz} L_j\left(A_{q_1} \ldots A_{q_d}\right) W\left(A_{q_1} \ldots A_{q_d}\right)}{\tilde{P}_j} \\ * \dfrac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \delta_{id}} \\ * \dfrac{1}{P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)} \end{array} \right) \right] \right) + \frac{\partial \ln\left[b\left(\delta_{id}\right)\right]}{\partial \delta_{id}}$$

$$= \left( \sum_{q_1}^{Q_1} \cdots \sum_{q_d}^{Q_D} \left[ \sum_{z=0}^{C} \left( \begin{array}{l} \dfrac{\bar{r}_{izq_1 \ldots q_d}}{P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)} \\ * \dfrac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \delta_{id}} \end{array} \right) \right] \right) + \frac{\partial \ln\left[b\left(\delta_{id}\right)\right]}{\partial \delta_{id}}. \qquad \text{(B.6)}$$

In a similar manner it is possible to approximate the derivative of Equation B.3 with respect to item discrimination and subjective response category threshold parameters in quadrature form leading to

$$\frac{\partial \ln(L)}{\partial \alpha_{id}} = \left( \sum_{q_1}^{Q_1} \cdots \sum_{q_d}^{Q_D} \left[ \sum_{z=0}^{C} \left( \begin{array}{l} \dfrac{\bar{r}_{izq_1 \ldots q_d}}{P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)} \\ * \dfrac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \alpha_{id}} \end{array} \right) \right] \right) + \frac{\partial \ln\left[a\left(\alpha_{id}\right)\right]}{\partial \alpha_{id}} \qquad \text{(B.7)}$$

and

$$\frac{\partial \ln(L)}{\partial \tau_{ik}} = \left( \sum_{q_1}^{Q_1} \cdots \sum_{q_d}^{Q_D} \left[ \sum_{z=0}^{C} \left( \frac{\bar{r}_{izq_1 \ldots q_d}}{P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)} * \frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \tau_{ik}} \right) \right] \right) + \frac{\partial \ln\left[t(\tau_{ik})\right]}{\partial \tau_{ik}}. \quad \text{(B.8)}$$

## APPENDIX C

The derivative of

$$P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right) \tag{C.1}$$

with respect to each item parameter requires the definition of

$$A = \exp\left[\left(z\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(A_{q_d}-\delta_{id})^2}\right) - \sum_{k=0}^{z}\psi_{ik}\right], \tag{C.2}$$

$$B = \exp\left[\left((M-z)\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(A_{q_d}-\delta_{id})^2}\right) - \sum_{k=0}^{z}\psi_{ik}\right], \tag{C.3}$$

and

$$G = \sum_{z=0}^{C}(A+B). \tag{C.4}$$

The derivative of C.1 with respect to item location is calculated for each item as

$$\frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \delta_{id}} = \frac{G\left(A'_\delta + B'_\delta\right) - G'_\delta\left(A+B\right)}{G^2} \tag{C.5}$$

where

$$A'_\delta = \frac{Az\left(\alpha_{id}\right)^2\left(\delta_{id} - A_{q_d}\right)}{\sqrt{\sum_{d=1}^{D}\left[\left(\alpha_{id}\right)^2\left(A_{q_d}-\delta_{id}\right)^2\right]}}, \tag{C.6}$$

$$B'_{\delta} = \frac{B(M-z)(\alpha_{id})^2\left(\delta_{id} - A_{q_d}\right)}{\sqrt{\sum_{d=1}^{D}\left[(\alpha_{id})^2\left(A_{q_d} - \delta_{id}\right)^2\right]}}, \tag{C.7}$$

and

$$G'_{\delta} = \sum_{z=0}^{C}\left(A'_{\delta} + B'_{\delta}\right). \tag{C.8}$$

The derivative of C.1 with respect to item discrimination also calculated for each item is

$$\frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \alpha_{id}} = \frac{G\left(A'_{\alpha} + B'_{\alpha}\right) - G'_{\alpha}\left(A + B\right)}{G^2} \tag{C.9}$$

where

$$A'_{\alpha} = \frac{Az\alpha_{id}\left(A_d - \delta_{id}\right)^2}{\sqrt{\sum_{d=1}^{D}\left[(\alpha_{id})^2\left(A_d - \delta_{id}\right)^2\right]}} - A\sum_{k=0}^{z}\tau_{ik}, \tag{C.10}$$

$$B'_{\alpha} = \frac{B(M-z)\alpha_{id}\left(A_d - \delta_{id}\right)^2}{\sqrt{\sum_{d=1}^{D}\left[(\alpha_{id})^2\left(A_d - \delta_{id}\right)^2\right]}} - B\sum_{k=0}^{z}\tau_{ik}, \tag{C.11}$$

and

$$G'_{\alpha} = \sum_{z=0}^{C}\left(A'_{\alpha} + B'_{\alpha}\right). \tag{C.12}$$

The derivative of C.1 with respect to subjective response category thresholds by item is

$$\frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \tau_{ik}} = \frac{G\left(A'_{\tau} + B'_{\tau}\right) - G'_{\tau}\left(A + B\right)}{G^2} \tag{C.13}$$

where

$$A'_{\tau} = -A\sum_{d=1}^{D}\alpha_{id}, \tag{C.14}$$

$$B'_{\tau} = -B\sum_{d=1}^{D}\alpha_{id}, \tag{C.15}$$

and

$$G'_{\tau} = \sum_{z=0}^{C}\left(A'_{\tau} + B'_{\tau}\right). \tag{C.16}$$

## APPENDIX D

Information with respect to MGGUM item location ($\delta_{id}$) and discrimination ($\alpha_{id}$) parameters using a dimension by dimension approach in Equation 69 is

$$\tilde{\mathbf{I}}_i = \begin{pmatrix} \tilde{I}_{\alpha_{id}\alpha_{id}} & \tilde{I}_{\alpha_{id}\delta_{id}} \\ \tilde{I}_{\alpha_{id}\delta_{id}} & \tilde{I}_{\delta_{id}\delta_{id}} \end{pmatrix} = -\mathrm{E}\begin{pmatrix} \dfrac{\partial^2 \ln L}{\partial \alpha_{id}^2} & \dfrac{\partial^2 \ln L}{\partial \alpha_{id}\ \partial \delta_{id}} \\ \dfrac{\partial^2 \ln L}{\partial \alpha_{id}\ \partial \delta_{id}} & \dfrac{\partial^2 \ln L}{\partial \delta_{id}^2} \end{pmatrix}. \tag{D.1}$$

Solving for one dimension of one item at a time and using Equation 45 in quadrature form, the likelihood is

$$L = \prod_{j=1}^{J}\prod_{i=1}^{I} P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right). \tag{D.2}$$

Next, take the log of equation D.2

$$\ln(L) = \sum_{j=1}^{J}\sum_{i=1}^{I} \ln\left[ P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)\right]. \tag{D.3}$$

Then, the first derivative of Equation D.3 with respect to the *i*th item location estimate on the *d*th dimension is

$$\frac{\partial \ln\left[ P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)\right]}{\partial \delta_{id}} = \sum_{i=1}^{I}\left( \frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \delta_{id}} * \frac{1}{P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)} \right) \tag{D.4}$$

and with respect to the *i*th item discrimination estimate on the *d*th dimension is

$$\frac{\partial \ln\left[P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)\right]}{\partial \alpha_{id}} = \sum_{i=1}^{I}\left(\frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \alpha_{id}} * \frac{1}{P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}\right) \quad \text{(D.5)}$$

where

$$\frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \delta_{id}} = \frac{G\left(A'_\delta + B'_\delta\right) - G'_\delta\left(A + B\right)}{G^2} \quad \text{(D.6)}$$

and

$$\frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \alpha_{id}} = \frac{G\left(A'_\alpha + B'_\alpha\right) - G'_\alpha\left(A + B\right)}{G^2}, \quad \text{(D.7)}$$

again using the definitions found in Appendix C.

Taking the second derivative of Equations D.6 and D.7 with respect to the *i*th item location estimate and item discrimination estimation on the *d*th dimension yields

$$\frac{\partial^2 P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \delta_{id}^2} = \frac{A''_\delta + B''_\delta}{G} - \frac{G''_\delta\left(A + B\right)}{G^2} - \frac{2G'_\delta\left(A'_\delta + B'_\delta\right)}{G^2} + \frac{2\left(G'_\delta\right)^2\left(A + B\right)}{G^3} \quad \text{(D.8)}$$

and

$$\frac{\partial^2 P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \alpha_{id}^2} = \frac{A''_\alpha + B''_\alpha}{G} - \frac{G''_\alpha\left(A + B\right)}{G^2} - \frac{2G'_\alpha\left(A'_\alpha + B'_\alpha\right)}{G^2} + \frac{2\left(G'_\alpha\right)^2\left(A + B\right)}{G^3} \quad \text{(D.9)}$$

where

$$A''_{\delta} = \frac{\partial^2 A}{\partial \delta_{id}^2}$$

$$= \frac{\left\{ z\left(\alpha_{id}\right)^2 \left( \sqrt{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2} \right) \left( A + A'_{\delta} \left[ \delta_{id} - A_{q_d} \right] \right) \right\}}{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2} \tag{D.10}$$

$$- \frac{\left\{ z\left(\alpha_{id}\right)^4 A(\delta_{id} - A_{q_d})^2 \right\}}{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2},$$

$$B''_{\delta} = \frac{\partial^2 B}{\partial \delta_{id}^2}$$

$$= \frac{\left\{ \left(M - z\right)\left(\alpha_{id}\right)^2 \left( \sqrt{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2} \right) \left( B + B'_{\delta} \left[ \delta_{id} - A_{q_d} \right] \right) \right\}}{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2} \tag{D.11}$$

$$- \frac{-\left\{ \left(M - z\right)\left(\alpha_{id}\right)^4 B(\delta_{id} - A_{q_d})^2 \right\}}{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2},$$

$$G''_{\delta} = \frac{\partial^2 G}{\partial \delta_{id}} = \sum_{w=0}^{C} \left( A''_{\delta} + B''_{\delta} \right), \tag{D.12}$$

$$A''_{\alpha} = \frac{\left[ z(A_{q_d} - \delta_{id})^2 \left( A + \alpha_{id} A'_{\alpha} \right) \sqrt{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2} \right]}{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2}$$

$$- \frac{\left[ \frac{z\left(\alpha_{id}\right)^2 A\left(A_{q_d} - \delta_{id}\right)^4}{\sqrt{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2}} \right]}{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2}, \tag{D.13}$$

$$B''_{\alpha} = \frac{\left[ (M-z)(A_{q_d} - \delta_{id})^2 \left( B + \alpha_{id} B'_{\alpha} \right) \sqrt{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2} \right]}{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2}$$

$$- \frac{\left[ \frac{(M-z)(\alpha_{id})^2 B \left( A_{q_d} - \delta_{id} \right)^4}{\sqrt{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2}} \right]}{\sum_{d=1}^{D} \alpha_{id}^2 (A_{q_d} - \delta_{id})^2}, \tag{D.14}$$

$$G''_{\alpha} = \frac{\partial^2 G}{\partial \alpha_{id}} = \sum_{w=0}^{C} \left( A''_{\alpha} + B''_{\alpha} \right). \tag{D.15}$$

Therefore, information for the *i*th item on the *d*th dimension for item location and item discrimination parameters is obtained via

$$\tilde{I}_{\delta_{id}\delta_{id}} = -\mathrm{E}\left( \frac{\partial^2 \ln(L)}{\partial \delta_{id}^2} \right)$$

$$-\mathrm{E}\left\{ \left[ \frac{\partial^2 P\left( Z_i = z \middle| A_{q_d} \right)}{\partial \delta_{id}^2} * \frac{1}{P\left( Z_i = z \middle| A_{q_d} \right)} \right] - \left[ \left( \frac{\partial P\left( Z_i = z \middle| A_{q_d} \right)}{\partial \delta_{id}} \right)^2 * \frac{1}{\left( P\left( Z_i = z \middle| A_{q_d} \right) \right)^2} \right] \right\} \tag{D.16}$$

and

$$\tilde{I}_{\alpha_{id}\alpha_{id}} = -\mathrm{E}\left(\frac{\partial^2 \ln(L)}{\partial \alpha_{id}^2}\right)$$
$$= -\mathrm{E}\left\{\begin{array}{l}\left[\frac{\partial^2 P\left(Z_i = z \middle| A_{q_d}\right)}{\partial \alpha_{id}^2} * \frac{1}{P\left(Z_i = z \middle| A_{q_d}\right)}\right] \\ -\left[\left(\frac{\partial P\left(Z_i = z \middle| A_{q_d}\right)}{\partial \alpha_{id}}\right)^2 * \frac{1}{\left(P\left(Z_i = z \middle| A_{q_d}\right)\right)^2}\right]\end{array}\right\}, \quad \text{(D.17)}$$

respectively.

Item information with respect to subjective response category thresholds is calculated separately from item location and discrimination parameters, as these values do not vary across dimensions

$$\tilde{\mathbf{I}}_{\tau_{ik}} = \begin{pmatrix} \tilde{I}_{\tau_{i1}\tau_{i1}} & \cdots & \tilde{I}_{\tau_{i1}\tau_{iC}} \\ \vdots & \ddots & \vdots \\ \tilde{I}_{\tau_{iC}\tau_{i1}} & \cdots & \tilde{I}_{\tau_{iC}\tau_{iC}} \end{pmatrix}. \quad \text{(D.18)}$$

The first derivative of Equation D.3 with respect to the *k*th subjective response category threshold for the *i*th item is

$$\frac{\partial \ln\left[P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)\right]}{\partial \tau_{ik}} = \sum_{i=1}^{I}\left(\frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \tau_{ik}} * \frac{1}{P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}\right) \quad \text{(D.19)}$$

where

$$\frac{\partial P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \tau_{ik}} = \frac{G\left(A'_\tau + B'_\tau\right) - G'_\tau\left(A + B\right)}{G^2} \quad \text{(D.20)}$$

can be determined using definitions from Appendix C.

The second derivative of Equation D.20 with respect to the *k*th subjective response category threshold is

$$\frac{\partial^2 P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \tau_{ik}^2} = \frac{A_\tau'' + B_\tau''}{G} - \frac{G_\tau''\left(A+B\right)}{G^2} - \frac{2G_\tau'\left(A_\tau' + B_\tau'\right)}{G^2} + \frac{2\left(G_\tau'\right)^2\left(A+B\right)}{G^3} \tag{D.21}$$

where

$$A_\tau'' = \frac{\partial^2 A}{\partial \tau_{ik}} = A\left(\sum_{d=1}^{D} \alpha_{id}\right)^2, \tag{D.22}$$

$$B'' = \frac{\partial^2 B}{\partial \tau_{ik}} = B\left(\sum_{d=1}^{D} \alpha_{id}\right)^2, \tag{D.23}$$

and

$$G_\tau'' = \frac{\partial^2 G}{\partial \tau_{ik}} = \sum_{w=0}^{C}\left(A_\tau'' + B_\tau''\right). \tag{D.24}$$

Therefore, the information for the *k*th subjective response category threshold parameter is obtained via

$$\begin{aligned} \tilde{I}_{\tau_{ik}\tau_{ik}} &= -\mathrm{E}\left(\frac{\partial^2 \ln\left(L\right)}{\partial \tau_{ik}^2}\right) \\ &= -\mathrm{E}\left\{\left[\frac{\partial^2 P\left(Z_i = z \middle| A_{q_d}\right)}{\partial \tau_{ik}^2} * \frac{1}{P\left(Z_i = z \middle| A_{q_d}\right)}\right] - \left[\left(\frac{\partial P\left(Z_i = z \middle| A_{q_d}\right)}{\partial \tau_{ik}}\right)^2 * \frac{1}{\left(P\left(Z_i = z \middle| A_{q_d}\right)\right)^2}\right]\right\}. \end{aligned} \tag{D.25}$$

The mixed derivative of Equation D.25 with respect to the *k*th and *k**th subjective response category threshold with $k \neq k^*$ is the foundation for computing joint information

$$\frac{\partial^2 P\left(Z_i = z \middle| A_{q_1} \ldots A_{q_d}\right)}{\partial \tau_{ik} \partial \tau_{ik^*}} = \frac{mA''_{\tau_k \tau_{k^*}} + mB''_{\tau_k \tau_{k^*}}}{G} - \frac{mG''_{\tau_k \tau_{k^*}}(A+B)}{G^2} - \frac{\frac{\partial G}{\partial \tau_{ik}}\left(\frac{\partial A}{\partial \tau_{ik}} + \frac{\partial B}{\partial \tau_{ik}}\right)}{G^2} - \frac{\frac{\partial G}{\partial \tau_{ik^*}}\left(\frac{\partial A}{\partial \tau_{ik^*}} + \frac{\partial B}{\partial \tau_{ik^*}}\right)}{G^2} + \frac{2\frac{\partial G}{\partial \tau_{ik}}\frac{\partial G}{\partial \tau_{ik^*}}(A+B)}{G^3} \tag{D.26}$$

where

$$mA''_{\tau_k \tau_{k^*}} = \frac{\partial^2 A}{\partial \tau_{ik} \partial \tau_{ik^*}} = \frac{\partial A}{\partial \tau_{ik^*}}\left(\sum_{d=1}^{D} \alpha_{id}\right)^2, \tag{D.27}$$

$$mB''_{\tau_k \tau_{k^*}} = \frac{\partial^2 B}{\partial \tau_{ik} \partial \tau_{ik^*}} = \frac{\partial B}{\partial \tau_{ik^*}}\left(\sum_{d=1}^{D} \alpha_{id}\right)^2, \tag{D.28}$$

and

$$mG''_{\tau_k \tau_{k^*}} = \frac{\partial^2 G}{\partial \tau_{ik} \partial \tau_{ik^*}} = \sum_{w=0}^{C}\left(mA''_{\tau_k \tau_{k^*}} + mB''_{\tau_k \tau_{k^*}}\right). \tag{D.29}$$

Joint information can now be calculated using

$$\tilde{I}_{\tau_{ik}\tau_{ik^*}} = -\mathrm{E}\left(\frac{\partial^2 \ln(L)}{\partial\tau_{ik}\partial\tau_{ik^*}}\right)$$

$$= -\mathrm{E}\left\{\left[\frac{\partial^2 P\left(Z_i = z \middle| A_{q_d}\right)}{\partial\tau_{ik}\partial\tau_{ik^*}} * \frac{1}{P\left(Z_i = z \middle| A_{q_d}\right)}\right] - \left[\frac{\partial P\left(Z_i = z \middle| A_{q_d}\right)}{\partial\tau_{ik}} * \frac{\partial P\left(Z_i = z \middle| A_{q_d}\right)}{\partial\tau_{ik^*}} * \frac{1}{\left[P\left(Z_i = z \middle| A_{q_d}\right)\right]^2}\right]\right\}. \quad \text{(D.30)}$$

## APPENDIX E

Using the standardized residuals found in Equation 74, the singular value decomposition of **S** involves the following iterative process as taken from ter Braak (1988) and Jongman, ter Braak, and van Tongeren (1995), adapted for an IRT framework

Step 1: Select arbitrary unique item locations, $\boldsymbol{\delta}^* = [\delta_1^*, \ldots, \delta_i^*]$.

Step 2: Calculate person locations based on weighted averages of item locations

$$\theta_j^* = \sum_{i=1}^{I} \frac{x_{ji}\delta_i^*}{x_{+i}} \tag{E.1}$$

Step 3: Calculate new item locations $\delta^*$ based on the new person locations

$$\delta_i^* = \sum_{j=1}^{J} \frac{x_{ji}\theta_j^*}{x_{j+}} \tag{E.2}$$

Step 4: For the first dimension/axis, skip to Step 6. For all other dimensions/axes, calculate $u$ and $v$

$$u = \sum_{j=1}^{J} \frac{x_{ji}\theta_j^*\theta_j^{*-1}}{x_{++}} \tag{E.3}$$

$$v = \sum_{i=1}^{I} \frac{x_{ji}\delta_i^*\delta_i^{*-1}}{x_{++}} \tag{E.4}$$

Step 5: Calculate the new person, $\theta^{*+1}$, and item, $\delta^{*+1}$, locations so that each dimension is orthogonal to previous dimensions

$$\theta_j^{*+1} = \theta_j^* - u\theta_j^{*-1} \tag{E.5}$$

$$\delta_i^{*+1} = \delta_i^* - v\delta_i^{*-1} \tag{E.6}$$

Step 6: Normalize the locations as detailed in Equations 77 and 78.

Step 7: Convergence is achieved when the new locations are sufficiently close to those of the previous iteration. An acceptable level of convergence is a difference of less than $10^{-10}$. If convergence has not been achieved return to Step 2 and repeat the process.

When detrending by polynomials, the extraction of additional dimensions requires not only orthogonality to previous dimensions, but orthogonality to polynomial functions of previous dimensions. Therefore, it is necessary to compute up to fourth degree polynomial values of item and person location scores (i.e. $\delta_{i2}^* = \left(\delta_i^*\right)^2$, $\delta_{i3}^* = \left(\delta_i^*\right)^3$, $\delta_{i4}^* = \left(\delta_i^*\right)^4$, etc.). These values are then used to in the computations found in Step 2 through Step 7.

## APPENDIX F

1. Abortion is unacceptable under any circumstances.
2. Abortion is the destruction of one life for the convenience of another.
3. Abortion is inhumane.
4. Abortion can be described as taking a life unjustly.
5. Abortion could destroy the sanctity of motherhood.
6. Abortion should not be made readily available to everyone.
7. Even if one believes that there may be some exceptions, abortion is still generally wrong.
8. Abortion is basically immoral except when the woman's physical health is in danger.
9. Abortion should be illegal except in extreme cases involving incest or rape.
10. My feelings about abortion are very mixed.
11. I cannot whole-heartedly support either side of the abortion debate.
12. Abortion should be a woman's choice, but should never be used simply due to its convenience.
13. Abortion should generally be legal, but should never be used as a conventional method of birth control.
14. Although abortion on demand seems quite extreme, I generally favor a woman's right to choose.
15. Regardless of my personal views about abortion, I do believe others should have the legal right to choose for themselves.
16. Society has no right to limit a woman's access to abortion.
17. A woman should retain the right to choose an abortion based on her own life circumstances.
18. Abortion should be legal under any circumstances.
19. Outlawing abortion violates a woman’s civil rights.

## REFERENCES


Ackerman, T.A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education, 7*, 255-278.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 105-113.

Andrich, D. (1988). The application of an unfolding model of the PIRT type to the measurement of attitude. *Applied Psychological Measurement, 12*, 33-51.

Andrich, D. (1989). A probabilistic IRT model for unfolding preference data. *Applied Psychological Measurement, 13*, 193-216.

Andrich, D. (1995). Hyperbolic cosine latent trait models for unfolding direct responses and pairwise preferences. *Applied Psychological Measurement, 19*, 269-290.

Andrich, D. (1996). A general hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology, 49*, 347-365.

Andrich, D., & Luo, G. (1993). A hyperbolic cosine latent trait model for unfolding dichotomous single-stimulus responses. *Applied Psychological Measurement, 17*, 253-276.

Baker, F.B. (1987). Methodology review: Item parameter estimation under the one-, two-, and three-parameter logistic models. *Applied Psychological Measurement, 11*, 111-141.

Béguin, A., & Glas, C. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika, 66*, 541-561.

Bennett, J.F., & Hays, W.L. (1960). Multidimensional unfolding: Determining the dimensionality of ranked preference data. *Psychometrika, 25*, 27-43.

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., & Mislevy, R.J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6*, 431-444.

Bolt, D.M., & Lall, V.F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. *Applied Psychological Measurement, 27*, 395-414.

Burdenski, T. (2000). Evaluating Univariate, bivariate, and multivariate normality using graphical and statistical procedures. *Multiple Linear Regression Viewpoints, 26*, 15-28.

Busing, F.M.T.A. (2010). *Advances in Multidimensional Unfolding.* Doctoral dissertation, Universiteit Leiden.

Busing, F.M.T.A., Groenen, P.J.F., & Heiser, W.J. (2005). Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika, 70,* 71-98.

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika, 75*, 33-57.

Carroll, J.D., & Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35*, 283-319.

Carter, N.T., & Dalal, D.K. (2010). An ideal point account of the JDI Work Satisfaction Scale. *Personality and Individual Differences, 49*, 743-748.

Carter, N.T., Lake, C.J., & Zickar, M.J. (2010). Toward understanding the psychology of unfolding. *Industrial and Organizational Psychology: Perspective on Science and Practice, 3*, 511-514.

Chernyshenko, O.S., Stark, S. Drasgow, F., & Roberts, B.W. (2007). Construction personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment, 19*, 88-106.

Coombs, C.H. (1950). Psychological scaling without a unit of measurement. *The Psychological Review, 57*, 145-158.

Coombs, C.H. (1960). A theory of data. *The Psychological Review, 67*, 143-159.

Coombs, C.H. (1964). *A Theory of Data*. Ann Arbor: Mathesis Press.

Cui, W., Roberts, J.S., & Bao, H. (April, 2004). *Data Demands for the Generalized Graded Unfolding Model*. Poster presented at the 2004 Annual Meeting of the National Council on Measurement in Education, San Diego, California.

de la Torre, J., Stark, S., & Chernyshenko, O.S. (2006). Markov Chain Monte Carlo estimation of item parameters for the generalized graded unfolding model. *Applied Psychological Measurement, 30*, 216-232.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, 39*, 1-38.

Doornik, J.A. (2003). *Object-oriented Matrix Programming Using Ox* (Version 3.1) [Computer software]. London, U.K.: Timberlake Consultants Press.

Drasgow, F., Chernyshenko, O.S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology: Perspectives on Science and Practice, 3*, 465-476.

Finkelman, M.D., Hooker, G., & Wang, Z. (2010). Prevalence and magnitude of paradoxical results in multidimensional item response theory. *Journal of Educational and Behavioral Statistics, 35*, 744-761.

Gao, F. & Chen, L. (2005). Bayesian or Non-Bayesian: A comparison study of item parameter estimation in the three-parameter logistic model. *Applied Measurement in Education, 18*, 351-380.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721-741.

Green, P.J., & Hastie, D.I. (2009). Reversible jump MCMC. *Genetics, 155*, 1391-1403.

Greenacre, M.J. (2007). *Correspondence Analysis in Practice. Second Edition.* London: Chapman & Hall/CRC.

Haberman, S.J. (1977). Maximum likelihood estimates in exponential response models. *Annals of Statistics, 5*, 815-841.

Harwell, M.R. & Baker, F.B. (1991). The use of prior distributions in marginalized Bayesian item parameter estimation: A didactic. *Applied Psychological Measurement, 15*, 375-389.

Harwell, M.R., Baker, F.B., & Zwarts, M. (1988). Item parameter estimation via marginal maximum likelihood and an EM algorithm: A didactic. *Journal of Educational Statistics, 13*, 243-271.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*, 97-109.

Hill, M.O. (1994). *DECORANA and TWINSPAN, for Ordination and Classification of Multivariate Species Data: A New Edition.* Oxfordshire, United Kingdom: Center for Ecology & Hydrology.

Hill, M.O., & Gauch, H.G. (1980). Detrended correspondence analysis: An improved ordination technique. *Vegetatio, 42*, 47-58.

Hoitjink, H. (1990). A latent trait model for dichotomous choice data. *Psychometrika, 55*, 641-656.

Javaris, K.N., & Ripley, B.D. (2007). An "unfolding" latent variable model for Likert attitude data: Drawing inferences adjusted for response style. *Journal of the American Statistical Association, 102*, 454-463.

Jongman, R.H.G., ter Braak, C.J.F., & van Tongeren, O.F.R. (1995). *Data Analysis in Community and Landscape Ecology*. Melbourne, Australia: Cambridge University Press.

Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*, 1-27.

Kruskal, J.B., & Carroll, J.D. (1969). Geometrical models and badness-of-fit functions. In P.R. Krishnaiah (Ed.), *Multivariate Analysis* (Vol. 2, pp. 639-671). New York: Academic Press.

Kruskal, J.B., & Wish, M. (1978). *Multidimensional Scaling*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-011. Beverly Hills, CA: Sage.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 44-53.

Lim, R.G., & Drasgow, F. (1990). Evaluation of two methods for estimating item response theory parameters when assessing differential item functioning. *Journal of Applied Psychology, 75*, 164-174.

Lord, F.M. (1986). Maximum likelihood and Bayesian parameter estimation in item response theory. *Journal of Educational Measurement, 23*, 157-162.

Lord, F.M., & Novick, M.R. (1968). *Statistical Theories of Mental Test Scores.* Reading, MA: Addison-Wesley.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique, and figure directions. *Statistics in Medicine, 28,* 3049-3067.

Maydeu-Olivares, A., Cai, L., & Hernandez, A. (2011). Comparing the fit of item response theory and factor analysis models. *Structural Equation Modeling, 18,* 333-356.

Maydeu-Olivares, A., & Garcia-Forero, C. (2010). Goodness-of-fit testing. *International Encyclopedia of Education, 7*, 190-196.

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

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics, 21*, 1087-1092.

Minchin, P.R. (1987). An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio, 67*, 1167-1179.

Mislevy, R.J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51*, 177-195.

Nader, I.W., Tran, U.S., & Formann, A.K. (2011). Sensitivity to initial values in full non-parametric maximum-likelihood estimation of the two-parameter logistic model. *British Journal of Mathematical and Statistical Psychology, 64*, 320-336.

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics, 31*, 705-767.

Noel, Y. (1999). Recovering unimodal latent patterns of change by unfolding analysis: Application to smoking cessation. *Psychological Methods, 4*, 173-191.

Patz, R.J., & Junker, B.W. (1999). A straightforward approach to Markov Chain Monte Carlo methods for item response theory. *Journal of Educational and Behavioral Statistics, 24*, 146-178.

Peet, R.K., Knox, R.G., Case, J.S., & Allen, R.B. (1988). Putting things in order: The advantages of detrended correspondence analysis. *The American Naturalist, 131*, 924-934.

Polak, M. (2011). *Item Analysis of Single-Peaked Response Data: The Psychometric Evaluation of Bipolar Measurement Scales*. Unpublished doctoral dissertation, Universiteit Leiden, Leiden, The Netherlands.

Rao, C.R. (1973). *Statistical Inference and Its Applications* (2nd ed.). New York: Wiley.

Rasch, G. (1960/1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: The University of Chicago Press.

Reckase, M.D. (2009). *Multidimensional Item Response Theory*. New York: Springer.

Reckase, M.D., & McKinley, R.L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement, 15*, 361-373.

Roberts, J.S. (2008). Modified likelihood-based item fit statistics for the generalized graded unfolding model. *Applied Psychological Measurement, 32*, 407-423.

Roberts, J.S., Donoghue, J.R., & Laughlin, J.E. (1998). *The Generalized Graded Unfolding Model: A General Parametric Item Response Model for Unfolding Graded Responses* (Research Report RR-98-32). Princeton, NJ: Educational Testing Service.

Roberts, J.S., Donoghue, J.R., & Laughlin, J.E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24*, 3-32.

Roberts, J.S., Donoghue, J.R., & Laughlin, J.E. (2002). Characteristics of MML/EAP parameter estimates in the generalized graded unfolding model. *Applied Psychological Measurement, 26*, 192-207.

Roberts, J.S., Fang, H., Cui, W., & Wang, Y. (2006). GGUM2004: A Windows-based program to estimate parameters in the generalized graded unfolding model. *Applied Psychological Measurement, 30*, 64-65.

Roberts, J.S., Jun, H., Thompson, V.M., & Shim, H. (March, 2009a). *A Distance-Based Multidimensional Extension of the Generalized Graded Unfolding Model*. Paper presented at the 2009 Annual Meeting of the National Council on Measurement in Education, San Diego, California.

Roberts, J.S., & Laughlin, J.E. (1996). A unidimensional item response model for unfolding responses from a graded disagree-agree scale. *Applied Psychological Measurement, 20*, 231-255.

Roberts, J.S., Laughlin, J.E., & Wedell, D.H. (1999). Validity issues in the Likert and Thurstone approaches to attitude measurement. *Educational and Psychological Measurement, 59*, 211-233.

Roberts, J.S., Lin, Y., & Laughlin, J.E. (2001). Computerized adaptive testing with the generalized graded unfolding model. *Applied Psychological Measurement, 25*, 177-196.

Roberts, J.S., & Shim, H. (July, 2010). *Multidimensional Unfolding with Item Response Theory: The Multidimensional Generalized Graded Unfolding Model*. Paper presented at the 2010 Annual Meeting of the Psychometric Society, Athens, Georgia.

Roberts, J.S., Shim, H., Jun, H., Thompson, V.M., & McIntyre, H. (April, 2009b). *A Comparison of the Multidimensional Generalized Graded Unfolding Model and the Classical Non-metric Multidimensional Unfolding Model for Analyzing Responses to Common Likert-type Questionnaires*. Paper presented at the 2009 Annual Meeting of the American Educational Research Association, San Diego, California.

Roberts, J.S., & Thompson, V.M. (June, 2008). *Accuracy of Alternative Parameter Estimation Methods with the Generalized Graded Unfolding Model*. Paper presented at the 2008 Annual Meeting of the Psychometric Society, Durham, New Hampshire.

Roberts, J.S., & Thompson, V.M. (2011). Marginal maximum a posteriori item parameter estimation for the generalized graded unfolding model. *Applied Psychological Measurement*.

Schilling, S., & Bock, R.D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika, 70*, 533-555.

Shaftel, J., Nash, B., & Gillmor, S. (2012). *Effects of the Number of Response Categories on Rating Scales.* Paper presented at the 2012 Annual Meeting of the American Educational Research Association, Vancouver, Canada.

Shepard, R.N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function (I). *Psychometrika, 27*, 125-139.

Shepard, R.N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function (II). *Psychometrika, 27*, 219-246.

Shepard, R.N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika, 39*, 373-421.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2007). *WinBUGS* (Version 1.4.2) [Computer software]. Cambridge, U.K.: MRC Biostatistics Unit, Institute of Public Health.

Stark, S., Chernyshenko, O.S., & Drasgow, F., & Williams, B.A. (2006). Item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology, 91*, 25-39.

Stout, W.F. (1987). A nonparametric approach to assessing latent trait unidimensionality. *Psychometrika, 52*, 589-617.

Sympson, J.B. (1978). A model for testing with multidimensional items. In D.J. Weiss (Ed.), *Proceedings of the 1977 Computerized Adaptive Testing Conference* (pp. 82-98). Minneapolis, MN: University of Minnesota.

Takane, Y., Young, F.W., & De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least-squares method with optimal scaling features. *Psychometrika, 42*, 7–67.

ter Braak, C.J.F. (1988). CANOCO: A FORTRAN program for canonical community ordination by [partial] [detrended] [canonical] correspondence analysis, principal components analysis, and redundancy analysis (Version 2.1). DLO-Agricultural Mathematics Group, Wageningen, The Netherlands.

ter Braak, C.J.F. (1995). Ordination. In R.H.G. Jongman, C.J.F. ter Braak, and O.F.R. van Tongeren (Eds.), *Data Analysis in Community and Landscape Ecology* (pp. 91-173). New York: Cambridge University Press.

ter Braak, C.J.F., & Prentice, I.C. (1988). A theory of gradient analysis. *Advances in Ecological Research, 18*, 271-317.

ter Braak, C.J.F., & Smilauer, P. (2002). *CANOCO Reference Manual and CANODRAW for Windows User's Guide: Software for Canonical Community Ordination (version 4.5)*. Ithaca, NY: Microcomputer Power.

Thompson, V.M., & Roberts, J.S. (July, 2010). *Item Information in the Multidimensional Generalized Graded Unfolding Model*. Poster presented at the 2010 Annual Meeting of the Psychometric Society, Athens, Georgia.

Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273-286.

Thurstone, L.L. (1928). Attitudes can be measured. *The American Journal of Sociology 33*, 529-553.

Weekers, A., & Meijer, R. (2008). Scaling response processes on personality items using unfolding and dominance models: An illustration with a Dutch dominance and unfolding personality inventory. *European Journal of Psychological Assessment, 24*, 65-77.

Whitely, S.E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika, 45*, 479-494.

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

Zhang, J., & Stout, W.F. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika, 64*, 213-249.