MULTIDIMENSIONAL ITEM RESPONSE THEORY: AN INVESTIGATION OF INTERACTION EFFECTS BETWEEN FACTORS ON ITEM PARAMETER RECOVERY USING MARKOV CHAIN MONTE CARLO

By

Jonghwan Lee

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Measurement and Quantitative Methods

2012

ABSTRACT

MULTIDIMENSIONAL ITEM RESPONSE THEORY: AN INVESTIGATION OF INTERACTION EFFECTS BETWEEN FACTORS ON ITEM PARAMETER RECOVERY USING MARKOV CHAIN MONTE CARLO

By

Jonghwan Lee

It has been more than 50 years since Lord (1952) published "A Theory of Test Scores (Psychometric Monograph No.7)" which is recognized as one of the most influential in Item Response Theory (IRT) history. Since then, there has been extensive research investigating several aspects of IRT such as: (1) Modeling; (2) Estimation of latent traits; and (3) Estimation of item parameters. There has also been extensive development of applications based on IRT such as (1) Equating; (2) Linking; (3) Differential Item Function (DIF); (4) Standard setting; and others. All those applications have the same assumption—that the item parameters are calibrated as accurately as possible. Nevertheless, there has been extensive research investigating the techniques to estimate the item and latent trait parameters. All previously developed estimation techniques are based on the uni-dimensional IRT model. However, estimation procedures have become more sophisticated because of the appearance of multidimensional item response theory models (MIRT). In MIRT, there are several factors that are influential in calibration procedures, such as (1) number of latent traits; (2) correlation between the latent traits; (3) non-normal distribution of latent traits; and (4) different types of configurations of latent traits (approximate simple structure and mixed structure).

In this study, the interaction effects of combined factors on item parameter recovery were investigated using the Markov Chain Monte Carlo simulation method. The findings show that a

higher number of dimensions require a bigger sample size than lower dimensions—2000 and 1000 sample sizes for 6-dimensions and 3-dimensions, respectively. That model does not consider correlation and skewness in the latent trait distribution, however. This study shows that if there is an additional factor introduced into the features of the latent trait such as correlation or skewness, increasing the sample size is not helpful in improving the accuracy of item parameter recovery. Rather, an alternative MIRT model should be considered in the case of correlated latent traits, and for transforming non-normal distributions of latent traits to normal distributions. *a*-parameters are more affected when there is correlation between latent traits. *d*-parameters have more influence when the latent trait distribution is skewed.

Overall, the more factors that influence the estimation of the parameters of the MIRT model, the higher the bias found in the item parameter calibration. If the latent structures are independent and normally distributed, then the higher the dimension is in the model specification, the less bias it will have in item parameter calibration. It is also true that if the latent structures have different types configuration, such as AS or MS, then increasing the number of dimensions may possibly decrease the bias created from different types of latent structures configuration. When the latent traits are suspected of having a skewed or non-normal distribution, then it bias is not improved by simply increasing the sample size, though it might be helpful to increase the number of items at the same time. Another way to fix this problem is to use a sample of examinees selected from a wide range of abilities. This is also true in the case of latent traits that are correlated with each other. Selecting the examinee group carefully greatly reduces the bias resulting from the item calibration procedure.

DEDICATION

To my brother, Koowhan Lee,

who supported me from the beginning of this long journey. Without his support, trust, and patience, I would not have been able to complete my doctoral program. I also would like to give my deep appreciation to my sister-in-law.

To my family,

who gave me endless support through my doctoral program. Their love and support made me a better person, and gave me unlimited energy to finish the program.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation to the committee chair and my advisor, Dr. Mark Reckase. Without his guidance, support and encouragement, I would not have been able to finish my degree. I remember when I took his class at the beginning of my doctoral studies. He offered me really deep insights, knowledge and understanding of what measurement is about. My deepest appreciation goes to the other committee members, Dr. Kimberley S. Maier, Dr. Spyros Konstanoupolos, and Dr. Ryan Bowles. In particular, Dr. Kimberley S. Maier's help and advice were essential to my finishing up the dissertation. Without her help in the last minutes, I would not have been able to complete it.

My sincere appreciation also goes to Dr. Barbara Schneider, who supported me financially and academically. She showed me the right direction for a scholar to go. Aside from her financial support, she trained me to be a better scholar. My appreciation also goes to all of the members of the College Ambition Program (CAP) project—Christina Mazuca, Justina Judy, and all of the others who supported me through hard times.

Also, while I have not named them all, my deep appreciation goes to all my friends whom I shared my life as a Ph.D. student at MSU. I would especially like to thank Eun Jeong Noh, who has been the best colleague during my doctoral program.

Last but most of all, my deepest appreciation goes to my family members. My father, SangYeon Lee, my mother, Dooyi Yoo, my brother,MaengHwan Lee, sister-in-law, Haesuk Kwon and sisters, nieces, and nephews; they all deserved to be called Doctor. Without their endless love and support, I would not be standing and hooding my doctoral gown.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This chapter presents a brief introduction to MIRT models, current issues in item parameter calibration procedures in MIRT models, the focus of this study, and research questions to be addressed.

1.1. Multi-dimensional Item Response Theory (MIRT)

Item Response Theory (IRT) has been recognized as one of the major developments in educational and psychological measurement during the 20th century. IRT is a mathematical expression that shows the relation between the characteristics of a person (e.g. a latent trait) and the characteristics of the test items. The history of IRT dates back to when Lawley (1944) and Tucker (1946) published their seminal articles. However, the most important contribution to the IRT literature occurred when Lord (1952) published "*A Theory of Test Score (Psychometric Monograph No.7)*". Lord, Novick, and Birnbaum (1968) published a book called "*Statistical Theories of Mental Test Scores*," which provides the basic assumptions of the IRT models such as 1) local independence, 2) uni-dimensionality of latent trait, and 3) monotonicity. All of these assumptions are crucial factors to the modeling of IRT.

During the last decade, IRT had been used as the primary tool in the educational and psychological measurement fields. Equating, linking, DIF, and computerized adoptive tests are just a few well-known IRT-based applications. Many of the uses of these applications depend on how accurately the item and person's parameters are estimated. In order to have stable and consistent estimation of parameters, all assumptions for item response theory should be fulfilled.

One of the most commonly violated assumptions is the uni-dimensionality in the latent trait structure implied by the item response data. In most instances, it is sufficient to assume that all the test items in a test are sensitive to differences in examinees along a single latent trait (Yanyan Sheng & Wikle, 2007). However, a large body of research has pointed out that this violation of uni-dimensionality leads to a certain degree of bias in parameter estimation, which has, therefore, led to the development of Multi-dimensional Item Response theory (MIRT) models (Bock & Aitkin, 1981; Reckase & McKinley, 1982; Samejima, 1974; Thissen & Steinberg, 1984; Whitely, 1980). Several multi-dimensional item response theory models are proposed such as: 1) multi-dimensional extension of the two-parameter logistic model (Reckase, 1985); 2) multi-dimensional extension of the three-parameter logistic model (Reckase, 2009); 3) multi-dimensional extension of the normal ogive model (McDonald, 1999; Samejima, 1974); 4) multi-dimensional partial credit model (Kelderman & Rijkes, 1994); 5) multi-dimensional extension of the generalized partial credit model (L. H. Yao & Schwarz, 2006); and 6) multi-dimensional extension of the graded response model (Muraki & Carlson, 1993).

Two different kinds of models are commonly referred to as MIRT models—compensatory models and non-compensatory models. In the framework of the compensatory model, the probability of answering correctly is influenced by a weighted linear combination of latent traits. In other words, the probability of answering correctly is influenced not just by one latent trait but by a weighted combination of latent traits. For example, mathematics tests are usually composed of more than two dimensions, such as understanding the problem. Which is comprehensive abilities in reading; translating the problem into an equation, which is the mathematic thinking; and solving the problem, which is analytic ability. All three dimensions should be combined to solve the problem correctly. In the framework of non-compensatory models, one needs to have sufficient levels of each of the measured latent traits in order to solve the question. That is, a deficiency in one latent trait cannot be offset by increasing another one (Bolt & Lall, 2003). As Bolt and Lall (2003) pointed out, the practical distinction between the two types of models is often based on the estimation techniques. In non-compensatory models, it is relatively hard to estimate parameters and to make inferences from them compared to compensatory models, because their estimation procedure requires sufficient variability in the relative difficulties of components across items to identify the dimensions (Maris, 1995).

Two statistical descriptions, MDIFF and MDISC, are commonly used to describe the characteristics of test items in MIRT models. Reckase (1985) described the multi-dimensional difficulty of the test item, often referred to as MDIFF. It has the same interpretation as the *b*-parameter in a uni-dimensional item response theory model that expresses the difficulty of the test item as a direction and a distance in the complete latent space. The equation to estimate the MDIFF is given below (see Reckase, 1985, for complete derivation):

$$\mathrm{MDIFF_i} = \frac{-\mathrm{d_i}}{\sqrt{\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathbf{a}_\mathrm{i}^2}} \tag{1.1}$$

Where $\mathbf{a}_\mathrm{i}$ is the vector of item discrimination parameter, and

$\mathrm{d_i}$ is a scalar parameter that is related to the difficulty of the item.

Reckase and McKinley (1991) developed the overall measure of multi-dimensional discrimination (MDISC), which is analogous to an *a*-parameter in the uni-dimensional model. Instead it represents a single *a*-parameter, it is an overall measure of the capability of an item to distinguish between individuals that are in different locations in the complete latent space. The equation for *m*-dimensional MDISC is given below. See Reckase and McKinley (1991) for the complete derivation.

$$\text{MDISC}=\sqrt{\sum_{k=1}^{m}\mathbf{a}_{ik}^{2}} \tag{1.2}$$

Where $\mathbf{a}_i$ is the vector of item discrimination parameters, and

m is number of dimension.

1.2. Current Issues in Item Parameter Calibration Procedures

Even though the development of estimation techniques in MIRT is still an on-going research topic, many of the estimation procedures used in uni-dimensional item response theory have been adopted into the estimation of parameters in MIRT models. A joint maximum likelihood (JML) (Birbaum, 1968) procedure implemented in LOGIST (Wingersky, Barton, & Lord, 1982), once the most popular computer program for estimating the parameters in unidimensional item response theory, was implemented in MIRTE (Carlson, 1987). The unweighted least squares estimation implemented in NOHARM (Fraser & McDonald, 1988), which is now being used to estimate parameters in both MIRT models and uni-dimensional IRT models. A marginal maximum likelihood estimation procedure is implemented in TESTFACT (Bock et al., 2003). There have been extensive comparative studies that compared the performance between NOHARM and TESTFACT (Béguin & Glas, 2001; Goaz & C.M, 2002; Stone & Yeh, 2006). These studies have shown that neither one of the programs is superior to the other. Rather, the accuracy of item recovery mainly depends on the specification of factors, such as the number of parameters to be estimated, sample size, dimensional structure, or number of items.

One limitation of using Maximum Likelihood Estimation (MLE) in an IRT framework is that, on occasion, parameters of some items cannot be estimated because of the data structure (Baker, 1987). For instance, when responses from an examinee are either all correct or incorrect, it cannot be used to estimate the parameters. Therefore, MLE removes these responses from the dataset. Dropping these non-usable responses causes a loss of information, and can decrease the sample size (e.g. number of response sets), which is a critical factor for the accurate estimation of parameters using the MLE procedure.

Even though there are several estimation procedures currently available, developing estimation procedures in MIRT models is still an active area of research. The major challenge in parameter estimation techniques in MIRT models is that it the relationship between parameter recovery and test specifications, such as number of dimensions, dimension structure, number of items, number of examinee, and selection of parameter distributions, is still not clear. Most estimation procedures are implemented in a computer program, and are used in both practical and research fields that require pre-specifications. Most of the computer programs require specification of the types of models, number, and structures of dimensions that could be estimated, as well as the specific types of algorithms being used to estimate parameters. So it is nearly impossible to explore all the possible relationships among all specifications under one estimation program. In addition, the problem with using several computer programs to investigate the relationship among several factors is that they do not always agree with each other.

Recently, Bayesian analysis, specifically the Markov Chain Monte Carlo (MCMC) methods, has received a great deal of attention from researchers (Béguin & Glas, 2001; Bolt & Lall, 2003; Patz & Junker, 1999a, 1999b; Wollack, Bolt, Cohen, & Lee, 2002). The history of MCMC dates back to 1953, when Metropolis first introduced the metropolis algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953). Prior to that time, MCMC was not recognized as one of the statistical methods because of the cost of computation. However, it has started to gain more attention now that high-speed computers have become available for less cost. However, this method has still suffered from criticism among statisticians, who infer that the MCMC method is subjective because of the biased selection of prior distribution. A. Gelman and Shalizi (2012) argued that Bayesian prior distribution is not a personal belief but a part of a hypothesized model. They also argued that the advantage of using Bayesian inference is deductive, meaning that Bayesian inferences could be made from the given data and some model assumptions. Nevertheless, Bayesian analysis has become the alternative estimation method when the model is complicated, so that it cannot be estimated analytically.

One of the advantages of using MCMC in the MIRT framework is that it could be used as an alternative method to estimate the parameters when a model is complex, so that it is difficult to estimate them analytically (Harwell, Stone, Hsu, & Kirisci, 1996). For example, the number of parameters to be estimated in a compensatory model is calculated by using N(M+1)+ M x Y. M is the number of dimensions, N is the number of items, and Y is the number of examinees (Reckase, 2009). If there are 50 items with five dimensions and 2000 examinees in the model estimation, then 10,300 parameters need to be estimated; it is nearly impossible to have stable estimates.

The other advantage of using MCMC is that it gives more control for researchers to examine the inter-relational effect of several factors at one time (M. Harwell et al., 1996). Since most current item parameter recovery software programs have their own specifications of modeling—such as types of MIRT models, number of dimensions, dimension structure, number of items, and number of examinees—MCMC gives more flexibility to estimate parameters under several factors simultaneously. This is one of the major motivations for using MCMC as the estimation technique for this study to examine the interaction effects of several factors in item parameter recovery in MIRT models.

1.3. Focus of This Study

The main focus of this study is to investigate the interaction effects between factors on item parameter recovery in the MIRT model using the MCMC simulation. Even though there are good existing estimation procedures (e.g. TESTFACT, NOHARM), those programs have limitations related to the relationship among several factors because of the way they are specified in the software. For example, TESTFACT does not provide the option of defining the correlated–dimension structures when running the calibration. MCMC provides a great deal of flexibility in estimating the parameter. In this study, MCMC is being used to investigate the interaction effect among several factors such as number of dimensions, number of items, number of examinees, and dimension structures.

1.4. Research Questions to be Addressed

In order to investigate the interaction between factors, the specifications for large scale assessments (e.g. ACT or NAEP) will be borrowed to guide the selection of number of dimensions and number of items. The number of items is fixed at 60. This number of items is the average number of items over different subject areas of the ACT college admissions tests—75 items for English, 60 items for Mathematics, 40 items for Reading, and 40 items for Science. Since this study explores a 6-dimension MIRT model, 60 items could be distributed evenly into 6-dimensions. Each dimension has 10 items. For a 3-dimension MIRT model, each dimension will have 20 items.

The specific research questions that will be answered from the MCMC simulation are:

1. Under two different numbers of dimensions, 3 and 6, what is the least sample size to obtain the most stable item parameter estimate?
2. Based on conditions from research question 1 plus two different dimension structures, approximate simple structure and mixed structure, what is the least sample size to obtain the most stable item parameter estimate?
3. Based on conditions from research questions 1 and 2 plus correlated latent traits (in mixed structure), then what is the least sample size to obtain the most stable item parameter estimate?
4. When a different shape of ability distribution (e.g. skewness) is imposed into the latent traits distribution, then what is the least sample size to obtain the most stable item parameters estimate?

CHAPTER 2

LITERATURE REVIEW

In this chapter, the theoretical foundations of this study are presented, such as types of latent traits structure configuration; approximate simple structure (AS) and mixed structure (MS); correlated latent traits; skewed latent traits distributions; and item parameter estimation techniques including Maximum likelihood and MCMC.

2.1 Uni-dimension to Multi-dimensions

Uni-dimensionality is one of the most commonly violated assumptions in the latent trait structure implied by the item response data. This violation might cause an increase in bias in the estimation of item parameters and latent traits. Dorans and Kingston (1985) examined the effect of uni-dimensionality violation on equating by using GRE verbal scores, and showed that the violation of uni-dimensionality increased the bias on item parameters estimation, and that it lead to an unsatisfactory equating result. Yanyan Sheng and Wikle (2007) also showed that using a uni-dimensional model returned unsatisfactory results when tests were composited with several distinct abilities; in addition, they showed that applying a multi-dimensional model into a uni-dimensional structure did not harm the estimation results. Therefore, using a multi-dimensional model is a safe way to get the sustainable estimates from the calibration procedure. The question that should then be asked is how many dimensions are needed to represent the latent traits adequately. Reise, Waller, and Comrey (2000) showed that it is better to have a larger number of dimensions when assessing dimensionality. However, as Reckase (2009) pointed out, a cost is paid if more dimensions than necessary are used in an analysis: having more item parameters that need to be estimated might increase the bias as well as the sample size. The most commonly used MIRT model has two- or three-dimensions, and research shows that it requires a sample size of at least 1000 in order to obtain a sustainable item calibration result. However, there is still a lack of research investigating the effect of high dimensions, with more than three-dimensions. Nevertheless, all aspects of latent traits need to be examined, such as the number of the sample size, the distributions of latent traits, the types of latent traits configuration, and the correlation between latent traits when data has a high dimension.

2.2 Types of Latent Configurations: Approximate Simple Structure (AS) and Mixed Structure (MS)

The structure of multidimensional tests is typically categorized into three kinds: 1) simple structure, 2) approximate simple structure, and 3) complex structure (e.g. mixed structure). Mixed structure will be used as interchangeable with complex structure from here on. Simple structure is the most restricted dimensional structure because it only has one nonzero *a*-parameter on one dimension, even though there are several dimensions (Thurstone, 1947). One nonzero *a*-parameter in a multidimensional structure does not often appear in practical test situations. Approximate simple structures have fewer restrictions on nonzero *a*-parameters. In the framework of an approximate simple structure, there are multiple dimensions, however, only one *a*-parameter has a meaningful interpretation, and additional nonzero *a*-parameters are somewhat trivial quantitatively (Walker, Azen, & Schmitt, 2006). Mixed structure may be the most realistic of the dimensional structures. It consists of multiple dimensions in which several *a*-parameters are nonzero. The definition of structure types is based on the weighting of *a*-parameters on dimensions clearly defined from previous research. However, the effect of different types of latent trait configuration on item parameter calibration has not been investigated when it is combined with other factors such as number of dimension, correlation between latent traits, and skewed latent traits distributions. Therefore, this needs to be explored.

2.3 Correlated Latent Traits

Unlike the effect of different structure types for latent traits configuration, the correlated latent traits influence on item parameter calibration has been investigated by several researchers (Batley & Boss, 1993; Finch, 2011; Robert L. McKinley & Reckase, 1984). Batley and Boss (1993) used a simulation study to identify the effect of correlated latent traits on item parameter calibration with two-dimensions. They found that the *d*-parameter does not get affected by correlated latent traits but *a*-parameters are more sensitive to correlated latent traits. Finch (2011) used a simulation study to examine the effect of correlated latent structures with two-dimensions and showed that correlated latent structures do have an influence on item parameter calibration. The magnitude of bias increasing with the magnitude of correlation among correlated latent traits for both *a*- and *d*-parameters. However, both studies only used a two-dimensional MIRT model, so they did not take into account a higher number of dimensions.

2.4 Skewed Latent Trait Distributions

Besides the number of dimensions, types of latent trait structures configuration, and correlated latent traits, non-normal distributions have been identified as being problematic to estimating accurate item parameters. De Ayala and Sava-Bolesta (1999) examined three situations—normal, positively skewed, and uniform distribution—and showed that skewed distributions contribute to high RMSE on item parameter estimates, and uniform distributions contributed to low RMSE. Either uniform or skewed latent traits distributions contributed to the bias in item parameter estimates. Finch (2011) used both positive and negative skewed distribution to identify the effect of non-normal distributions. It was shown that *a*-parameters were consistently underestimated and the bias of *d*-parameters was associated with the direction of skewness. So it is clear that non-normal distributions in the latent traits distributions cause bias in item parameter estimates. Yet, it is not clear how skewness in the latent traits distributions contributes to bias in item parameter estimates if there are several factors combined together, especially with a high number of dimensions.

2.5 Item Parameter Estimation Techniques: MLE and MCMC

There are several parameter estimation techniques suggested by previous studies in the MIRT framework, and implemented in commonly available estimation programs. For example, the marginal maximum likelihood estimation procedure (Bock & Aitkin, 1981) is being implemented in the well-known estimation program TESTFACT (Bock et al., 2003); the unweighted least square method is implemented in NOHARM (Fraser & McDonald, 1988); and the Markov chain Monte Carlo method with Metropolis-Hastings sampling is implemented in BMIRT (L. Yao, 2003).

TESTFACT uses a marginal maximum likelihood estimation (MMLE) procedure to estimate item parameters, and then uses a Bayesian estimation method to estimate the latent traits. TESTFACT specifically uses this MMLE procedure based on the expectation/maximization (EM) algorithm developed by Dempster, Laird, and Rubin (1977). In general, the EM algorithm is an iterative computation of maximum likelihood estimates in the presence of unobserved random variables. Suppose we have a joint probability density function, $f(U, \theta|\xi)$, where U is observed incomplete data and ξ represents the item parameters to be estimated. For the two- and three-parameter logistic IRT models, the distribution of $f(U, \theta|\xi)$ is unknown, so sufficient statistics are not available. Then, the expected values of the $\log f(U, \theta|\xi)$, conditional on some observed representation of θ, are taken and treated as if they were known. This is called the expectation step. These expected values are then used to find the estimated item parameters that are maximizing the log of the likelihood function. This is called the maximization step. See Dempster et al. (1977) and Baker and Kim (2004) for more complete mathematical derivations of the EM algorithm. While this MMLE procedure has been shown to have a consistent performance of parameter recovery in both unidimensional and multidimensional IRT models, it has certain limitations. First, it requires eliminating the response strings that are perfectly correct or incorrect before running the estimation. That might result in loss of information about some examinees. Second, sometimes it returns infinity estimates for discrimination parameters, or zero estimates for discrimination that have an effect on the estimates of other parameters.

In order to overcome the limitations from the MMLE estimation technique, many researchers turned their attention to a Bayesian method, particularly the Markov Chain Monte Carlo (MCMC) procedure. The use of MCMC in an IRT framework is relatively new. Albert (1992) used MCMC with Gibbs sampling with a two-parameter normal ogive model to estimate both the item and person parameters. To run the analysis, he used both simulated data with 30 items and 100 subjects, and real data from the Mathematic placement test from the Department of Mathematics and Statistics at Bowling Green State University. He showed that MCMC with Gibbs sampling gave compatible estimates of item and person parameters to those from the maximum likelihood estimation procedure. Patz and Junker (1999a) showed the potential benefit of MCMC in an IRT framework. In their paper, they reviewed MCMC methods, including two different sampling techniques, Gibbs sampling and Metropolis-Hasting sampling. They suggested that Metropolis-Hasting with Gibbs sampling is more appropriate in an IRT framework than pure Metropolis-Hasting sampling. When the number of parameters increases, it becomes very difficult to maintain reasonable acceptance probabilities in a pure Metropolis-Hasting sampling method while thoroughly exploring the parameter space. In a subsequent paper (Patz & Junker, 1999b), they examined several variations such as multiple item types, missing data, and a rating response IRT model. They showed that MCMC is a good alternative to parameter estimation techniques when the model increases in complexity in terms of the number of parameters that need to be estimated. Wollack et al. (2002) did a comparable study of the effectiveness of MCMC with MML estimation at recovering the underlying parameters of a complex IRT model; the nominal response model. They showed that a greater sample size (300 to 500) retuned better recovery in both MCMC and MLE. They found that the advantage of using MCMC in an IRT framework is its ease of implementation with complex IRT models. However, all these studies focused on the uni-dimensional item response theory models.

Later, MCMC methods were implemented for multidimensional item response theory models (Béguin & Glas, 2001; Bolt & Lall, 2003; Fu, Tao, & Shi, 2009; Y. Sheng, 2008). Béguin and Glas (2001) also found it easier to implement the estimation procedure with more complicated high-dimensional models using the MCMC method. They used a five-dimensional, three-parameter logistic model to examine the recovery of MCMC. They showed that MCMC recovered the true item parameters better than TESTFACT and NOHARM. Even though there are several studies using MCMC in multidimensional item response theory models, there is not extensive research to investigate the effect of a variation of factors under MIRT models. That is the motivation for this study. Equipped with the advantage of easy implementation of a complex model, the interaction effects between factors on item parameters recovery in MIRT model are investigated using MCMC.

CHAPTER 3

RESEARCH DESIGN

In this chapter, the specifications of the research design are presented, including model specification, data generation, specification of MCMC simulation, likelihood function of prior distributions, assessing convergence of MCMC simulation, and evaluation criteria of simulation results.

Simulated data is used in this study rather than a real dataset. The rationale behind using simulated data instead of real data is that it is often suggested in order to separate the effect of model misfit and calibration errors (Bolt, 1999; Davey, Nering, & Thompson, 1997). The model specification and data generation procedure are explained in the following sections.

3.1. Model Specification

Let X_{ij} denote the response of person *j* on item *i* (1 if correct, 0 if not correct). Then the probability of answering correctly is given as follows (Reckase, 1985):

$$P\left(X_{ij}=1 \middle| \boldsymbol{\theta}_j, \mathbf{a}_i, d_i\right)= \frac{\exp[\,\mathbf{a}_i\boldsymbol{\theta}_j + d_i]}{1+\exp[\,\mathbf{a}_i\boldsymbol{\theta}_j + d_i]} \tag{3.1}$$

where $P\left(X_{ij}=1 \middle| \boldsymbol{\theta}_j, \mathbf{a}_i, d_i\right)$ is the probability of a correct response to item *i* by person *j*,

$\mathbf{a}_i$ is a vector of discrimination parameters,

d_i is a scalar parameter that is related to difficulty of the item, and

θ_j is a vector of ability parameters.

The most commonly used multidimensional item response theory models are multidimensional two- and three-parameter logistic models (M2PL and M3PL respectively). The only difference between the two models is that the M3PL model has a pseudo-guessing parameter. The problematic role of the pseudo-guessing parameter in recovering *a*- and *b*-parameters was investigated by several studies (Baker, 1987; Hulin, Lissak, & Drasgow, 1982; Kolen, 1981; Robert L. McKinley & Reckase, 1980; Thissen & Wainer, 1982). Despite all the difficulties estimating the *c*-parameter, whether or not to include *c*-parameter into the model is still being debated. Yen (1981) used the simulation study to show that the data set generated by a three-parameter model fits very well with a two-parameter model. R. L. Mckinley and Mills (1985) showed the same result as Yen (1981). Since this study does not focus on the model-fit, a two-parameter model will be used to investigate the effect of variation of a variety of factors.

3.2 Data Generation

Several factors are considered for generating true item parameters: 1) number of dimensions; 2) different types of latent traits configuration; 3) number of items; and 4) correlated latent traits

First, two different numbers of dimensions—3- and 6-dimensions—are considered. Reise et al. (2000) summarized that it would be better to overestimate the number of dimensions than to underestimate them, in order to accurately represent the major relationships in the item response data. Most previous studies in MIRT have had three dimensions, so this study expands the number of dimensions up to six dimensions.

Second, Thurstone (1947) suggested that the number of variables needed to run the analysis with *m* factors is two or three times greater. Holzinger and Harman (1941) gave a formula to estimate the required number of variables in *m* factor analysis:

$$n \geq \frac{(2m+1)+\sqrt{8m+1}}{2} \tag{3.2}$$

Since this study examines 3-6 dimensions, the minimum numbers needed are 6 items for 3 dimensions, 8 items for 4 dimensions, and 10 items for 5 dimensions. However, Thurstone (1947) also suggested having more than five- or six-times more items than *m* factors. In practice, most tests have composited with more than 50 items. For example, the ACT exam has 75 items for English, 60 items for Math, and 40 items for Science. For this study, 60 items will be used because the average number of items in the ACT is around 60, and this will allow an evenly distributed item number across dimensions. So in 3-dimensions, each dimension has 20 items. In 6-dimensions, each dimension has 10 items.

Third, two different types of latent trait structures will be examined—approximate simple structure and complex structure. Typically, items that lie within 20° of the x, y, or other axis (Froelich, 2001) are called an approximate simple structure. Items that lie within 40° of the x, y, or other axis are called a complex structure. Two different angles, 20 and 40, are used to generate two different dimension structures, the approximate simple structure and complex structure, respectively.

Fourth, in order to examine the correlated latent traits (only for the complex structure), correlations of 0, 0.3, and 0.6, are considered. Those correlations were selected to provide a broad range of potential conditions, from low to high. Previous studies covered a broad range of correlations. Walker et al. (2006) used 0.3, 0.6, and 0.9. Finch (2010) used 0, 0.3, 0.5, and 0.8. Tate (2003) used 0.6.

With specifications on each factor, the true item parameters in *k*-dimensions corresponding to each dimension are generated using the following equations:

$$a_k = \text{MDIS} * \text{Cos}(\alpha_k) \tag{3.3}$$

$$d = -\text{MDISC} * \text{MDIFF} \tag{3.4}$$

First, MDISCs and MDIFFs were randomly drawn from specific distributions: the former from lognormal distribution with $\mu = -0.15$ and $\sigma = 0.35$, which resulted in an MDISC mean of 0.92 and a standard deviation of 0.33; the latter was from a normal distribution with mean of 0 and standard deviation of 0.7. Second, directional angles, αs (i.e. angles between the item vector and ability axes), were generated using uniform distribution with a range specified by dimensional structure (i.e. either approximate simple structure or complex structure). Finally, item parameters were calculated from MDISC, MDIFF, and α values, using the above formulas. True item parameters for 3- and 6-dimensions with AS and MS structures are given in table 3.2.1, 3.2.2, and 3.2.3, respectively.

The M2PL model specified above is being used to generate the response dataset. Multivariate normal distribution with a mean of 0, and a covariance matrix based on the specification of correlations above, are used for ability distribution. More details about skewed multivariate normal distribution are given in section 3.2.1. In order to make interpretation simple for this study, it is assumed that all dimensions have the same distribution, with a mean of zero and a standard deviation of one, to effect identification. 10 replications of each test forms will be generated by using the specifications described above. The probability of a correct response for each examinee to each item is calculated using the M2PL model. If a random number drawn from a uniform distribution U(0,1) is less than the model-based probability, the item response is coded correct. Otherwise, it is coded wrong.

Table 3.2.1. True item parameters for 3-dimensions with AS and MS

	AS				MS			
Item	a_1	a_2	a_3	d	a_1	a_2	a_3	d
1	0.6020	0.1407	0.1385	0.7506	1.1640	0.6381	0.6250	2.1288
2	0.7295	0.2057	0.2030	0.7883	0.4481	0.2773	0.2723	0.7329
3	0.7213	0.1874	0.1848	0.7644	0.8766	0.4355	0.4255	1.1914
4	0.5356	0.1472	0.1452	0.5175	0.5042	0.2912	0.2855	0.3389
5	0.9726	0.2906	0.2871	0.9289	0.6011	0.3417	0.3349	0.3980
6	0.8741	0.2391	0.2359	0.7038	0.5029	0.3414	0.3357	0.2235
7	0.7222	0.1721	0.1694	0.4040	1.0557	0.5804	0.5685	0.3750
8	1.0943	0.3078	0.3037	-0.1223	0.9356	0.4877	0.4771	0.2451
9	1.4082	0.3596	0.3544	-0.2112	0.3315	0.1730	0.1692	0.0841
10	0.5542	0.1316	0.1296	-0.1009	0.5715	0.3527	0.3462	0.0672
11	0.7781	0.2492	0.2463	-0.1730	0.8342	0.4447	0.4353	-0.0475
12	0.5478	0.1437	0.1416	-0.1277	0.3442	0.2557	0.2518	-0.0616
13	1.0639	0.2491	0.2451	-0.3469	0.5922	0.3810	0.3743	-0.1198
14	0.9555	0.2656	0.2621	-0.5162	0.5573	0.3460	0.3397	-0.1700
15	0.6593	0.2031	0.2007	-0.3835	0.7906	0.4591	0.4502	-0.3780
16	0.6599	0.1826	0.1802	-0.4514	0.4770	0.2959	0.2905	-0.3180
17	0.6149	0.1736	0.1713	-0.5285	0.6218	0.4454	0.4384	-0.6780
18	1.1168	0.2937	0.2896	-1.1527	0.6143	0.2615	0.2544	-0.6647
19	0.7429	0.1899	0.1871	-0.8571	0.3699	0.2631	0.2590	-0.5872
20	0.7586	0.2159	0.2132	-1.5381	0.5889	0.3311	0.3244	-0.8884

Table 3.2.1 (cont'd)

Item	AS				MS			
	a_1	a_2	a_3	d	a_1	a_2	a_3	d
21	0.1721	0.6725	0.1813	0.8524	0.5397	0.7587	0.5164	2.2377
22	0.1754	0.6294	0.1839	0.4900	0.2355	0.3713	0.2241	0.5308
23	0.2008	0.7295	0.2107	0.5229	0.2308	0.4026	0.2184	0.3592
24	0.1333	0.5351	0.1406	0.3036	0.5580	0.7269	0.5356	0.7211
25	0.2229	0.8836	0.2350	0.3803	0.2857	0.4936	0.2705	0.3081
26	0.2168	0.8370	0.2282	0.3132	0.2523	0.3444	0.2417	0.1273
27	0.1878	0.8051	0.1989	0.2636	0.4029	0.5241	0.3867	0.0848
28	0.0919	0.3694	0.0970	0.1049	0.2958	0.6376	0.2759	0.0673
29	0.1713	0.6772	0.1806	0.1689	0.3286	0.3545	0.3174	0.0202
30	0.2030	0.8934	0.2153	0.2120	0.3054	0.4096	0.2927	-0.2209
31	0.2563	0.9089	0.2686	0.2164	0.4887	0.6831	0.4677	-0.3759
32	0.1957	0.8729	0.2078	-0.0192	0.1972	0.3526	0.1863	-0.1892
33	0.1899	0.8053	0.2009	-0.0235	0.6358	1.0602	0.6032	-0.6289
34	0.1045	0.4679	0.1109	-0.0803	0.4602	0.8952	0.4325	-0.6906
35	0.2565	0.8979	0.2686	-0.2928	0.3343	0.6409	0.3145	-0.6367
36	0.1843	0.7790	0.1950	-0.4160	0.3775	0.6870	0.3563	-0.7641
37	0.2035	0.9778	0.2172	-0.6982	0.3112	0.4629	0.2970	-0.6938
38	0.2747	0.9867	0.2881	-0.9177	0.3221	0.4046	0.3095	-0.8181
39	0.2282	0.7695	0.2386	-0.9011	0.5129	0.6889	0.4917	-1.5367
40	0.2465	0.8983	0.2587	-1.4694	0.5399	0.9294	0.5113	-1.8720
41	0.1498	0.1270	0.6262	0.9049	0.3150	0.3232	0.4964	1.1908
42	0.2540	0.2070	1.2731	1.5969	0.1904	0.1967	0.3764	0.5311
43	0.3129	0.2694	1.2082	1.1065	0.4930	0.5051	0.7298	0.9977
44	0.1911	0.1683	0.6405	0.5277	0.4709	0.4845	0.8213	1.0370
45	0.2758	0.2372	1.0717	0.8405	0.5749	0.5887	0.8336	1.0558
46	0.2268	0.1954	0.8709	0.5912	0.3476	0.3577	0.6133	0.5801
47	0.1811	0.1555	0.7099	0.4600	0.3760	0.3847	0.5225	0.3747
48	0.1573	0.1355	0.6031	0.3737	0.3956	0.4051	0.5758	0.3425
49	0.2482	0.2166	0.8838	0.3402	0.3766	0.3868	0.6125	0.3032
50	0.1114	0.0960	0.4269	0.0877	0.4140	0.4260	0.7238	0.2654

Table 3.2.1 (cont'd)

Item	AS				MS			
	a_1	a_2	a_3	d	a_1	a_2	a_3	d
51	0.1280	0.1115	0.4617	0.0557	0.7243	0.7436	1.1708	0.3762
52	0.4395	0.3812	1.6269	0.0449	0.4038	0.4153	0.6951	0.1581
53	0.2044	0.1760	0.7874	0.0033	0.4908	0.5046	0.8318	-0.1574
54	0.5334	0.4606	2.0224	-0.5475	0.8608	0.8825	1.3079	-0.2834
55	0.1719	0.1445	0.7516	-0.2949	0.3260	0.3361	0.6119	-0.1262
56	0.1997	0.1680	0.8692	-0.4760	0.4737	0.4885	0.8899	-0.5125
57	0.2855	0.2452	1.1175	-0.6721	0.4358	0.4479	0.7311	-0.5060
58	0.2011	0.1739	0.7563	-0.4799	0.2744	0.2829	0.5100	-0.5520
59	0.1574	0.1326	0.6811	-0.5184	0.1413	0.1463	0.2938	-0.3710
60	0.3683	0.3210	1.3208	-1.2562	0.1831	0.1876	0.2716	-0.8392

Table 3.2.2. True item parameters for 6-dimensions with AS structures

Item	a_1	a_2	a_3	a_4	a_5	a_6	d
1	0.9520	0.1509	0.1481	0.1235	0.1297	0.1437	1.8397
2	0.5181	0.0880	0.0865	0.0731	0.0764	0.0841	0.0031
3	0.9168	0.1827	0.1801	0.1564	0.1624	0.1759	-0.1967
4	0.9747	0.1658	0.1630	0.1378	0.1441	0.1585	-0.5534
5	0.6815	0.1202	0.1183	0.1007	0.1051	0.1151	-0.3904
6	0.8702	0.1735	0.1710	0.1485	0.1541	0.1670	-0.5404
7	0.6559	0.1229	0.1211	0.1041	0.1084	0.1180	-0.4196
8	0.8031	0.1656	0.1633	0.1425	0.1477	0.1596	-0.5817
9	0.9516	0.1559	0.1532	0.1285	0.1347	0.1487	-0.8593
10	1.2195	0.2386	0.2351	0.2036	0.2115	0.2294	-1.1938
11	0.1372	0.7207	0.1263	0.1405	0.1267	0.1179	1.1547
12	0.1130	0.6605	0.1030	0.1161	0.1034	0.0954	0.5270
13	0.1115	0.5830	0.1026	0.1142	0.1030	0.0959	0.3405
14	0.1461	0.8875	0.1327	0.1502	0.1332	0.1224	0.4820
15	0.0836	0.4798	0.0763	0.0858	0.0766	0.0708	0.2475
16	0.1617	0.8979	0.1481	0.1659	0.1486	0.1377	0.2265
17	0.0757	0.3922	0.0698	0.0775	0.0700	0.0652	-0.0450
18	0.1777	0.9654	0.1630	0.1822	0.1636	0.1519	-0.2064
19	0.1537	0.9154	0.1398	0.1580	0.1404	0.1292	-0.4622
20	0.1599	1.0718	0.1436	0.1648	0.1442	0.1312	-0.7069

Table 3.2.2 (cont'd)

Item	a_1	a_2	a_3	a_4	a_5	a_6	d
21	0.2030	0.2372	0.7504	0.2303	0.2581	0.2417	1.1806
22	0.2696	0.3050	0.7318	0.2979	0.3266	0.3096	1.1368
23	0.4085	0.4637	1.1559	0.4526	0.4974	0.4709	1.2722
24	0.4984	0.5586	1.2067	0.5465	0.5953	0.5664	1.0743
25	0.2642	0.3044	0.8661	0.2963	0.3289	0.3096	0.3194
26	0.1351	0.1543	0.4089	0.1505	0.1661	0.1569	0.1412
27	0.2346	0.2704	0.7712	0.2633	0.2923	0.2751	0.1326
28	0.1911	0.2161	0.5173	0.2111	0.2314	0.2194	-0.0865
29	0.1669	0.1879	0.4270	0.1837	0.2006	0.1906	-0.0959
30	0.2918	0.3309	0.8164	0.3230	0.3548	0.3360	-0.3722
31	0.1242	0.1303	0.1223	0.3513	0.1167	0.1086	0.5930
32	0.2778	0.2884	0.2744	0.5595	0.2647	0.2505	0.8166
33	0.1562	0.1636	0.1539	0.4213	0.1471	0.1373	0.3586
34	0.1755	0.1850	0.1724	0.5616	0.1637	0.1510	0.3673
35	0.1793	0.1878	0.1766	0.4892	0.1688	0.1574	0.0287
36	0.2276	0.2387	0.2241	0.6449	0.2139	0.1989	0.0046
37	0.3478	0.3639	0.3426	0.9250	0.3278	0.3061	-0.0168
38	0.4200	0.4400	0.4135	1.1599	0.3951	0.3681	-0.1859
39	0.2725	0.2862	0.2680	0.8027	0.2554	0.2370	-0.1443
40	0.1766	0.1865	0.1734	0.5929	0.1643	0.1509	-0.3328
41	0.3069	0.3063	0.3151	0.2989	0.7750	0.2895	1.3252
42	0.3111	0.3106	0.3189	0.3036	0.7132	0.2948	1.0745
43	0.3164	0.3158	0.3268	0.3064	1.0268	0.2945	0.6668
44	0.2639	0.2634	0.2722	0.2559	0.8103	0.2465	0.2626
45	0.3705	0.3698	0.3816	0.3598	1.0755	0.3471	0.1008
46	0.2365	0.2361	0.2430	0.2303	0.6116	0.2229	-0.0393
47	0.2883	0.2879	0.2953	0.2815	0.6329	0.2736	-0.1310
48	0.2275	0.2271	0.2327	0.2224	0.4598	0.2165	-0.3131
49	0.2023	0.2019	0.2075	0.1972	0.4829	0.1913	-0.3752
50	0.2153	0.2150	0.2208	0.2100	0.5090	0.2037	-0.8940

Table 3.2.2 (cont'd)

Item	a_1	a_2	a_3	a_4	a_5	a_6	d
51	0.0925	0.0843	0.0961	0.0933	0.0862	0.4523	0.4311
52	0.1231	0.1118	0.1280	0.1241	0.1145	0.6205	0.5341
53	0.1036	0.0936	0.1078	0.1045	0.0959	0.5480	0.2823
54	0.2041	0.1833	0.2131	0.2060	0.1882	1.1438	0.5113
55	0.2574	0.2305	0.2690	0.2599	0.2368	1.4803	0.5318
56	0.1547	0.1355	0.1629	0.1564	0.1400	1.0549	0.2146
57	0.1035	0.0920	0.1083	0.1045	0.0947	0.6258	-0.0509
58	0.1510	0.1351	0.1579	0.1525	0.1388	0.8765	-0.2256
59	0.1318	0.1182	0.1376	0.1330	0.1214	0.7458	-0.3648
60	0.0942	0.0831	0.0989	0.0952	0.0857	0.6073	-0.4395

Table 3.2.3. True item parameters *for* 6-dimensions with MS

Item	a_1	a_2	a_3	a_4	a_5	a_6	d
1	0.5618	0.2124	0.2371	0.2319	0.2238	0.2178	0.9419
2	0.6429	0.2158	0.2433	0.2375	0.2284	0.2218	0.7950
3	0.2921	0.1039	0.1166	0.1139	0.1098	0.1067	0.3548
4	0.4081	0.1479	0.1657	0.1620	0.1561	0.1518	0.2712
5	0.3061	0.1455	0.1599	0.1569	0.1521	0.1487	0.2168
6	0.9760	0.3461	0.3883	0.3794	0.3655	0.3553	0.5893
7	0.5915	0.2430	0.2695	0.2639	0.2552	0.2488	-0.1134
8	0.9778	0.3464	0.3887	0.3798	0.3658	0.3556	-0.1765
9	0.8383	0.2666	0.3022	0.2947	0.2829	0.2743	-0.2088
10	0.7652	0.2870	0.3205	0.3134	0.3024	0.2942	-1.0466
11	0.1760	0.4991	0.1674	0.1632	0.1690	0.1680	0.4335
12	0.2174	0.5049	0.2083	0.2038	0.2099	0.2089	0.3587
13	0.2892	0.8794	0.2742	0.2669	0.2769	0.2752	0.4488
14	0.1517	0.4334	0.1442	0.1406	0.1456	0.1447	0.0275
15	0.3612	0.9773	0.3442	0.3359	0.3473	0.3453	0.0268
16	0.3578	0.8802	0.3422	0.3345	0.3451	0.3432	-0.0135
17	0.2362	0.4645	0.2274	0.2230	0.2290	0.2279	-0.0619
18	0.3304	0.8253	0.3158	0.3087	0.3185	0.3168	-0.6525
19	0.3958	1.0617	0.3773	0.3682	0.3807	0.3785	-0.9367
20	0.3017	0.7136	0.2889	0.2826	0.2912	0.2897	-0.9932

Table 3.2.3 (cont'd)

Item	a_1	a_2	a_3	a_4	a_5	a_6	d
21	0.2030	0.2372	0.7504	0.2303	0.2581	0.2417	1.1806
22	0.2696	0.3050	0.7318	0.2979	0.3266	0.3096	1.1368
23	0.4085	0.4637	1.1559	0.4526	0.4974	0.4709	1.2722
24	0.4984	0.5586	1.2067	0.5465	0.5953	0.5664	1.0743
25	0.2642	0.3044	0.8661	0.2963	0.3289	0.3096	0.3194
26	0.1351	0.1543	0.4089	0.1505	0.1661	0.1569	0.1412
27	0.2346	0.2704	0.7712	0.2633	0.2923	0.2751	0.1326
28	0.1911	0.2161	0.5173	0.2111	0.2314	0.2194	-0.0865
29	0.1669	0.1879	0.4270	0.1837	0.2006	0.1906	-0.0959
30	0.2918	0.3309	0.8164	0.3230	0.3548	0.3360	-0.3722
31	0.1242	0.1303	0.1223	0.3513	0.1167	0.1086	0.5930
32	0.2778	0.2884	0.2744	0.5595	0.2647	0.2505	0.8166
33	0.1562	0.1636	0.1539	0.4213	0.1471	0.1373	0.3586
34	0.1755	0.1850	0.1724	0.5616	0.1637	0.1510	0.3673
35	0.1793	0.1878	0.1766	0.4892	0.1688	0.1574	0.0287
36	0.2276	0.2387	0.2241	0.6449	0.2139	0.1989	0.0046
37	0.3478	0.3639	0.3426	0.9250	0.3278	0.3061	-0.0168
38	0.4200	0.4400	0.4135	1.1599	0.3951	0.3681	-0.1859
39	0.2725	0.2862	0.2680	0.8027	0.2554	0.2370	-0.1443
40	0.1766	0.1865	0.1734	0.5929	0.1643	0.1509	-0.3328
41	0.3069	0.3063	0.3151	0.2989	0.7750	0.2895	1.3252
42	0.3111	0.3106	0.3189	0.3036	0.7132	0.2948	1.0745
43	0.3164	0.3158	0.3268	0.3064	1.0268	0.2945	0.6668
44	0.2639	0.2634	0.2722	0.2559	0.8103	0.2465	0.2626
45	0.3705	0.3698	0.3816	0.3598	1.0755	0.3471	0.1008
46	0.2365	0.2361	0.2430	0.2303	0.6116	0.2229	-0.0393
47	0.2883	0.2879	0.2953	0.2815	0.6329	0.2736	-0.1310
48	0.2275	0.2271	0.2327	0.2224	0.4598	0.2165	-0.3131
49	0.2023	0.2019	0.2075	0.1972	0.4829	0.1913	-0.3752
50	0.2153	0.2150	0.2208	0.2100	0.5090	0.2037	-0.8940

Table 3.2.3 (cont'd)

Item	a_1	a_2	a_3	a_4	a_5	a_6	d
51	0.4085	0.4467	0.4182	0.4495	0.4382	0.9276	2.2550
52	0.2182	0.2408	0.2239	0.2424	0.2358	0.5720	1.0999
53	0.2610	0.2877	0.2677	0.2897	0.2818	0.6773	0.7342
54	0.2968	0.3279	0.3047	0.3302	0.3210	0.7924	0.8207
55	0.1851	0.2060	0.1904	0.2076	0.2014	0.5460	0.1620
56	0.2181	0.2464	0.2253	0.2485	0.2402	0.7587	-0.1287
57	0.3034	0.3305	0.3102	0.3325	0.3245	0.6449	-0.1592
58	0.2349	0.2665	0.2429	0.2688	0.2595	0.8504	-0.4674
59	0.1894	0.2096	0.1946	0.2111	0.2052	0.5170	-0.4735
60	0.3743	0.4144	0.3844	0.4174	0.4055	1.0314	-1.5128

3.2.1 Skewed Multivariate Normal Distribution

To generate the multivariate skew normal distribution, this study follows the proposal of alternative parameterization defined by Arellano-Valle and Azzalini (2008). The *d*-dimensional skew normal density function is defined as follows:

$$f_d(x; \xi, \Omega, \alpha) = 2\phi_d(x - \xi; \Omega)\, \Phi\left\{\alpha^T \omega^{-1}(x - \xi)\right\}, \qquad \left(x \in R^d\right) \tag{3.5}$$

where $\phi_d(x; \Omega)$ is the $N_d(0, \Omega)$ density function for *d* x *d* positive definite symmetric matrix Ω, ξ is a vector location parameter, α is a vector shape parameter $\left(\xi,\ \alpha \in R^d\right)$, and ω is a diagonal matrix formed by the standard deviation of Ω.

This procedure uses centered parameters (CP), which are transformed from direct parameters (DP) such as mean, covariance matrix, and skewness. Under a certain choice of $CP(\mu,\ \Sigma,\ \Upsilon)$ that belongs to the admissible CP sets corresponds to $DP(\xi,\ \Omega,\ \alpha)$. The CP is defined as:

$$\mu = E(Y) = \xi + \omega\mu_Z$$

$$\Sigma = var(Y) = \Omega - \omega\mu_Z\mu_Z^T\omega = \omega\Sigma_Z\omega$$

After some algebra, DP is calculated from CP as:

$$\xi = \mu - \sigma\sigma_Z^{-1}\mu_Z, \quad \omega = \sigma\sigma_Z^{-1}, \quad \Omega = \Sigma + \omega\mu_Z\mu_Z^T\omega \quad ,$$

where

$$\mu_Z = \frac{c}{\sqrt{1+c^2}}, \quad c = \left(\frac{2Y}{4-\pi}\right)^{1/3}$$

More detail about re-parameterization between CP and DP can be found in the study by Arellano-Valle and Azzalini (2008).

3.3 MCMC Simulation

MCMC methods have become a familiar method for estimating the parameters of complex statistical models with the rapid decrease in computing costs. Despite the fact that MCMC methods can be implemented for complicated statistical models, the basic idea underlying MCMC methods is extremely simple. The main idea is to sample from the posterior distribution (e.g. target distribution), and use those samples to make inferences about parameters of interest. Suppose we have a joint distribution $p(\theta, \beta)$, where θ is the ability parameter and β is a vector of item parameters. Ultimately, our goal is to find the joint posterior distribution, such as $p(\theta, \beta|X) \propto p(X|\theta, \beta) * p(\theta, \beta)$. In order to find such a joint distribution, we run the Markov Chain with a transition kernel, the probability of moving to a new state $\left(\theta^{k+1}, \beta^{k+1}\right)$, given the current state of the chain $\left(\theta^{k}, \beta^{k}\right)$. There are two well-known transition kernels, Gibbs sampling (Geman & Geman, 1984) and Metropolis-Hasting sampling (Hastings, 1970; Metropolis et al., 1953).

Gibbs sampling uses the full conditional distribution to sample from the sequence of parameters. Suppose that we have a vector of parameter $\Theta(\theta_1,\theta_2,\ldots,\theta_k)$. Then the Gibbs sampling algorithm is defined as follows:

1. Set the starting values for the vector of parameter
2. Set $j = j + 1$
3. Sample $\left(\theta_1^j \middle| \theta_2^{j-1}, \theta_3^{j-1}, \cdots, \theta_k^{j-1}\right)$.
4. Sample $\left(\theta_2^j \middle| \theta_1^{j}, \theta_3^{j-1}, \cdots, \theta_k^{j-1}\right)$.

$\vdots$

k. Sample $\left(\theta_k^j \middle| \theta_1^{j}, \theta_2^{j}, \cdots, \theta_{k-1}^{j}\right)$.

k+1. Return to step 2 and repeat until convergence.

Metropolis-Hasting sampling is used when it is difficult to use the full conditional distribution. The Metropolis-Hasting sampling algorithm is defined as follows:

1. Establish the starting value for parameter θ.
2. Specify a proposal density $r(\theta^j, \theta^{(j+1)})$, which defines the proposal density from the current state θ^j to the next state $\theta^{(j+1)}$.
3. Given the current state θ^j, draw the candidate parameter θ^* from the proposal density.
4. Compute probability

$$\alpha\left(\theta^j,\theta^*\right)=\min\left\{1, \frac{g(\theta^*)r(\theta^*, \theta^j)}{g(\theta^j)r(\theta^j, \theta^*)}\right\} ,$$

   where $g(.)$ is the density of the target distribution.

5. Compare $\alpha\left(\theta^j,\theta^*\right)$ with a U(0,1) random draw u. If $\alpha\left(\theta^j,\theta^*\right) > u$, then set $\theta^j = \theta^*$.

Otherwise, set $\theta^{j+1} = \theta^{j}$.

In this study, the Gibbs sampling built in the OpenBUGS program (v. 3.11) is used to run the simulation. The High Performance Computing Center (HPCC) at Michigan State University was used to run the simulation. HPCC provides seven clusters, which are composed with various numbers of nodes. The system runs on Red Hat Enterprise Linux 6.1. HPCC also provides various computational software, such as OpenBUGS, Matlab, and R (HPCC, 2012).

In OpenBUGS, there are several sampling algorithms provided and systematically implemented into the program. Once OpenBUGS starts to run the simulation, then the sampling algorithm stored in the program is automatically loaded and the most appropriate sampling algorithm is identified. These sampling algorithms include a proposal distribution with normal distribution, univariate distribution, and multivariate normal distribution.

Once the sampling algorithm and prior distribution are specified, there are a number of decisions that need to be made to make sure that the inference from the MCMC result is meaningful and useful. First, the initial values should be set to run the simulation. Second, is deciding whether multiple chains or a single long chain should be run. Third, the length of iterations needs to be set. Fourth, the convergence of the Markov chain needs to be diagnosed to make sure it has reached the target distribution (i.e. stationary distribution).

Brooks (1998) showed that starting values do not have a serious impact on any inference from MCMC because the sample used to make inferences is chosen after the chain has reached a stable stage (i.e. stationary distribution). However, the choice of starting values may affect the performance of the Markov chain. Several methods have been suggested by researchers. Andrew Gelman and Rubin (1992) suggested starting values be sampled from a high density distribution of a mixture t-distribution, which is called a simple mode-finding algorithm. Brooks and Morgan (1994) suggested the use of an annealing algorithm to sample initial values.

Once starting values are set, the next step is to decide whether to run multiple Markov chains in parallel or a single long chain. Geyer (1992) suggested a long single chain because running multiple chains does not guarantee that each short chain will have reached a stationary distribution. Even though multiple chains give a diagnostic value about the length of iterations, that inference is not valid if multiple chains do not give an agreeable result; the other side of this argument is that this result of agreement does not confirm that each multiple chain has been reached at the stationary distribution. Geyer (1992) also argued that one very long run could give a valuable diagnostic on the convergence of Markov chains. If the run does not seem to reach stationary distribution, then it is too short and a longer chain needs to be run. Even though multiple chains have a small advantage in the diagnostic of convergence, this small advantage could not be critical because a single long run could also provide a useful diagnostic value.

The next step is to set up the length of burn-in (i.e. warm-up). Several researches have shown the formal analysis to calculate how many iterations should be thrown away (Kelton & Law, 1984; Raftery & Lewis, 1992; Ripley & Kirkland, 1990). However, Geyer (1992) argued that this formal analysis does not seem necessary in a practical running of Markov chains. He suggested that it would suffice to throw away 1 or 2% of a run. More can be thrown away later if the auto-covariance calculations or time-series plots showed the slow of mixing.

3.4 Accessing Convergence of MCMC Simulation

The crucial part of using MCMC in parameter estimation is to assess how well the MCMC algorithm (e.g. Gibbs sampling and Metropolis-Hasting sampling) performs. Without evidence of having reached the target distribution (i.e. stationary distribution), the inferences made from the MCMC method should be questioned. Several studies suggest different diagnostic methods for convergence (Andrew Gelman & Rubin, 1992; Geweke, 1992; Heidelberger & Welch, 1983; Raftery & Lewis, 1992).

Andrew Gelman and Rubin (1992) used multiple sequences of chains to estimate the variance, called the potential scale reduction factor (PSRF). See Gelman and Rubin (1992) for details. If the value of the PSRF is large, then the convegerence of the Markov chain is suspicious and more iterations need to be run. If the value of PSRF is close to 1, then the Markov chain is close to stationary distribution. Geweke (1992) used the spectral density to estimate the variance from a single long chain. The basic idea is that there are no discontinuties at frequency zero for the times series of spectral density. The diagnostics are performed by dividing the iterations into two parts, the first part (10% and more) and the last part (50% and more), and taking the difference of the means of each part and dividing it by its standard error, which becomes the Z-score test statistic. See Geweke (1992) for technical details. Heidelberger and Welch (1983) developed a method that was initially used to estimate the length of the iteration. However, it is feasable for checking the diagnostics of convergence in a Gibbs simulation. The basic idea is to use the test statistics, which are based on the Cramer-von statistics, in the sequence of iterations. The first step is to set up the initial check-point and estimate the confidential interval, if it passes the testing procedure. This step is repeated until either the iteration reaches the end or a confidential interval meets the accuracy criteria. For more technical details, the reader is encouraged to see their paper. In this study, Heidelberger's and Welch's diagnostic is used to to diagnose the convergence of Gibbs sampling.

Cowles and Carlin (1996) reviewed several MCMC convergence diagnotic procedures, and showed that none of the procedures is superior to other. They also suggested using multiple procedures to diagnose the convergence because each procedure has its own properties for accessing the convergence. In this study, since a single long chain was used, both the Geweke (1992) and Heidelberger and Welch (1983) were used to examinee the convergence of iterations.

In addition to this, graphical methods will be used to access the convergence of the chain, such as time-series, autocorrelation function, and posterior density plots.

3.5 Prior Distributions and Likelihood Functions

3.5.1 Prior Distributions

Suppose that the probability of a correct response for examinee *j* on item *i* is given as follows:

$$P_i(\boldsymbol{\theta}_j) = P(X_{ij}=1 \mid \boldsymbol{\theta}_j, \mathbf{a}_i, d_i) = \frac{\exp[\,\mathbf{a}_i\boldsymbol{\theta}_j + d_i]}{1+\exp[\,\mathbf{a}_i\boldsymbol{\theta}_j + d_i]} \tag{3.6}$$

and suppose that the M2PL model holds local independence assumption. Then the joint probability of a response is given as follows:

$$P[\mathbf{X} \mid \boldsymbol{\Theta}, \boldsymbol{\Sigma}, \mathbf{a}, \mathbf{d}] = \prod_{j=1}^{N}\prod_{i=1}^{n}\left[P_i(\boldsymbol{\theta}_j)\right]^{x_{ij}}\left[1-P_i(\boldsymbol{\theta}_j)\right]^{1-x_{ij}} \tag{3.7}$$

where, X_{ij} is the observed response of *j*th examinee on *i*th item, $\boldsymbol{\Theta}$ is N x *k* vector of ability. *k* is the number of dimensions in the M2PL model.

This study assumes that the ability parameter is considered to be mutually independent, and to follow the multivariate normal with mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ (Béguin &

Glas, 2001; Bolt & Lall, 2003). Then prior distribution $\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}_j)$ for ability distribution is specified as:

$$\pi_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_j\right)=\mathrm{N}_k\left(\boldsymbol{\theta}_j \middle| \boldsymbol{\mu}, \Sigma^2\right)=(2\pi)^{-\frac{k}{2}}|\Sigma|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\boldsymbol{\theta}_j-\boldsymbol{\mu}\right)'\Sigma^{-1}\left(\boldsymbol{\theta}_j-\boldsymbol{\mu}\right)\right] \tag{3.8}$$

where, $\mu_{\boldsymbol{\theta}}=[0_1, \ldots, 0_k]$,

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}=\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}$$

.

The mean vector follow all zeros, and the covariance matrix $\boldsymbol{\Sigma}$ is equal to identity matrix. The reason for using the identity matrix as the covariance in the prior multivariate normal distribution is that the purpose of this study is to investigate the mis-specified model effect in item calibration. In practical situations, it is nearly impossible to know the structure of the latent constructs. So, most item parameter calibration procedures use the standard covariance matrix as prior distribution. By using the standard covariance matrix as the prior distribution, it is possible to see the effect of the mis-specification effect in item calibration procedures.

Prior distribution for discrimination parameter a_{ik} takes normal distribution as $a_{ik} \sim \mathrm{N}_a(\mu,\sigma^2)\mathrm{I}(a_{ik}>0)$, where $I(.)$ is an indicator function used to make sure that the a_{ik} parameter is sampled from a positive area of normal distribution. Fu et al. (2009) used and showed that the truncated normal distribution for a-parameter is the appropriate choice, and works very well in an MCMC simulation setting as prior distribution. Mean 0 and variance 2 are used as the hyper-parameter. Béguin and Glas (2001) also used mean 0 and variance 1, and showed that these prior specifications worked very well in item parameter calibration in an MCMC simulation. So, while the specification for this study is less informative than their specification, Bolt and Lall (2003) used only slightly more informative priors (mean=0, variance=2). They also showed this prior distribution works very well in an MCMC simulation for MIRT framework.

Prior distribution for difficulty parameter d_i takes normal distribution $d_i \sim \mathrm{N}_d(\mu,\sigma^2)$. Mean 0 and variance 20 are used as prior distribution for d-parameter. This prior is a non-informative prior. Most of the previous studies used a broad range of variance, such as from 1 to 5 (Baker & Kim, 2004; Bolt & Lall, 2003; Finch, 2011; Harwell et al., 1996; Maydeu-Olivares, 2001; Y. Sheng, 2008). However, the large variance, 20, is used for this study because the purpose of this study is to evaluate the estimates, given incorrect assumptions about the latent structures. In order to evaluate the estimates, it is necessary to not rely on the prior specifications. Using a noninformative prior would enable the estimation to rely more heavily on the data rather than the prior (Baker & Kim, 2004).

For a- and d-parameters, this study uses a univariate normal distribution. This is very common practice in an MCMC simulation study in the MIRT framework. e.g. (Baker & Kim, 2004; Bolt & Lall, 2003; Fu et al., 2009).

3.5.2 Likelihood Functions

The overall posterior for item and ability parameters distribution is expressed in the following manner. Let us assume that MIRT model $\mathrm{P}(\mathrm{X}_{\mathrm{ij}})$ holds the local independence assumption, then the likelihood function of the M2PL models with response data X_{ij}, $N \times n$ matrix, can be expressed as:

$$L\left(X_{ij}|\,\boldsymbol{\Theta},\boldsymbol{\Sigma},\mathbf{A},d\right) = \prod_{j=1}^{N}\prod_{i=1}^{n} P_i\left(\boldsymbol{\theta_j}\right)^{x_{ij}} * Q_i\left(\boldsymbol{\theta_j}\right)^{1-x_{ij}} \tag{3.9}$$

where

X_{ij} is the response of *j*th examinee on item *i*,

$\boldsymbol{\Theta}$ is *N* x *k* vector of ability,

$\boldsymbol{A}$ is $n \times p$ item discrimination vector, and

d is $n \times 1$ item difficulty vector.

Since we define the prior distributions for proficiency and the item parameters as multivariate normal distribution $\pi_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_j, \mathbf{A}, d\right)$, then the full joint posterior distribution is given,

$$p\left(\boldsymbol{\theta_j}, \mathbf{A}, d \,\middle|\, \boldsymbol{\Theta}, \boldsymbol{\Sigma}, \mathbf{A}, d\right) \propto L\left(X_{ij}|\, \boldsymbol{\Theta}, \boldsymbol{\Sigma}, \mathbf{A}, d\right)\ \pi_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_j, \mathbf{A}, d\right)$$

For the proficiency, the joint posterior distribution could be obtained as;

$$p\left(\boldsymbol{\theta_j}\middle|\boldsymbol{\Theta}, \boldsymbol{\Sigma}, \mathbf{A}, d, X\right) \propto \prod_{i=1}^{n}\left[P_i\left(\boldsymbol{\theta}_j\right)^{u_{ij}} Q_i\left(\boldsymbol{\theta}_j\right)^{1-u_{ij}} * \pi_\theta\left(\boldsymbol{\theta_j}\right)\right]$$

where

$$\pi_\theta\left(\boldsymbol{\theta_j}\right) = (2\pi)^{-\frac{k}{2}}|\Sigma|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\boldsymbol{\theta}_j-\boldsymbol{\mu}\right)' \Sigma^{-1}\left(\boldsymbol{\theta}_j-\boldsymbol{\mu}\right)\right]$$

For the item discrimination parameter, the joint posterior distribution can be expressed as,

$$p\left(\boldsymbol{a_i}\middle|\boldsymbol{\Theta}, \boldsymbol{\Sigma}, d, X\right) \propto \prod_{i=1}^{n}\left[P_i\left(\boldsymbol{\theta}_j\right)^{u_{ij}} Q_i\left(\boldsymbol{\theta}_j\right)^{1-u_{ij}} * \pi_a\left(\boldsymbol{a}_i\right)\right]$$

where

$$\pi_a(\boldsymbol{a}_i) = (2\pi)^{-\frac{k}{2}}|\Sigma|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{a}_i-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{a}_i-\boldsymbol{\mu})\right]$$

For the difficulty parameter, the joint posterior distribution can be expressed as,

$$p(d_i|\boldsymbol{\Theta}, \boldsymbol{\Sigma}, \mathbf{A}, \mathrm{X}) \propto \prod_{i=1}^{n}\left[\mathrm{P_i}\left(\boldsymbol{\theta}_j\right)^{u_{ij}}\mathrm{Q_i}\left(\boldsymbol{\theta}_j\right)^{1-u_{ij}} * \pi_d(d_i)\right]$$

where

$$\pi_d(d_i) = (2\pi)^{-\frac{k}{2}}|\Sigma|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(d_i-\boldsymbol{\mu})'\Sigma^{-1}(d_i-\boldsymbol{\mu})\right]$$

3.6 Evaluation Criteria for Simulation Results

The inference from a Bayesian framework is different from the perspective of frequentist statistics. The inference of frequentist statistics is based on the sampling distribution of a population parameter, so that the summary statistics could be used to make inference. The inference from a Bayesian framework is based on the posterior distribution, which is the combined knowledge of prior distribution and likelihood function, views the data as fixed, and generates a distribution for population parameters. The evaluation of frequentist estimates is often based on its performance as measured by characteristics such as bias and mean square error. While the Bayesian paradigm does not use these measures because of their underlying view of the nature of a population parameter, these characteristics will also be considered in this study. Although it is acknowledged that the application of these measures conflicts with the Bayesian paradigm, it is constructive to consider them in addition to the Bayesian measures as a means of providing evaluation measures that might be more familiar to most researchers.

Three summary statistics are used to evaluate the simulation results: BIAS, Mean Absolute Deviation (MAD), and Root Mean Squared Error (RMSE). The magnitude of those criteria is compared based on the factors implemented into the MIRT model. For example, to evaluate the a_1- parameter through the simulation, those criteria are computed by

$$\text{BIAS}=\sum_{i=1}^{I}\frac{(\hat{a}_{1i}\ -\ a_{1i})}{I} \tag{3.10}$$

$$\text{MAD}=\sum_{i=1}^{I}\frac{|\hat{a}_{1i}\ -\ a_{1i}|}{I} \quad 3.11)$$

$$\text{MSE}=\sqrt{\sum_{i=1}^{I}\frac{(\hat{a}_{1i}\ -\ a_{1i})^2}{I}} \tag{3.12}$$

where a_{1i} is the true value of a_1 for item i, $\hat{a}_{1i}$ is the corresponding estimate, and I is number of items. The reported criteria are the averaged value across replications for each parameter.

CHAPTER 4

RESULTS

This section covers the simulation results, including convergence diagnostics for the MCMC chains, and item parameter recovery in the MIRT model when different factors are imposed into the model, such as the number of dimensions (3 and 6), types of latent traits configuration (AS and MS), correlated latent structures (.3 and .6), and skewed latent traits distributions (-.9 for the negative and +.9 for the positive skew) with different sample sizes (1000, 1500, 2000, and 3000). Each combination of simulation has 10 replications, which make a total of 840 sets of simulations.

4.1. Convergence Diagnostic

Patz and Junker (1999a) emphasized that two things must be determined. One is that the MCMC has to reach into stationary distribution, meaning that it has converged. The other one is that the MCMC standard error associated with the point estimates should be small. In order to examine the convergence of the MCMC chains, two convergence diagnostic procedures were used, the Heidelberger and Welch diagnostic, and Geweke. Graphical diagnostics, including autocorrelation, trace, and posterior density plots, are also conducted. Due to the large number of item parameters from the different sets of simulations, only a selected set of items is presented to show the convergence of the MCMC chain. The set of items used to show the convergence was chosen to show the worst convergence from all of the sets of simulations. Therefore, it is assumed that the rest of items are naturally in the acceptable range of convergence. In order to estimate the standard error of point estimates, which is the mean for this study, the batch standard error built-in function in CODA is used.

4.1.1. Heidelberger and Welch diagnostic

Heidelberger and Welch (1983) is one of the MCMC convergence diagnostic methods implemented into the CODA package in R program. The technical details of the method are covered in the research design section. The set of simulations chosen is the worst scenario case: 6 dimensions, mixed structures, correlated latent structures, and skewed latent structures. Among 60 items across 6 *a*-parameters, only a few items do not pass the convergence test, which requires an extended length of chain. All items for the *d*-parameter passed the convergence test. Even though extended length of chain should improve the convergence, 25,000 iterations with 3,000 burn-ins are enough to have a satisfactory convergence. See the tables in the appendix for a full description of which items did not pass the convergence test.

4.1.2. Geweke diagnostics

In addition to Heidelberger's and Welch's convergence diagnostic, Geweke's convergence is used to access the MCMC convergence. The same item parameter as used for Heidelberger's and Welch's test is used for Geweke's convergence test. If $|Z| < 2$, then it is assumed that the iteration reached stationary distribution and has met the convergence. See Table 4.1.1. The Z-score for the first 20 items are shown. Even though a few of the items do not seem to reach to convergence, the Z-score for most of the items is less than 2, and it confirms that the MCMC iteration has reach to convergence. The Z-score for the 60 items is given in the appendix.

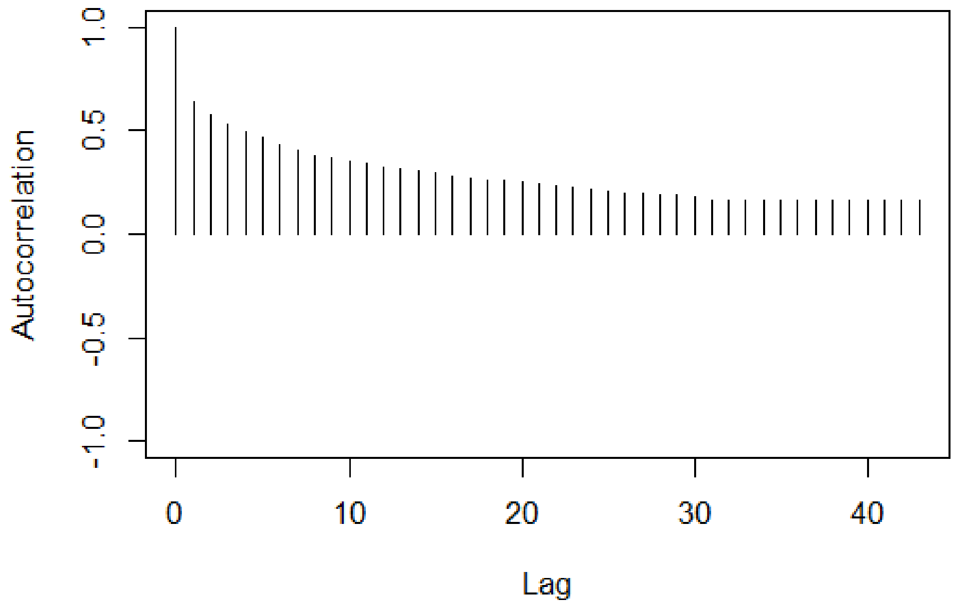
Table 4.1.1: Geweke's Z-score

Item	a_1	a_2	a_3	a_4	a_5	a_6	d
1	1.2684	-2.2061	0.5387	0.2543	0.7505	1.6152	0.5404
2	0.8198	-1.1491	1.3394	0.6109	-1.4409	0.4359	-0.0797
3	1.1488	-2.3534	0.8052	1.5895	-1.4047	1.8357	-0.2742
4	1.3278	-2.5092	0.7329	0.5589	-0.1756	2.0876	0.5249
5	0.7640	-3.5548	0.8827	0.7436	0.0197	2.3901	0.0838
6	2.3268	-1.6575	0.6066	-0.1853	-1.6516	1.2716	-0.0478
7	1.5009	-1.9750	1.2068	0.5158	-1.5500	2.3976	-0.6697
8	1.3062	-3.2312	0.3156	0.5285	-2.9689	2.1849	1.1820
9	0.8921	-2.0617	0.3080	-0.3810	-0.3362	3.0078	0.1300
10	1.2711	-3.6032	0.6145	0.7171	0.4523	2.9258	0.0248
11	-0.4740	-0.8094	2.1629	1.7458	-1.2479	-0.1922	1.0436
12	-0.3858	0.2128	1.1112	-0.1536	-0.3504	-0.3221	-0.6892
13	0.2475	1.7524	-0.1811	-0.7205	-0.1001	0.5925	-0.7560
14	0.3892	-0.0242	-2.0751	1.2241	-2.3642	0.3590	-1.0243
15	-0.9393	0.9163	-1.2904	1.1393	-1.2072	-0.1175	-0.4807
16	1.0436	1.3765	-1.9252	1.3216	-2.5827	0.1225	-0.9919
17	0.5820	0.3020	-1.4122	0.5939	-1.1244	0.3885	-0.6820
18	0.5422	0.7274	1.1886	0.1160	-2.4029	0.3778	-2.1434
19	0.4158	0.6139	-1.9429	1.0971	-2.1644	0.1458	0.2930
20	-0.3669	0.7525	1.5577	0.9327	-1.5352	-0.4552	-1.8313

4.1.3. Graphical diagnostics: Autocorrelation, posterior density, and trace plots

In addition to the two convergence diagnostics, Geweke, and Heidelberger and Welch, three graphical diagnostic plots are used to show the convergence of the MCMC iteration. Due to the large number of item parameters, only some selected items' autocorrelation plot, trace plot and density plot are presented. Figure 4.1 shows the autocorrelation plot for a_1-parameter of item number 6 in 6-dimension with MS structures. It shows that the MCMC iteration reaches the stationary distribution.

Figure 4.1. Autocorrelation plot for a_1 of item 6



Figures 4.2 and 4.3 show the trace and posterior density plot, respectively, for the same item in the autocorrelation plot. It does not look like MCMC has a good convergence. However, it does take a sample from the range that is close to the true item parameter (a_1 = .9760). If MCMC chains have more iterations, the convergence will improve. Since the autocorrelation plot shows a promising convergence, 25000 iterations with 3000 burn-ins should be enough to get a satisfactory convergence.

Figure 4.2. Trace plot for a_1 of item 6

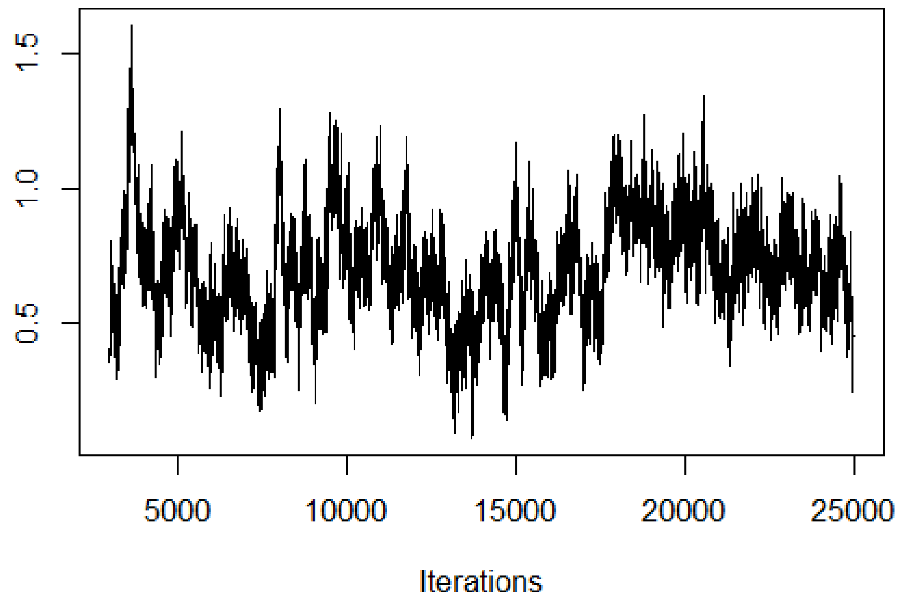
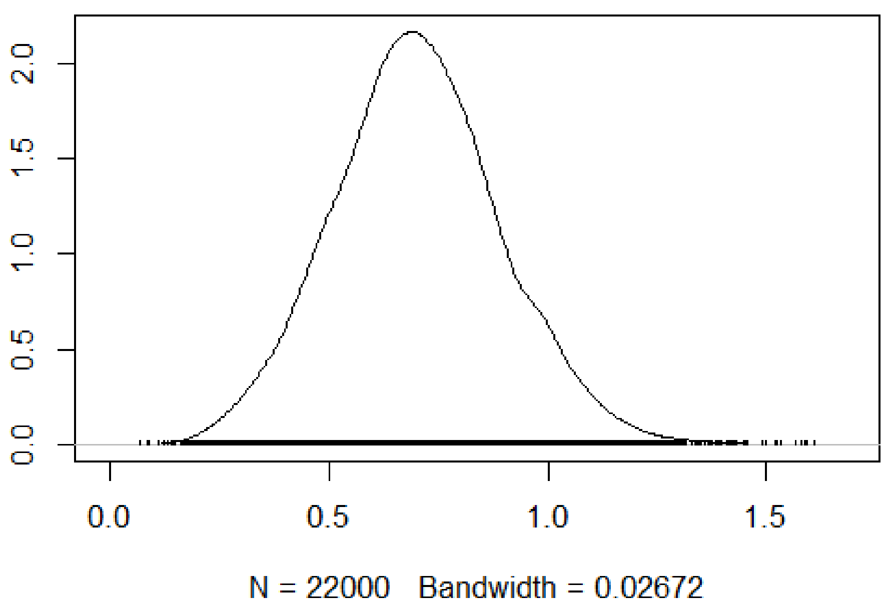


Figure 4.3. Posterior density plot for a_1 of item 6



4.1.4. MCMC standard error

As specified in the beginning of this chapter, MCMC simulation requires two diagnostics, a convergence test and MCMC standard error. While the convergence test shows how well the MCMC simulation has reached into stationary distribution, an MCMC standard error shows whether the number of iterations used in the simulation are enough or not. Table 4.1.2 shows the batched standard error for the first 20 items. All standard errors are in three digits, which confirms that the number of iterations of MCMC used in this study are enough. The batch standard error for the full 60 items is given in the appendix.

Table 4.1.2: MCMC standard error

Item	a_1	a_2	a_3	a_4	a_5	a_6	d
1	0.0045	0.0016	0.0030	0.0011	0.0018	0.0037	0.0013
2	0.0016	0.0010	0.0007	0.0017	0.0013	0.0018	0.0008
3	0.0039	0.0028	0.0026	0.0010	0.0030	0.0034	0.0010
4	0.0041	0.0020	0.0021	0.0022	0.0028	0.0046	0.0012
5	0.0019	0.0022	0.0012	0.0015	0.0026	0.0036	0.0015
6	0.0042	0.0012	0.0027	0.0027	0.0019	0.0029	0.0015
7	0.0029	0.0028	0.0020	0.0021	0.0019	0.0026	0.0008
8	0.0046	0.0027	0.0033	0.0021	0.0014	0.0030	0.0016
9	0.0041	0.0019	0.0025	0.0026	0.0018	0.0052	0.0015
10	0.0048	0.0050	0.0044	0.0024	0.0033	0.0075	0.0030
11	0.0004	0.0029	0.0019	0.0009	0.0026	0.0036	0.0021
12	0.0013	0.0030	0.0018	0.0015	0.0013	0.0040	0.0012
13	0.0006	0.0025	0.0009	0.0013	0.0014	0.0019	0.0013
14	0.0007	0.0017	0.0031	0.0014	0.0018	0.0056	0.0016
15	0.0012	0.0011	0.0021	0.0008	0.0018	0.0018	0.0011
16	0.0015	0.0035	0.0022	0.0020	0.0018	0.0049	0.0017
17	0.0008	0.0008	0.0014	0.0014	0.0009	0.0024	0.0022
18	0.0012	0.0018	0.0018	0.0017	0.0030	0.0054	0.0028
19	0.0006	0.0040	0.0030	0.0019	0.0017	0.0069	0.0021
20	0.0011	0.0039	0.0034	0.0024	0.0018	0.0065	0.0039

4.1.5. Item parameter recovery diagnostic

Once the convergence of the MCMC simulation is confirmed, it is necessary to diagnose how well the item parameters are recovered. In order to assess item parameter recovery, a credible interval is constructed. The credible interval shows how well the estimated item parameters are in the range of standard error. For this study, the Highest posterior intervals (HPD) are assessed for each estimated item parameter to ensure that they are in the range of satisfactory standard error. HPD is also implemented in the CODA package in R. For this study, a probability of .95 is set. Tables 4.1.3, 4.1.4, and 4.1.5 show HPD for the a-parameters and the d-parameter for 6-dimensions with correlated and skewed latent structures. All of the estimated means of item parameters are between the lower and upper bounds of HPD, which shows that the item parameters are recovered in the acceptance area.

Table 4.1.3: Highest posterior interval for a_1, a_2, and a_3 parameters

	a_1			a_2			a_3		
Item	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper
1	0.4391	0.0954	0.7284	0.4089	0.1175	0.7038	0.2820	0.0633	0.4875
2	0.4231	0.2224	0.6228	0.2595	0.0001	0.5307	0.1491	0.0002	0.3045
3	0.2637	0.0000	0.4801	0.1915	0.0001	0.4614	0.0934	0.0000	0.2142
4	0.2947	0.0465	0.5199	0.2285	0.0001	0.4742	0.1479	0.0002	0.2919
5	0.2678	0.0754	0.4488	0.2495	0.0861	0.4126	0.2260	0.0558	0.3832
6	0.6398	0.1883	1.0030	0.4364	0.0211	0.9389	0.2927	0.0433	0.5433
7	0.4619	0.1718	0.7242	0.3278	0.0180	0.6745	0.1568	0.0000	0.3094
8	0.7090	0.1330	1.1650	0.5096	0.0422	1.0880	0.2635	0.0009	0.5114
9	0.6421	0.1152	1.0620	0.4505	0.0195	0.9843	0.2702	0.0112	0.5044
10	0.5545	0.2008	0.8570	0.3781	0.0189	0.7737	0.1971	0.0035	0.3775

Table 4.1.4: Highest posterior interval for a_4, a_5, and a_6 - parameters

	a_4			a_5			a_6		
Item	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper
1	0.2835	0.0004	0.5377	0.2550	0.0001	0.5782	0.2077	0.0136	0.3922
2	0.2468	0.0165	0.4621	0.2548	0.0002	0.4930	0.3048	0.0237	0.5645
3	0.1031	0.0000	0.2407	0.1021	0.0000	0.2418	0.0845	0.0000	0.1962
4	0.1736	0.0000	0.3575	0.1893	0.0000	0.3624	0.1669	0.0001	0.3263
5	0.1818	0.0301	0.3305	0.1767	0.0202	0.3268	0.1756	0.0317	0.3211
6	0.3494	0.0501	0.6744	0.3547	0.0715	0.6439	0.3651	0.1037	0.6280
7	0.3361	0.1379	0.5467	0.3274	0.1133	0.5305	0.2476	0.0020	0.4594
8	0.3800	0.0192	0.7268	0.3895	0.0633	0.7610	0.3041	0.0001	0.5794
9	0.2873	0.0028	0.5907	0.2910	0.0174	0.5629	0.3047	0.0269	0.5469
10	0.3540	0.1244	0.5991	0.3254	0.1014	0.5519	0.2775	0.0486	0.5020

Table 4.1.5: Highest posterior interval for d- parameters

	d		
Item	Mean	Lower	Upper
1	0.995	0.897	1.085
2	0.762	0.680	0.855
3	0.353	0.277	0.431
4	0.283	0.209	0.364
5	0.151	0.073	0.224
6	0.552	0.463	0.646
7	-0.095	-0.182	-0.017
8	-0.138	-0.232	-0.050
9	-0.163	-0.250	-0.080
10	-1.029	-1.129	-0.926

4.2. 3-Dimensions

In this section, 3-dimensions of latent structures are explored, with factors such as types of latent traits configuration (AS vs. MS), correlated latent traits (.3 and .6), and skewed latent traits distributions (+.9 and -.9).

4.2.1. Approximate Simple Structure (AS) and Mixed Structures (MS)

The potential effect of structure types of the latent construct—approximate simple (AS) and mixed structures (MS)—on item parameter recovery in the MIRT model is examined with three different sample sizes, 1000, 1500, and 2000.

Table 4.2.1 shows the BIAS when the different types of latent structures were imposed into item recovery in the MIRT model. When approximate simple structure (AS) is imposed into item parameter recovery, the item parameters are overestimated for *a*-parameters, compared to MS. However, the *d*-parameter shows a different pattern. *d*-parameter is underestimated when AS is imposed into item recovery, compared to MS.

Table 4.2.1: BIAS for different types of latent traits configuration (AS vs. MS)

Item Parameters	Structures	N		
		1000	1500	2000
a_1	AS	0.0941	0.0837	0.0784
	MS	-0.0007	-0.0204	-0.0131
a_2	AS	0.0649	0.0615	0.0415
	MS	0.0083	-0.0061	-0.0095
a_3	AS	0.1048	0.0922	0.0901
	MS	-0.0022	-0.0089	-0.0191
d	AS	-0.3348	-0.3471	-0.3420
	MS	0.0069	-0.0057	0.0028

Table 4.2.2. MAD for different types of latent traits configuration (AS vs. MS)

Item parameters	Structures	N		
		1000	1500	2000
a_1	AS	0.1951	0.1847	0.1883
	MS	0.1115	0.0940	0.1086
a_2	AS	0.3018	0.3060	0.2940
	MS	0.1236	0.0980	0.1159
a_3	AS	0.3506	0.3397	0.3510
	MS	0.1103	0.0924	0.0939
d	AS	0.6351	0.6376	0.6349
	MS	0.0642	0.0527	0.0462

Table 4.2.3. RMSE for different types of latent traits configuration (AS vs. MS)

Item parameters	Structures	N		
	N	1000	1500	2000
a_1	AS	0.1951	0.1847	0.1883
	MS	0.1115	0.0940	0.1086
a_2	AS	0.3018	0.3060	0.2940
	MS	0.1236	0.0980	0.1159
a_3	AS	0.3506	0.3397	0.3510
	MS	0.1103	0.0924	0.0939
d	AS	0.6351	0.6376	0.6349
	MS	0.0642	0.0527	0.0462

Tables 4.2.2 and 4.2.3 show the mean absolute difference (MAD) and root mean square error (RMSE), respectively. MAD and RMSE show that MS has better item recovery in terms of the magnitude of MAD and RMSE than does AS. The range of RMSE for *a*-parameters for AS is from .1847 to .3510, but is .0924 to .1159 for MS. AS has a bigger and wider range of estimated *a*-parameters than MS. The range of RMSE for the *d*-parameter shows the same pattern. AS has a range from .6349 to .6376, and MS has a range from .0462 to .0642. There is no significant improvement as the sample size is being increased from 1000 to 2000, which tells us that the

sample size (N=1,000) has an enough power to estimate the stable item recovery whether latent traits structures have different configuration of types such as AS or MS.

4.2.2. Correlated Latent Traits

Only MS is used to see the interaction effect of correlated latent traits. The AS is not used for correlated latent structures because it has a dominated dimension, which has superior power to make the student able to answer correctly. On other hand, the MS has several dimensions that could contribute to students choosing the correct answer. Thus, the relationship among dimensions in MS has a more serious impact on students' getting a correct answer.

Table 4.2.4 shows the BIAS when different correlations (.3 and .6) are imposed into the latent structures. With a .3 correlation between latent structures, *a*-parameters are underestimated, ranging from -.0544 to -.0285. When a .6 correlation is imposed into latent structures, *a*-parameters are overestimated, ranging from -.0064 to .0394. The *d*-parameter does not show any noticeable pattern in item recovery, whether latent structures have a high or low correlation. Tables 4.2.5 and 4.2.6 show the MAD and RMSE, respectively. The magnitude of MAD and RMSE show that a 0.6 correlation has more impact on the recovery of *a*-parameters than a .3 correlation. The *d*-parameter does not show any influence from any correlation, in terms of the magnitude of MAD and RMSE. As sample size increases from 1000 to 1500 and to 2000, the item recovery does not seem to improve for the *a*-parameters. On other hand, the *d*-parameter does improve when the sample size increases, even though the magnitude of improvement is really small; the difference from the biggest to the smallest is just .002.

Table 4.2.4. BIAS for correlated latent traits (MS only)

Item parameters	Correlation	N		
		1000	1500	2000
a_1	0	-0.0007	-0.0204	-0.0131
	0.3	-0.0401	-0.0411	-0.0544
	0.6	0.0184	0.0028	0.0147
a_2	0	0.0083	-0.0061	-0.0095
	0.3	-0.0285	-0.0345	-0.0392
	0.6	0.0312	0.0394	0.0091
a_3	0	-0.0022	-0.0089	-0.0191
	0.3	-0.0291	-0.0377	-0.0436
	0.6	0.0193	-0.0064	0.0084
d	0	0.0069	-0.0057	0.0028
	0.3	0.0032	0.0071	-0.0047
	0.6	-0.0101	0.0039	0.0036

Table 4.2.5. MAD for correlated latent traits (MS only)

Item parameters	Structures	N		
		1000	1500	2000
a_1	0	0.11152	0.09399	0.10859
	0.3	0.11307	0.10851	0.11819
	0.6	0.12724	0.12279	0.1218
a_2	0	0.0083	0.09801	0.11592
	0.3	-0.0285	0.09715	0.1105
	0.6	0.0312	0.11997	0.11018
a_3	0	0.11025	0.09238	0.09389
	0.3	0.11554	0.11273	0.12013
	0.6	0.12942	0.12417	0.12348
d	0	0.06419	0.0527	0.04618
	0.3	0.06427	0.05234	0.04592
	0.6	0.06262	0.05008	0.04719

Table 4.2.6. RMSE for correlated latent traits (MS only)

Item Parameters	Structures	N		
		1000	1500	2000
a_1	0	0.1301	0.1122	0.1300
	0.3	0.1280	0.1220	0.1340
	0.6	0.1459	0.1430	0.1367
a_2	0	0.1451	0.1149	0.1389
	0.3	0.1251	0.1119	0.1250
	0.6	0.1486	0.1407	0.1281
a_3	0	0.1324	0.1120	0.1109
	0.3	0.1313	0.1255	0.1338
	0.6	0.1511	0.1417	0.1425
d	0	0.0785	0.0661	0.0554
	0.3	0.0785	0.0653	0.0561
	0.6	0.0785	0.0622	0.0575

4.2.3. Skewed Latent Trait Distributions

When skew is imposed on the latent traits distributions, it does not have as much influence on item recovery for the *a*-parameters, whether they have AS or MS. It also does not have any noticeable influence whether there is a positive or a negative skew on the latent structures. In all cases, the *a*-parameters were overestimated. However, the *d*-parameters showed an interesting pattern of item recovery. When the latent structures have an AS, the *d*-parameter was underestimated, ranging in BIAS from -.4810 to -.2048. However, a positive skew, which means that a low ability sample is used, slightly underestimated the *d*-parameter, compared with no skew on the latent traits distributions. When a negative skew is imposed into the latent traits distributions, it slightly overestimates, when compared with no skew on the latent traits distributions. When the latent traits have an MS, both *a*-parameters and *d*-parameters showed the same pattern as an AS. However, the magnitude of RMSE and MAD showed that AS had more trouble estimating the *d*-parameter when skew was imposed on the latent traits distributions. See tables 4.2.8 and 4.2.9.

Table 4.2.7. BIAS when skew is imposed on the latent traits distributions (+.9 and -.9)

Item Parameters	Structures	Skew	N		
			1000	1500	2000
a_1	AS	No	0.0941	0.0837	0.0784
		Positive	0.1530	0.1394	0.1342
		Negative	0.1505	0.1389	0.1303
	MS	No	-0.0007	-0.0204	-0.0131
		Positive	0.0553	0.0382	0.0589
		Negative	0.0524	0.0395	0.0316
a_2	AS	No	0.0649	0.0615	0.0415
		Positive	0.1257	0.1107	0.1111
		Negative	0.1216	0.1178	0.1078
	MS	No	0.0083	-0.0061	-0.0095
		Positive	0.0777	0.0701	0.0417
		Negative	0.0672	0.0605	0.0432
a_3	AS	No	0.1048	0.0922	0.0901
		Positive	0.1529	0.1500	0.1518
		Negative	0.1569	0.1411	0.1373
	MS	No	-0.0022	-0.0089	-0.0191
		Positive	0.0581	0.0416	0.0398
		Negative	0.0686	0.0494	0.0524
d	AS	No	-0.3348	-0.3471	-0.3420
		Positive	-0.4635	-0.4810	-0.4698
		Negative	-0.2102	-0.2048	-0.2108
	MS	No	0.0069	-0.0057	0.0028
		Positive	-0.1651	-0.1484	-0.1493
		Negative	0.1439	0.1521	0.1428

Table 4.2.8. MAD when skew is imposed on the latent traits distributions (+.9 and -.9)

Item parameters	Structures	Skew	N		
			1000	1500	2000
a_1	AS	No	0.1951	0.1847	0.1883
		Positive	0.2181	0.2071	0.2012
		Negative	0.2138	0.2070	0.2030
	MS	No	0.1115	0.0940	0.1086
		Positive	0.1230	0.1099	0.1102
		Negative	0.1222	0.1094	0.1073
a_2	AS	No	0.3018	0.3060	0.2940
		Positive	0.3109	0.3075	0.2985
		Negative	0.3145	0.3090	0.2996
	MS	No	0.1236	0.0980	0.1159
		Positive	0.1448	0.1226	0.1066
		Negative	0.1329	0.1279	0.1182
a_3	AS	No	0.3506	0.3397	0.3510
		Positive	0.3800	0.3632	0.3668
		Negative	0.3813	0.3645	0.3654
	MS	No	0.1103	0.0924	0.0939
		Positive	0.1252	0.1110	0.1000
		Negative	0.1240	0.1219	0.1043
d	AS	No	0.6351	0.6376	0.6349
		Positive	0.6822	0.6872	0.6833
		Negative	0.6034	0.5982	0.6048
	MS	No	0.0642	0.0527	0.0462
		Positive	0.1673	0.1500	0.1501
		Negative	0.1485	0.1525	0.1431

Table 4.2.9. RMSE when skew is imposed on the latent traits distributions (+.9 and -.9)

Item parameters	Structures	Skew	N		
			1000	1500	2000
a_1	AS	No	0.1951	0.1847	0.1883
		Positive	0.2181	0.2071	0.2012
		Negative	0.2138	0.2070	0.2030
	MS	No	0.1301	0.1122	0.1300
		Positive	0.1470	0.1301	0.1264
		Negative	0.1441	0.1262	0.1267
a_2	AS	No	0.3383	0.3432	0.3017
		Positive	0.3238	0.3170	0.3084
		Negative	0.3274	0.3189	0.3077
	MS	No	0.1236	0.0980	0.1159
		Positive	0.1448	0.1226	0.1066
		Negative	0.1329	0.1279	0.1182
a_3	AS	No	0.3825	0.3681	0.3598
		Positive	0.3930	0.3736	0.3756
		Negative	0.3935	0.3753	0.3745
	MS	No	0.1324	0.1120	0.1109
		Positive	0.1480	0.1370	0.1200
		Negative	0.1457	0.1414	0.1218
d	AS	No	0.6425	0.6421	0.6384
		Positive	0.6893	0.6925	0.6874
		Negative	0.6116	0.6037	0.6088
	MS	No	0.0785	0.0661	0.0554
		Positive	0.1827	0.1616	0.1593
		Negative	0.1673	0.1640	0.1518

4.2.4. Correlated Latent Traits and Skewed Latent Traits Distributions

In this section, results are presented for when both correlation and skew are imposed on the latent traits. When both correlation and skew are imposed into the latent traits, high correlation (.6) with skewed distributions contributes to a higher magnitude of bias for *a*-parameters than low correlations. The mixed factors of both correlation and skew overestimate the *a*-parameters. However, the *d*-parameter is not influenced by whether latent traits have high or low correlations. Imposing correlation with a skewed latent trait distribution does increase the magnitude of bias. Table 4.2.10 shows the bias when both correlation and skew are imposed on the latent traits.

Table 4.2.10. BIAS when both correlation and skew are imposed on the latent traits

Item parameters	Correlation	Skew	N		
			1000	1500	2000
a_1	0	No	-0.0007	-0.0204	-0.0131
	0.3	Positive	0.0108	0.0150	0.0116
		Negative	0.0138	0.0063	0.0094
	0.6	Positive	0.0667	0.0683	0.0625
		Negative	0.0802	0.0575	0.0672
a_2	0	No	0.0083	-0.0061	-0.0095
	0.3	Positive	0.0233	0.0212	0.0107
		Negative	0.0270	0.0150	0.0062
	0.6	Positive	0.0753	0.0721	0.0501
		Negative	0.0902	0.0696	0.0614
a_3	0	No	-0.0022	-0.0089	-0.0191
	0.3	Positive	0.0177	0.0111	-0.0008
		Negative	0.0154	0.0073	0.0008
	0.6	Positive	0.0710	0.0533	0.0691
		Negative	0.0677	0.0624	0.0584
d	0	No	0.0069	-0.0057	0.0028
	0.3	Positive	-0.0982	-0.0981	-0.1173
		Negative	0.1072	0.1052	0.1137
	0.6	Positive	-0.1053	-0.0938	-0.1001
		Negative	0.1213	0.0850	0.0888

When both correlation and skew are imposed on the latent traits, the *d*-parameter was overestimated when a negative skew was applied, and underestimated when a positive skew was applied. Increasing the sample size from 1000 to 1500 and to 2000 did not improve the item recovery. The magnitude of BIAS, MAD, and RMSE stay in the small range, except for *d*-parameters with a .6 correlation and skewed trait distributions. When the sample size was increased from 1000 to 1500, it improves the estimate of the *d*-parameter. However, the estimates do not improve when the sample size increased from 1500 to 2000. Table 4.2.11 and 4.2.12 show MAD and RMSE when both correlation and skew are imposed on the latent traits, respectively.

Table 4.2.11. MAD when both correlation and skew are imposed on the latent traits

Item parameters	Correlation	Skew	N 1000	1500	2000
a_1	0	No	0.1115	0.0940	0.1086
	0.3	Positive	0.1213	0.1233	0.1152
		Negative	0.1214	0.1242	0.1193
	0.6	Positive	0.1473	0.1473	0.1363
		Negative	0.1490	0.1398	0.1431
a_2	0	No	0.1236	0.0980	0.1159
	0.3	Positive	0.1173	0.1242	0.1136
		Negative	0.1267	0.1177	0.1083
	0.6	Positive	0.1445	0.1381	0.1245
		Negative	0.1459	0.1348	0.1363
a_3	0	No	0.1103	0.0924	0.0939
	0.3	Positive	0.1220	0.1241	0.1185
		Negative	0.1212	0.1103	0.1305
	0.6	Positive	0.1378	0.1405	0.1336
		Negative	0.1477	0.1398	0.1439
d	0	No	0.0642	0.0527	0.0462
	0.3	Positive	0.1110	0.1039	0.1187
		Negative	0.1192	0.1101	0.1157
	0.6	Positive	0.1163	0.0994	0.1034
		Negative	0.1301	0.0951	0.0938

Table 4.2.12. RMSE when both correlation and skew are imposed into the latent structures

Item parameters	Correlation	Skew	N		
			1000	1500	2000
a_1	0	No	0.1301	0.1122	0.1300
	0.3	Positive	0.1418	0.1447	0.1314
		Negative	0.1440	0.1422	0.1352
	0.6	Positive	0.1697	0.1668	0.1507
		Negative	0.1679	0.1565	0.1644
a_2	0	No	0.1451	0.1149	0.1389
	0.3	Positive	0.1343	0.1470	0.1323
		Negative	0.1439	0.1370	0.1274
	0.6	Positive	0.1670	0.1587	0.1403
		Negative	0.1639	0.1534	0.1566
a_3	0	No	0.1324	0.1120	0.1109
	0.3	Positive	0.1398	0.1398	0.1360
		Negative	0.1451	0.1289	0.1478
	0.6	Positive	0.1579	0.1596	0.1480
		Negative	0.1684	0.1550	0.1665
d	0	No	0.0785	0.0661	0.0554
	0.3	Positive	0.1286	0.1177	0.1293
		Negative	0.1379	0.1248	0.1277
	0.6	Positive	0.1361	0.1135	0.1153
		Negative	0.1493	0.1096	0.1056

4.3. 6-Dimensions

This section presents 6-dimensions of the latent trait, with factors such as types of latent trait configuration (AS vs. MS), correlated latent traits (.3 and .6), and skewed latent traits distributions (+.9 and -.9).

4.3.1. Approximate Simple Structures and Mixed Structures

When there are different types of latent trait configuration such as AS and MS on 6-dimensions, the magnitude of BIAS for the *a*-parameters for AS is bigger than for MS, except the *a1*-parameter, which has a smaller BIAS with AS. Table 4.3.1 shows the BIAS when AS and MS are imposed into 6-dimension latent traits. The *d*-parameter shows no significantly different pattern between AS and MS in terms of BIAS. Both the AS and MS structures give a satisfactory item recovery, ranging from -.0086 to +.0093. When sample sizes increase from 1000 to 1500, 2000, and 3000, the magnitude of BIAS of *a*-parameters gets smaller for both AS and MS structures, which shows that a sample size of at least 2000 is required for satisfactory item parameter recovery in 6-dimension latent traits.

Table 4.3.1. BIAS for different types of latent trait configuration (AS vs. MS)

Item parameters	Structures	No			
		1000	1500	2000	3000
a_1	AS	0.0462	0.0181	0.0117	-0.0024
	MS	0.0369	0.0226	0.0159	0.0003
a_2	AS	0.0462	0.0234	0.0113	0.0038
	MS	0.0466	0.0046	-0.0016	-0.0177
a_3	AS	0.0462	0.0117	-0.0002	-0.0147
	MS	0.0363	0.0052	-0.0015	-0.0127
a_4	AS	0.0462	0.0117	-0.0002	0.0026
	MS	0.0237	0.0057	-0.0058	-0.0144
a_5	AS	0.0543	0.0162	0.0106	0.0046
	MS	0.0183	0.0059	-0.0040	-0.0078
a_6	AS	0.0522	0.0268	0.0129	0.0029
	MS	0.0276	0.0142	0.0051	0.0012
d	AS	0.0045	-0.0086	-0.0044	0.0035
	MS	-0.0025	0.0093	0.0040	0.0012

Magnitude of MAD and RMSE in Tables 4.3.2 and 4.3.3 show that the AS structures give a better item recovery for *a*-parameters, compared to MS structures. It shows that *a*-

parameters are more problematic for both AS and MS, compared to d-parameter in 6-dimensions. In particular, AS structures show an irregular pattern in some of the a-parameters such as a_3, a_4, and a_5. When the sample size increases, the RMSE goes higher than with a smaller sample size.

Overall, MS shows a higher RMSE, compared to AS. However, MS shows a stable and gradual decrease when sample size increases. There is no noticeable influence from the different types of latent structures configuration. Both the AS and MS structures have almost identical magnitudes of RMSE and MAD.

Table 4.3.2. MAD for different types of latent trait configuration (AS vs. MS)

Item parameters	Structures	N			
		1000	1500	2000	3000
a_1	AS	0.0818	0.0590	0.0543	0.0452
	MS	0.1465	0.1149	0.1075	0.0933
a_2	AS	0.0805	0.0624	0.0540	0.0659
	MS	0.1517	0.1053	0.1145	0.0860
a_3	AS	0.0805	0.0833	0.0528	0.0696
	MS	0.1443	0.1087	0.1119	0.0782
a_4	AS	0.0805	0.1043	0.0564	0.0717
	MS	0.0965	0.1044	0.1040	0.0743
a_5	AS	0.0805	0.0806	0.0527	0.0458
	MS	0.1044	0.0920	0.1201	0.0959
a_6	AS	0.0994	0.0622	0.0542	0.0466
	MS	0.1478	0.1108	0.1154	0.0868
d	AS	0.0741	0.0516	0.0503	0.0393
	MS	0.0692	0.0594	0.0462	0.0392

4.3.2. Correlated Latent Traits

Only MS is used to see the interaction effect of correlated latent traits. The AS has not been used for correlated latent traits because it has a dominated dimension, which has superior power for making a student able to answer correctly. On other hand, MS has several dimensions that could contribute to making students give a correct answer. So the relationship among dimensions in MS has a more serious impact on students' getting a correct answer.

Table 4.3.3. RMSE for different types of latent trait configuration (AS vs. MS)

Item parameters	Structures	N			
		1000	1500	2000	3000
a_1	AS	0.1044	0.0741	0.0673	0.0549
	MS	0.1687	0.1388	0.1260	0.1108
a_2	AS	0.1022	0.0790	0.0670	0.1069
	MS	0.1702	0.1247	0.1391	0.1008
a_3	AS	0.1022	0.0790	0.0670	0.1153
	MS	0.1676	0.1303	0.1332	0.0960
a_4	AS	0.1399	0.1761	0.0707	0.1207
	MS	0.1627	0.1250	0.1249	0.0904
a_5	AS	0.1399	0.1761	0.0660	0.0563
	MS	0.1579	0.1271	0.1450	0.1152
a_6	AS	0.1473	0.0784	0.0682	0.0576
	MS	0.1710	0.1347	0.1350	0.1058
d	AS	0.0898	0.0628	0.0619	0.0473
	MS	0.0852	0.0734	0.0554	0.0476

Table 4.3.4 shows the BIAS when different magnitudes of correlations (.3 and .6) are imposed into the latent traits in 6-dimensions. It clearly shows that the magnitude of BIAS for *a*-parameters increases when the correlation increases. In the case of no correlation and a sample size of 1000, the BIAS for the a_1 parameter is .0369. It goes up to .0778 when correlation .3 is used, which is almost twice that of no correlation. It goes up to .01661 when the correlation .6 is used, which is two times larger than the .3 correlation. On the other hand, the *d*-parameter is not influenced by the correlation between the latent traits. The magnitude of BIAS does not change significantly, whether it has a high correlation (.6) or a low correlation (.3) between the latent structures. When the sample size increases from 1000 to 1500, 2000, and 3000, the magnitude of

BIAS is gradually decreased for the correlated latent traits. However, the change of BIAS does not seem significant.

Table 4.3.4. BIAS for correlated latent traits (MS only)

Item Parameters	Correlation	N			
		1000	1500	2000	3000
a_1	0	0.0369	0.0226	0.0159	0.0003
	0.3	0.0778	0.0599	0.0523	0.0482
	0.6	0.1661	0.1454	0.1391	0.1279
a_2	0	0.0207	0.0046	-0.0016	-0.0177
	0.3	0.0591	0.0428	0.0352	0.0278
	0.6	0.1518	0.1275	0.1303	0.0913
a_3	0	0.0209	0.0052	-0.0015	-0.0127
	0.3	0.0570	0.0426	0.0354	0.0275
	0.6	0.1476	0.1274	0.1137	0.1079
a_4	0	0.0237	0.0057	-0.0058	-0.0144
	0.3	0.0663	0.0513	0.0389	0.0283
	0.6	0.1581	0.1379	0.1253	0.1129
a_5	0	0.0183	0.0059	-0.0040	-0.0078
	0.3	0.0584	0.0405	0.0347	0.0194
	0.6	0.1481	0.1298	0.1242	0.1115
a_6	0	0.0276	0.0142	0.0051	0.0012
	0.3	0.0682	0.0527	0.0401	0.0391
	0.6	0.1555	0.1374	0.1258	0.1213
d	0	-0.0025	0.0093	0.0040	0.0012
	0.3	0.0054	0.0073	0.0009	0.0024
	0.6	0.0155	-0.0029	0.0095	0.0016

Tables 4.3.5 and 4.3.6 show the MAD and RMSE when the latent traits are being correlated. When the sample size is 1000, the magnitude of RMSE for the *a*-parameters is almost same to no correlation and to .3 correlation. However, .6 has a much higher RMSE than .3 and no correlation. When the sample size is being increased to 1500, 2000, and 3000, the magnitude of RMSE for no correlation drops more rapidly than for the .3 or .6 correlations. However, the

d-parameter has almost the same rate of change in terms of magnitude of RMSE and MAD when 0, .3 and .6 correlations are imposed into the latent traits. It also gradually decreases when the sample size increases. Even though the size of change is not significant, a bigger sample size contributes to decreasing the RMSE and MAD of the d-parameter.

Table 4.3.5. MAD for correlated latent traits (MS only)

Item Parameters	Correlation	N			
		1000	1500	2000	3000
a_1	0	0.1465	0.1149	0.1075	0.0933
	0.3	0.1662	0.1528	0.1402	0.1389
	0.6	0.2294	0.2126	0.2047	0.2026
a_2	0	0.1517	0.1053	0.1145	0.0860
	0.3	0.1629	0.1457	0.1400	0.1331
	0.6	0.2238	0.2069	0.2045	0.1850
a_3	0	0.1443	0.1087	0.1119	0.0782
	0.3	0.1614	0.1532	0.1404	0.1322
	0.6	0.2229	0.2027	0.1943	0.1899
a_4	0	0.1423	0.1044	0.1040	0.0743
	0.3	0.1646	0.1551	0.1422	0.1285
	0.6	0.2259	0.2133	0.2017	0.1962
a_5	0	0.1445	0.0920	0.1201	0.0959
	0.3	0.1605	0.1484	0.1403	0.1209
	0.6	0.2173	0.2089	0.2000	0.1922
a_6	0	0.1478	0.1108	0.1154	0.0868
	0.3	0.1670	0.1507	0.1501	0.1412
	0.6	0.2250	0.2127	0.2043	0.2044
d	0	0.0692	0.0594	0.0462	0.0392
	0.3	0.0696	0.0578	0.0469	0.0407
	0.6	0.0719	0.0659	0.0499	0.0385

Table 4.3.6. RMSE for correlated latent traits (MS only)

Item Parameters	Correlation	N			
		1000	1500	2000	3000
a_1	0	0.1687	0.1388	0.1260	0.1108
	0.3	0.1771	0.1661	0.1520	0.1489
	0.6	0.2386	0.2219	0.2126	0.2128
a_2	0	0.1702	0.1247	0.1391	0.1008
	0.3	0.1732	0.1570	0.1501	0.1431
	0.6	0.2337	0.2139	0.2117	0.1925
a_3	0	0.1676	0.1303	0.1332	0.0960
	0.3	0.1742	0.1652	0.1478	0.1425
	0.6	0.2335	0.2117	0.2010	0.1983
a_4	0	0.1627	0.1250	0.1249	0.0904
	0.3	0.1758	0.1640	0.1527	0.1388
	0.6	0.2369	0.2215	0.2075	0.2021
a_5	0	0.1684	0.1142	0.1450	0.1152
	0.3	0.1727	0.1604	0.1484	0.1311
	0.6	0.2279	0.2185	0.2090	0.1987
a_6	0	0.1710	0.1347	0.1350	0.1058
	0.3	0.1795	0.1608	0.1590	0.1510
	0.6	0.2371	0.2181	0.2122	0.2105
d	0	0.0852	0.0734	0.0554	0.0476
	0.3	0.0854	0.0714	0.0577	0.0508
	0.6	0.0901	0.0799	0.0596	0.0471

4.3.3. Skewed Latent Traits Distributions

This section presents the results for 6-dimensions with skewed latent traits distributions on both AS and MS. Table 4.3.7 shows the BIAS when negative (-.9) and positive (+.9) skews are imposed into the latent traits distributions. The magnitude of BIAS for the *a*-parameters shows that a skew on the latent traits distributions increases the BIAS, whether it has AS or MS. However, there is not any significant difference between the AS and MS structures. It does not show any significant difference between a negative or positive skew.

Table 4.3.7. BIAS when skew is imposed on the latent traits distributions (+.9 and -.9)

Item Para-meters	Skew	AS				MS			
		1000	1500	2000	3000	1000	1500	2000	3000
a_1	No	0.046	0.018	0.012	-0.002	0.037	0.023	0.016	0.000
	Positive	0.110	0.090	0.080	0.068	0.127	0.110	0.098	0.092
	Negative	0.104	0.088	0.072	0.072	0.129	0.096	0.102	0.094
a_2	No	0.047	0.023	0.011	0.004	0.021	0.005	-0.002	-0.018
	Positive	0.110	0.084	0.076	0.065	0.110	0.085	0.077	0.080
	Negative	0.106	0.084	0.073	0.064	0.104	0.087	0.084	0.069
a_3	No	0.036	0.012	0.000	-0.015	0.021	0.005	-0.001	-0.013
	Positive	0.094	0.078	0.075	0.060	0.106	0.091	0.086	0.072
	Negative	0.099	0.076	0.069	0.062	0.110	0.088	0.089	0.073
a_4	No	0.044	0.019	0.015	0.003	0.024	0.006	-0.006	-0.014
	Positive	0.105	0.086	0.075	0.068	0.115	0.092	0.088	0.076
	Negative	0.105	0.093	0.076	0.069	0.114	0.095	0.084	0.085
a_5	No	0.054	0.016	0.011	0.005	0.018	0.006	-0.004	-0.008
	Positive	0.104	0.085	0.081	0.069	0.109	0.099	0.087	0.079
	Negative	0.106	0.084	0.077	0.068	0.107	0.086	0.088	0.076
a_6	No	0.052	0.027	0.013	0.003	0.028	0.014	0.005	0.001
	Positive	0.112	0.088	0.086	0.074	0.120	0.101	0.096	0.092
	Negative	0.110	0.089	0.085	0.066	0.115	0.096	0.091	0.081
d	No	0.005	-0.009	-0.004	0.003	-0.003	0.009	0.004	0.001
	Positive	-0.177	-0.185	-0.184	-0.178	-0.228	-0.209	-0.221	-0.213
	Negative	0.168	0.172	0.165	0.172	0.227	0.238	0.220	0.224

The *d*-parameter shows a different pattern for negative or positive. When the positive skew is imposed into the latent traits distributions, it underestimates the *d*-parameter. When the negative skew is imposed into the latent traits distributions, it overestimates the *d*-parameter. MS has a bigger BIAS when skew is implied into the latent structures than does AS.

Table 4.3.8. MAD when skew is imposed on the latent traits distributions (+.9 and -.9)

Item Para-meters	Skew	AS				MS			
		1000	1500	2000	3000	1000	1500	2000	3000
a_1	No	0.082	0.059	0.054	0.045	0.146	0.115	0.126	0.093
	Positive	0.124	0.103	0.090	0.079	0.201	0.170	0.162	0.135
	Negative	0.121	0.102	0.083	0.080	0.203	0.153	0.148	0.142
a_2	No	0.059	0.062	0.054	0.066	0.152	0.105	0.114	0.086
	Positive	0.103	0.101	0.091	0.075	0.189	0.151	0.133	0.133
	Negative	0.102	0.096	0.087	0.073	0.189	0.161	0.146	0.126
a_3	No	0.073	0.083	0.053	0.070	0.144	0.109	0.112	0.078
	Positive	0.106	0.093	0.088	0.071	0.187	0.149	0.145	0.119
	Negative	0.112	0.090	0.080	0.071	0.186	0.170	0.147	0.132
a_4	No	0.096	0.104	0.056	0.072	0.142	0.104	0.104	0.074
	Positive	0.117	0.098	0.090	0.076	0.188	0.162	0.154	0.131
	Negative	0.117	0.103	0.088	0.079	0.203	0.176	0.150	0.138
a_5	No	0.104	0.081	0.053	0.046	0.144	0.092	0.120	0.096
	Positive	0.119	0.120	0.092	0.078	0.197	0.167	0.145	0.143
	Negative	0.117	0.094	0.088	0.078	0.194	0.160	0.148	0.140
a_6	No	0.099	0.062	0.054	0.047	0.148	0.111	0.115	0.087
	Positive	0.123	0.120	0.095	0.081	0.200	0.174	0.157	0.145
	Negative	0.120	0.104	0.095	0.076	0.202	0.165	0.158	0.147
d	No	0.074	0.052	0.050	0.039	0.069	0.059	0.046	0.039
	Positive	0.178	0.185	0.184	0.178	0.230	0.210	0.222	0.213
	Negative	0.170	0.173	0.166	0.172	0.227	0.238	0.220	0.224

Tables 4.3.8 and 4.3.9 show the MAD and RMSE, which confirms the pattern of effect of skew in the latent traits distributions in AS and MS. Whether it has a negative or positive skew, the effect of skew is almost identical for *a*-parameter recovery. Whether it has AS or MS, the magnitude of MAD and RMSE for *a*-parameters increase almost the same amount when skew is implied into the latent traits distributions. The *d*-parameter shows a different story; MS has a bigger MAD and RMSE than AS when there is skew. Increasing the sample size decreases the

RMSE for the *a*-parameters, even though the change is not significantly large. The *d*-parameter does not improve when sample size increases.

Table 4.3.9. RMSE when skew is imposed on the latent traits distributions (+.9 and -.9)

Item Para-meters	Skew	AS				MS			
		1000	1500	2000	3000	1000	1500	2000	3000
a_1	No	0.082	0.059	0.054	0.045	0.146	0.115	0.108	0.093
	Positive	0.124	0.103	0.090	0.079	0.201	0.170	0.146	0.135
	Negative	0.121	0.102	0.083	0.080	0.203	0.153	0.167	0.142
a_2	No	0.102	0.079	0.067	0.107	0.170	0.125	0.139	0.101
	Positive	0.152	0.121	0.107	0.088	0.208	0.166	0.149	0.152
	Negative	0.145	0.120	0.106	0.086	0.207	0.177	0.164	0.144
a_3	No	0.093	0.133	0.065	0.115	0.168	0.130	0.133	0.096
	Positive	0.130	0.113	0.105	0.083	0.204	0.170	0.161	0.137
	Negative	0.135	0.108	0.097	0.085	0.201	0.186	0.162	0.151
a_4	No	0.140	0.176	0.071	0.121	0.163	0.125	0.125	0.090
	Positive	0.141	0.117	0.107	0.091	0.206	0.176	0.167	0.149
	Negative	0.144	0.124	0.106	0.093	0.217	0.191	0.169	0.159
a_5	No	0.158	0.127	0.066	0.056	0.168	0.114	0.145	0.115
	Positive	0.142	0.168	0.110	0.092	0.215	0.185	0.166	0.159
	Negative	0.140	0.113	0.104	0.093	0.210	0.176	0.167	0.160
a_6	No	0.147	0.078	0.068	0.058	0.171	0.135	0.135	0.106
	Positive	0.150	0.165	0.113	0.095	0.217	0.189	0.174	0.167
	Negative	0.146	0.123	0.112	0.090	0.214	0.180	0.173	0.171
d	No	0.090	0.063	0.062	0.047	0.085	0.073	0.055	0.048
	Positive	0.196	0.198	0.194	0.183	0.245	0.221	0.229	0.219
	Negative	0.187	0.185	0.176	0.179	0.242	0.248	0.228	0.229

4.3.4. Correlated and Skewed Latent Traits Distributions

This section explores the influence when both factors, correlation and skew, are imposed into the latent traits distributions. Table 4.3.10 shows the BIAS for item parameters, when both factors are implemented into item recovery procedures. As with 3-dimension structures, only MS is examined. For *a*-parameters, the magnitude of BIAS almost doubles, compared to having just

Table 4.3.10. BIAS when both correlation and skew are imposed on the latent traits distributions

Item parameters	Correlation	Skew	1000	1500	2000	3000
a_1	0	No	0.0369	0.0226	0.0159	0.0003
	0.3	Positive	0.1818	0.1623	0.1491	0.1469
		Negative	0.1855	0.1751	0.1628	0.1497
	0.6	Positive	0.3336	0.3250	0.3206	0.2972
		Negative	0.3475	0.3136	0.3445	0.2994
a_2	0	No	0.0207	0.0046	-0.0016	-0.0177
	0.3	Positive	0.1600	0.1397	0.1357	0.1480
		Negative	0.1692	0.1515	0.1398	0.1250
	0.6	Positive	0.3194	0.3237	0.3009	0.2878
		Negative	0.3284	0.2969	0.3154	0.2810
a_3	0	No	0.0209	0.0052	-0.0015	-0.0127
	0.3	Positive	0.1551	0.1436	0.1356	0.1279
		Negative	0.1649	0.1507	0.1393	0.1227
	0.6	Positive	0.3099	0.2945	0.2946	0.2947
		Negative	0.3125	0.2953	0.3005	0.2562
a_4	0	No	0.0237	0.0057	-0.0058	-0.0144
	0.3	Positive	0.1680	0.1526	0.1290	0.1268
		Negative	0.1743	0.1436	0.1392	0.1329
	0.6	Positive	0.3176	0.3195	0.3042	0.2778
		Negative	0.3306	0.3067	0.2961	0.3281
a_5	0	No	0.0183	0.0059	-0.0040	-0.0078
	0.3	Positive	0.1577	0.1349	0.1352	0.1254
		Negative	0.1716	0.1457	0.1441	0.1239
	0.6	Positive	0.3143	0.3046	0.2840	0.2946
		Negative	0.3324	0.3036	0.2747	0.2737
a_6	0	No	0.0276	0.0142	0.0051	0.0012
	0.3	Positive	0.1725	0.1549	0.1277	0.1301
		Negative	0.1788	0.1536	0.1454	0.1315
	0.6	Positive	0.3235	0.3196	0.2794	0.2876
		Negative	0.3314	0.3175	0.3105	0.3059
d	0	No	-0.0025	0.0093	0.0040	0.0012
	0.3	Positive	-0.1990	-0.2172	-0.2298	-0.2296
		Negative	0.2257	0.2212	0.2180	0.2367
	0.6	Positive	-0.3665	-0.3576	-0.3484	-0.3524
		Negative	0.3533	0.3825	0.3774	0.3083

one factor such as correlation or skew. For example, the a_1-parameter has a .778 with a sample of 1000 and the correlation .3. a_1-parameter has a .1271 with a sample of 1000 and a positive skew.

However, when the a_1-parameter has factors, .3 correlation and a positive skew, then BIAS becomes .1818, which is almost double, compared to the model with just one factor. The same thing happens when a .6 correlation and negative skew are implemented into the latent traits distributions. Even though the magnitude of BIAS decreases slightly with an increasing sample size when there is only correlation between latent traits, the magnitude of BIAS does not change significantly when two factors are incorporated at the same time.

For the *d*-parameter, if there is only one factor, such as correlation or skew, then the BIAS is smaller than the model with two factors incorporated together. For example, when the sample size is 1000, the BIAS for the *d*-parameter with a .6 correlation is .0155. The BIAS for the *d*-parameter with a negative skew is .2269. It becomes .3533 with two factors, a .6 correlation and a negative skew together. However, if it has a low correlation like .3, then the BIAS does not get bigger with two factors together. For example, the BIAS is .0054 with a .3 correlation and .2269 with a negative skew. But it does not get bigger with two factors, .3 correlation and negative skew; it becomes .2257, which is almost the same as the one-factor model. Increasing the sample size does not help to improve item recovery in terms of BIAS for the *d*-parameters.

Table 4.3.11. MAD when both correlation and skew are imposed on the latent traits distributions

Item parameters	Correlation	Skew	1000	1500	2000	3000
a_1	0	No	0.1465	0.1149	0.1075	0.0933
	0.3	Positive	0.2430	0.2229	0.2110	0.1999
		Negative	0.2456	0.2354	0.2173	0.2122
	0.6	Positive	0.3515	0.3429	0.3363	0.3239
		Negative	0.3622	0.3384	0.3576	0.3210
a_2	0	No	0.1517	0.1053	0.1145	0.0860
	0.3	Positive	0.2277	0.2194	0.2058	0.2061
		Negative	0.2329	0.2242	0.2147	0.2004
	0.6	Positive	0.3442	0.3440	0.3263	0.3146
		Negative	0.3490	0.3283	0.3385	0.3128
a_3	0	No	0.1443	0.1087	0.1119	0.0782
	0.3	Positive	0.2265	0.2168	0.2035	0.1847
		Negative	0.2219	0.2158	0.2072	0.1942
	0.6	Positive	0.3365	0.3226	0.3173	0.3236
		Negative	0.3307	0.3199	0.3279	0.2951
a_4	0	No	0.1423	0.1044	0.1040	0.0743
	0.3	Positive	0.2331	0.2262	0.2038	0.1974
		Negative	0.2434	0.2162	0.2086	0.2044
	0.6	Positive	0.3447	0.3474	0.3311	0.3121
		Negative	0.3648	0.3408	0.3288	0.3588
a_5	0	No	0.1423	0.0920	0.1201	0.0959
	0.3	Positive	0.2331	0.2185	0.2026	0.2027
		Negative	0.2434	0.2109	0.2199	0.2016
	0.6	Positive	0.3397	0.3279	0.3148	0.3250
		Negative	0.3535	0.3335	0.3024	0.3032
a_6	0	No	0.1478	0.1108	0.1154	0.0868
	0.3	Positive	0.2405	0.2241	0.2059	0.1984
		Negative	0.2366	0.2188	0.2139	0.2036
	0.6	Positive	0.3464	0.3405	0.3076	0.3205
		Negative	0.3545	0.3477	0.3341	0.3295
d	0	No	0.0692	0.0594	0.0462	0.0392
	0.3	Positive	0.2013	0.2181	0.2299	0.2296
		Negative	0.2275	0.2219	0.2180	0.2368
	0.6	Positive	0.3672	0.3576	0.3484	0.3524
		Negative	0.3533	0.3826	0.3774	0.3693

Table 4.3.12. RMSE when both correlation and skew are imposed into the latent traits distributions

Item parameters	Correlation	Skew	1000	1500	2000	3000
a_1	0	No	0.16872	0.13883	0.12603	0.11077
	0.3	Positive	0.25402	0.23006	0.21845	0.20997
		Negative	0.25749	0.24566	0.22746	0.22123
	0.6	Positive	0.36246	0.35228	0.34555	0.33322
		Negative	0.37556	0.34777	0.36783	0.32881
a_2	0	No	0.17021	0.12465	0.13907	0.10079
	0.3	Positive	0.23887	0.22802	0.21277	0.21702
		Negative	0.24379	0.23363	0.22521	0.20858
	0.6	Positive	0.35497	0.35256	0.33727	0.32443
		Negative	0.35974	0.33733	0.34732	0.32279
a_3	0	No	0.16764	0.13034	0.13316	0.09597
	0.3	Positive	0.23695	0.22517	0.21237	0.19643
		Negative	0.23287	0.22581	0.21561	0.20698
	0.6	Positive	0.34726	0.33189	0.326	0.33579
		Negative	0.34624	0.33159	0.33718	0.30498
a_4	0	No	0.16272	0.12503	0.12491	0.09044
	0.3	Positive	0.24128	0.23584	0.21435	0.20556
		Negative	0.2536	0.22742	0.2177	0.21572
	0.6	Positive	0.35652	0.36047	0.33894	0.32107
		Negative	0.37697	0.34968	0.33705	0.36794
a_5	0	No	0.16838	0.11416	0.14503	0.11516
	0.3	Positive	0.23629	0.22647	0.21165	0.21
		Negative	0.24238	0.22185	0.22668	0.20988
	0.6	Positive	0.34988	0.33614	0.3235	0.33836
		Negative	0.36593	0.34331	0.31133	0.31198
a_6	0	No	0.17101	0.13471	0.13501	0.10583
	0.3	Positive	0.24892	0.23503	0.21329	0.20637
		Negative	0.24713	0.22909	0.21924	0.21408
	0.6	Positive	0.35638	0.35053	0.31489	0.33113
		Negative	0.36351	0.35942	0.34556	0.33867
d	0	No	0.08515	0.07341	0.05538	0.04755
	0.3	Positive	0.21524	0.22885	0.23666	0.23601
		Negative	0.24315	0.23492	0.22548	0.24242
	0.6	Positive	0.38004	0.36935	0.35513	0.35789
		Negative	0.36803	0.3905	0.38492	0.37408

CHAPTER 5

SUMMARY AND DISCUSSION

This chapter will give a brief overview of the study, followed by a summary and a detailed discussion of the results. Finally, the implications and limitations of the study will be presented.

5.1. Overview of the Study

This study investigates the influence of multiple factors—such as sample size, the types of latent trait configuration, number of dimensions, correlation between latent traits, and skew in the latent traits distributions—on item parameter recovery in a multi-dimensional item response theory model. In order to examine the influence of combinations of factors, the Markov Chain Monte Carlo method is used, specifically, with a Gibbs sampling technique used to run the simulation study. 60 items and different sample sizes of 1000, 1500, and 2000 for a 3-dimension model, and 1000, 1500, 2000, and 3000 for a 6-dimension model, are used. Two different types of latent trait configuration are used: Approximate simple and Mixed traits. .3 and .6 correlation are used to generate the correlated latent traits. Skew given to the latent traits distributions follows the Pearson's skewness index: +.9 and -.9 for positive and negative skewness, respectively. A total of 84 sets of combination of factors, plus 10 replications for each set, resulted in 840 simulation sets being run.

5.2. Summary of Results

5.2.1. Sample Size

When different sample sizes are used to calibrate the item parameters, it improves the item parameter calibration process. For 3-dimensions, increasing the sample size from 1000 to 1500 and 2000 does not improve the item calibration, which shows that a sample size of 1000 is large enough for 3-dimensions. For 6-dimensions, increasing the sample size from 1000 to 1500, 2000, and 3000 does improve the item calibration. However, the improvement from 2000 to 3000 does not seem significant, which shows that a sample size of 2000 is enough to have an adequate item calibration for 6-dimensions.

5.2.2. Types of Latent Trait Configuration: Approximate Simple (AS) and Mixed Traits (MS)

Different patterns are shown, dependent on the number of dimensions and item parameters. AS has a higher error for 3-dimensions and item parameters compared to MS. AS overestimates for *a*-parameters, but underestimates for the *d*-parameter, compared to MS. However, the types of latent traits do not influence item calibration, whether it's *a*-parameters or the *d*-parameter in the 6-dimension model.

5.2.3. Correlated Latent Traits

Correlated latent traits are only examined in mixed traits. When correlation is implemented into the latent traits, it shows that different behaviors depend on the number of dimensions. When there are 3-dimensions, a high correlation, like a .6 correlation, overestimates but a low correlation, like a .3 correlation, underestimates for *a*-parameters. The *d*-parameter is not influenced by correlated latent traits. When there are 6-dimensions, both low and high correlations overestimate for *a*-parameters. The *d*-parameter is not influenced by the correlated latent traits. In terms of magnitude of bias, the higher the correlation, the bigger the error. Increasing sample size does not improve the item calibration accuracy.

5.2.4. Skewed Latent Traits Distributions

When skewness is implemented into the model, it overestimates for *a*-parameters, regardless of the types of latent traits and number of dimensions. It does not matter whether the skew is negative or positive, they all overestimate the *a*-parameters. The *d*-parameter shows different behaviors, depending on the number of dimensions. In 3-dimensions, both negative and positive skew underestimate the *d*-parameter for an AS structure. However, whereas the positive skew underestimates the *d*-parameter for MS traits, the negative skew overestimates it. In 6-dimensions, the positive skew underestimates and the negative skew overestimates, regardless of the types of latent traits. Increasing the sample size does not improve the item calibration.

5.2.5. Correlated and Skewed Latent Traits Distributions

Only mixed structure is examined. When both correlation and skewness are implemented into the model together, the pattern of influence is similar to when only skewness is implemented into the model. However, it increases the magnitude of bias compared to the model with only one factor included. As the size of correlation increases from .3 to .6, it doubles the size of bias regardless of the number of dimensions and skewed latent traits. Increasing the sample size does not improve the item calibration.

5.3. Discussion

This study explores the interaction effect on item parameter recovery when there are multiple factors combined in the MIRT model. The primary purpose of this study is to find the appropriate sample size in order to have accurate item parameters from the calibration procedures, when there more than one factor is involved in the model.

First, it is clear that an MIRT model with higher dimensions needs to have a larger sample size to get better results on the calibration of item parameters. However, that is only the case for a MIRT model without any other factors such as correlation or skewness imposed on the latent traits distributions. Traditionally, it is common to have a sample size of 1000 when an IRT model has a uni-dimension. From the results of this study, this is enough for a 3-dimension MIRT model to get a satisfactory item parameter recovery. However, a sample size of more than 1000 is required if the number of dimensions increases to 6. This requires a sample size of 2000 to get a satisfactory item parameter recovery. Thus, a larger sample size is recommended if the number of dimensions increases.

Second, the types of latent trait configuration show different behaviors, depending on the number of dimensions. This study shows that the latent trait types have more trouble in approximate simple traits in 3-dimensions. When number of dimensions goes up 6-dimensions, AS and MS show almost identical behavior. For a MIRT model of 3-dimensions, AS shows an overestimated bias for *a*-parameters, and an underestimated bias for the *d*-parameter. It is interesting that MS has a lower bias than AS in 3-dimensions, when AS and MS have almost the same bias in 6-dimensions. Results show that the interaction effect of latent traits types is cancelled when the number of dimensions is higher. This finding suggests that if researchers consider using a MIRT model with AS, using high dimensions will give better results in item parameters recovery, rather than increasing the sample size.

Third, when the correlation between latent traits is in the MIRT model, the result shows that a combination of high correlations and a high number of dimensions in latent traits contributes to a high magnitude of bias. The interaction effect from combining correlation and number of dimensions is more troublesome than with the number of dimensions and types of latent traits combined. It appears that when researchers suspect a correlation between the latent traits, it is not helpful to just increase the sample size. Rather, researchers should find alternative MIRT models that account for the correlated latent traits.

Fourth, when skewed latent traits distributions are in the MIRT model with different types of latent traits configuration, results show that the bias increases. The amount of increased bias is almost the same whether there are 3-dimensions or 6-dimensions. There is also the same amount of increase in bias whether it has AS or MS. The improvement of item parameters calibration in terms of bias is not influenced by increasing sample size. This finding suggests that researchers should correct the latent traits distribution if he/she is suspicious about non-normal distribution on the latent traits distribution.

Fifth, when all factors are combined—correlation, skewness, and number of dimensions—the model with a low correlation and a small number of dimensions has a lower bias than the model with a high correlation and a high number of dimensions.

Overall, increasing the sample size helps to improve the accuracy of item parameter recovery when the latent traits have different types of structures configuration, AS and MS, and a high number of dimensions. However, if the latent traits are correlated, then solely increasing sample size does not improve the accuracy of item parameter estimates. Rather, the number of items should be increased along with increasing the sample size. It is also possible to get a normal distribution of latent traits distributions by selecting the sample group carefully. This is also true for skewed latent traits distributions. The sample group must be selected with careful consideration. Based on the test specifications, the sample group should be selected from a wide range of abilities, from low to high. Having test-takers with a wide range of abilities will prevent the latent trait distribution from being skewed.

5.4. Implications and Limitations

5.4.1. Implications

With the popularity of item response theory (IRT) in the field of measurement, its use is not just limited to measurement but expanded to almost all fields of behavior science research. Since the research in behavioral sciences is getting complicated, it requires IRT models with more than just one dimension. That is where Multi-dimensional item response theory (MIRT) models come in. Most applications based on MIRT models require assumptions that the item parameters are accurately estimated in advance. Traditionally, it is known that a larger sample size gives a better item recovery result. However, there is a lack of research on how the item parameter recovery is influenced if other factors are included in the model, such as correlated latent traits or skewed latent traits distributions. It could be just one factor, or it could be a combination of more than one factor. In this study, the interaction effect of combined factors on item recovery is explored. It clearly shows that increasing sample size does not improve item parameter calibration if there is more than one factor involved. Rather, correlated or skewed latent traits distribution should be corrected before running a calibration program, in order to have more accurate results for item parameter calibration. This finding is helpful for researchers in that it will save costs associated with recruiting a larger sample size than is necessary.

5.4.2. Limitations

This study has several limitations. First, due to computing and time resources, the replications for each condition is limited to 10, which might contribute to some estimation errors of the MCMC simulation. An MCMC simulation suggests having 50 replications to have a stable estimation, if it is necessary. A future study with more replications would yield more affirmative results. Second, it is assumed that all the latent traits in the MIRT models have the same distribution in order to make interpretation clear. However, it might be not practical to have such an assumption in a real situation. Considering different distributions on each latent trait will be the next step for a future study.

APPENDIX

APPENDIX

Table A.1.1. Heidelberger and Welch's Convergence Diagnostic: a_1-parameter

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
1	passed	1	0.1442	passed	0.1639	0.0067
2	passed	8801	0.3837	passed	0.0502	0.0021
3	passed	1	0.1708	passed	0.1289	0.0058
4	passed	1	0.1972	passed	0.2258	0.0077
5	passed	1	0.2051	passed	0.0650	0.0032
6	passed	1	0.0822	passed	0.2272	0.0076
7	passed	1	0.0704	passed	0.1701	0.0045
8	passed	4401	0.0520	passed	0.1576	0.0069
9	passed	1	0.1214	passed	0.1415	0.0061
10	passed	1	0.1676	passed	0.1408	0.0082
11	passed	1	0.7937	passed	0.0543	0.0017
12	passed	1	0.8795	passed	0.2147	0.0038
13	passed	1	0.9896	passed	0.2221	0.0040
14	passed	1	0.4052	passed	0.0441	0.0014
15	passed	1	0.5056	passed	0.0709	0.0019
16	passed	1	0.8939	passed	0.2111	0.0052
17	passed	1	0.8323	passed	0.0704	0.0022
18	passed	1	0.9746	passed	0.2887	0.0055
19	passed	1	0.9890	passed	0.1642	0.0064
20	passed	1	0.9834	passed	0.1682	0.0063
21	passed	1	0.8823	passed	0.1950	0.0083
22	passed	1	0.5622	passed	0.1570	0.0070
23	passed	1	0.8627	passed	0.2056	0.0088
24	passed	1	0.8851	passed	0.2555	0.0103
25	passed	1	0.9247	passed	0.2024	0.0125
26	passed	1	0.9384	passed	0.1585	0.0048
27	passed	1	0.5585	passed	0.0764	0.0052
28	passed	1	0.9005	passed	0.1347	0.0086
29	passed	1	0.9085	passed	0.1105	0.0064
30	passed	1	0.8991	passed	0.2837	0.0153

Table A.1.1 (cont'd)

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
31	passed	1	0.1405	passed	0.1080	0.0050
32	passed	1	0.1368	passed	0.3248	0.0147
33	passed	1	0.6697	passed	0.0779	0.0020
34	passed	1	0.0722	passed	0.1757	0.0062
35	passed	1	0.1455	passed	0.0602	0.0032
36	passed	1	0.1289	passed	0.1385	0.0071
37	passed	4401	0.0676	passed	0.1523	0.0056
38	passed	4401	0.1081	passed	0.1934	0.0077
39	passed	1	0.0541	passed	0.1052	0.0047
40	passed	1	0.1469	passed	0.2060	0.0072
41	passed	6601	0.2188	passed	1.2414	0.0067
42	passed	1	0.1921	passed	0.7931	0.0033
43	passed	1	0.0592	passed	0.4075	0.0025
44	passed	1	0.1926	passed	1.2614	0.0064
45	passed	1	0.8881	passed	0.7710	0.0035
46	passed	1	0.4568	passed	0.8599	0.0039
47	passed	1	0.8898	passed	0.6345	0.0027
48	passed	1	0.2415	passed	1.0436	0.0041
49	passed	1	0.5545	passed	0.3498	0.0021
50	passed	1	0.0824	passed	1.1286	0.0065
51	passed	2201	0.0644	passed	0.1981	0.0039
52	passed	2201	0.1180	passed	0.1256	0.0040
53	passed	1	0.5720	passed	0.0372	0.0012
54	passed	2201	0.0770	passed	0.1985	0.0082
55	passed	1	0.0509	passed	0.3401	0.0137
56	passed	4401	0.0500	passed	0.1273	0.0083
57	passed	2201	0.2062	passed	0.1021	0.0043
58	passed	2201	0.1348	passed	0.1106	0.0061
59	passed	1	0.0552	passed	0.1467	0.0065
60	failed	NA	0.0144	<NA>	NA	NA

Table A.1.2. Heidelberger and Welch's Convergence Diagnostic: a_2-parameter

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
1	passed	1	0.4900	passed	0.1398	0.0051
2	passed	1	0.3567	passed	0.0449	0.0018
3	passed	1	0.0847	passed	0.2098	0.0060
4	passed	1	0.2916	passed	0.1834	0.0057
5	passed	1	0.1021	passed	0.1714	0.0054
6	passed	1	0.2743	passed	0.0600	0.0030
7	passed	2201	0.0680	passed	0.1779	0.0043
8	passed	1	0.0546	passed	0.2037	0.0056
9	passed	1	0.1991	passed	0.1110	0.0049
10	passed	1	0.0974	passed	0.2746	0.0103
11	passed	1	0.8854	passed	0.1397	0.0048
12	passed	1	0.5049	passed	0.1468	0.0043
13	passed	1	0.6684	passed	0.1864	0.0039
14	passed	1	0.8327	passed	0.0837	0.0040
15	passed	1	0.5037	passed	0.0659	0.0022
16	passed	1	0.4983	passed	0.1433	0.0051
17	passed	1	0.7507	passed	0.0763	0.0019
18	passed	1	0.7597	passed	0.0650	0.0033
19	passed	1	0.9326	passed	0.2022	0.0076
20	passed	1	0.8574	passed	0.1883	0.0070
21	passed	1	0.9903	passed	0.8620	0.0035
22	passed	2201	0.3543	passed	0.7216	0.0032
23	passed	1	0.0525	passed	0.9079	0.0031
24	passed	1	0.7157	passed	1.0763	0.0034
25	passed	1	0.8079	passed	1.3329	0.0063
26	passed	1	0.6680	passed	0.5539	0.0026
27	passed	1	0.6465	passed	0.8473	0.0033
28	passed	1	0.1866	passed	0.9922	0.0052
29	passed	1	0.1764	passed	0.8136	0.0029
30	passed	1	0.8469	passed	1.7251	0.0121

Table A.1.2 (cont'd)

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
31	passed	1	0.2422	passed	0.2146	0.0075
32	passed	1	0.2488	passed	0.1892	0.0134
33	passed	1	0.4696	passed	0.0543	0.0017
34	passed	1	0.2085	passed	0.1909	0.0065
35	passed	1	0.1949	passed	0.1151	0.0061
36	passed	1	0.3144	passed	0.1575	0.0082
37	passed	1	0.1547	passed	0.1309	0.0046
38	passed	1	0.3931	passed	0.1528	0.0075
39	passed	1	0.2314	passed	0.0653	0.0035
40	passed	1	0.3368	passed	0.1920	0.0080
41	passed	1	0.9508	passed	0.2529	0.0121
42	passed	1	0.7858	passed	0.1380	0.0069
43	passed	1	0.6484	passed	0.0722	0.0028
44	passed	1	0.9602	passed	0.2707	0.0126
45	passed	1	0.7523	passed	0.2023	0.0079
46	passed	1	0.9605	passed	0.1594	0.0071
47	passed	1	0.9031	passed	0.1614	0.0063
48	passed	1	0.9542	passed	0.1245	0.0076
49	passed	1	0.9107	passed	0.0698	0.0028
50	passed	1	0.9583	passed	0.1381	0.0095
51	passed	1	0.9925	passed	0.0456	0.0015
52	passed	1	0.1668	passed	0.1254	0.0034
53	passed	1	0.7793	passed	0.0856	0.0024
54	passed	1	0.1752	passed	0.2110	0.0052
55	passed	1	0.4989	passed	0.2251	0.0085
56	passed	1	0.1174	passed	0.1266	0.0051
57	passed	1	0.9240	passed	0.1584	0.0034
58	passed	1	0.3858	passed	0.1622	0.0051
59	passed	1	0.4570	passed	0.0754	0.0032
60	passed	1	0.5378	passed	0.0862	0.0028

Table A.1.3. Heidelberger and Welch's Convergence Diagnostic: a_3-parameter

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
1	passed	1	0.3273	passed	0.1843	0.0049
2	passed	1	0.1127	passed	0.0310	0.0008
3	passed	1	0.2428	passed	0.2680	0.0043
4	passed	1	0.3839	passed	0.1292	0.0040
5	passed	1	0.3759	passed	0.0641	0.0026
6	passed	1	0.8066	passed	0.2080	0.0052
7	passed	1	0.7666	passed	0.2092	0.0039
8	passed	1	0.2356	passed	0.2039	0.0051
9	passed	1	0.2503	passed	0.1300	0.0041
10	passed	1	0.2929	passed	0.1998	0.0075
11	passed	1	0.0906	passed	0.6696	0.0035
12	passed	1	0.5316	passed	0.6718	0.0037
13	passed	1	0.7671	passed	0.5099	0.0029
14	passed	1	0.2323	passed	0.7219	0.0044
15	passed	4401	0.2848	passed	0.4731	0.0034
16	passed	1	0.3261	passed	0.8160	0.0039
17	passed	1	0.1415	passed	0.4161	0.0026
18	passed	1	0.7838	passed	0.8923	0.0043
19	passed	1	0.4407	passed	1.1336	0.0050
20	passed	1	0.3758	passed	1.0691	0.0043
21	passed	1	0.6809	passed	0.1312	0.0056
22	passed	1	0.1968	passed	0.1310	0.0048
23	passed	1	0.6898	passed	0.1795	0.0063
24	passed	1	0.8118	passed	0.1196	0.0064
25	passed	1	0.6931	passed	0.1750	0.0094
26	passed	1	0.5298	passed	0.0947	0.0036
27	passed	1	0.8980	passed	0.1751	0.0060
28	passed	1	0.7576	passed	0.1575	0.0072
29	passed	1	0.7461	passed	0.0940	0.0045
30	passed	1	0.9191	passed	0.3664	0.0121

Table A.1.3 (cont'd)

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
31	passed	1	0.7805	passed	0.1207	0.0067
32	passed	1	0.3288	passed	0.4486	0.0178
33	passed	1	0.0688	passed	0.0517	0.0017
34	passed	1	0.3841	passed	0.0795	0.0045
35	passed	1	0.4351	passed	0.2063	0.0071
36	passed	1	0.6514	passed	0.1878	0.0089
37	passed	1	0.6334	passed	0.1695	0.0055
38	passed	1	0.5502	passed	0.1831	0.0090
39	passed	1	0.1899	passed	0.1535	0.0065
40	passed	1	0.5580	passed	0.2825	0.0096
41	passed	1	0.6049	passed	0.1525	0.0067
42	passed	1	0.9888	passed	0.2754	0.0041
43	passed	1	0.6072	passed	0.0791	0.0023
44	passed	1	0.8013	passed	0.1710	0.0068
45	passed	1	0.8216	passed	0.0935	0.0038
46	passed	1	0.8315	passed	0.1783	0.0053
47	passed	1	0.8043	passed	0.0477	0.0017
48	passed	1	0.6360	passed	0.1670	0.0067
49	passed	1	0.3797	passed	0.0423	0.0010
50	passed	1	0.9174	passed	0.1648	0.0062
51	passed	1	0.6593	passed	0.1320	0.0033
52	passed	1	0.2957	passed	0.0590	0.0016
53	passed	1	0.1351	passed	0.0658	0.0021
54	passed	1	0.8000	passed	0.2865	0.0054
55	passed	1	0.8433	passed	0.1534	0.0075
56	passed	1	0.6246	passed	0.2531	0.0057
57	passed	1	0.5458	passed	0.0855	0.0023
58	passed	1	0.9998	passed	0.0796	0.0033
59	passed	1	0.8554	passed	0.1059	0.0039
60	passed	1	0.9950	passed	0.0777	0.0025

Table A.1.4. Heidelberger and Welch's Convergence Diagnostic: a_4-parameter

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
1	passed	1	0.8384	passed	0.0857	0.0042
2	passed	1	0.4861	passed	0.1386	0.0045
3	passed	1	0.9591	passed	0.1648	0.0061
4	passed	1	0.9535	passed	0.1311	0.0069
5	passed	1	0.9088	passed	0.1973	0.0067
6	passed	1	0.3075	passed	0.1082	0.0059
7	passed	1	0.4810	passed	0.1132	0.0047
8	passed	1	0.8041	passed	0.1796	0.0066
9	passed	1	0.7099	passed	0.2218	0.0077
10	passed	1	0.8828	passed	0.2129	0.0107
11	passed	1	0.2346	passed	0.0457	0.0013
12	passed	1	0.4755	passed	0.0783	0.0026
13	passed	1	0.7504	passed	0.0864	0.0023
14	passed	1	0.5600	passed	0.1215	0.0039
15	passed	1	0.2157	passed	0.0494	0.0016
16	passed	1	0.5655	passed	0.2358	0.0046
17	passed	1	0.3445	passed	0.0848	0.0020
18	passed	1	0.7231	passed	0.1373	0.0041
19	passed	1	0.9106	passed	0.1711	0.0059
20	passed	1	0.2778	passed	0.1132	0.0045
21	passed	1	0.9017	passed	0.1405	0.0050
22	passed	1	0.7041	passed	0.2745	0.0048
23	passed	1	0.4705	passed	0.1718	0.0057
24	passed	1	0.3635	passed	0.1171	0.0053
25	passed	1	0.9047	passed	0.2765	0.0078
26	passed	1	0.1473	passed	0.0947	0.0030
27	passed	1	0.2132	passed	0.0983	0.0045
28	passed	1	0.2948	passed	0.1237	0.0052
29	passed	1	0.8598	passed	0.0761	0.0033
30	passed	1	0.7905	passed	0.1751	0.0085

Table A.1.4. (cont'd)

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
31	passed	1	0.6487	passed	0.1878	0.0070
32	passed	1	0.2903	passed	0.2787	0.0157
33	passed	1	0.2784	passed	0.0706	0.0024
34	passed	1	0.2801	passed	0.1119	0.0058
35	passed	1	0.6944	passed	0.1777	0.0072
36	passed	1	0.8458	passed	0.1107	0.0065
37	passed	1	0.2277	passed	0.0798	0.0036
38	passed	1	0.3503	passed	0.1410	0.0069
39	passed	1	0.5789	passed	0.1025	0.0049
40	passed	1	0.2582	passed	0.1466	0.0073
41	passed	2201	0.1445	passed	0.2391	0.0120
42	passed	1	0.1182	passed	0.1018	0.0051
43	passed	2201	0.1314	passed	0.0923	0.0030
44	passed	1	0.1352	passed	0.3369	0.0097
45	passed	1	0.1254	passed	0.0832	0.0045
46	passed	2201	0.1085	passed	0.1250	0.0063
47	passed	1	0.3517	passed	0.0934	0.0047
48	passed	1	0.1384	passed	0.2634	0.0082
49	failed	NA	0.0209	<NA>	NA	NA
50	passed	2201	0.0792	passed	0.2110	0.0104
51	passed	1	0.5177	passed	0.3994	0.0026
52	passed	1	0.7741	passed	0.5317	0.0031
53	passed	1	0.2188	passed	0.5265	0.0026
54	passed	1	0.3363	passed	1.0391	0.0039
55	passed	1	0.7575	passed	1.5751	0.0093
56	passed	1	0.3442	passed	1.0872	0.0046
57	passed	1	0.1699	passed	0.5742	0.0027
58	passed	1	0.1757	passed	0.8852	0.0039
59	passed	1	0.4223	passed	0.7819	0.0031
60	failed	NA	0.0053	<NA>	NA	NA

Table A.1.5. Heidelberger and Welch's Convergence Diagnostic: a_5-parameter

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
1	passed	1	0.3111	passed	0.9233	0.0051
2	passed	1	0.0617	passed	0.4825	0.0029
3	passed	1	0.1048	passed	0.8559	0.0032
4	passed	8801	0.7956	passed	0.9417	0.0043
5	passed	1	0.2455	passed	0.7335	0.0036
6	passed	1	0.2840	passed	0.8501	0.0032
7	passed	1	0.2068	passed	0.6322	0.0026
8	passed	1	0.1675	passed	0.8743	0.0049
9	passed	1	0.1783	passed	0.8883	0.0036
10	passed	1	0.6803	passed	1.3362	0.0063
11	passed	1	0.6104	passed	0.1136	0.0035
12	passed	1	0.4848	passed	0.0640	0.0023
13	passed	1	0.3505	passed	0.0710	0.0020
14	passed	1	0.2855	passed	0.0763	0.0029
15	passed	1	0.2674	passed	0.1098	0.0030
16	passed	1	0.6172	passed	0.1134	0.0036
17	passed	1	0.8342	passed	0.1104	0.0022
18	passed	1	0.4870	passed	0.1225	0.0042
19	passed	1	0.7277	passed	0.0783	0.0042
20	passed	1	0.7492	passed	0.0766	0.0038
21	passed	1	0.4253	passed	0.0934	0.0042
22	passed	1	0.6548	passed	0.1479	0.0044
23	passed	1	0.4247	passed	0.1476	0.0058
24	passed	1	0.2799	passed	0.1271	0.0065
25	passed	1	0.3366	passed	0.1203	0.0066
26	passed	2201	0.0962	passed	0.0846	0.0036
27	passed	1	0.3677	passed	0.1345	0.0055
28	passed	1	0.7694	passed	0.0839	0.0048
29	passed	1	0.5978	passed	0.1935	0.0061
30	passed	1	0.7086	passed	0.3203	0.0109

Table A.1.5. (cont'd)

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
31	passed	4401	0.0607	passed	0.2183	0.0096
32	passed	2201	0.0524	passed	0.2524	0.0195
33	passed	1	0.1119	passed	0.0316	0.0011
34	passed	4401	0.0751	passed	0.2844	0.0085
35	passed	6601	0.1056	passed	0.1603	0.0085
36	passed	4401	0.0685	passed	0.2305	0.0103
37	passed	1	0.0672	passed	0.1937	0.0065
38	passed	6601	0.1083	passed	0.2315	0.0110
39	passed	4401	0.0685	passed	0.2543	0.0087
40	passed	4401	0.1258	passed	0.1578	0.0099
41	passed	1	0.2125	passed	0.2331	0.0091
42	passed	1	0.3345	passed	0.0936	0.0042
43	passed	1	0.6440	passed	0.1624	0.0033
44	passed	1	0.3572	passed	0.1022	0.0070
45	passed	1	0.3983	passed	0.0806	0.0045
46	passed	1	0.2448	passed	0.2372	0.0071
47	passed	1	0.3762	passed	0.0950	0.0034
48	passed	1	0.1572	passed	0.2000	0.0082
49	passed	1	0.9481	passed	0.1178	0.0026
50	passed	1	0.3442	passed	0.1580	0.0081
51	passed	1	0.0670	passed	0.0953	0.0033
52	passed	1	0.6455	passed	0.2184	0.0044
53	passed	1	0.7212	passed	0.1691	0.0048
54	passed	1	0.8314	passed	0.2338	0.0078
55	passed	1	0.6786	passed	0.2056	0.0116
56	passed	1	0.9270	passed	0.2319	0.0093
57	passed	1	0.9462	passed	0.0524	0.0023
58	passed	1	0.9435	passed	0.1342	0.0066
59	passed	1	0.9124	passed	0.0896	0.0050
60	passed	1	0.6614	passed	0.1320	0.0047

Table A.1.6. Heidelberger and Welch's Convergence Diagnostic: a_6-parameter

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
1	passed	1	0.0775	passed	0.1321	0.0068
2	passed	1	0.0519	passed	0.0855	0.0036
3	passed	4401	0.1454	passed	0.0973	0.0067
4	passed	4401	0.0583	passed	0.2228	0.0088
5	passed	4401	0.0609	passed	0.1148	0.0075
6	passed	1	0.0762	passed	0.1157	0.0066
7	passed	4401	0.1225	passed	0.1266	0.0051
8	passed	2201	0.0743	passed	0.0859	0.0058
9	failed	NA	0.0417	<NA>	NA	NA
10	passed	8801	0.1337	passed	0.1666	0.0148
11	passed	1	0.7935	passed	0.1572	0.0053
12	passed	1	0.3321	passed	0.1326	0.0056
13	passed	1	0.6442	passed	0.0600	0.0024
14	passed	1	0.4623	passed	0.2844	0.0064
15	passed	1	0.6926	passed	0.0569	0.0022
16	passed	1	0.5607	passed	0.1230	0.0068
17	passed	1	0.6998	passed	0.1128	0.0030
18	passed	1	0.5527	passed	0.1418	0.0068
19	passed	1	0.5657	passed	0.1532	0.0090
20	passed	1	0.2704	passed	0.1088	0.0078
21	passed	1	0.5207	passed	0.1212	0.0057
22	passed	1	0.1000	passed	0.2039	0.0064
23	passed	1	0.5022	passed	0.2133	0.0074
24	passed	1	0.4065	passed	0.2556	0.0090
25	passed	1	0.5171	passed	0.3417	0.0112
26	passed	1	0.5130	passed	0.0359	0.0014
27	passed	1	0.2682	passed	0.1036	0.0060
28	passed	1	0.8106	passed	0.3651	0.0083
29	passed	1	0.1690	passed	0.1472	0.0061
30	passed	1	0.4291	passed	0.4453	0.0143

Table A.1.6. (cont'd)

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
31	passed	1	0.0522	passed	0.8905	0.0033
32	passed	2201	0.1351	passed	1.9137	0.0152
33	passed	1	0.2599	passed	0.3255	0.0024
34	passed	1	0.2116	passed	0.8337	0.0040
35	passed	4401	0.0913	passed	0.7995	0.0035
36	passed	1	0.9670	passed	0.9977	0.0040
37	passed	1	0.3390	passed	0.6473	0.0029
38	passed	1	0.6223	passed	1.0091	0.0031
39	passed	1	0.1249	passed	0.7710	0.0039
40	passed	1	0.5696	passed	1.0333	0.0065
41	passed	1	0.9036	passed	0.3593	0.0087
42	passed	1	0.3645	passed	0.2061	0.0058
43	passed	1	0.9814	passed	0.0831	0.0023
44	passed	1	0.8682	passed	0.1466	0.0080
45	passed	1	0.2253	passed	0.0792	0.0035
46	passed	1	0.8282	passed	0.2029	0.0053
47	passed	1	0.3333	passed	0.1226	0.0039
48	passed	1	0.8901	passed	0.2383	0.0072
49	passed	1	0.8399	passed	0.0403	0.0011
50	passed	1	0.7208	passed	0.1279	0.0069
51	passed	1	0.6865	passed	0.1132	0.0031
52	passed	1	0.3310	passed	0.1632	0.0044
53	passed	1	0.1952	passed	0.0911	0.0040
54	passed	1	0.7648	passed	0.2428	0.0079
55	passed	1	0.4198	passed	0.2921	0.0134
56	passed	1	0.2595	passed	0.1943	0.0087
57	passed	1	0.7476	passed	0.0378	0.0016
58	passed	1	0.3425	passed	0.1708	0.0076
59	passed	1	0.4455	passed	0.1377	0.0061
60	passed	1	0.5871	passed	0.0772	0.0036

Table A.1.7. Heidelberger and Welch's Convergence Diagnostic: *d*-parameter

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
1	passed	1	0.9248	passed	1.8052	0.0040
2	passed	1	0.7610	passed	-0.0532	0.0014
3	passed	1	0.7914	passed	-0.2077	0.0019
4	passed	1	0.2363	passed	-0.5214	0.0023
5	passed	1	0.2884	passed	-0.4003	0.0018
6	passed	1	0.3306	passed	-0.5679	0.0023
7	passed	1	0.7550	passed	-0.4760	0.0020
8	passed	1	0.1794	passed	-0.7079	0.0040
9	passed	1	0.3384	passed	-0.9386	0.0031
10	passed	1	0.3201	passed	-1.3475	0.0046
11	passed	1	0.0621	passed	1.1477	0.0023
12	passed	1	0.7794	passed	0.5408	0.0017
13	passed	1	0.6429	passed	0.3588	0.0020
14	passed	1	0.4652	passed	0.4011	0.0023
15	passed	1	0.8312	passed	0.2872	0.0017
16	passed	1	0.3772	passed	0.1790	0.0022
17	passed	1	0.3167	failed	-0.0419	0.0044
18	passed	1	0.1589	passed	-0.1822	0.0026
19	passed	1	0.0851	passed	-0.4961	0.0036
20	passed	1	0.2886	passed	-0.6791	0.0041
21	passed	1	0.3598	passed	1.1359	0.0028
22	passed	1	0.3049	passed	0.9683	0.0042
23	passed	1	0.7184	passed	0.7617	0.0038
24	passed	1	0.4188	passed	0.6249	0.0026
25	passed	1	0.6544	passed	-0.7869	0.0050
26	passed	1	0.7085	passed	-0.2982	0.0018
27	passed	1	0.5311	passed	-0.5098	0.0024
28	passed	1	0.0803	passed	-0.6942	0.0055
29	passed	1	0.5646	passed	-0.7061	0.0034
30	passed	1	0.6575	passed	-2.1652	0.0103

Table A.1.7. (cont'd)

Items	Stationarity Test	Start	P-value	Halfwidth Test	Mean	Halfwidth
31	passed	1	0.1513	passed	1.1994	0.0026
32	passed	2201	0.1029	passed	1.9186	0.0104
33	failed	NA	0.0021	<NA>	NA	NA
34	passed	1	0.1325	passed	0.2784	0.0022
35	passed	8801	0.1823	passed	0.0436	0.0025
36	passed	1	0.0728	passed	0.0395	0.0029
37	passed	1	0.4140	passed	-0.1569	0.0021
38	passed	1	0.0562	passed	-0.8579	0.0026
39	passed	1	0.3011	passed	-0.8446	0.0017
40	passed	4401	0.1305	passed	-1.0531	0.0096
41	passed	1	0.1420	passed	1.4298	0.0039
42	passed	1	0.1577	passed	0.7137	0.0021
43	passed	1	0.8701	passed	0.3740	0.0016
44	passed	1	0.4816	passed	0.5017	0.0059
45	passed	1	0.8048	passed	0.3044	0.0021
46	passed	1	0.9666	failed	-0.0189	0.0020
47	passed	1	0.7360	passed	-0.1252	0.0016
48	passed	1	0.1475	passed	-0.1840	0.0023
49	passed	1	0.2165	passed	-0.0847	0.0012
50	passed	1	0.6451	passed	-1.1910	0.0063
51	passed	1	0.4431	passed	0.4207	0.0014
52	passed	1	0.5711	passed	0.4594	0.0022
53	passed	1	0.6277	passed	0.3294	0.0014
54	passed	1	0.6573	passed	0.4527	0.0032
55	passed	1	0.4995	passed	0.5587	0.0056
56	passed	1	0.1381	passed	0.2048	0.0021
57	passed	1	0.6640	passed	-0.0334	0.0018
58	passed	1	0.2890	passed	-0.2808	0.0021
59	passed	1	0.6563	passed	-0.3581	0.0019
60	passed	1	0.3434	passed	-0.3982	0.0023

Table A.2. 1. Geweke's Z-score

Item	a_1	a_2	a_3	a_4	a_5	a_6	d
1	1.2684	-2.2061	0.5387	0.2543	0.7505	1.6152	0.5404
2	0.8198	-1.1491	1.3394	0.6109	-1.4409	0.4359	-0.0797
3	1.1488	-2.3534	0.8052	1.5895	-1.4047	1.8357	-0.2742
4	1.3278	-2.5092	0.7329	0.5589	-0.1756	2.0876	0.5249
5	0.7640	-3.5548	0.8827	0.7436	0.0197	2.3901	0.0838
6	2.3268	-1.6575	0.6066	-0.1853	-1.6516	1.2716	-0.0478
7	1.5009	-1.9750	1.2068	0.5158	-1.5500	2.3976	-0.6697
8	1.3062	-3.2312	0.3156	0.5285	-2.9689	2.1849	1.1820
9	0.8921	-2.0617	0.3080	-0.3810	-0.3362	3.0078	0.1300
10	1.2711	-3.6032	0.6145	0.7171	0.4523	2.9258	0.0248
11	-0.4740	-0.8094	2.1629	1.7458	-1.2479	-0.1922	1.0436
12	-0.3858	0.2128	1.1112	-0.1536	-0.3504	-0.3221	-0.6892
13	0.2475	1.7524	-0.1811	-0.7205	-0.1001	0.5925	-0.7560
14	0.3892	-0.0242	-2.0751	1.2241	-2.3642	0.3590	-1.0243
15	-0.9393	0.9163	-1.2904	1.1393	-1.2072	-0.1175	-0.4807
16	1.0436	1.3765	-1.9252	1.3216	-2.5827	0.1225	-0.9919
17	0.5820	0.3020	-1.4122	0.5939	-1.1244	0.3885	-0.6820
18	0.5422	0.7274	1.1886	0.1160	-2.4029	0.3778	-2.1434
19	0.4158	0.6139	-1.9429	1.0971	-2.1644	0.1458	0.2930
20	-0.3669	0.7525	1.5577	0.9327	-1.5352	-0.4552	-1.8313
21	-2.1450	-0.1815	-0.8239	0.5877	1.4657	1.5479	-0.5681
22	-1.2775	-3.2277	-1.2714	0.9894	0.6789	2.9973	-2.7610
23	-2.7402	-2.1957	-0.2052	2.4467	1.7355	1.6357	-1.0770
24	-1.5634	0.2541	-0.7805	-1.3888	2.2053	2.4522	0.9058
25	-1.6566	0.4288	-0.8126	0.8734	2.0408	2.0909	-1.2165
26	-1.6212	-0.9506	-0.4867	0.6011	2.3581	1.2245	-0.8022
27	-2.8851	-0.3538	-0.8161	1.5896	1.8637	1.5609	-0.5495
28	-3.0024	-1.6878	0.0708	1.7112	0.6105	1.3291	1.5138
29	-2.1789	-1.3104	-0.9396	1.3913	1.6996	1.7909	-0.1070
30	-2.2778	0.9394	0.3624	1.0308	1.6574	2.1224	-1.4205

Table A.2.1. (cont'd)

Item	a_1	a_2	a_3	a_4	a_5	a_6	d
31	0.1678	-2.5702	-0.0885	-0.8382	-0.9783	1.6324	1.1548
32	-0.0277	-2.1961	1.3086	-1.2008	-1.8658	2.3903	2.2688
33	-0.3205	-0.9302	0.8677	-2.4352	-1.0420	2.3969	2.6859
34	-0.8227	-1.2366	0.6951	-1.4195	-1.1550	0.6767	1.3804
35	0.4982	-2.3341	0.6859	-1.1272	-0.9521	1.0472	1.6419
36	-0.4855	-2.0135	0.0016	-0.4144	-1.4494	0.3119	1.4489
37	1.1050	-2.3846	-1.2996	-1.9396	-1.2063	1.6314	0.7137
38	0.8087	-0.8641	-0.1113	-1.5132	-2.2510	0.7292	0.0170
39	0.5295	-2.4461	0.8458	-1.1646	-2.1291	1.5168	-0.2428
40	-0.1855	-1.0243	0.4129	-1.8310	-1.2711	0.0369	0.9766
41	-0.6586	1.7794	-0.6834	-4.3514	-0.9686	1.7381	-0.7062
42	-0.2370	2.6316	-0.2186	-3.4110	-1.1005	0.6683	-0.6826
43	1.2948	1.6047	-1.3724	-2.9975	0.3032	0.9923	-0.5413
44	0.4015	1.9269	-1.3193	-2.7121	-1.5692	1.1988	0.2489
45	-0.9935	2.2310	-1.1292	-3.1834	-0.4129	0.7945	-0.6069
46	-0.4188	2.1789	-0.4423	-3.6156	-0.8354	1.3713	0.5700
47	-0.0810	0.5557	-0.3994	-2.6835	-0.0723	1.3062	-1.2262
48	1.0662	1.7195	-0.4714	-3.0514	-0.8465	0.5978	1.5022
49	0.3858	0.8214	-1.2955	-2.7006	-0.1919	1.0801	-0.0473
50	1.7297	2.0107	-0.1287	-3.3321	-0.6108	0.9165	-0.9695
51	2.7123	0.3227	-0.3417	-0.1434	-2.9107	0.7698	-0.7996
52	3.0361	-1.3009	-1.6603	0.4125	-1.6686	2.0707	-0.3658
53	0.6232	-0.5849	0.9672	1.1769	-0.0656	0.6545	0.1518
54	3.1589	-1.3031	-0.7525	0.8369	-1.3537	1.1618	-1.1000
55	3.2102	-1.0141	-0.9114	-1.0415	-1.3377	1.8209	-1.6240
56	4.3559	-2.8829	-0.2860	-0.2098	-1.4810	1.9488	-1.6734
57	3.3913	0.2152	-0.7293	1.5084	-0.4950	0.8239	0.7482
58	4.1752	-1.9060	-0.1184	-1.9989	-1.2315	1.7296	-0.7038
59	3.4294	-0.4988	-0.6279	-0.3578	-0.5170	1.4703	-1.8652
60	3.8606	0.0203	-0.2590	-0.9981	-2.0552	1.1665	-1.4288

Table A.3. 1. MCMC standard error

Item	a_1	a_2	a_3	a_4	a_5	a_6	d
1	0.0045	0.0016	0.0030	0.0011	0.0018	0.0037	0.0013
2	0.0016	0.0010	0.0007	0.0017	0.0013	0.0018	0.0008
3	0.0039	0.0028	0.0026	0.0010	0.0030	0.0034	0.0010
4	0.0041	0.0020	0.0021	0.0022	0.0028	0.0046	0.0012
5	0.0019	0.0022	0.0012	0.0015	0.0026	0.0036	0.0015
6	0.0042	0.0012	0.0027	0.0027	0.0019	0.0029	0.0015
7	0.0029	0.0028	0.0020	0.0021	0.0019	0.0026	0.0008
8	0.0046	0.0027	0.0033	0.0021	0.0014	0.0030	0.0016
9	0.0041	0.0019	0.0025	0.0026	0.0018	0.0052	0.0015
10	0.0048	0.0050	0.0044	0.0024	0.0033	0.0075	0.0030
11	0.0004	0.0029	0.0019	0.0009	0.0026	0.0036	0.0021
12	0.0013	0.0030	0.0018	0.0015	0.0013	0.0040	0.0012
13	0.0006	0.0025	0.0009	0.0013	0.0014	0.0019	0.0013
14	0.0007	0.0017	0.0031	0.0014	0.0018	0.0056	0.0016
15	0.0012	0.0011	0.0021	0.0008	0.0018	0.0018	0.0011
16	0.0015	0.0035	0.0022	0.0020	0.0018	0.0049	0.0017
17	0.0008	0.0008	0.0014	0.0014	0.0009	0.0024	0.0022
18	0.0012	0.0018	0.0018	0.0017	0.0030	0.0054	0.0028
19	0.0006	0.0040	0.0030	0.0019	0.0017	0.0069	0.0021
20	0.0011	0.0039	0.0034	0.0024	0.0018	0.0065	0.0039
21	0.0023	0.0009	0.0031	0.0018	0.0013	0.0026	0.0011
22	0.0011	0.0028	0.0034	0.0020	0.0015	0.0035	0.0023
23	0.0020	0.0023	0.0041	0.0021	0.0027	0.0028	0.0021
24	0.0035	0.0018	0.0035	0.0019	0.0029	0.0039	0.0010
25	0.0024	0.0015	0.0061	0.0027	0.0028	0.0045	0.0017
26	0.0011	0.0009	0.0025	0.0016	0.0023	0.0007	0.0006
27	0.0011	0.0009	0.0027	0.0022	0.0014	0.0030	0.0014
28	0.0026	0.0027	0.0052	0.0023	0.0017	0.0028	0.0028
29	0.0014	0.0016	0.0020	0.0006	0.0015	0.0032	0.0019
30	0.0043	0.0014	0.0084	0.0033	0.0046	0.0067	0.0024

Table A.3.1. (cont'd)

Item	a_1	a_2	a_3	a_4	a_5	a_6	d
31	0.0025	0.0047	0.0045	0.0042	0.0044	0.0025	0.0016
32	0.0069	0.0067	0.0143	0.0071	0.0076	0.0107	0.0074
33	0.0006	0.0007	0.0014	0.0014	0.0005	0.0012	0.0013
34	0.0029	0.0039	0.0039	0.0026	0.0041	0.0018	0.0012
35	0.0013	0.0031	0.0049	0.0033	0.0038	0.0023	0.0014
36	0.0030	0.0040	0.0064	0.0027	0.0047	0.0011	0.0020
37	0.0026	0.0028	0.0049	0.0020	0.0025	0.0016	0.0017
38	0.0036	0.0035	0.0069	0.0028	0.0052	0.0010	0.0015
39	0.0025	0.0018	0.0053	0.0023	0.0049	0.0021	0.0008
40	0.0030	0.0033	0.0078	0.0033	0.0043	0.0019	0.0047
41	0.0041	0.0030	0.0024	0.0067	0.0049	0.0029	0.0030
42	0.0010	0.0018	0.0006	0.0028	0.0023	0.0018	0.0012
43	0.0015	0.0012	0.0010	0.0023	0.0020	0.0006	0.0008
44	0.0027	0.0028	0.0027	0.0052	0.0051	0.0022	0.0031
45	0.0010	0.0023	0.0009	0.0026	0.0030	0.0014	0.0009
46	0.0011	0.0021	0.0022	0.0041	0.0047	0.0020	0.0003
47	0.0010	0.0014	0.0005	0.0018	0.0020	0.0018	0.0007
48	0.0018	0.0023	0.0023	0.0049	0.0054	0.0028	0.0013
49	0.0006	0.0009	0.0004	0.0009	0.0011	0.0005	0.0007
50	0.0032	0.0024	0.0020	0.0067	0.0050	0.0027	0.0028
51	0.0024	0.0004	0.0017	0.0014	0.0013	0.0014	0.0008
52	0.0026	0.0021	0.0011	0.0005	0.0019	0.0021	0.0010
53	0.0004	0.0011	0.0008	0.0017	0.0023	0.0019	0.0005
54	0.0059	0.0028	0.0020	0.0016	0.0027	0.0026	0.0014
55	0.0095	0.0040	0.0023	0.0031	0.0036	0.0064	0.0031
56	0.0049	0.0032	0.0019	0.0014	0.0038	0.0039	0.0021
57	0.0029	0.0011	0.0009	0.0015	0.0008	0.0007	0.0005
58	0.0039	0.0027	0.0006	0.0020	0.0015	0.0038	0.0013
59	0.0041	0.0017	0.0017	0.0025	0.0018	0.0023	0.0010
60	0.0043	0.0014	0.0006	0.0026	0.0021	0.0016	0.0014

Table A.4. 1. Highest posterior density(HPD) interval for a_1, a_2, and a_3-parameters

	a_1			a_2			a_3		
Item	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper
1	0.4391	0.0954	0.7284	0.4089	0.1175	0.7038	0.2820	0.0633	0.4875
2	0.4231	0.2224	0.6228	0.2595	0.0001	0.5307	0.1491	0.0002	0.3045
3	0.2637	0.0000	0.4801	0.1915	0.0001	0.4614	0.0934	0.0000	0.2142
4	0.2947	0.0465	0.5199	0.2285	0.0001	0.4742	0.1479	0.0002	0.2919
5	0.2678	0.0754	0.4488	0.2495	0.0861	0.4126	0.2260	0.0558	0.3832
6	0.6398	0.1883	1.0030	0.4364	0.0211	0.9389	0.2927	0.0433	0.5433
7	0.4619	0.1718	0.7242	0.3278	0.0180	0.6745	0.1568	0.0000	0.3094
8	0.7090	0.1330	1.1650	0.5096	0.0422	1.0880	0.2635	0.0009	0.5114
9	0.6421	0.1152	1.0620	0.4505	0.0195	0.9843	0.2702	0.0112	0.5044
10	0.5545	0.2008	0.8570	0.3781	0.0189	0.7737	0.1971	0.0035	0.3775
11	0.1686	0.0000	0.3611	0.4741	0.0165	0.8374	0.4556	0.0001	0.8314
12	0.1848	0.0191	0.3351	0.3416	0.0395	0.5908	0.2955	0.0171	0.5266
13	0.2615	0.0237	0.4697	0.5800	0.0640	1.0610	0.6120	0.0987	1.0560
14	0.1323	0.0000	0.3359	0.2015	0.0000	0.4248	0.2786	0.0924	0.4711
15	0.4396	0.2154	0.6607	0.6190	0.2644	0.9730	0.6072	0.1872	0.9761
16	0.3100	0.1039	0.5053	0.5817	0.1493	0.9736	0.6299	0.2528	0.9931
17	0.2081	0.0300	0.3853	0.2808	0.0584	0.5050	0.3113	0.1002	0.5185
18	0.3610	0.1401	0.5742	0.5326	0.1412	0.8522	0.4333	0.0448	0.7542
19	0.3489	0.1246	0.5729	0.6200	0.1501	1.0270	0.6585	0.2524	1.0050
20	0.2767	0.0703	0.4730	0.5061	0.1364	0.8542	0.5335	0.1652	0.8535
21	0.3273	0.0482	0.5954	0.3737	0.0718	0.6560	0.6078	0.2977	0.9319
22	0.3709	0.1186	0.6453	0.2244	0.0010	0.4357	0.4046	0.0449	0.7408
23	0.4475	0.1017	0.7802	0.3632	0.0482	0.6782	0.7089	0.1893	1.2030
24	0.5029	0.1270	0.8894	0.4468	0.1236	0.7812	0.8776	0.3027	1.3520
25	0.3087	0.0702	0.5427	0.2788	0.0456	0.5022	0.5294	0.1766	0.8402
26	0.1467	0.0000	0.3085	0.1303	0.0000	0.2996	0.2461	0.0034	0.4647
27	0.2187	0.0000	0.4096	0.2319	0.0156	0.4361	0.4698	0.1489	0.7695
28	0.2722	0.0474	0.4843	0.2648	0.0843	0.4579	0.3387	0.1477	0.5366
29	0.1234	0.0000	0.3330	0.1158	0.0000	0.2656	0.2487	0.0199	0.4855
30	0.3013	0.0021	0.5559	0.2987	0.0252	0.5448	0.6055	0.1879	0.9877

Table A.4.1 (cont'd)

	a_1			a_2			a_3		
Item	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper
31	0.1522	0.0047	0.2842	0.1603	0.0001	0.3323	0.1500	0.0005	0.3019
32	0.3570	0.1577	0.5593	0.3348	0.1026	0.5948	0.2731	0.0718	0.4737
33	0.2137	0.0000	0.4061	0.2126	0.0005	0.4315	0.1932	0.0007	0.3828
34	0.1648	0.0001	0.3237	0.2142	0.0000	0.5014	0.1675	0.0000	0.3547
35	0.2751	0.0370	0.5158	0.2200	0.0000	0.5371	0.1799	0.0005	0.4348
36	0.2169	0.0093	0.3917	0.2506	0.0012	0.5891	0.2267	0.0002	0.4736
37	0.3929	0.1567	0.6657	0.4359	0.1055	0.8444	0.4205	0.1471	0.7060
38	0.5042	0.2097	0.8409	0.5910	0.2139	1.1360	0.4643	0.1417	0.8278
39	0.4053	0.1762	0.6415	0.3408	0.0028	0.7085	0.2378	0.0000	0.5132
40	0.2997	0.1066	0.4915	0.2671	0.0541	0.4905	0.1311	0.0000	0.2837
41	0.2293	0.0015	0.4439	0.3296	0.0280	0.5968	0.2998	0.0007	0.5660
42	0.2147	0.0001	0.4189	0.3180	0.0448	0.5826	0.3374	0.0559	0.6078
43	0.3591	0.1057	0.6021	0.3474	0.0868	0.6372	0.3528	0.0888	0.6014
44	0.1824	0.0001	0.4058	0.2222	0.0000	0.4880	0.2804	0.0000	0.5656
45	0.2999	0.0550	0.5765	0.3013	0.0050	0.5861	0.3434	0.0306	0.6602
46	0.2635	0.0924	0.4518	0.2218	0.0373	0.4192	0.2498	0.0593	0.4512
47	0.2594	0.0657	0.4476	0.2495	0.0379	0.4583	0.2585	0.0361	0.4601
48	0.1476	0.0001	0.2800	0.1172	0.0000	0.2624	0.1627	0.0000	0.3496
49	0.1664	0.0088	0.3095	0.1501	0.0000	0.2953	0.2080	0.0138	0.4060
50	0.1538	0.0001	0.2953	0.2102	0.0030	0.3933	0.2519	0.0613	0.4418
51	0.6727	0.2409	1.1290	0.4785	0.2135	0.7612	0.4100	0.0800	0.7388
52	0.3490	0.0002	0.7552	0.1677	0.0000	0.3440	0.2427	0.0001	0.5124
53	0.3859	0.0005	0.8374	0.3027	0.0004	0.5594	0.3254	0.0330	0.5962
54	0.4791	0.2725	0.6785	0.3733	0.1680	0.5928	0.3602	0.1511	0.5845
55	0.3799	0.0432	0.7382	0.1656	0.0002	0.3393	0.1450	0.0000	0.3408
56	0.4183	0.0626	0.8133	0.2143	0.0185	0.3955	0.2104	0.0000	0.4654
57	0.4119	0.1147	0.7182	0.3469	0.1424	0.5369	0.2875	0.0586	0.5086
58	0.4040	0.0418	0.8260	0.2692	0.0521	0.4725	0.2798	0.0638	0.5194
59	0.2662	0.0002	0.5714	0.1525	0.0000	0.3045	0.1725	0.0000	0.3640
60	0.6179	0.1506	1.1200	0.3280	0.0687	0.5845	0.3308	0.0008	0.6453

Table A.4. 2. Highest posterior density(HPD) interval for a_4, a_5, and a_6-parameters

	a_4			a_5			a_6		
Item	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper
1	0.2835	0.0004	0.5377	0.2550	0.0001	0.5782	0.2077	0.0136	0.3922
2	0.2468	0.0165	0.4621	0.2548	0.0002	0.4930	0.3048	0.0237	0.5645
3	0.1031	0.0000	0.2407	0.1021	0.0000	0.2418	0.0845	0.0000	0.1962
4	0.1736	0.0000	0.3575	0.1893	0.0000	0.3624	0.1669	0.0001	0.3263
5	0.1818	0.0301	0.3305	0.1767	0.0202	0.3268	0.1756	0.0317	0.3211
6	0.3494	0.0501	0.6744	0.3547	0.0715	0.6439	0.3651	0.1037	0.6280
7	0.3361	0.1379	0.5467	0.3274	0.1133	0.5305	0.2476	0.0020	0.4594
8	0.3800	0.0192	0.7268	0.3895	0.0633	0.7610	0.3041	0.0001	0.5794
9	0.2873	0.0028	0.5907	0.2910	0.0174	0.5629	0.3047	0.0269	0.5469
10	0.3540	0.1244	0.5991	0.3254	0.1014	0.5519	0.2775	0.0486	0.5020
11	0.1222	0.0000	0.3054	0.1821	0.0000	0.4141	0.1007	0.0000	0.2582
12	0.2088	0.0008	0.3806	0.2310	0.0181	0.4378	0.1470	0.0000	0.2870
13	0.2744	0.0000	0.5738	0.3650	0.0285	0.6871	0.2843	0.0335	0.5376
14	0.1193	0.0000	0.2549	0.1325	0.0005	0.2714	0.2422	0.0552	0.4182
15	0.2856	0.0407	0.5105	0.3341	0.0447	0.6257	0.3733	0.1377	0.6148
16	0.3676	0.1380	0.5967	0.4319	0.1977	0.6687	0.3311	0.0911	0.5816
17	0.2456	0.0552	0.4348	0.2712	0.0703	0.4660	0.2531	0.0797	0.4220
18	0.4480	0.1309	0.7226	0.4662	0.1824	0.7217	0.2917	0.0371	0.5201
19	0.4237	0.1591	0.6793	0.4791	0.2287	0.7170	0.3769	0.1374	0.6429
20	0.3582	0.1247	0.6077	0.4087	0.1875	0.6338	0.2804	0.0732	0.4986
21	0.1848	0.0000	0.4534	0.1927	0.0000	0.4253	0.4498	0.0622	0.8637
22	0.3644	0.0969	0.6468	0.3170	0.0900	0.5683	0.5480	0.2509	0.7975
23	0.5774	0.1825	1.0100	0.5581	0.1912	0.9691	0.7314	0.1784	1.2470
24	0.5668	0.1525	1.0400	0.5433	0.1763	0.9609	0.8712	0.2701	1.4220
25	0.3727	0.0965	0.6913	0.3684	0.1085	0.6543	0.5068	0.1410	0.8788
26	0.1922	0.0000	0.4181	0.1928	0.0000	0.3921	0.2241	0.0001	0.4684
27	0.3276	0.0313	0.6284	0.3374	0.0405	0.6318	0.4209	0.0632	0.7981
28	0.2724	0.0625	0.4872	0.2766	0.0900	0.4650	0.2715	0.0403	0.5102
29	0.2065	0.0137	0.3825	0.1953	0.0027	0.3550	0.2819	0.0490	0.4899
30	0.3309	0.0053	0.6782	0.3192	0.0213	0.6025	0.5282	0.0336	0.9640

Table A.4.2. (cont'd)

	a_4			a_5			a_6		
Item	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper
31	0.2315	0.0204	0.4286	0.2048	0.0084	0.3978	0.1365	0.0000	0.2789
32	0.3687	0.0990	0.6434	0.3256	0.0640	0.6256	0.2532	0.0582	0.4514
33	0.2678	0.0046	0.4834	0.2316	0.0002	0.4907	0.1532	0.0000	0.3337
34	0.3725	0.0612	0.6457	0.3308	0.0572	0.6152	0.1561	0.0000	0.3338
35	0.3932	0.0357	0.7333	0.3256	0.0160	0.6432	0.2673	0.0428	0.5019
36	0.4119	0.0433	0.7303	0.3581	0.0149	0.7049	0.2032	0.0000	0.4233
37	0.5198	0.0836	0.9124	0.4435	0.0662	0.9204	0.3382	0.0176	0.6104
38	0.6925	0.1501	1.2180	0.6159	0.1196	1.2230	0.3403	0.0001	0.6501
39	0.4355	0.0094	0.8144	0.3532	0.0001	0.7837	0.2722	0.0132	0.5119
40	0.2400	0.0015	0.4833	0.2079	0.0001	0.4755	0.1391	0.0000	0.3016
41	0.4999	0.0077	0.9106	0.5692	0.0675	0.9438	0.2372	0.0005	0.4805
42	0.5684	0.1043	0.9903	0.6169	0.1509	0.9729	0.2643	0.0002	0.5175
43	0.5240	0.1021	0.9244	0.5663	0.1037	0.9333	0.3555	0.1043	0.6148
44	0.5913	0.0438	1.0930	0.6649	0.0207	1.1130	0.2968	0.0015	0.5717
45	0.6445	0.0850	1.1410	0.7042	0.0881	1.1610	0.3872	0.0885	0.6939
46	0.3687	0.0659	0.6378	0.3952	0.0533	0.6619	0.3096	0.1269	0.4953
47	0.3803	0.0423	0.6772	0.4220	0.0339	0.7147	0.2613	0.0538	0.4728
48	0.3253	0.0831	0.5482	0.3269	0.0638	0.5617	0.2193	0.0291	0.4036
49	0.3336	0.1069	0.5563	0.3376	0.0824	0.5729	0.2251	0.0247	0.4327
50	0.2754	0.0001	0.5310	0.3261	0.0049	0.5575	0.2285	0.0336	0.4119
51	0.4254	0.0559	0.8094	0.3958	0.0641	0.7377	0.6865	0.1266	1.1060
52	0.2087	0.0002	0.4690	0.1705	0.0000	0.3709	0.5125	0.1348	0.8197
53	0.2556	0.0001	0.5437	0.2351	0.0000	0.5069	0.5249	0.0376	0.8535
54	0.4559	0.2397	0.6772	0.4276	0.2307	0.6223	0.4754	0.2766	0.6707
55	0.1703	0.0000	0.4026	0.1513	0.0000	0.3628	0.4313	0.0014	0.7206
56	0.2713	0.0001	0.5568	0.2352	0.0019	0.4501	0.5093	0.0874	0.8352
57	0.3599	0.1188	0.6000	0.3625	0.1297	0.6261	0.4348	0.1058	0.7186
58	0.3263	0.0520	0.5949	0.2978	0.0604	0.5506	0.5050	0.1282	0.8171
59	0.2808	0.0547	0.4967	0.2540	0.0609	0.4470	0.3626	0.0888	0.6082
60	0.3860	0.0799	0.7282	0.3464	0.0389	0.6580	0.7296	0.1819	1.1510

Table A.4. 3. Highest posterior density(HPD) interval for *d*-parameter

		d	
Item	Mean	Lower	Upper
1	0.995	0.897	1.085
2	0.762	0.680	0.855
3	0.353	0.277	0.431
4	0.283	0.209	0.364
5	0.151	0.073	0.224
6	0.552	0.463	0.646
7	-0.095	-0.182	-0.017
8	-0.138	-0.232	-0.050
9	-0.163	-0.250	-0.080
10	-1.029	-1.129	-0.926
11	0.364	0.285	0.450
12	0.358	0.276	0.434
13	0.336	0.240	0.440
14	0.011	-0.064	0.089
15	0.006	-0.090	0.090
16	-0.008	-0.098	0.081
17	-0.115	-0.189	-0.046
18	-0.673	-0.767	-0.578
19	-0.947	-1.059	-0.848
20	-0.960	-1.045	-0.857
21	1.139	1.032	1.244
22	1.115	1.007	1.205
23	1.293	1.164	1.428
24	1.154	1.026	1.291
25	0.297	0.216	0.385
26	0.104	0.028	0.182
27	0.113	0.036	0.201
28	-0.099	-0.174	-0.010
29	-0.124	-0.198	-0.049
30	-0.446	-0.537	-0.356

Table A.4.3. (cont'd)

	d		
Item	Mean	Lower	Upper
31	0.558	0.481	0.636
32	0.854	0.764	0.944
33	0.340	0.265	0.422
34	0.328	0.249	0.411
35	0.034	-0.046	0.111
36	-0.027	-0.106	0.056
37	0.028	-0.060	0.118
38	-0.251	-0.350	-0.150
39	-0.230	-0.312	-0.139
40	-0.399	-0.480	-0.327
41	1.236	1.119	1.350
42	1.077	0.967	1.176
43	0.625	0.530	0.720
44	0.196	0.106	0.278
45	0.018	-0.073	0.115
46	-0.070	-0.147	0.014
47	-0.164	-0.245	-0.076
48	-0.386	-0.466	-0.310
49	-0.414	-0.493	-0.333
50	-0.873	-0.967	-0.790
51	2.374	2.197	2.572
52	1.131	1.031	1.232
53	0.783	0.690	0.878
54	0.789	0.698	0.885
55	0.195	0.117	0.267
56	-0.155	-0.239	-0.077
57	-0.127	-0.213	-0.045
58	-0.456	-0.545	-0.368
59	-0.495	-0.576	-0.415
60	-1.522	-1.658	-1.406

REFERENCES

REFERENCES

Albert, J. H. (1992). Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling. *Journal of Educational Statistics, 17*(3), 251-269.

Arellano-Valle, R. B., & Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis, 99*(7), 1362-1382. doi: 10.1016/j.jmva.2008.01.020

Baker, F. B. (1987). Methodology Review: Item Parameter Estimation Under the One-, Two-, and Three-Parameter Logistic Models. *Applied Psychological Measurement, 11*(2), 111-141. doi: 10.1177/014662168701100201

Baker, F. B., & Kim, S.-H. (2004). *Item Response Theory: Parameter estimation techniques* (Second Edition, Revised and Expanded ed.). New York, NY: Marcel Dekker.

Batley, R.-M., & Boss, M. W. (1993). The Effects on Parameter Estimation of Correlated Dimensions and a Distribution-Restricted Trait in a Multidimensional Item Response Model. *Applied Psychological Measurement, 17*(2), 131-141. doi: 10.1177/014662169301700203

Béguin, A., & Glas, C. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika, 66*(4), 541-561. doi: 10.1007/bf02296195

Birbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. L. M. R. Novick (Ed.), *Statistical theories of mental test scores* (pp. 397-472). Reading, MA: Addison-Wesley.

Bock, R. D., & Aitkin, M. (1981). Marginal Maximum-Likelihood Estimation of Item Parameters - Application of an Em Algorithm. *Psychometrika, 46*(4), 443-459.

Bock, R. D., Gibbons, R., SChillings, S. G., Muraki, E., Wilson, D. T., & Woods, R. (2003). Testfact(version 4). Chicago, IL: Scientific Software International.

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*(6), 395-414. doi: 10.1177/0146621603258350

Brooks, S. P. (1998). Markov chain Monte Carlo method and its application. *Journal of the Royal Statistical Society Series D-the Statistician, 47*(1), 69-100.

Brooks, S. P., & Morgan, B. J. T. (1994). Automatic starting point selection for function optimization. *Statistics and Computing, 4*(3), 173-177. doi: 10.1007/bf00142569

Carlson, J. E. (1987). Multidimensional Item Response Theory Estimation: A Computer Program. IOWA, IA: ACT.

Cowles, M. K., & Carlin, B. P. (1996). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association, 91*(434), 883-904.

De Ayala, R. J., & Sava-Bolesta, M. (1999). Item Parameter Recovery for the Nominal Response Model. *Applied Psychological Measurement, 23*(1), 3-19. doi: 10.1177/01466219922031130

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. Series B (Methodological), 39*(1), 1-38.

Dorans, N. J., & Kingston, N. M. (1985). The Effects of Violations of Unidimensionality on the Estimation of Item and Ability Parameters and on Item Response Theory Equating of the GRE Verbal Scale. *Journal of Educational Measurement, 22*(4), 249-262.

Finch, H. (2010). Item Parameter Estimation for the MIRT Model. *Applied Psychological Measurement, 34*(1), 10-26. doi: 10.1177/0146621609336112

Finch, H. (2011). Multidimensional Item Response Theory Parameter Estimation With Nonsimple Structure Items. *Applied Psychological Measurement, 35*(1), 67-82. doi: 10.1177/0146621610367787

Fraser, C., & McDonald, R. P. (1988). NOHARM II : A FORTRAN program for fitting unidimensional and multidimensional normal ogive models of latent trait theory. Armidale, Australia: University of New England, Centre for Behavioral Studies.

Froelich, A. G. (2001). *Assessing the uni-dimensionality of test items and some asymptotics of parametric item response theory*. Unpublished doctoral dissertation. Department of Statistics. University of Illinois at Urbana-Champaign.

Fu, Z.-H., Tao, J., & Shi, N.-Z. (2009). Bayesian estimation in the multidimensional three-parameter logistic model. *Journal of Statistical Computation and Simulation, 79*(6), 819-835. doi: 10.1080/00949650801966876

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science, 7*(4), 457-472.

Gelman, A., & Shalizi, C. R. (2012). Philosophy and the practice of Bayesian statistics in the social sciences. In H. Kincaid (Ed.), *The Oxford Handbook of Philosophy of Social Science*: Oxford University Press.

Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-6*(6), 721-741.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In J. M. Bernardo, J. Berger, A. P. Dawid & A. F. M. Smith (Eds.), *Bayesian Statistics* (4th ed., pp. 169-193). Oxford, U.K: Oxford University Press.

Geyer, C. J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science, 7*(4), 473-483.

Goaz, J. K., & C.M, W. (2002). *An Empirical comparison of multidimensional item response theory data using TESTFACT and NOHARM*. Paper presented at the the annual meeting of the National Council for Measurement in Education, New Orleans.

Harwell, M., Stone, C. A., Hsu, T.-C., & Kirisci, L. (1996). Monte Carlo Studies in Item Response Theory. *Applied Psychological Measurement, 20*(2), 101-125. doi: 10.1177/014662169602000201

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika, 57*(1), 97-109.

Heidelberger, P., & Welch, P. D. (1983). Simulation Run Length Control in the Presence of an Initial Transient. *Operations Research, 31*(6), 1109-1144.

Holzinger, K., & Harman, H. (1941). *Factor analysis: A synthesis of factorial methods*. Chicago, IL: The University of Chicago Press.

HPCC. (2012). HPCC Retrieved December 11, 2012, from WIKI: https://wiki.hpcc.msu.edu/display/hpccdocs/Documentation+and+User+Manual#DocumentationandUserManual-Overview

Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of Two- and Three-Parameter Logistic Item Characteristic Curves: A Monte Carlo Study. *Applied Psychological Measurement, 6*(3), 249-260. doi: 10.1177/014662168200600301

Kelderman, H., & Rijkes, C. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika, 59*(2), 149-176. doi: 10.1007/bf02295181

Kelton, W. D., & Law, A. M. (1984). An Analytical Evaluation of Alternative Strategies in Steady-State Simulation. *Operations Research, 32*(1), 169-184.

Kolen, M. J. (1981). Comparison of Traditional and Item Response Theory Methods for Equating Tests. *Journal of Educational Measurement, 18*(1), 1-11.

Lawley, D. N. (1944). *The factorial analysis of multiple item tests* Paper presented at the The Royal Society of Edinburgh.

Lord, F. (1952). A Theory of Test Score(Psychometric Monograph No. 7). Richimond, WA: Psychometric Corporation.

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Oxford, England: Addison-Wesley.

Maris, E. (1995). Psychometric Latent Response Models. *Psychometrika, 60*(4), 523-547.

Maydeu-Olivares, A. (2001). Multidimensional Item Response Theory Modeling of Binary Data: Large Sample Properties of NOHARM Estimates. *Journal of Educational and Behavioral Statistics, 26*(1), 51-71. doi: 10.3102/10769986026001051

McDonald, R. P. (1999). *Test theory: A unified treatment*: Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Mckinley, R. L., & Mills, C. N. (1985). A Comparison of Several Goodness-of-Fit Statistics. *Applied Psychological Measurement, 9*(1), 49-57.

McKinley, R. L., & Reckase, M. D. (1980). A comparison of the ANCILLES and LOGIST parameter estimation procedure for the three-parameter logistic model using goodness of fit as a criterion. Columbia MO: University of Missouri Tailored Testing Laboratory.

McKinley, R. L., & Reckase, M. D. (1984). An investigation of the effect of correlated abilities on observed test characteristics (T. D. Division, Trans.). Iowa City, IA: American College Testing Program.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics, 21*(6), 1087-1092.

Muraki, E., & Carlson, J. E. (1993). *Full-information factor analysis for polytomous item responses*. Paper presented at the American Educational Research Association, Atlanta.

Patz, R. J., & Junker, B. W. (1999a). A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models. *Journal of Educational and Behavioral Statistics, 24*(2), 146-178.

Patz, R. J., & Junker, B. W. (1999b). Applications and Extensions of MCMC in IRT: Multiple Item Types, Missing Data, and Rated Responses. *Journal of Educational and Behavioral Statistics, 24*(4), 342-366.

Raftery, A. E., & Lewis, S. (1992). How many iterations in the Gibbs Sampler? In J. M. bernardo, J. berger, A. P. Dawid & A. F. M. Smith (Eds.), *Bayesian Statistics* (4th ed., pp. 763-773). Oxford, U.K: Oxford University Press.

Reckase, M. D. (1985). The Difficulty of Test Items That Measure More Than One Ability. *Applied Psychological Measurement, 9*(4), 401-412. doi: 10.1177/014662168500900409

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

Reckase, M. D., & McKinley, R. L. (1982). *Some Latent Trait Theory in a Multidimensional Latent Space*. S.l.: Distributed by ERIC Clearinghouse.

Reckase, M. D., & McKinley, R. L. (1991). The Discriminating Power of Items That Measure More Than One Dimension. *Applied Psychological Measurement, 15*(4), 361-373. doi: 10.1177/014662169101500407

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*(3), 287-297.

Ripley, B. D., & Kirkland, M. D. (1990). Iterative simulation methods. *Journal of Computational and Applied Mathematics, 31*(1), 165-172. doi: 10.1016/0377-0427(90)90347-3

Samejima, F. (1974). Normal Ogive Model on Continuous Response Level in Multidimensional Latent Space. *Psychometrika, 39*(1), 111-121.

Sheng, Y. (2008). A MATLAB package for Markov chain Monte Carlo with a multi-unidimensional IRT model. *Journal of Statistical Software, 28*(10).

Sheng, Y., & Wikle, C. K. (2007). Comparing Multiunidimensional and Unidimensional Item Response Theory Models. *Educational and Psychological Measurement, 67*(6), 899-919. doi: 10.1177/0013164406296977

Stone, C. A., & Yeh, C.-C. (2006). Assessing the Dimensionality and Factor Structure of Multiple-Choice Exams. *Educational and Psychological Measurement, 66*(2), 193-214. doi: 10.1177/0013164405282483

Tate, R. (2003). A Comparison of Selected Empirical Methods for Assessing the Structure of Responses to Test Items. *Applied Psychological Measurement, 27*(3), 159-203. doi: 10.1177/0146621603027003001

Thissen, D., & Steinberg, L. (1984). A Response Model for Multiple-Choice Items. *Psychometrika, 49*(4), 501-519.

Thissen, D., & Wainer, H. (1982). Some Standard Errors in Item Response Theory. *Psychometrika, 47*(4), 397-412.

Thurstone, L. L. (1947). *Multiple-factor analysis; a development and expansion of The Vectors of Mind*: Chicago, IL, US: University of Chicago Press.

Tucker, L. (1946). Maximum validity of a test with equivalent items. *Psychometrika, 11*(1), 1-13. doi: 10.1007/bf02288894

Walker, C. M., Azen, R., & Schmitt, T. (2006). Statistical Versus Substantive Dimensionality. *Educational and Psychological Measurement, 66*(5), 721-738. doi: 10.1177/0013164405285907

Whitely, S. E. (1980). Multicomponent Latent Trait Models for Ability Tests. *Psychometrika, 45*(4), 479-494.

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). LOGIST user's guide. Princeton NJ: Educational Testing Service.

Wollack, J. A., Bolt, D. M., Cohen, A. S., & Lee, Y.-S. (2002). Recovery of Item Parameters in the Nominal Response Model: A Comparison of Marginal Maximum Likelihood Estimation and Markov Chain Monte Carlo Estimation. *Applied Psychological Measurement, 26*(3), 339-352. doi: 10.1177/0146621602026003007

Yao, L. (2003). BMIRT: Bayesian multivariate item response theory. Monterey, CA: CTB/McGraw-Hill.

Yao, L. H., & Schwarz, R. D. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. *Applied Psychological Measurement, 30*(6), 469-492. doi: Doi 10.1177/0146621605284537

Yen, W. M. (1981). Using Simulation Results to Choose a Latent Trait Model. *Applied Psychological Measurement, 5*(2), 245-262.