

## ABSTRACT

Title of Document: A MIXTURE RASCH MODEL WITH A COVARIATE:  
A SIMULATION STUDY VIA BAYESIAN MARKOV  
CHAIN MONTE CARLO ESTIMATION

Yunyun Dai, Doctor of Philosophy, 2009

Directed By: Professor, Robert J. Mislevy,  
Department of Measurement, Statistics and Evaluation

Mixtures of item response theory models have been proposed as a technique to explore response patterns in test data related to cognitive strategies, instructional sensitivity, and differential item functioning (DIF). Estimation proves challenging due to difficulties in identification and questions of effect size needed to recover underlying structures. In particular, the impact of auxiliary variables, or covariates, for examinees in estimation has not been systematically explored. The goal of this dissertation is to carry out a systematically designed simulation study to investigate the performance of mixture Rasch model (MRM) under Bayesian estimation using Markov Chain Monte Carlo (MCMC) method. The dependent variables in this study are (1) the proportion of cases in which the generating mixture structure is recovered, and (2) among those cases in which the structure is recovered, the bias and root mean squared error of parameter estimates. The foci of the study are to use a flexible logistic regression model to parameterize the relation between latent class membership and the examinee covariate, to study MCMC estimation behavior in light of effect size, and to provide insights and suggestions on model application and model estimation.

A MIXTURE RASCH MODEL WITH A COVARIATE:  
A SIMULATION STUDY VIA BAYESIAN MCMC ESTIMATION

By

Yunyun Dai

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2009

Advisory Committee:  
Professor Robert J. Mislevy, Chair  
Professor Gregory R. Hancock  
Professor Jeffrey Harring  
Professor Hong Jiao  
Professor Michel Wedel

© Copyright by  
Yunyun Dai  
2009

## Dedication

This dissertation is dedicated to:

My dear mother and beloved father.

## Acknowledgements

I would like to thank the following individuals and the Department of Measurement, Statistics and Evaluation (EDMS). To Prof. Lissitz, for giving me the once in a lifetime opportunity to study in this great program; to Prof. Hancock, for his amazing statistical courses and insightful suggestions on my dissertation research; to Prof. Haring and Prof. Jiao for their constructive comments on this research; to Prof. Wedel for his thoughtful suggestions for this study.

I am especially thankful to my advisor, Prof. Mislevy, for his continuous supports throughout my doctoral study and his experienced guidance through this dissertation study.

I would also like to thank Association of American Medical Colleges (AAMC) for their summer graduate student research program in which I initially studied the mixture IRT models.

# Table of Contents

List of Tables.....	vi
List of Figures.....	vii
Chapter 1: Introduction.....	1
1.1 General Introduction of Basic Models.....	3
1.2 General Introduction of Mixture IRT model.....	4
1.3 Early Development in MRM.....	5
1.4 The Rationale to Include Auxiliary Person’s Variable.....	6
1.5 Review of Other Recent Studies about MRM.....	7
1.6 Description of the Necessity of This Research.....	7
1.7 Contribution from this Dissertation Study.....	8
1.8 Brief Summary of Following Chapters.....	9
Chapter 2: Literature Review.....	11
2.1 Rasch Model and Latent Class Model.....	11
2.2 Detailed Description of MRM.....	15
2.3 Studies Related to the Development of Mixture Rasch Model.....	17
2.4 Model Estimation.....	22
2.5 Identification Difficulty in Bayesian method.....	25
Chapter 3: Methodology.....	28
3.1 MRM with covariate.....	28
3.2 Primary design of the simulation study.....	37
3.3 Evaluation of Estimation Outcomes.....	43
3.3.1 Convergence.....	44
3.3.2 Bias.....	44
3.4 Computer Programs.....	45
Chapter 4: Results Summary.....	47
4.1.1 Review of History Plots from WinBUGS.....	48
4.1.2 Review of the Correlations between Estimated and Generated values.....	53
4.1.3 Review of the Percentage of Significant DIF Identified.....	55
4.1.4 Comparison of Signs of DIF Parameters with Generating DIF Parameters.....	57
4.1.5 Examination of Ill-Recovered Replications.....	57
4.1.6 Solution to Label-Switching Runs.....	59
4.2.1 Recovery of Underlying Structures.....	61
4.2.2 Bias and RMSE: Regression Analysis.....	71
4.2.3 Bias and RMSE: Graphic Presentation.....	77
Chapter 5: Discussion.....	83
5.1 Discussion.....	83
5.2 Summary of Results.....	85

5.3 Insights on Model Estimation.....	88
5.4 Future Research.....	90
Appendix A: Inference on $\exp[\text{sl}(2)]$ .....	93
Appendix B: Annotated SAS Code.....	95
Appendix C: Annotated WINBUGS Code.....	98
Bibliography.....	99

## List of Tables

Table 3.1	Relation between parameters in logistic regression and membership probabilities, as a function of the person's covariate.....	37
Table 3.2	Factors varied in this simulation study.....	37
Table 3.3	Factors kept constant in this simulation study.....	40
Table 4.1a	Denotation of result categories, result types and simulation factors in Table 4.1b and Table 4.1c.....	63
Table 4.1b	Result types with corresponding result categories and result summary decision....	64
Table 4.1c	Distribution of replication for each result category and result type.....	65
Table 4.1d:	List of poor mixing chains by simulation cell.....	68
Table 4.1e:	List of within-chain label-switching chains by simulation cell.....	68
Table 4.1f:	List of simulation cells with recovery rate less than 50%.....	68
Table 4.2	Denotation of simulation conditions in the regression analysis.....	69
Table 4.3	Regression analysis of recovered replications by simulation factors.....	70
Table 4.4a	Regression analysis of bias by simulation factors for the mixing proportion.....	72
Table 4.4b	Regression analysis of RMSE by simulation factors for the mixing proportion....	72
Table 4.5a	Regression analysis of bias by simulation factors for the mean of latent ability distributions.....	73
Table 4.5b	Regression analysis of RMSE by simulation factors for the mean of latent ability distributions.....	73
Table 4.6a	Regression analysis of bias by simulation factors for the precision of latent ability distributions.....	73
Table 4.6b	Regression analysis of RMSE by simulation factors for the precision of latent ability distributions.....	73
Table 4.7a	Regression analysis of bias by simulation factors for the intercept and slope in the logit link .....	75
Table 4.7b	Regression analysis of RMSE by simulation factors for the intercept and slope in the logit link .....	76
Table 4.8	Regression analysis of RMSE by simulation factors for average item difficulty parameters in each latent class .....	77
Table 4.9	Simulation factors and their denotation .....	78
Table 4.10a	Pattern observed in Figure 15-Figure 18 for the mixing proportion .....	82
Table 4.10b	Pattern observed in Figure 15-Figure 18 for the mixing proportion .....	82
Table 4.10c	Pattern observed in Figure 15-Figure 18 for the mixing proportion .....	82
Table 4.10d	Pattern observed in Figure 15-Figure 18 for the mixing proportion .....	82

## List of Figures

Figure 2.1	The latent continuum for the Rasch model.....	13
Figure 2.2	The latent continuum for the mixture Rasch model.....	17
Figure 4.1	Example of a MCMC run with stationary distribution after burn-in cycles.....	49
Figure 4.2	Example of a MCMC run with between-chain label-switching problem.....	49
Figure 4.3	Example of a collapsed MCMC run with between-chain label-switching problem.....	51
Figure 4.4	Example of a MCMC run with within-chain label-switching problem.....	51
Figure 4.5	Example of a MCMC run with poor mixing.....	52
Figure 4.6a	Example of a MCMC run with poor mixing fluctuation in mixing proportion.....	52
Figure 4.6b	Example of a MCMC run with stable estimation of item difficulty.....	53
Figure 4.7	Histogram of correlation index.....	55
Figure 4.8	Histogram of correlation index.....	56
Figure 4.9	Histogram of correlation index between excluded and included output.....	58
Figure 4.10	Histogram of percentage of significant item between excluded and included output.....	59
Figure 4.11a:	Bias of Average Large Size DIF Parameters by Latent Group Distribution.....	79
Figure 4.11b:	RMSE of Average Large Size DIF Parameters by Latent Group Distribution.....	79
Figure 4.12a:	Bias of Average Large Size DIF Parameters by Covariate Distribution.....	79
Figure 4.12b:	RMSE of Average Large Size DIF Parameters by Covariate Distribution.....	79
Figure 4.13a:	Bias of Average Large Size DIF Parameters by Levels of Large DIF Effect.....	80
Figure 4.13b:	RMSE of Average Large Size DIF Parameters by Levels of Large DIF Effect.....	80
Figure 4.14a:	Bias of Average Large Size DIF Parameters by Subjects' Latent Abilities.....	80
Figure 4.14b:	RMSE of Average Large Size DIF Parameters by Subjects' Latent Abilities.....	80
Figure 4.15a:	Bias of Latent Proportion One by Latent Group Distribution .....	81
Figure 4.15b:	RMSE of Latent Proportion One by Latent Group Distribution.....	81
Figure 4.16a:	Bias of Latent Proportion One by Covariate Distribution .....	81
Figure 4.16b:	RMSE of Latent Proportion One by Covariate Distribution.....	81
Figure 4.17a:	Bias of Latent Proportion One by Levels of Large DIF Effect.....	81
Figure 4.17b:	RMSE of Latent Proportion One by Levels of Large DIF Effect.....	81
Figure 4.18a:	Bias of Latent Proportion One by Subjects' Latent Abilities .....	81
Figure 4.18b:	RMSE of Latent Proportion One by Subjects' Latent Abilities.....	81

# Chapter 1: Introduction

Nowadays, with the improvement in computing capability and applications of quantitative methods to analyze collected data, there is increasing interest in looking for important pieces of information from test items in the field of education and survey items in the context of social science. Within the context of educational assessment, an obvious and common example is ‘score’ from various tests and evaluations. The familiar scenario is to have a group of people take a certain number of items and, after analyzing the subjects’ response data, assign a score for each person. There are many ways to process data based on the purpose of the assessment. Among the available analytical tools is one which aims to estimate both the subjects’ latent abilities and test items’ characteristics. These underlying features of subjects and items can be modeled by the means of latent variable analysis.

The group of models, entitled ‘latent variable models’ were fully developed within the last twenty years and were actively applied in research studies, especially in the areas of education and behavioral science. Basically, these groups of statistical models relate a set of observed variables to unobserved latent variables. These models presume that the latent factors account for important characteristics in the observed response data.

The assumption of unobserved quantities is a common feature for models from the family of latent variable models. Within the latent variable family, two types of models are of special interest in social science. They are the item response theory (IRT) models and the latent class models (LCM). The hypothesized latent variable in IRT model is assumed to be continuous and usually is regarded as the ‘latent trait’ in a quantitative

sense. In contrast, for models from the LCM family, the latent variable is regarded as categorical in a qualitative sense. In this simulation study, the model of interest is the mixture Rasch model (MRM). The MRM is a synthesis of two main popular statistical models—one is the simple Rasch model which belongs to the family of item response theory and the other one is the unconstrained LCM from the latent class analysis (LCA) family. The book edited by Hancock and Samuelsen (2008) contains various chapters about latent variable mixture models studied by leading scholars in the field of latent variable methods, such as the book chapter by Muthén (2008) and the applications of these methods in assessment and diagnosis, such as the book chapter by Mislevy, et al. (2008) and the book chapter by von Davier (2008).

The mixture model investigated in this study is a member from the family of mixture IRT models in which the component of mixture is the IRT model. There are other types of mixture models in which the latent class model is combined with other type of models, such as the growth curve model and factor models. For example, in Muthén (2001a, 2001b, & 2004), the author proposed and studied the growth mixture model where the component of mixture is the growth curve model. In this model, the intercept parameter and slope parameter from the growth model are allowed to vary across latent classes of subjects. Another mixture model is the factor mixture model described and discussed by Lubke & Muthén (2005). As a combination of the latent class model and factor model, the factor mixture model can be used to analyze the unobserved heterogeneity in the population. In studying different mixture models, researchers are especially interested in the estimation procedures for the corresponding mixture model and the recovery of latent

structures. Similarly, the main foci in this simulation study are the estimation procedure of MRM with a covariate and the recovery of generating latent structures.

### **1.1 General Introduction of Basic Models**

IRT, also referred to as latent trait theory, is a modern psychometric theory that studies relations among observed responses, items and hypothesized latent abilities. For IRT models, probability of observed correct response to each item is modeled as the function of item parameter and person parameter. To be more specific, the standard IRT model estimates probabilities of correct responses and puts item parameters, along with subjects' proficiency parameters, in a latent continuum. In the context of educational assessment, items are questions in an actual test.

Generally, item parameters are regarded as the quantities of certain characteristics for an item, such as items' difficulty level, items' discrimination power among subjects and items pseudo-guessing level from subjects. Importantly, in a standard IRT model, item parameters are assumed to be the same for each examinee. A person's latent ability parameter, in general, is conceptualized as an unobserved trait of that person who takes the test. This hypothesized property of a person is quantified via IRT models.

In the context of social science, normally the response data are discrete, either dichotomous or polytomous. For dichotomous response data, based on different characteristics involved in the IRT models, there are three commonly used models—the one-parameter logistic model (1PLM), the two-parameter logistic model (2PLM) and three-parameter logistic model (3PLM). The 1PLM, which is also called the Rasch model, estimates both subjects' ability parameters and item difficulty parameters. The other two

IRT models allow for additional item parameters, such as item discrimination parameters and item pseudo-guessing parameters, to model a wide variety of response functions. In the current study, item difficulty is the main item characteristic involved in the proposed model. And this research focuses on analyzing dichotomous scored response data using the Rasch model's basic modeling idea which is one of the main components in mixture Rasch model introduced in the subsequent section.

The LCA is another statistical method to analyze discrete response data and classify subjects into different unobserved qualitative classes called 'latent classes'. The unobserved latent class membership is defined by conditional independence. That is, after conditioning on the subjects' latent class membership, subjects' observed responses are independent from each other. Based on the observed data, the main focus for LCA is to classify subjects into different latent groups. This is different from IRT in that the IRT model aims to estimate quantitative parameters for both items and persons.

## **1.2 General Introduction of Mixture IRT Model**

The mixture IRT models are the combination of IRT models and LCA models introduced in previous sections. To be more specific, IRT models with different values of item parameters are assumed for different latent classes of subjects or persons. Two lines of research in IRT suggest the usefulness of this idea. One finds that sometimes a given test item would either be relatively easier or more difficult for students from different demographic groups. This is the so-called differential item functioning (DIF) (Camilli & Shepard, 1994) in the testing theories. Another line of research is the observation that in

some educational examinations, different examinees use different strategies to solve test items (Kyllonen, Lohman Snow, 1984). Both of these phenomena would mean that the same IRT parameter may have different values across various examinee groups.

There are two goals for a mixture IRT model. One is to calculate item parameters and person parameters (i.e., examinees' latent traits) and the other one is to estimate subjects' latent class memberships. It should be noted that in mixture IRT models, a set of item parameters are estimated for each latent class. For the person parameter, although it is presumed that a subject belongs to only one class, class membership is not known with certainty, so during estimation a person parameter is estimated conditioning upon membership in each class. In terms of class membership, each subject is assumed to belong to exactly one latent class, but because class membership is not observed, what are estimated are the probabilities of being in each of the classes. For example, assuming that there are two latent classes for a particular dataset, for a hypothesized subject 'A' the estimated probabilities for subject 'A' belonging to latent class one and latent class two could be 80% and 20%, respectively. In short, Mixture IRT models posit multiple groups of subjects in which different IRT models hold, although group membership is not directly observed.

### **1.3 Early Development in MRM**

There were several important developments to mixture IRT models in the early 1990's. Rost (1990) proposed the mixed Rasch model by estimating a set of item parameters for each latent class and therefore the items with different effects for each latent class could be identified. A more general framework to detect differential item

functioning (DIF) was proposed in Kelderman & Macready (1990). By combining the ideas of latent class models and latent trait models, Kelderman & Macready (1990) used the loglinear latent class model to model interaction effects between the grouping variables (either manifest or latent) and item parameters. A mixture linear logistic test model was proposed by Mislevy & Verhelst (1990) under the assumption that different groups of subjects adopted different strategies.

#### **1.4 The Rationale to Include Auxiliary Person's Variable**

As Embretson (2006) pointed out, a disadvantage of using the mixture IRT model was the difficulty in interpreting the qualitative meaning of the latent groups identified by the models. In addition, the observed response data by itself might not be sufficient to generate precise estimations of parameters, especially for complicated statistical models. Thus, when examinees' background variables are available, the author recommended including in the model the background variable as the collateral information. It has been shown by other research, such as Smit et al. (1999), that the addition of auxiliary variables helped to lower the standard error and provided more precisely estimated parameters. A main feature for this proposed research is the inclusion of the covariate which is assumed to be related to the latent class distribution. In the context of mixture models, researchers call the size of latent classes 'mixing proportions' (Fruhwirth-Schnatter, 2006).

### **1.5 Review of Other Recent Studies about MRM**

In the study by Smit et al. (1999), the authors applied the mixture Rasch model (MRM) to simulated data to evaluate the gain of incorporating collateral information. Instead of using maximum likelihood estimation as in the Smit et al. (1999) study, this proposed study implements Markov chain Monte Carlo (MCMC) algorithm to measure mixture Rasch model under Bayesian estimation. MCMC estimation is a resampling algorithm that provides an empirical approximation of the full posterior distribution of model parameters. An advantage of Bayesian MCMC is that the problem of local maximums encountered in maximum likelihood (ML) estimation can be avoided.

A more recent application by Cohen & Bolt (2005) uses MRM to try to locate the cause of DIF. Additionally, the dissertation study by Samuelsen (2005) is relevant to DIF detection using MRM. In Samuelsen (2005), the author carried out a simulation study to compare the performance of the traditional manifest DIF method and the performance of MRM when these two methods were used to detect the latent structure of DIF.

### **1.6 Description of the Necessity of This Research**

In this study, I use the MRM to simulate datasets where test items are constructed to perform differently across latent classes. Similar work has been done in the other studies introduced above. However, this research builds upon several previous research studies and differs from them in the sense of the manipulated factors in the Monte Carlo study and the direct inclusion of the covariate to estimate mixing proportions. In the previously conducted simulation studies, researchers used mixture Rasch model to distinguish among very distinct subpopulations. But in real world situations, the unknown but true

underlying pattern might be more complicated and/or more subtle than the simply hypothesized pattern in the simulated data. It is therefore important to analyze various sets of simulated data which differ in the degree of differences among mixture components so that this current study can provide advices and cautions about to what extent researchers can obtain successful model estimation.

Additionally, I evaluate the performances of MRM with a covariate and MRM without a covariate so that results from the comparison can demonstrate the advantages and possible drawbacks of incorporating collateral information with evidence from the Monte Carlo study. It is believed that the comparison of MRM with a covariate and MRM without a covariate can serve as the grounding to provide supports and evidence about the value of including the auxiliary variable in the MRM.

### **1.7 Contribution from this Dissertation Study**

It is expected that with a careful design of this simulation study, a contribution can be made towards understanding how the MRM performs under different simulated conditions, and caveats for using this complex model can be provided for other researchers, especially those who are interested in applying MRM for DIF detection under the latent class framework. The contributions of the current study are to model the recovery of simulated structures as a function of various effect sizes on item difficulty across latent groups, to explore the effects of connections between the covariate and latent classes on the mixture Rasch model under these conditions, and to provide advices for dealing with the label-switching problem in this particular form of mixture modeling under MCMC estimation.

When researchers try to apply complex statistical models on real data, they always first consider the capacity and accuracy of the model as two important aspects. This research study aims at providing concrete suggestions for the practical application of this particular mixture Rasch model with a covariate. The current simulation study mainly focuses on the effects of various DIF effect sizes, latent mixing proportions and the magnitudes of connections between latent classes and the covariate. This study seeks to answer the following research questions: (1) How do these main simulation factors affect the performance of MRM with a covariate?; and (2) Under some of the extreme simulation conditions, which are small DIF effect size, weak links between the distributions of covariate and the distributions of latent group membership, how does MCMC estimation recover the generating true parameters?

### **1.8 Brief Summary of Following Chapters**

In following chapters, the MRM with a covariate is proposed to model the scenario of different sets of item parameters for different latent classes. Chapter 2 is a detailed literature review of mixture Rasch model. Specifically, I describe the historical background and relevant recent studies employing MRM. After the literature review, perspectives on drawbacks and possible improvement for those research studies are discussed. The last part of Chapter 2 presents model estimation, difficulty in identifying this complex statistical model and the corresponding solutions to model convergence. By doing this, I show that the complexity of MRM brings both challenges and promises to representing complex item characteristics and data structures.

Chapter 3 is a thorough description of this Monte Carlo study in modeling MRM with a covariate. Mainly, I describe the manipulating simulation factors and compare these factors to relevant research studies. After the description of the stimulation design is the presentation of detailed model set-up and a unique way to include the covariate in the MRM.

Chapter 4 is the description of my findings and result summaries. There are two major parts in my result summary section. The first part is to understand how each simulation factor affects the recovery of underlying structures and the percentage of stable estimation for each simulation cell. The second part is to study how each simulation factor influences the accuracy of parameter estimations. I apply multiple regression methods to evaluate the relative important effects from the five manipulated simulation factors. In addition, I use charts to graphically show the effects from these simulation factors upon the accuracies of parameter estimations.

In the final chapter, Chapter 5, I restate the research questions and discuss how the main findings from the current study answer the proposed research questions. This is followed by the descriptions of future directions for researchers with similar interests in model development to expand this proposed mixed IRT model.

## Chapter 2: Literature Review

In this chapter, the parameters of the mixture Rasch model are defined and reviews of relevant research studies are presented. The main content for this chapter is divided into three sections. In the first section, I describe the Rasch model and latent class model. The next section explains the formulation of the mixture Rasch model with detailed notation. In the third section, I review a number of studies on MRM, regarding the early development, relevant simulation studies and recent applications of MRM. Additionally, I compare parameter estimation techniques and review solutions for solving identification problems in model estimation. This chapter emphasizes the necessity of a complex model with the possibility to measure sophisticated data structures, the flexibility to suit various modeling hypotheses, and the capacity to satisfy empirical explanations of analytical results.

### **2.1 Rasch Model and Latent Class Model**

As mentioned in Chapter 1, the Rasch model is one source of the modeling ideas for the proposed MRM with covariate. Danish statistician Georg Rasch first introduced this model in 1960 (Rasch, 1960, 1961). The simple Rasch model estimates only one item characteristic—the item difficulty. Summarized in the book by Lord (1980), this model is considered a special case of the item response theory models.

In general, researchers use the Rasch model to analyze dichotomous choice data, which are normally coded as 0 and 1 for further model estimation. Within the educational context, ‘0’ represents a subject’s incorrect answer to an item and ‘1’ represents a

subject's correct answer to an item. As a basic function of the Rasch model, this model calculates the conditional probability of the correct answer for each item. This is called the item characteristic function in measurement theory. The Rasch model estimates the conditional probability through the logistic function conditioning of item parameters and person parameters. The main purpose for this fundamental IRT model is to estimate the quantitative traits processed by each subject and each item.

Usually in IRT, a particular response pattern is the responses from a person to all items in a test or survey. In this study, a subject's response pattern is denoted as 'X' which represents responses (i.e.,  $x_1, \dots, x_j$ ) from a subject to all the items in a test form. The mathematical form of the Rasch model is given below:

$$P(X | b_1, \dots, b_N, \theta_1, \dots, \theta_M) = \prod_{i=1}^N \prod_{j=1}^M \left( \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)} \right)^{x_{ij}} \left( 1 - \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)} \right)^{1-x_{ij}} \quad (2.1)$$

In the Eq. (2.1), I use 'j' to represent a subject and 'i' to represent an item. I denote 'b<sub>i</sub>' as the item difficulty parameter for item 'i'. In addition, 'θ<sub>j</sub>', which is usually referred to as the person parameter in the Rasch model, stands for the latent ability of a person 'j'. I use X<sub>ij</sub> to represent the observed response from the subject 'j' to item 'i'. A defining feature for the Rasch model is the intention to estimate both the item difficulty and the person's latent ability and put them in the same latent continuum accordingly. The graphic representation for this effort is shown in Figure 2.1 below.

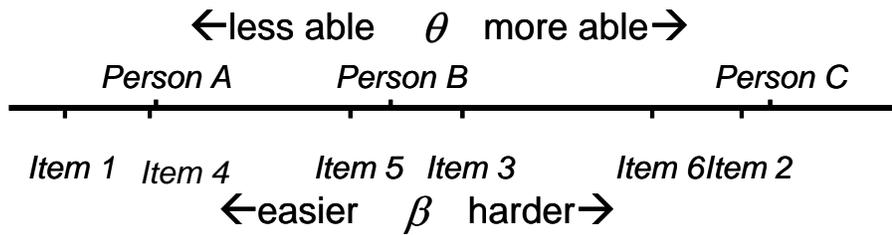


Figure 2.1: The latent continuum for the Rasch model

Ordering each item and each person in the latent scale shows the level of difficulty for each item and latent ability for each subject easily. For example, in Figure 2.1, person C has a higher latent ability than person A or person B because person C's ability is the closest to the right hand side of the latent continuum, the area where the most difficult items are located. Thus, it is expected that person C has the highest probability to answer the test items correctly, followed by person B and then by person A. By the same token, item 2 would be the most difficult item for all the subjects because this item is located on the hardest end of the hypothesized latent scale.

Another source of modeling features for MRM is the latent class model (LCM). It has been decades since Lazarsfeld & Henry (1968) first developed the LCM. This statistical method primarily aims to analyze dichotomous response data and identify homogenous groups within heterogeneous samples. This data format and modeling goal are popularly applied in behavioral sciences. Therefore, methodologists regard LCM as a very useful statistical tool to analyze quantitative social science data (Dayton, 1999; McCutcheon, 1987).

Though both Rasch models and latent class models are used for analyzing discrete data, they are quite different in their assumptions of underlying data structures. For the

Rasch model, the goal is to obtain the measurement of item parameters and person parameters. Through conditioning on person parameters and item parameters, researchers assume local independence among subjects. This assumption means the probability of each unique response pattern is equal to multiplying probabilities for each item after conditioning on person parameters and item parameters.

For the latent class model, the modeling goal is to identify unobserved qualitative groups among subjects, such as an honest group of examinees and cheating group of examinees in an educational setting. Though the assumption of local independence is also assumed for this line of latent class analysis, the application scenario is different from that of the Rasch model. For the latent class model, the probability of each unique response pattern is equal to the weighted average of the  $G$  latent class-specific probabilities. Under the assumption of the existence of  $G$  latent classes, the basic model for response patterns  $(x_1, \dots, x_j)$  is denoted as  $X$  and is given in Eq. (2.2) below:

$$\begin{aligned}
 P(X) &= \sum_{g=1}^G P(g = g_i) P(X | g = g_i) \\
 &= \sum_{g=1}^G P(g = g_i) \prod_j P(x_j | g = g_i), \tag{2.2}
 \end{aligned}$$

where  $i = 1, G$ .

As illustrated in the Eq. (2.2) above, I show that the Rasch model is more restrictive than the latent class model because the Rasch model assumes items to be homogenous for all subjects. On the contrary, for a particular response pattern in the latent class analysis, the model assumes subjects to be homogenous given their estimated latent class memberships. To combine the distinguishing modeling features from both the Rasch

model and the latent class analysis, researchers use MRM to suit less restricted model assumptions and estimate more complicated data structures.

## 2.2 Detailed Description of MRM

An important advantage of MRM is that this model allows for the estimations of different sets of item parameters for each latent class. Conditioning on item difficulty parameter, subject's latent trait and the corresponding latent class membership, the probability of a response (i.e.,  $X_{ij}$ ) is shown in the following Eq. (2.3).

$$P(X_{ij} | \theta_j, b_{ig}, g) = \sum_g \pi_g \left( \frac{\exp(\theta_j - b_{ig})}{1 + \exp(\theta_j - b_{ig})} \right)^{x_{ij}} \left( 1 - \frac{\exp(\theta_j - b_{ig})}{1 + \exp(\theta_j - b_{ig})} \right)^{1-x_{ij}} \quad (2.3)$$

Conditioning on item difficulty parameters, subjects' latent traits and latent class memberships, the probability of the response pattern (i.e.,  $X$ ) for  $N$  subjects is:

$$\begin{aligned} & P(X | \theta_1, \dots, \theta_J, b_{1g}, \dots, b_{Jg}, g) \\ &= \prod_{i=1}^I \sum_g \pi_g \prod_{j=1}^J \left( \frac{\exp(\theta_j - b_{ig})}{1 + \exp(\theta_j - b_{ig})} \right)^{x_{ij}} \left( 1 - \frac{\exp(\theta_j - b_{ig})}{1 + \exp(\theta_j - b_{ig})} \right)^{1-x_{ij}} \end{aligned} \quad (2.4)$$

Thus, the likelihood function of the mixture Rasch model is as follows:

$$\begin{aligned} & L(\theta_1, \dots, \theta_J, b_{1g}, \dots, b_{Jg}, g | X) \\ &= \prod_{i=1}^I \prod_{j=1}^J \sum_g [\pi_g P(X_{ig}=1 | \theta_j, b_{ig}, g)^{x_{ij}} (1 - P(X_{ig}=1 | \theta_j, b_{ig}, g))^{1-x_{ij}}] \zeta_{ij} \end{aligned} \quad (2.5)$$

In the Eq. (2.5) above, ‘j’ represents a particular subject, ‘i’ denotes a particular item, and ‘g’ stands for an assumed latent class in the model. As defined, ‘ $b_{ig}$ ’ represents the item difficulty parameters for item ‘i’ in latent class ‘g’. The observed response to item ‘i’ is represented by  $X_i$ . I denote latent mixing proportion for each latent class as ‘ $\pi_g$ ’. It should be noted that MRM differs from the regular Rasch model in that MRM specifies and estimates ‘g’ sets of item parameters instead of calculating only one set of item parameters for all subjects in the simple Rasch model. As shown in the likelihood function, the responses from subject ‘j’ to item ‘i’ (i.e.,  $x_{ij}$ ), are equal to 0 for an incorrect answer and 1 for a correct response.  $\zeta_{ij}$ , which is the latent class indicator for subjects, is equal to 1 for a particular latent class ‘g’ and 0 for all other latent classes.

In Figure 2.2 below, a graphic representative of MRM is shown to illustrate its basic modeling idea. As described previously, this model allows for the same set of test items being estimated differently for each latent class. For example, in latent class one, item 2 is the hardest item according to its location on the latent continuum. In contrast, in latent class 2, this item is an easy item since it is on the left hand side of the latent scale. In addition, MRM also calculates subjects’ probabilities of belonging to each latent class. Generally speaking, a subject’s latent group membership is decided based on his or her highest probability for a latent class. As shown in Figure 2.2, persons A, B, and C are categorized in latent class one while persons D, E, and F are categorized in latent class two. It is a characteristic feature of the MRM, as for latent class models, which assumes that any given person belongs to a particular latent class exclusively, but it is not known a priori which latent class it is.

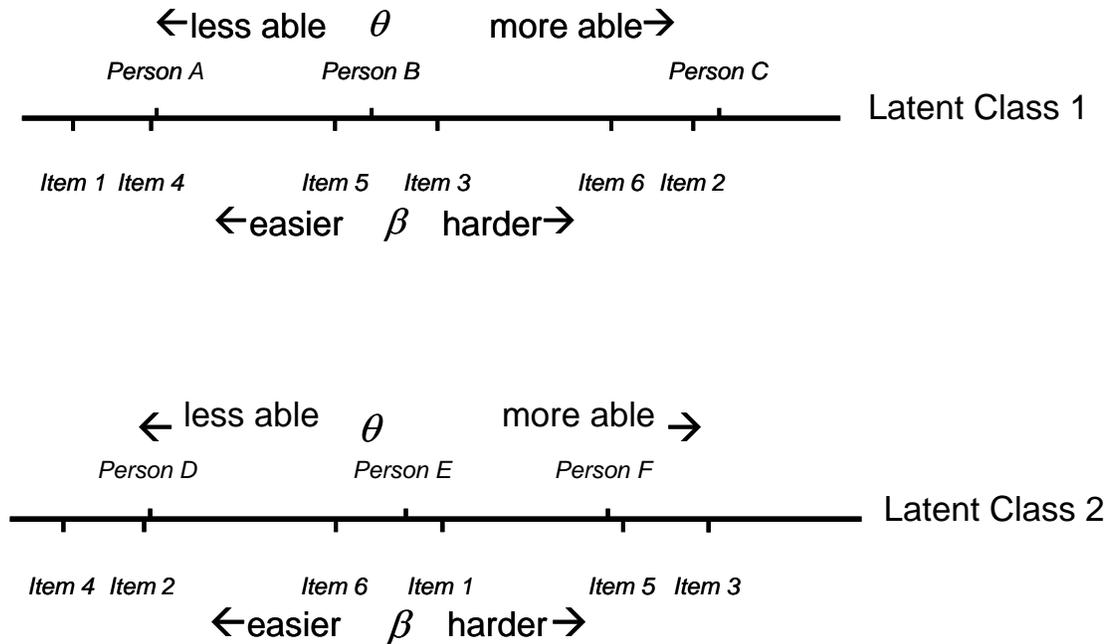


Figure 2.2 The latent continuum for the mixture Rasch model

### 2.3 Studies Related to the Development of Mixture Rasch Model

In the early 1990's, there were three important articles contributing to the initial development of the MRM. They are the Rost (1990), Kelderman & Macready (1990) and Mislevy & Verhelst (1990). These articles studied the MRM from different perspective with varying response functions and assumptions of underlying structures.

Rost (1990) proposed a model, named the 'mixed Rasch model' which combined the Rasch model with the latent class model. The aim of integrating these two modeling techniques is to combine the theoretical strengths of both approaches. The author first described the response probabilities and the latent class structure of the model and demonstrated how to use an EM algorithm to estimate parameters involved. After that, the MRM was used to analyze real data from a study on physical education as an application example. Data analysis showed that the simple Rasch model can't scale this

particular test data because the Rasch model assumed constant item difficulty parameters for all subjects.

The application example in Rost (1990) aroused interest in employing the latent class approach to further analyze subjects' response data when a simple Rasch model is not sufficient to account for complexity in the data. Therefore, the author applied the mixed Rasch model to this empirical data to identify structural differences in subjects' abilities and item parameter profiles with a two-class solution. Without assuming the response probabilities to be the same for all individuals in a latent class and allowing for different sets of item parameters to be estimated for each latent class, the mixed Rasch model proves to be a better analytical tool than a simple latent class analysis or the Rasch model alone. The formulation of MRM by Rost (1990) is the initial building block of the MRM with covariate for this study.

By combining the ideas of the latent class model and the latent trait model, Kelderman & Macready (1990) used the loglinear latent class model to analyze the interaction effects between the grouping variables (either manifest or latent) and item parameters. As mentioned previously, the MRM can be used in the DIF study in which researchers presume the existence of difference in item characteristics as the items don't function the same way for subjects with similar latent ability. In Kelderman & Macready (1990), the authors showed that it was possible while using the loglinear latent class models to explain DIF from different sources of data. By providing a general modeling framework, difference in item difficulties or error rate could be detected across levels of grouping variables. In addition, the occurrence of DIF across observed or unobserved groups could be estimated. Furthermore, Kelderman & Macready (1990) showed that it

was possible to model the interaction effects between the grouping variables (either manifest or latent) and item parameters. Though the framework under the loglinear model provides flexibility in modeling DIF, this dissertation study focuses on the model within the framework of IRT in which subjects' responses and the corresponding latent factors are modeled by using the logit link function. I am aware that a probit link function is also a very convenient link in MCMC. In the context of IRT, researchers have proven that the logit link is equivalent to the probit link after a simple linear transformation.

Another important work contributing to the early development of the mixture Rasch model is the paper by Mislevy & Verhelst (1990). The authors proposed the mixture linear logistic test model to explore the prior assumption that different strategies were adopted by different groups of subjects. By applying this particular mixed model, certain hypotheses were made about item features and the numbers of items and latent classes were restricted to some degree (Rost, 1990). The current study targets a general IRT framework without the pre-specification of certain item features and hypothesized response patterns.

A recent simulation study about MRM including collateral information from examinees was done by Smit et al. (1999). The authors pointed out that the Rasch model was too restrictive for social science data. To adapt to more complex data from social science studies, the authors suggested using either a more complicated IRT model or a hybrid of the Rasch model and latent class model. The hybrid approach was adopted in Smit et al. (1999) to identify homogenous Rasch scalable groups. This research study focused on evaluating the usefulness of incorporating collateral information in the MRM. Generally speaking, MRM by itself require either substantial differences between latent

groups or relatively large sample sizes to obtain good estimation results. The incorporating of a collateral variable helps to relieve the rigid requirement of latent class structure or sample size. The authors found out that standard error and the assignments of latent groups benefit substantially from the inclusion of a collateral variable related to latent groups.

There are recent studies that applied MRM to identifying items which function differently across groups of subjects. An important study conducted by Cohen & Bolt (2005) included two main studies. In the first study, the authors compared the DIF items identified through the traditional manifest groups approach and latent group approach. The analytical results showed that by using manifest group characteristics, such as gender or ethnicity, the cause of DIF items was difficult to explain. Additionally, Cohen & Bolt (2005) found that manifest characteristics, which were weakly related to latent groups, were actually disadvantaged or advantaged by test items.

In the second part of their article, Cohen and Bolt (2005) used a two-step procedure to identify DIF items in a college mathematics exam. The first step in Cohen & Bolt (2005) was to locate DIF item using mixture IRT model as an exploratory approach to identifying latent classes. After the latent classes were formed, to the authors compared characteristics of examinees for each latent class and formulated a plausible explanation of the qualitative differences among the identified latent classes. The limitation of their study is the usage of a two-step process to detect DIF. It should be noted that the identification of the cause of DIF in Cohen & Bolt (2005) is an 'after the fact' analysis while I directly include the covariate in the MRM in the proposed study.

In the dissertation study by Samuelsen (2005) and the corresponding chapter in the book about mixture models within latent variable framework by Hancock & Samuelsen (2008), the author argued that manifest characteristics, such as gender or ethnicity, could be poor proxies for educationally advantaged or disadvantaged groups. In Samuelsen's (2005) study, the author generated datasets based on different levels of overlap among manifest groups and latent groups. Estimations of the DIF effect through both the manifest DIF approach and latent class DIF approach using MRM were conducted. The results showed that as the overlaps between manifest groups and latent groups decreased the power to correctly identify DIF items decreased. Since it is always unknown whether the manifest characteristics are the true reasons for DIF, manifest DIF analysis might be problematic. Forming the latent classes of subjects instead is a better approach to determine educationally advantaged or disadvantaged groups.

A recent study, Cho et al. (2007), analyzed the influence of different kinds of prior information for detecting latent classes in the MRM. The main focus in Cho et al. (2007) was to examine the effect of various types of prior information on the MCMC estimation results. Including a covariate in the MRM served as one source of prior information in Cho et al. (2007) and the effect from the covariate on the estimation of MRM was briefly touched upon. In this proposed study, I systematically study the MRM with covariate and the influence of the covariate on estimating latent classes and item parameters is the main emphasis.

## 2.4 Model Estimation

From a modeling standpoint, the ultimate goal is to use a parsimonious mathematical model to well represent the underlying data structure. In other words, researchers try to use a smaller set of parameters to account for the complexity in observed data. By choosing a statistical model to fit the observed data, it is always assumed that the observed data is a random sample from corresponding probability distributions. There are two major mathematical functions in statistical inference—one is the probability function and the other is the likelihood function. The probability function is used to predict values based on statistical model with known parameters. In contrast, the likelihood function is used to estimate parameters conditioning on observed data and the random sample assumption. This implies that the statistical inference is based on the observed data and is used to compute the value of model parameters.

The maximum likelihood (ML) estimation has long served as a useful approach to estimate statistical models. In this estimation approach, parameters are calculated by seeking the largest likelihood for the observed data. For fitting finite mixture models, several early research studies contributed to this estimation approach. They were Hasselblad (1966, 1969), Day (1969) and Wolfe (1970). However, this approach has several drawbacks. Among those are the possibility of multiple local maximums, unbounded likelihood functions, and the difficulty in choosing suitable start values for the EM algorithm. On the contrary, the Bayesian estimation method is more suited for measuring the finite mixture model. Early research studies, such as Diebolt and Robert (1994) estimated the mixture model via posterior simulation by applying the Bayesian technique.

With the technical improvement in computer computing speed, Bayesian estimation has become more and more popular to estimate statistical models. Comparing to the more traditional estimation method (i.e., the maximum likelihood estimation), the Bayesian approach has the following advantages. First, the interpretation of confidence intervals is more meaningful under the context of the Bayesian approach than the frequentist approach. Second, incorporation of prior information in the mathematic form makes it possible to update belief about parameters. Third, the comparison of the nested and the non-nested model is more convenient when using the model fit index from Bayesian estimation. Fourth, in Bayesian estimation, no asymptotic properties are required and finite sample inference is possible. The requirement of asymptotic assumptions for ML estimation is a downside for using this estimation method to fit statistical models to data.

The following are the reasons why the Bayesian approach is regarded as a useful approach to estimate finite mixture models in particular. First, this approach allows for the inclusion of prior information into the model estimation process. With a structural exploration of existing literature, relevant studies would always provide information about parameter estimations and their corresponding probability distributions. Aside from the observed data, the additional source of information for other related studies helps to obtain better estimated model parameters and smooth the likelihood function for the finite mixture model (Fruhirth-Schnatter 2006). Additionally, with the availability of complete posterior distribution, the issue of parameter uncertainty is more easily addressed. Finally, different from MLE which relies on the assumption of normality or asymptotic normality, the Bayesian approach suits situations where the normality assumption is not robust or is heavily violated.

Fruhwirth-Schnatter (2006) also discussed the fact that practical use of Bayesian estimation presented difficulties in estimating the finite mixture model. This is because no natural conjugate prior is available for the mixture likelihood function which leads to the problem that the posterior density obtained via the Bayesian theorem does not belong to any tractable distribution family. This difficulty is solved with the advent of Markov chain Monte Carlo methods (MCMC).

The MCMC algorithm is essentially Monte Carlo integration using Markov chains. When there are no closed mathematic forms and analytical solutions, MCMC techniques, such as Gibbs sampling, allow the forming of a stationary distribution, which approximates the posterior distribution. Throughout multiple iterations, the MCMC technique draws random samples from the stationary posterior distribution and estimates parameters of interest accordingly.

In an important article by Patz & Junker (1999), the authors showed that the MCMC technique was a straightforward way to estimate complex IRT models. An important benefit of using the MCMC method is the incorporation of uncertainty about item parameters into trait inference and the incorporation of uncertainty about latent traits into estimating the item parameters. Moreover, there are several research studies on mixture IRT models adopting this Bayesian MCMC approach. These include the study by Bolt et al. (2001) which applied mixture models for analyzing multiple-choice items, Cohen & Bolt (2005) which adopted the concept from mixture 3PLM to identify differentiate functioning items and the simulation study by Samuelsen (2005) which compared the accuracy and power of the manifest DIF approach and latent class DIF approach to address the issues of DIF.

This proposed study is differed from the Smit, et al. (1999) study in which the MLE approach was adopted. Instead, the current study implements the MCMC algorithm on the MRM under Bayesian estimation. The advantage of Bayesian estimation is that the problem of the local maximum encountered in ML estimation could be avoided. In addition, typical issues such as the label-switching problem in Bayesian estimation might occur in the model estimation. I describe this issue in the following section and summarize the findings about this issue based on results from this simulation study.

Another important advantage of Bayesian method over maximum likelihood estimation is the flexibility to estimate complex models, especially when there is no closed-form mathematic expression of the probability function. Detailed description and application of the Bayesian method can be found in Congdon (2003, 2005, and 2007). As a member of the finite mixture models, the MRM is complex in its nature especially because it is difficult to identify the latent mixing proportions. Thus, it is important to adopt the Bayesian approach to examine the performance of the complicated MRM with covariate.

## **2.5 Identification Difficulty in Bayesian method**

It is inevitable that the flexibility of model estimation under the Bayesian MCMC approach comes with the hindrance of specifying prior distributions and identifying the model. Estimation of the finite mixture model is subject to some model identification issues, one of which is the so-called label-switching problem. As described in Jasra et al. (2005), “if a symmetric (exchangeable) prior is placed on component-related parameter, then the posterior distribution will be invariant to permutations in labeling of the

parameters. The marginal posterior distribution for the parameter will be identical for each mixture component.” In other words, it is quite possible that there is more than one way to label the mixing proportions, especially when the prior information is non-informative and identifiable constraints are missing. For example, in Figure 2.2, identical solutions might result if latent class two is relabeled as latent class one. This problem causes difficulty in identifying those parameters when their values are related to the mixing proportions.

In the literature, researchers suggest several solutions to the label-switching problem. There are artificial identification constraints (e.g., Diebolt & Robert 1994), relabeling of algorithms to perform a k-means type clustering of the MCMC samples (Stephens 1997, Celeux 1998), label invariant loss functions (Celeux, Hurn & Robert 2000, Hurn et al. 2003), and random permutation samplers of Fruhwirth-Schnatter (2001). In Jasra et al. (2005), the authors provide a detailed review of available solutions.

These solutions to the label-switching problem relied heavily on the mathematic computations and usually caused ambiguous explanations of estimated parameters. In a recent article by Chung et al. (2004), a simple solution to tackle the label-switching problem is suggested. The authors implemented this approach by supplying a small amount of estimated mixing proportions as known prior information to break the symmetry in posterior distribution. This can take the form of fixing the class memberships of a small number of examinees—usually one or two per class—to fixed values based on preliminary runs that suggest examinees for which these assignments have very high posterior probabilities under a given permutation. In my preliminary analyses, I adopted this simple solution introduced by Chung et al. (2004) to try to solve

the label-switching problem occurred in MCMC updating. I am aware that there are other feasible solutions, one of which is to impose an order restriction on the Dirchlet prior of the membership probabilities.

## Chapter 3: Methodology

In this chapter, I provide a detailed explanation for incorporating the covariate in the MRM. Also, I explain the factors manipulated in the current study. In addition to the description of varied simulation conditions, I illustrate how to analyze outcomes from this simulation study. Additionally, I indicate the computer programs used to generate and analyze the simulated data. Finally, I describe relevant graphic functions in the software which relate to evaluating the simulation study results.

### **3.1 MRM with covariate**

In the dissertation study by Samuelsen (2005), the author investigated the performance of MRM with covariates. Samuelsen (2005) incorporated covariates in the MRM by creating dichotomous variables as indicators of memberships within manifest groups and calculated proportions of every manifest group in each latent class individually and separately. In the proposed study, I define the manifest group proportions with a more flexible modeling technique. In this technique, these proportions are imbedded in the form of a logistic model. This approach makes it possible to model the relation between latent groups and distribution of manifest groups and make it possible to include more than one predictor in the logistic equation.

In the current study, I chose a dichotomous variable as the covariate incorporated in the MRM model. Within the context of DIF operational procedures, researchers from the testing industry are generally interested in comparing groups of subjects. Thus, categorical variables are of interest and one is used as a covariate in this study. As

mentioned in the previous paragraph, I use logistic regression model to include the covariate directly in the model estimation. This is a flexible way to incorporate additional covariates. Within this framework, either categorical and/or continuous or both types of variables can be added as a covariate in this proposed model.

In this simulation study, the specifications of probabilities are similar to Eq. (2.4) and Eq. (2.5) described in Chapter 2. To estimate MRM with a covariate, I specify the probabilities for N subjects after conditioning on item difficulty parameters, subjects' latent traits, and latent class memberships as:

$$\begin{aligned}
& P(X | \theta_1, \dots, \theta_J, b_{1g}, \dots, b_{Jg}, \mathbf{g}) \\
&= \prod_{i=1}^I \sum_g \pi_g \int \prod_{j=1}^J \left( \frac{\exp(\theta_j - b_{ig})}{1 + \exp(\theta_j - b_{ig})} \right)^{x_{ij}} \left( 1 - \frac{\exp(\theta_j - b_{ig})}{1 + \exp(\theta_j - b_{ig})} \right)^{1-x_{ij}} \\
&\quad \times P(\theta_j, b_{ig}, \mathbf{g}) d(\theta_j, b_{ig}, \mathbf{g}) \quad (3.1)
\end{aligned}$$

And the likelihood function of the mixture Rasch model is as follows:

$$\begin{aligned}
& L(\theta_1, \dots, \theta_J, b_{1g}, \dots, b_{Jg}, \mathbf{g} | X) \\
&= \prod_{i=1}^I \int \prod_{j=1}^J \sum_g [\pi_g P(X_{ig}=1 | \theta_j, b_{ig}, \mathbf{g})^{x_{ij}} (1 - P(X_{ig}=1 | \theta_j, b_{ig}, \mathbf{g}))^{1-x_{ij}}]^\zeta_{ij} \\
&\quad \times P(\theta_j, b_{ig}, \mathbf{g}) d(\theta_j, b_{ig}, \mathbf{g}) \quad (3.2)
\end{aligned}$$

The difference between the MRM with covariate and MRM without covariate is the extra layer of modeling technique from  $\pi_g$  (i.e., the mixing proportion) on the subject's

covariate. In this proposed study, the  $\pi_g$  is modeled by the logistic regression function with the covariate as the predictor in the regression equation.

As previously specified, I model a two-class solution in the latent class framework. Thus in a multivariate logit model, this means that the dependent variable  $\pi_{jg}$  has two outcome categories (i.e.,  $g = 1$  or  $2$  for person ‘j’). Furthermore,  $\pi_{j1}$  stands for the probability of a person ‘j’ belonging to latent class one and  $\pi_{j2}$  stands for the probability of a person ‘j’ belonging to latent class two. Also,  $y_j$  is used to represent the covariate for subject ‘j’. In the multinomial logit model shown in Eq. (3.3) below, this means that:

$$\text{logit}(\pi_{jg}) = \beta_{0g} + \beta_{1g} y_j \quad (3.3)$$

$$\pi_{jg} = \frac{\exp(\beta_{0g} + \beta_{1g} y_j)}{\sum_{g=1}^2 \exp(\beta_{0g} + \beta_{1g} y_j)} \quad (3.4)$$

$$\pi_{j1} = P(g=1 | y_j) = \frac{\exp(\beta_{01} + \beta_{11} y_j)}{\sum_{g=1}^2 \exp(\beta_{0g} + \beta_{1g} y_j)} \quad (3.5)$$

$$\pi_{j2} = P(g=2 | y_j) = \frac{\exp(\beta_{02} + \beta_{12} y_j)}{\sum_{g=1}^2 \exp(\beta_{0g} + \beta_{1g} y_j)} \quad (3.6)$$

As specified in Eq. (3.5) and Eq. (3.6), these expressions are not identified without further restrictions. One way of doing this is to require  $\sum_g \beta_{0g} = 0$  and  $\sum_g \beta_{1g} = 0$ , as illustrated in Bock (1972). An alternative solution is to fix  $\beta_{01} = 0$  and  $\beta_{11} = 0$ . In this

logistic regression set-up, I adopt the second restriction, which is  $\beta_{01}=0$  and  $\beta_{11}=0$ . In this way, Eq. (3.5) and Eq. (3.6) become:

$$\begin{aligned}\pi_{j1} &= P(g=1 | y_j) = \frac{\exp(0 + 0 \times y_j)}{\exp(0 + 0 \times y_j) + \exp(\beta_{02} + \beta_{12} y_j)} \\ &= \frac{1}{1 + \exp(\beta_{02} + \beta_{12} y_j)}\end{aligned}\quad (3.7)$$

$$\begin{aligned}\pi_{j2} &= P(g=2 | y_j) = \frac{\exp(\beta_{02} + \beta_{12} y_j)}{\exp(0 + 0 \times y_j) + \exp(\beta_{02} + \beta_{12} y_j)} \\ &= \frac{\exp(\beta_{02} + \beta_{12} y_j)}{1 + \exp(\beta_{02} + \beta_{12} y_j)}\end{aligned}\quad (3.8)$$

In this study, I define the covariate as a dichotomous variable with values corresponding to either ‘0’ or ‘1’. This means for a particular subject ‘j’, his or her  $y_j$  is equal to either ‘0’ or ‘1’. To be more specific, I lay out the detailed equations regarding the two possible and exclusive situations for subject ‘j’:

- i. when the subject’s covariate  $y_j$  is equal to 0;
- ii. when subject’s covariate  $y_j$  is equal to 1.

For situation (i) in which  $y_j = 0$ , Eq. (3.7) and Eq. (3.8) can be further simplified to:

$$\begin{aligned}\pi_{j1} &= P(g=1 | y_j = 0) = \frac{\exp(0 + 0 \times 0)}{\exp(0 + 0 \times 0) + \exp(\beta_{02} + \beta_{12} \times 0)} \\ &= \frac{1}{1 + \exp(\beta_{02} + \beta_{12} \times 0)} \\ &= \frac{1}{1 + \exp(\beta_{02})}\end{aligned}\quad (3.9)$$

$$\begin{aligned}
\pi_{j2} = P(g=2 | y_j = 0) &= \frac{\exp(\beta_{02} + \beta_{12} \times 0)}{\exp(0 + 0 \times 0) + \exp(\beta_{02} + \beta_{12} \times 0)} \\
&= \frac{\exp(\beta_{02} + \beta_{12} \times 0)}{1 + \exp(\beta_{02} + \beta_{12} \times 0)} \\
&= \frac{\exp(\beta_{02})}{1 + \exp(\beta_{02})} \tag{3.10}
\end{aligned}$$

For situation (ii) in which  $y_j = 1$ , Eq. (3.7) and Eq. (3.8) can be further simplified to:

$$\begin{aligned}
\pi_{j1} = P(g=1 | y_j = 1) &= \frac{\exp(0 + 0 \times 1)}{\exp(0 + 0 \times 1) + \exp(\beta_{02} + \beta_{12} \times 1)} \\
&= \frac{1}{1 + \exp(\beta_{02} + \beta_{12} \times 1)} \\
&= \frac{1}{1 + \exp(\beta_{02} + \beta_{12})} \tag{3.11}
\end{aligned}$$

$$\begin{aligned}
\pi_{j2} = P(g=2 | y_j = 1) &= \frac{\exp(\beta_{02} + \beta_{12} \times 1)}{\exp(0 + 0 \times 1) + \exp(\beta_{02} + \beta_{12} \times 1)} \\
&= \frac{\exp(\beta_{02} + \beta_{12} \times 1)}{1 + \exp(\beta_{02} + \beta_{12} \times 1)} \\
&= \frac{\exp(\beta_{02} + \beta_{12})}{1 + \exp(\beta_{02} + \beta_{12})} \tag{3.12}
\end{aligned}$$

For the Eq. (3.9), Eq. (3.10), Eq. (3.11), and Eq. (3.12) shown above, the quantities in these equations are equal to:

$$\pi_{j1} = P(g=1 | y_j = 0) = \frac{1}{1 + \exp(\beta_{02})} \tag{3.13}$$

$$\pi_{j2} = P(g=2 | y_j = 0) = \frac{\exp(\beta_{02})}{1 + \exp(\beta_{02})} \tag{3.14}$$

$$\pi_{j1} = P(g=1 | y_j = 1) = \frac{1}{1 + \exp(\beta_{02} + \beta_{12})} \quad (3.15)$$

$$\pi_{j2} = P(g=2 | y_j = 1) = \frac{\exp(\beta_{02} + \beta_{12})}{1 + \exp(\beta_{02} + \beta_{12})} \quad (3.16)$$

Thus the odds ratio between different covariate groups and latent classes is equal to the following:

$$\text{Odds ratio} = \frac{(P(g=1 | y_i = 0) / (P(g=2 | y_i = 0)))}{(P(g=1 | y_i = 1) / (P(g=2 | y_i = 1)))} \quad (3.17)$$

Based on Eq. (3.13), Eq. (3.14), Eq. (3.15), and Eq. (3.16), the odds ratio is equal to

$$\text{Odds ratio} = \frac{(P(g=1 | y_i = 0) / (P(g=2 | y_i = 0)))}{(P(g=1 | y_i = 1) / (P(g=2 | y_i = 1)))} = \exp(\beta_{12}) \quad (3.18)$$

The above inference about  $\exp(\beta_{02})$  is also shown in Appendix A.

As mentioned before, the form of the logistic regression function is used to model the relation between the covariate with latent classes. For this proposed study, I use computer program WinBUGS (Spiegelhalter, Thomas & Best, 2000) to measure the proposed MRM model with covariate using the Bayesian MCMC estimation. In WinBUGS, I specify the corresponding mathematical expression in Eq. (3.3) and Eq. (3.4) as the following:

$$P[i,g] = \text{PHI}[i,g] / \text{sum}(\text{PHI}[i,]) \quad (3.19)$$

$$\log(\text{PHI}[i,g]) = \text{int}[g] + \text{sl}[g] \times \text{cov}[i] \quad (3.20)$$

$$\text{PHI}[i,g] = \exp(\text{int}[g] + \text{sl}[g] \times \text{cov}[i]) \quad (3.21)$$

In this study,  $g$  is given the value of 1 or 2. And I indicate in WinBUGS the restriction required to identify the multinomial logit model as:

$$\text{int}[1] = 0, \quad (3.22)$$

$$\text{sl}[1] = 0. \quad (3.23)$$

In WinBUGS code shown in Eq. (3.19), Eq. (3.20), Eq. (3.21) Eq. (3.22), and Eq. (3.23), I use slightly different denotations.  $P[i,g]$  stands for the conditional probability for subjects ‘i’ being in latent class ‘g’.  $\text{int}[g]$  is the intercept parameter and  $\text{sl}[g]$  is the slope parameter in the logit model.

When  $g=1$  and applying Eq. (3.22) and Eq. (3.23), we have the following:

$$\begin{aligned}
P[i,1] &= \frac{PHI[i,1]}{PHI[i,1] + PHI[i,2]} \\
&= \frac{\exp(\text{int}[1] + \text{sl}[1] \times \text{cov}[i])}{\exp(\text{int}[1] + \text{sl}[1] \times \text{cov}[i]) + \exp(\text{int}[2] + \text{sl}[2] \times \text{cov}[i])} \\
&= \frac{\exp(0 + 0 \times \text{cov}[i])}{\exp(0 + 0 \times \text{cov}[i]) + \exp(\text{int}[2] + \text{sl}[2] \times \text{cov}[i])} \\
&= \frac{1}{1 + \exp(\text{int}[2] + \text{sl}[2] \times \text{cov}[i])} \tag{3.24}
\end{aligned}$$

When  $g=2$  and applying Eq. (3.22) and Eq. (3.23), we have the following:

$$\begin{aligned}
P[i,2] &= \frac{PHI[i,2]}{PHI[i,1] + PHI[i,2]} \\
&= \frac{\exp(\text{int}[2] + \text{sl}[2] \times \text{cov}[i])}{\exp(\text{int}[1] + \text{sl}[1] \times \text{cov}[i]) + \exp(\text{int}[2] + \text{sl}[2] \times \text{cov}[i])} \\
&= \frac{\exp(\text{int}[2] + \text{sl}[2] \times \text{cov}[i])}{\exp(0 + 0 \times \text{cov}[i]) + \exp(\text{int}[2] + \text{sl}[2] \times \text{cov}[i])} \\
&= \frac{\exp(\text{int}[2] + \text{sl}[2] \times \text{cov}[i])}{1 + \exp(\text{int}[2] + \text{sl}[2] \times \text{cov}[i])} \tag{3.25}
\end{aligned}$$

By comparing Eq. (3.24) & Eq. (3.25) with Eq. (3.7) and Eq. (3.8), I show that the logistic regression function is correctly specified in the WinBUGS code shown in Appendix C. Based on findings from the preliminary analyses, I chose more informative

prior distributions so that the model could be estimated in WinBUGS and these prior distributions are similar to those used in other researchers' studies, such as those prior distributions specified in Samuelsen (2005).

The prior distributions for the estimated parameters in WinBUGS are as follows:

- i. Distribution of item difficulty parameters for latent class  $g$ :  $b_{ig} \sim \text{Normal}(0, 1)$
- ii. Distribution of differences in item difficulty between two latent classes (DIF effect):  $b_{i2} - b_{i1} \sim \text{Normal}(0, 0.5)$
- iii. Distribution of subjects' latent ability:  $\theta_{jg} \sim \text{Normal}(\mu_g, \tau_g)$
- iv. Means in the distribution of subjects' latent ability:  $\mu_g \sim \text{Normal}(0, 1)$
- v. Precision in the distribution of subjects' latent ability:  $\tau_g \sim \text{Gamma}(1, 1)$
- vi. Distribution of intercept in the logistic regression:  $\beta_{02} \sim \text{Normal}(0, 1)$
- vii. Distribution of slope in the logistic regression:  $\beta_{12} \sim \text{Normal}(0, 1)$

In the WinBUGS setup shown in Eq. (3.19),  $P[i, g]$  is used to represent the probability that subject 'i' belongs to latent class 'g'. The value of  $P[i, g]$  is set to be equal to a fraction with the  $\text{PHI}[i, g]$  as the numerator divided by the denominator, sum of  $\text{PHI}[i, \cdot]$  across latent classes. Since  $\text{PHI}[i, g]$  could be larger than one, in this way, the conditional probability,  $P[i, g]$ , is always smaller or equal to one.

Eq. (3.20) and Eq. (3.21) above show how I model an individual's covariate in WinBUGS so that the connection between the covariate and examinee's probability of belonging to certain latent classes could be investigated. To include various kinds of covariates, either dichotomous or continuous, the logarithm of  $\text{PHI}[i, g]$  is set as the criteria variable and the covariate is entered as the predictor variable in the regression

function. Again in Eq. (3.20) and Eq. (3.21),  $\text{int}[g]$  stands for the intercept for latent class 'g' and  $\text{sl}[g]$  is the slope for latent class 'g'. As noted above, for each latent class, there is one separate set of intercept and slope to be estimated. For the purpose of model identification,  $\text{int}[1]$  and  $\text{sl}[1]$ , which are the intercept and slope for latent class one, are set at zero. As shown in the following section, adding these two constraints does not change the flexibility to model distribution of manifest groups in each corresponding latent class.

I provide the above mentioned WinBUGS code for estimating the MRM with covariate in Appendix C. For modeling item difficulty parameters, it should be noted that one set of item difficulty is set up and also the DIF parameters are modeled and estimated. For the other latent class, I calculate item difficulty parameters by adding the DIF parameters to the set of item difficulty parameters for the first latent class.

The following example is provided to illustrate the method. In this illustration, taking advanced math classes in high school is used as a hypothesized dichotomous covariate for an individual student. The variable is dummy-coded, with one representing taking advanced math classes and zero as not taking the classes. To ease interpretation, only one covariate is involved in the logistic regression equation. However, having more than one covariate as predictors would be desirable in future study. Under the logistic regression framework, Table 3.1 gives the formulas to calculate the distributions of the covariate based on the estimated intercept and slope from the logistic regression functions.

Table 3.1: Relation between parameters in logistic regression and membership probabilities, as a function of the person's covariate

	Not taking advanced math class (nt)=0	Taking advanced math class (t)=1
Latent class 1	$P[i,1]_{nt} = 1/(1+\exp(\text{int}[2]))$	$P[i,1]_t = 1/(1+\exp(\text{int}[2]+\text{sl}[2]))$
Latent class 2	$P[i,2]_{nt} = \exp(\text{int}[2])/(1+\exp(\text{int}[2]))$	$P[i,2]_t = \exp(\text{int}[2]+\text{sl}[2])/(1+\exp(\text{int}[2]+\text{sl}[2]))$
	100%	100%

### 3.2 Primary Design of the Simulation Study

This section focuses on factors manipulated or kept constant in the proposed simulation study. An important contribution of this methodology study is to provide suggestions or cautions for researchers who apply theoretical models to the analysis of practical problems. To serve this purpose, I make levels for each manipulated factor consistent with real world test scenario or similar to relevant research studies in the literature. Table 3.2 (below) briefly describes the manipulated factors and their corresponding levels.

Table 3.2: Factors varied in this simulation study

Factors varied in this simulation study	Corresponding values
distribution of subjects' latent ability	$N(0,1), N(0,1); N(0,1), N(1,1)$
distribution of two latent classes	15%:85%, 30%:70%, 50%:50%
the number of items with large DIF effect size	20%, 40%,
magnitude of $\exp(\text{sl}[2])$	1, 10, 50
Proportions for the covariate	50%:50%, 30%:70%

In this research, I have two levels in the simulation study for the distributions of subjects' latent abilities. The first level is to have the normal distribution with mean equal to zero and variance equal to one for both latent classes. This situation ensures that the distinction of latent groups in the mixture Rasch model is decided only from the perspective of item difficulties, as the distribution of examinees' latent abilities is

assumed to be the same. However, as shown in results from the preliminary analysis (discussed further in chapter 4), when the proportion of different item difficulties is small, there are great burdens in model estimation. In some extreme cases, the model can't be identified or the estimation is not converged. Thus, the second level for distributions of subjects' latent abilities is set to be Normal (0, 1) for one latent group and Normal (1, 1) for another group. In this way, more distinct latent groups are generated and it is expected that estimation of MRM with one covariate would be more easily converged.

Based on the total sample size, I vary the distribution of subjects between two latent classes across cells. This factor in the study is manipulated systematically and is differed from other relevant research studies (Smit et al. 1999 and Cho et al. 2007), which only had a 50%:50% split for the latent proportions. To mimic possible types of splits in a practical set-up, I vary the distributions of the two latent classes with three conditions, 50%:50%, 30%:70%, 15%:85%. These three conditions should be able to mimic most basic situations in real testing practices.

An important characteristic of the MRM is that this model allows for the estimations of different sets of item difficulties for each latent class. However, as discussed in Rost (1990), the sum of the item difficulty in each latent class should be made equal to zero. This restriction is equivalent to having the sum of item difficulty parameters,  $\sum_i b_{ig}$ , equal to zero within each latent class. This constraint is used to make sure to resolve the usual indeterminacy encountered in IRT models.

In the current study, I set up item difficulty parameters after two distinctive conditions have been satisfied to meet the constraint described above and make sure the differences in item parameters for each latent class are distinguishable for the model to

identify different latent groups of subjects. The first condition is to separate or place items with relatively large DIF effect size across two latent classes, so that subjects could be classified into different latent groups. The second condition is to have items with relatively small DIF effect size with opposite sign to the large DIF item. In this way, the sum of item difficulties within each latent class is equal to zero. In this current study, I set the large DIF effect size equal to values around 1.3. For the relatively small difference in item difficulty, the value is around 0.3. As an important factor of interest in the simulation study, the proportion of DIF items would be 20% and 40%. Correspondingly, with the test length equal to 30, the number of items with large DIF effect size is equal to 6 and 12.

The most important factor varied in the simulation study is the magnitude of connection between the covariate and the distribution of latent classes. It is expected that the stronger the covariate related to the classification of subjects into different latent groups, the better estimate of parameter it would be. This means there is less standard error and more accurate estimates. Thus, this factor consists of three different levels: non-connection (i.e., independence), moderate connection, and strong connection between the covariate the latent classes.

To express the effect of the covariate on the latent group classification, I use the odds ratio as a summary statistic to represent the effect mathematically. In the Appendix A, I illustrated that the  $\exp(\text{slope}[2])$  is equal to the odds ratio of comparing the subjects from one manifest group against the other manifest group for the covariate. The value of this single index clearly demonstrates the magnitude of the connection between the covariate and the latent classes. The different levels of  $\exp(\text{slope}[2])$  are set equal to 1, 10 and 50,

which correspondingly represent zero, medium, and strong relations between the covariate and latent groups.

In this simulation study, a final but equally important factor manipulated is the distribution of manifest groups for the covariate. As mentioned in the previous section, to put the MRM model with covariate in a hypothesized context, the covariate can be hypothetically constructed as examinees who take advanced math classes in high school versus examinees who don't take advance math classes. Generally, in this real world example, the composition of this variable is a small manifest group versus a large manifest group. In another scenario for the DIF analysis, the manifest groups for the covariate are of similar sizes, such as males and females. Thus to represent popular contexts for practical DIF analysis, the two levels varied for the composition of subjects' covariates are set to be 30%:70% and 50%:50%.

After describing conditions varied in the simulation study, it should be noted that there are some design characteristics that are kept constant. Because it is not possible in any simulation study to investigate all factors of potential interest, in light of constraints on time and computation, this ensures that the main efforts focus on the important factors of interest. These unvaried conditions are the number of latent classes, the number of covariates involved and the number of items. These unvaried factors are listed in Table 3.3 below.

Table 3.3: Factors kept constant in this simulation study

Factors kept constant in this simulation study	Corresponding value
number of latent classes	2
number of covariate	1
test length	30
Total number of subjects	1000

To start from the most common situation in applied research studies, the number of latent classes is not varied and kept equal to two. This is consistent with the simulation design in Smit et al. (1999) and Samuelsen (2005). Additionally, to study the interaction between identification of latent classes and distribution of manifest variables, this current study includes one covariate in modeling MRM. I include a single covariate because the introduction of the covariate inevitably introduces relative complexity to the parameter estimations. To systematically study the effect of the covariate on overall model estimation and control the complexity of the model, I start the investigation from the case of a single covariate.

The third factor that is kept constant is the test length, which is equal to 30 items. This is larger than the test length, 20, used in Samuelsen (2005) and is triple the test length, 10, used in Smit et al. (1999). It is intended that results from the present simulation study would be able to provide suggestions for practical test situations. And 30 items for a test is a relatively reasonable test length and is often seen in educational examinations such as sections in standardized achievement tests and sections of the SAT and GRE examinations.

The last factor that remains unchanged is the number of subjects for each cell. Usually, researchers from relevant studies varied the number of subjects to compare the impact of a large number of subjects on the estimated parameters. However, in this study, more attention is focused on the impact of the covariate on the capacity to distinguish between latent groups of subjects. Additionally, although more subjects might provide more information for estimating complex models, a larger sample size also requires a heavier computing effort. Thus, the size of subjects is kept unchanged to simplify the

design of the study. The total sample size for each replication was kept at 1000, to complete this simulation study within a reasonable time frame and involving a moderate number of resources. The sample size of 1000 is often seen in real-world testing practice. Future research might find it useful to study the model with varying sample sizes.

Based on the experience from my preliminary simulation study, each single simulation run takes about six hours. Because of the nature and complexity of the simulation outcomes, I conduct the processing of simulation results manually, instead of output extractions by the computer program. The number of replications for each cell is set equal to 10 to account for possible modification of the computer program. In some instances, the computer program may require further modifications after the initial runs and rerunning the analysis for certain types of outcomes (e.g. such as the runs with the occurrence of the between-chain label-switching problem). Ten replications for each cell would be enough to obtain a general idea of how MRM with a covariate performs under different simulation conditions.

For each cell in the simulation design, there is one extra replication added to the ten replications. I estimate this extra replication using the simple Rasch model on response data generated from the MRM. The reason for this extra replication is that in the preliminary study, there are occasions where the estimates of mixing proportions turn out to be a very small and a very large group while the generating parameters are far away from the estimated ones. Usually, for this type of result, the estimated item difficulty parameters for the large latent class are the average of two groups of generated item difficulties from both latent classes and the estimated item difficulty parameters for the small classes are the random combination of two groups of generated item parameters.

Thus, for each cell in the design, if this type of result appears, I use the additional replication for comparison.

In summary, I manipulate five simulation factors and keep the other four factors constant in the simulation study. Taking into account of the varied factors, there are  $3 \times 2 \times 3 \times 2 \times 2 = 72$  conditions in total. With 11 replications for each condition in the design, there are 792 simulation runs for the entire study.

### **3.3 Evaluation of Estimation Outcomes**

Based on findings from the preliminary analyses, an initial step for each simulation run should be the classification of results into different categories. Simulation outputs are classified due to various types of outcomes. Some of them might require additional modifications in the computer program and need to be followed by additional runs. For example, if the label-switching problem occurs during the estimation process, additional information will be added to the program. That is to say, based on a subject's posterior probabilities of belonging to each latent class, a few of them, who have the highest probability of belonging to a particular latent class, will be selected and their corresponding latent class membership will be made known and the program rerun. If the label-switching problem is solved by providing extra information, then the output from the converged run will proceed to the next step, which is the summary of estimated parameters.

### **3.3.1 Convergence**

Based on the experience from the preliminary study, the burn-in cycle for each simulation run is around 4000 iterations. After discarding the first 4000 iterations, I request an additional 6000 iterations to achieve convergence if possible. To identify the converged or non-converged MCMC chains, WinBUGS provides some useful graphic tools to visually identify whether multiple MCMC chain converged to a single solution or multiple solutions. This function is accomplished by looking at the history graph, which plots the values for requested parameter across all MCMC iterations. I use the mixing proportions as the first category of parameters to check for the issue of convergence since occurrence of the label-switching problem is intuitively shown for these parameters. In this study, I consider a replication properly converged if two requested MCMC chains, which start with different random values, merge into one stable chain and also the estimated parameters are within a reasonable range. First, I report numbers of converged replications. Second, I present a summary of the simulation outcomes.

### **3.3.2 Bias and RMSE**

To compare the parameters from the simulated data and the recovered data, root mean square error (RMSE) is frequently used to measure the accuracy of parameter estimates. In this simulation study, this index is used to quantify the difference between the generating parameters and their estimated values from WinBUGS. For each cell, RMSE is computed across the converged replications. The formula for RMSE is shown below. Basically, researchers compute RMSE as the average square root of the squared difference between estimated parameters from converged replications and the generated

parameters across simulation cells. Additionally, I calculate bias for each estimated parameter across simulation cells. Bias differs from RMSE in that, instead of calculating squared difference between estimated and true values in RMSE, this index indicates the average difference between the estimated value and the true value.

$$\text{RMSE}(\theta_{\text{estimated}}, \theta_{\text{true}}) = \sqrt{\frac{\sum_{i=1}^n (\theta_{i,\text{estimated}} - \theta_{\text{true}})^2}{n}} \quad (3.26)$$

$$\text{bias}(\theta_{\text{estimated}}, \theta_{\text{true}}) = \frac{\sum_{i=1}^n (\theta_{i,\text{estimated}} - \theta_{\text{true}})}{n} \quad (3.27)$$

It is of special interest to check two groups of estimated parameters against the generating parameters. The first are the estimated proportions of latent classes and proportions of manifest groups for covariates. And the second are the specified items identified to be significantly different across two latent classes (i.e., the large DIF items). In Chapter 4, I use charts to graphically show the effects of various simulation factors on parameter accuracy indexes (i.e., bias and RMSE).

### 3.4 Computer Programs

In this study, each replication dataset is processed in three stages, using three different software packages. In stage one, I use SAS (version 9.0, 2002) to generate dataset based on the hypothesized distribution and corresponding true parameters. In stage two, I use WinBUGS 1.4 to obtain estimated parameters for the MRM, item difficulties, mixing proportions, and group memberships for each subject. In the final stage, I extract estimates from the WinBUGS program into Microsoft Excel and compare with the true values to decide whether a simulation run is converged and if converged, I

compute the accuracy of estimated parameters. The program code for SAS and WinBUGS are provided in Appendix B and Appendix C.

## Chapter 4: Results Summary

In this chapter, I analyze and summarize output from the Bayesian MCMC estimation of simulated data. There are replications which accurately identify the underlying DIF structure while some other replications are unable to recover the generating DIF structure. Another phenomenon is the stability of MCMC estimation. As described in Chapter 3, the MCMC estimation might be converged or non-converged after 4000 burn-in cycles and 6000 additional MCMC cycles.

On one hand, the complexity of the simulation results provides rich information for researchers to study in terms of different possible results from the estimation of MRM with a covariate, especially under known but extreme simulation conditions. On the other hand, the complexity of the results adds a complication to identifying and categorizing each replication into different outcome groups for the result summary. I analyze the results in two steps:

- (1) I categorize the output into different qualitative outcome groups that reflect whether a single estimation run converged and if so, whether it recovered the generating two-class latent structure and the DIF structure;

- (2) From among those simulation runs that did recover the true generating structure, I present the summary statistics for evaluating parameter estimations. In these two steps, I study the influence of the simulation factors as the independent variables for the outcome analyses—category proportions in the first case and quantitative parameter recovery statistics in the second case.

In this research study, the result classification process is a multi-faceted process during which I decide certain cut-off points. Thus in this chapter, there are two major foci. The first focus is on the detailed description of result classification process. The second focus is on the evaluation of the estimated parameters.

#### **4.1.1: Review of History Plots from WinBUGS**

The first step in the result review process is to look at the history plots from WinBUGS and label each replication from the simulation study. In this step, I mainly focus on the estimation of mixing proportions after 4000 burn-in cycles. The goal for this step is to differentiate between converged replications and non-converged replications. The estimations of mixing proportion serve as good indicators when deciding whether a replication was converged or non-converged, because there is much information about them in the simulated data and the chains of converged solutions are visually distinct in WinBUGS. I examine the item parameter estimates and DIF parameter estimates in some ambiguous cases to help resolve the question of convergence.

The converged runs are those runs in which after 4000 burn-in cycles, the MCMC chains achieve stationary distributions. For a single replication, I classify the converged run into three possible groups of output. The first one is when two requested MCMC chains in WinBUGS converged to one solution (see Figure 4.1). The second one is when the two chains converged to two distinct solutions but after switching the labels of parameters from one latent class to the other, the estimations of parameters are essentially the same. These are the so called between-chain label-switching runs described in Chapter 3 (see Figure 4.2). Figure 4.2 shows an example of the WinBUGS history plot of

two chains from a simulation run in which there are two latent classes, with mixing proportions equal to 0.6 and 0.4. As presented in Figure 4.2, one MCMC chain has converged to a solution in which Class one is the latent class with 60% of the subjects and the other MCMC chain has converged to a solution in which Class one is the solution with 40% of the subjects. I verify this interpretation by examining the item parameters associated with the two solutions.

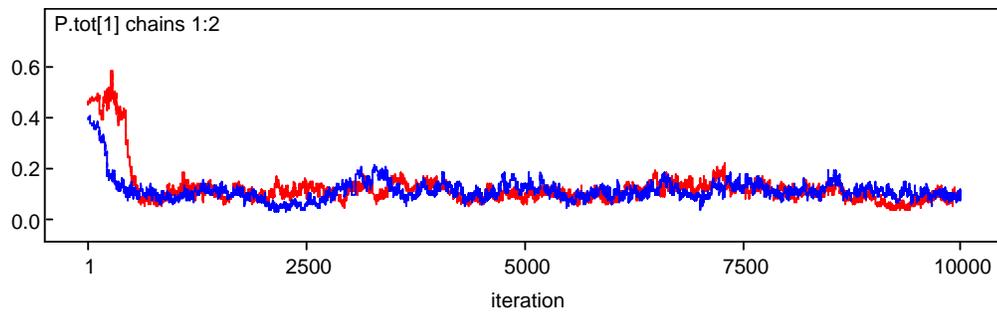


Figure 4.1: Example of a MCMC run with stationary distribution after burn-in cycles

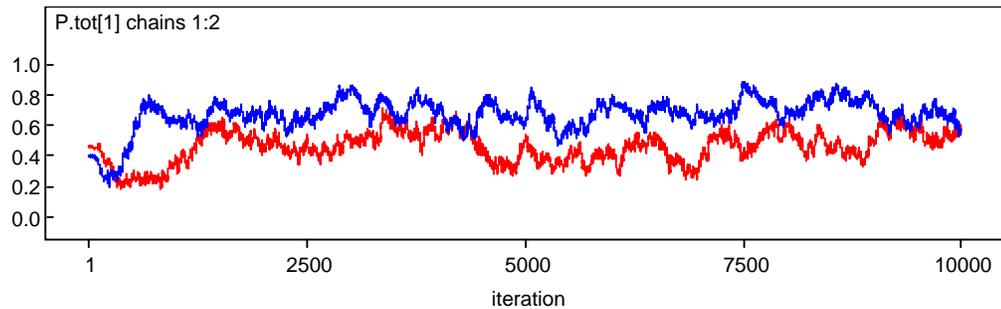


Figure 4.2: Example of a MCMC run with between-chain label-switching problem

The third group of the converged runs is a collapsed run with the mixing proportions close to zero or one. This type of result is essentially those simulation runs collapsed to one latent class solution. I exclude these collapsed runs from the final result summary of MRM parameter estimates. From the perspective of the MCMC estimation, these

collapsed runs are properly converged in the sense that the history plots exhibit convergence to a stationary distribution. However, from the perspective of the substance of that solution, they have not recovered a two-class structure but instead produced a solution with essentially one latent class.

As described in Chapter 3 regarding the design of this simulation study and simulation conditions, there are 10 replications of data simulated for each of the 72 combinations of simulation conditions. In addition, there is one more simulated dataset for each of the 72 cells and I use the simple Rasch model to estimate these datasets. By doing this, I am able to compare item parameters estimated from the MRM with a covariate and from the Rasch model, especially for the simulation runs that collapsed to one class solution.

I find that the correlation between the item difficulty parameters from the Rasch model runs and those from the dominant class in the collapsed runs from the MRM with a covariate is very high. That is to say the two-class solution has returned values that are essentially a single class, with item difficulties that correspond to those in the homogeneous population Rasch model analysis. This means that there was not enough information from the data to separate subjects into two latent groups. I identify the ‘collapsed runs’ phenomenon by locating those runs with class proportions close to 0% or 100%. Figure 4.3 shows an example of the collapsed runs. This example also illustrates the between-chain label-switching, with Class one being the class with near 100% in one MCMC chain and Class one being the class with near 0% in the other MCMC chain.

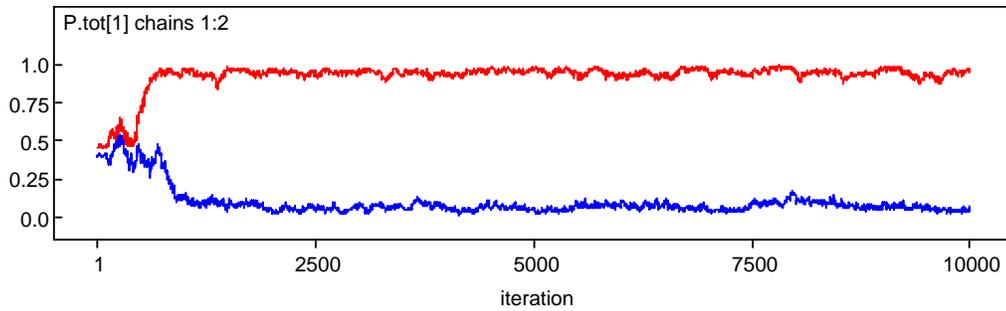


Figure 4.3: Example of a collapsed MCMC run with between-chain label-switching problem

The non-converged runs are those runs where after 4000 burn-in cycles, the WinBUGS history plot shows strong evidence that the MCMC chains has not attained a stationary position (see Figure 4.4). There are two types of runs belonging to the non-converged category. The first is the within-chain label-switching run. I identify this type of label-switching when the estimated value for a parameter switches between values from different latent classes within a MCMC chain. The other type of non-converged run occurs when MCMC chains fluctuate non-systematically within wide range of possible values for the parameter (see Figure 4.5). The occurrence of this type of fluctuation is called “poor mixing” in the MCMC literature.

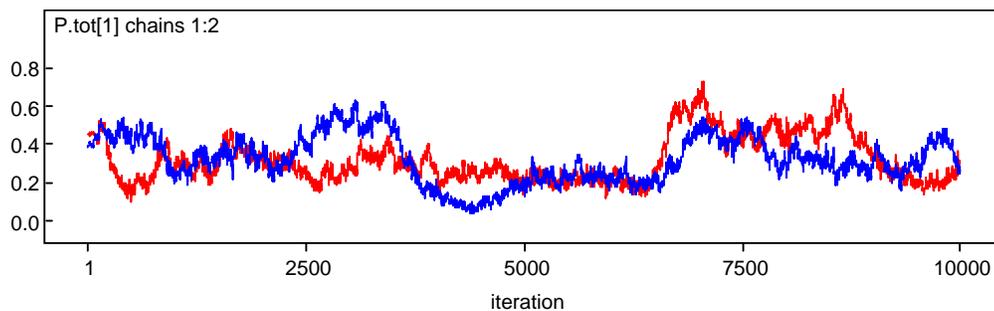


Figure 4.4: Example of a MCMC run with within-chain label-switching problem

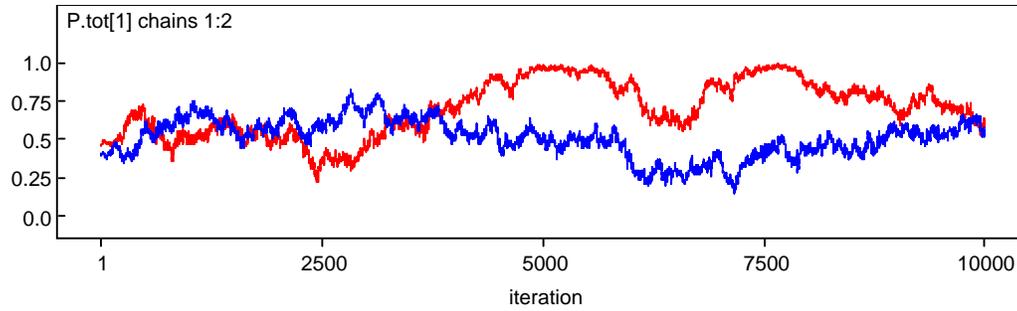


Figure 4.5: Example of a MCMC run with poor mixing

Most of the time, the mixing proportions are able to indicate the overall condition of an entire group of parameters for a single simulation replication. However, there are a small number of cases where I identify the ‘poor mixing’ issue in the mixing proportion parameters, but I find out that the estimations of item parameters are nevertheless stable. To be conservative in summarizing the simulation results, the output from this result category is excluded from the calculation of bias and RMSE. Below is an example of this type of results with Figure 4.6a for mixing proportion and Figure 4.6b for item difficulty.

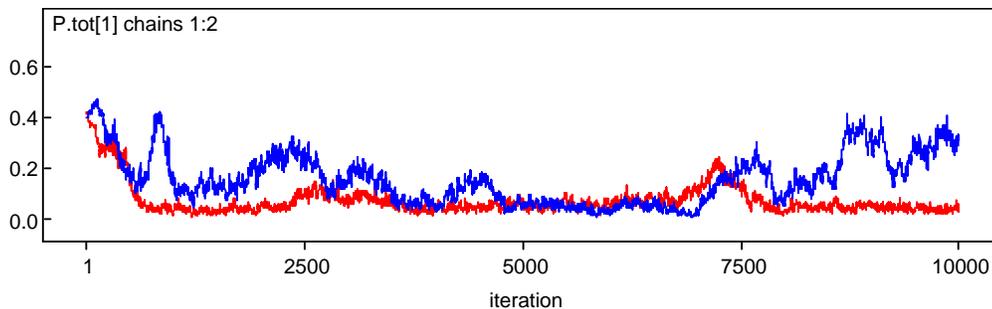


Figure 4.6a: Example of a MCMC run with poor mixing fluctuation in mixing proportion

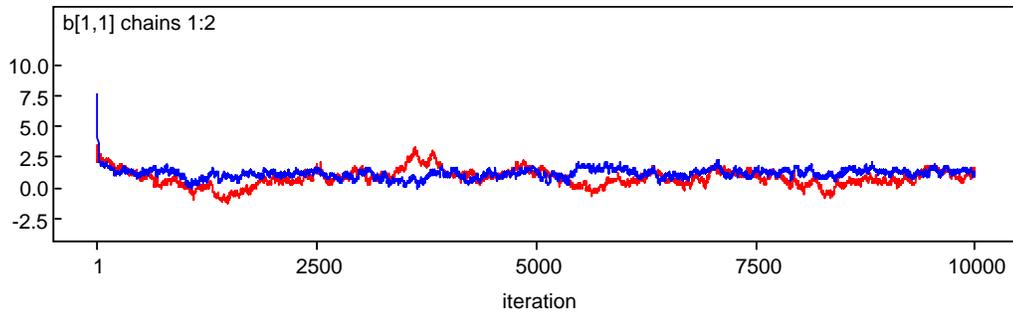


Figure 4.6b: Example of a MCMC run with stable estimation of item difficulty

#### 4.1.2 Review of the Correlations between Estimated and Generated values

After the initial review of all trace plots from WinBUGS, the next step in the classification process is to compare the estimated parameters with the generating parameters. I calculate the correlation between estimated and generating values for all DIF parameters because this correlation is a very good indicator for deciding whether a replication recovered the generating DIF structure. To be more specific, if the correlation is close to one then the correlation provides strong evidence that the generating DIF structure is correctly identified. Those items generated to have large DIF values did in fact exhibit large DIF estimates in the estimated solution, and those items for which the item parameters were generated similarly in both classes did in fact exhibit DIF estimates near zero. Conversely, if the correlation is close to zero, it means the generating DIF structure has not been well recovered.

The correlations between estimated DIF parameters and the generating structure have another important function in the result summary process. It is also a good indicator of whether parameters from a particular replication need to be relabeled or not. If the correlation between estimated DIF parameters and generating values is high and positive, this means that they are very close to true values of DIF parameters. If correlation

between estimated DIF parameters and generating values is high and negative, this means that the estimated DIF parameters are very close to the opposite sign of true DIF values. In that case, these parameters need to be relabeled. Figure 4.7 below is the histogram of the calculated Pearson correlations. In both sides of Figure 4.7, the correlation is very high, close to -1.0 or +1.0. For these runs, it is also easy to categorize them into the result category with recovered DIF structure. In the middle of Figure 4.7, the correlation index is somewhere close to zero. This means that the estimated DIF structure is very different from the generating DIF structure. Thus for these runs, I can easily identify them as the simulations runs without recovery of the DIF structure.

In addition, as shown in Figure 4.7, the values in the correlation analysis range from -1.0 to 1.0 and the scale of correlations is continuous. There are high correlations, medium high correlations, low correlations, and very low correlations. Though correlations between estimated DIF and generating DIF provide very useful information about how far the estimated DIF values are from the true values, there are replications with medium correlations for which it can be hard to decide whether to include or exclude these cases in the final result summary. Thus additional information is needed for deciding to which result category a particular replication should belong.

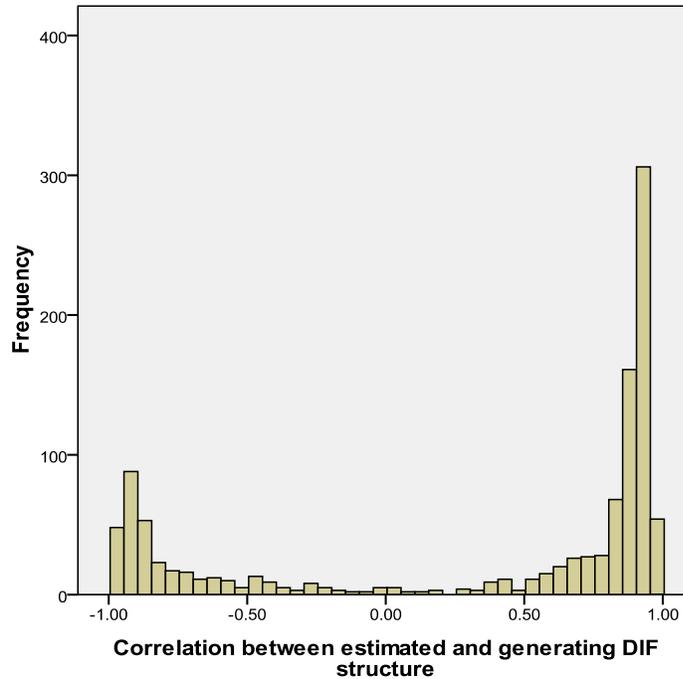


Figure 4.7: Histogram of correlation index

#### 4.1.3 Review of the Percentage of Significant DIF Identified

In the following step of output classification, I calculate the percentage of statistically significant DIF parameters which were generated with large DIF effect size. In 50% of the 72 simulation cells, 6 out of 30 items have large DIF effect. In the other 50% of the 72 simulation cells, 12 out of 30 items have large DIF effect. Thus I use percentage instead of frequency count of the statistically significant DIF parameters as a measure to show the portion of significant DIF parameters.

Different from step two in which I measure how close the estimated DIF values are to the generating values, step three focuses on how accurately the large DIF parameters were estimated. The logic behind this step is that a replication with a recovered DIF structure should be able to identify most of the large DIF items as statistically significant. For some replications, the mean of estimated DIF values might be close to the generating

values, but if the standard deviation of the parameter is too large, the estimated DIF parameters might not be significantly different from zero. If the majority of large DIF items are not statistically significant from zero, then the DIF structure can't be considered as well recovered. Figure 4.8 is the histogram of percentage of statistically significant DIF parameters out of total number of large DIF parameters. In this step, I decide that only replications with 50% of the large DIF items identified as statistically significant are selected to be part of the final result summary.

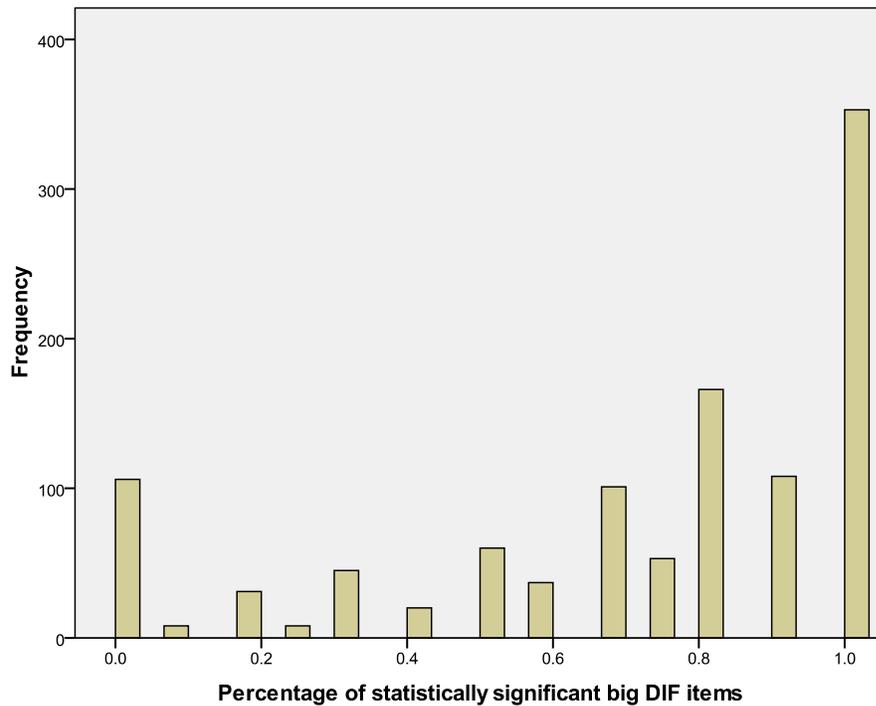


Figure 4.8: Histogram of correlation index

#### **4.1.4 Comparison of Signs of DIF Parameters with Generating DIF Parameters**

The last step in the result review process is to compare the signs of DIF parameters with the generated one. It is expected that well recovered a DIF structure should also identify the direction of DIF between two latent classes correctly. In this process, 9 replications out of the total 720 replications are identified as problematic and excluded from the calculation of bias and RMSE. In fact, these replications correspond well to findings from the previous review steps. Some of them are the runs collapsed to extreme values while some of them have the poor mixing issues. In the ‘collapsed’ runs, the mixing proportions have values very close to either zero or one. The MCMC chains with non-systematical fluctuation across iterations are identified as ‘poor mixing’ runs.

After careful examination of the simulation output using the four-step result review process described above, I include a major part of the output results in the final result summary and exclude small amounts of the output. The entire simulation output is closely screened and each reviewing step eliminates part of the output based on strong logical thinking. By categorizing output into different qualitative groups, I assure that the calculation of bias estimates are based on well-recovered parameters.

#### **4.1.5 Examination of Ill-Recovered Replications**

By going through the four steps in the result screen process, there are 80% of replications out of 720 are selected in the calculation of bias and RMSE for parameter estimations. Before I further describe of the results, it is necessary to study those replications where the underlying structure is not well recovered. Below, I show two charts, Figure 4.9 and Figure 4.10, in which I compare the indices of interest between

included output and excluded output. It is noticeable that there are significant differences in the aspect of correlation and percentage of large DIF item with statistical significance. In Figure 4.9, those replications with low correlation with generating DIF structure are excluded from the final result summary of MRM parameter estimates.

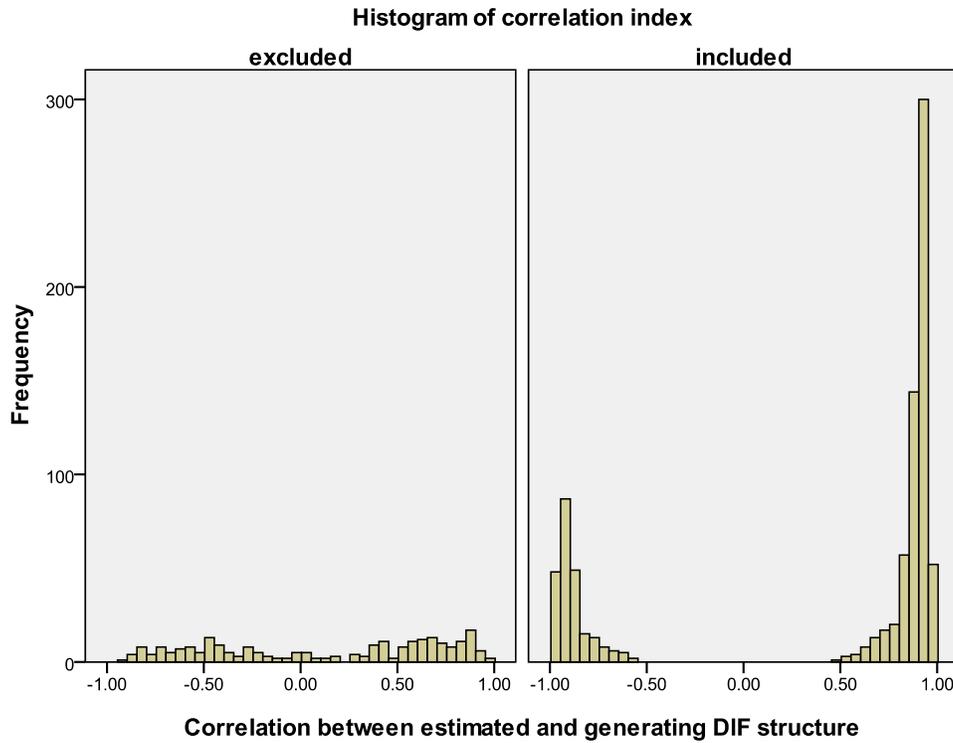


Figure 4.9: Histogram of correlation index between excluded and included output

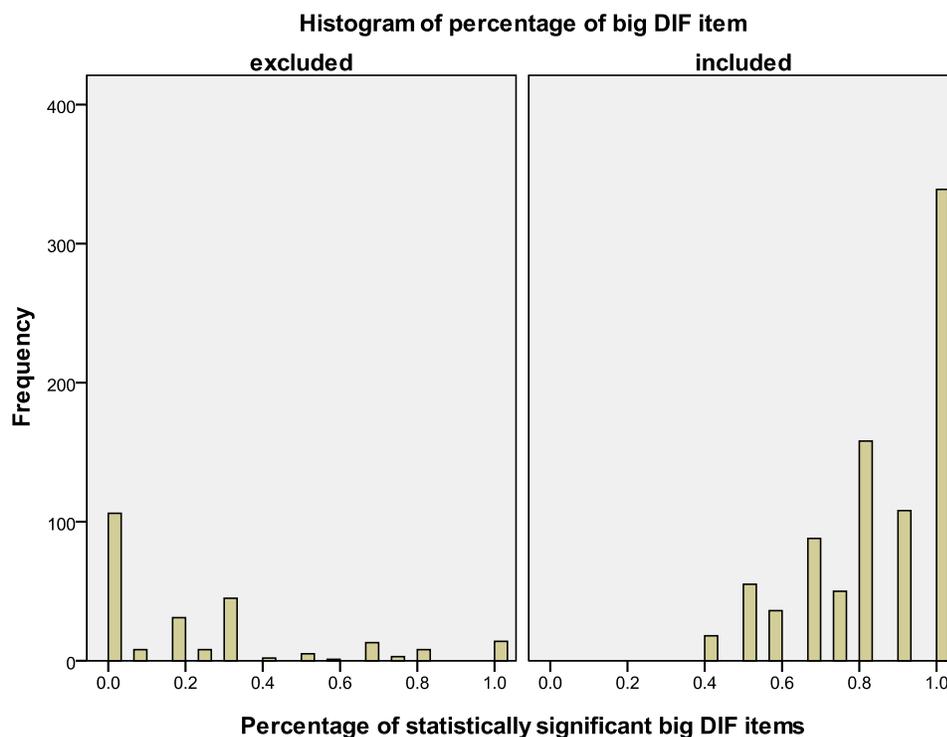


Figure 4.10: Histogram of percentage of significant item between excluded and included output

It is the same for Figure 4.10 above, in which the ill-estimated runs are those where only a small amount of large DIF items are estimated as statistically different from zero. By closely examining the excluded output and comparing with included output, I am confident that the remaining outputs are the ones for which the generating structure is well-recovered and the estimation is stable.

#### 4.1.6 Solution to Label-Switching Runs

In this simulation study, the solution employed to deal with the label-switching problem is to relabel estimated parameters from each latent class in the replications where between-chain label-switching occurs. First, this issue is inevitable in both Bayesian estimation and ML estimation: In ML estimation it is manifested as multiple

maxima that are identical except for labeling, and in MCMC estimation it is manifested as between-chain or within-chain label-switching. Second, in the preliminary analysis I examined the simple solution recommended by Chung, et al (2003).

However, this solution was not sufficient to resolve the label-switching problem in the proposed MRM with a covariate model. In total, only 21% of the label-switching runs were resolved by adding additional group membership information to simulated subjects. It may be the case that continuing to fix more examinees' latent class membership parameters would eventually resolve the labeling issue, but there are 720 replications and each of them takes ten hours for 10,000 iterations in WinBUGS to finish. It is unnecessary to rerun replications with an ineffective approach under reasonable time constraint when alternatives are available. Thus it was decided in this simulation study to relabel parameters after runs were complete, so that all runs would have the same labeling for result summary.

It should be noted that switching labels between latent classes is not sufficient to fix the within-chain label-switching runs. In total, there are 9 MCMC chains from 7 simulation replications belonging to this result category. They only represent 0.97% of total 720 simulation replications and 0.63% of the total 1440 MCMC chains for this simulation study. More details about these within-chain label-switching runs are presented in next section.

To be conservative in summarizing the simulation results, these replications are excluded from the final result summary. In practical work, an analyst might use the Chung, et al. (2003) strategy as follows: Identify a stretch of cycles that appeared to conform to a given labeling, obtain posteriors for examinees' class memberships based

on that stretch, fix the values for additional subjects in classes for which their posterior means were near one, and carry out a new analysis with these stronger label-identification constraints.

#### **4.2.1 Recovery of Underlying Structures**

Based on the five-step result reviewing process, the simulation output is classified into different result categories and shown in Table 4.1c. To better understand Table 4.1c, I provide the detailed denotations and descriptions of result categories, result types, and simulation factors in Table 4.1a below. In addition, Table 4.1b below is used to show the corresponding result summary decision for each result category in the three specified result types.

As shown in Table 4.1b and Table 4.1c, there are three major result types in the simulation output:

- i. Two MCMC chains converged to a single solution and this type of result consists of 48% of 720 simulation replications;
- ii. Two MCMC chains are non-converged and the result category for both MCMC chains is the same. This type of result consists of 37% of 720 simulation replications;
- iii. Two MCMC chains are non-converged and the result category for each MCMC chains is different. This type of result consists of 16% of 720 simulation replications.

Table 4.1c below shows the classification of the results for each simulation cell and the frequency count for each category of results. The values in Table 4.1c are the

numbers of replication for each simulation cell belonging to each result category per major result types. In this study, there are ten replications for each of the 72 simulation cells. And for each replication, two MCMC chains are requested for each WinBUGS run. Thus there are 720 simulation replications and 1440 MCMC chains for the entire simulation study.

Overall, 80% of MCMC runs recover the underlying latent class structure and DIF structure correctly. 20% of the simulation output did not recover the underlying latent class and DIF structure well. For those simulation outputs where the underlying structure is not recovered well, there are 17.29% of total simulation results belonging to the result category of ‘collapsed solution’. This category of outcome is essentially the single class solution with collapsed latent classes. In addition, there are about 1.94% of total simulation outcomes identified as the ‘poor mixing’ category.

Table 4.1d below shows which simulation cells these ‘poor mixing’ MCM chains belong to. In addition, there are 0.90% of the 720 replications (i.e., 1440 MCMC chains) identified as having within-chain label-switching issues. In addition, Table 4.1e below shows which simulation cells these ‘within-chain label-switching’ MCMC chains belong to. In Table 4.1f below, I list the simulation cells where the recovery rate of the latent structure and DIF structure is less than 50%. In Table 4.1d and Table 4.1e, we can see that for simulation cells with combinations of extreme simulation conditions (i.e., no connections between covariate groups, only 20% of items with large DIF effect size and when the distributions of latent ability are the same across latent classes), the within-chain label-switching or ‘poor mixing’ occurred. As shown in Table 4.1f, these extreme simulation conditions also lead to the low recovery rate of the latent class structure and

DIF structure. In Table 4.1f, 10 out of 11 simulation cells with less than 50% recovery rate are those simulation cells with no connection between the covariate and latent classes (i.e., odds ratio equal to 1). As will be seen in the following section, the proportions of structure recovery depend strongly on the independent variables (i.e., simulation factors) in the study.

Table 4.1a: Denotation of result categories, result types and simulation factors in Table 4.1b to Table 4.1f

id	Detailed description	denotation
1	within-chain label-switching	WCLS
2	latent structure recovered	LSR
3	between-chain label-switching	BCLS
4	collapsed	C
5	collapsed + latent structure recovered	C+LSR
6	collapsed + poor mixing	C+PM
7	collapsed + between-chain label-switching	C+BCLS
8	latent structure recovered + poor mixing	LSR+PM
9	latent structure recovered + within-chain label-switching	LSR+WCLS
10	number of replication recovered the latent structure	R
11	recovery rate	R%
12	non-converged	NC
13	two chains converged	2 chains converged
14	two chains non-converged same result category	2 chains NC same
15	two chains non-converged different result category	2 chains NC different
16	percent of items with large DIF effect size= 20%	DIF20
17	percent of items with large DIF effect size:=40%	DIF40
18	Distribution of subjects in covariate groups=30%:70%	cov30
19	Distribution of subjects in covariate groups= 50%:50%	cov50
20	Distribution of subjects between two latent classes=15%:85%	LC15
21	Distribution of subjects between two latent classes=30%:70%	LC30
22	Distribution of subjects between two latent classes=50%:50%	LC50
23	Odds ratio between latent proportions and covariate groups=1	es1
24	Odds ratio between latent proportions and covariate groups=10	es10
25	Odds ratio between latent proportions and covariate groups=50	es50
26	Prior distribution of subjects' latent ability= $N(0,1),N(0,1)$	m0
27	Prior distribution of subjects' latent ability= $N(0,1),N(1,1)$	m1

Table 4.1b: Result types with corresponding result categories and result summary decision

result types	result category	result included or excluded	percent	latent structure recovered	PM	WCLS	Collapsed
two MCMC chains converged to a single solution	WCLS	excluded	0.56%			0.56%	
	C	excluded	5.42%				5.42%
	LSR	included	41.39%	41.39%			
	PM	excluded	0.42%		0.42%		
two MCMC chains are non-converged and the result category for both MCMC chains is the same.	C	excluded	4.86%				4.86%
	LSR	included	6.67%	6.67%			
	BCLS	included	24.31%	24.31%			
	PM	excluded	0.42%		0.42%		
two MCMC chains are non-converged and the result category for each MCMC chains is different.	WCLS	excluded	0.28%			0.28%	
	C+LSR	50% counted and included	12.36%	6.18%			6.18%
	C+PM	excluded	0.69%		0.35%		0.35%
	C+BCLS	50% counted and included	0.97%	0.49%			0.49%
	LSR+PM	50% counted and included	1.53%	0.76%	0.76%		
LSR+WCLS	50% counted and included	0.14%	0.07%		0.07%		
Result summary Total			100.00%	79.86%	1.94%	0.90%	17.29%

Table 4.1c: Distribution of replication for each result category and result type

simulation factor description	2 chains converged				2 chains NC same					2 chains NC different					R	R%	total
	WCLS	C	LSR	PM	C	LSR	BCLS	PM	WCLS	C+LSR	C+PM	C+BCLS	LSR+PM	LSR+WCLS			
dif20-cov30-LC15-es1-m0	0	3	4	0	0	0	0	0	0	3	0	0	0	0	5.5	55%	10
dif20-cov30-LC15-es1-m1	0	2	1	0	1	1	0	0	0	5	0	0	0	0	4.5	45%	10
dif20-cov30-LC15-es10-m0	0	0	2	0	2	0	0	0	0	6	0	0	0	0	5	50%	10
dif20-cov30-LC15-es10-m1	0	0	5	0	0	3	0	0	0	2	0	0	0	0	9	90%	10
dif20-cov30-LC15-es50-m0	0	0	3	0	0	1	0	0	0	6	0	0	0	0	7	70%	10
dif20-cov30-LC15-es50-m1	0	0	4	0	0	0	6	0	0	0	0	0	0	0	10	100%	10
dif20-cov30-LC30-es1-m0	0	1	2	0	1	0	0	0	0	5	0	0	1	0	5	50%	10
dif20-cov30-LC30-es1-m1	0	0	1	0	0	1	5	0	0	2	0	1	0	0	8.5	85%	10
dif20-cov30-LC30-es10-m0	0	0	1	1	0	1	1	0	0	6	0	0	0	0	6	60%	10
dif20-cov30-LC30-es10-m1	0	0	6	0	0	0	4	0	0	0	0	0	0	0	10	100%	10
dif20-cov30-LC30-es50-m0	0	0	6	0	0	0	4	0	0	0	0	0	0	0	10	100%	10
dif20-cov30-LC30-es50-m1	0	0	7	0	0	0	3	0	0	0	0	0	0	0	10	100%	10
dif20-cov30-LC50-es1-m0	0	4	0	0	4	0	0	1	0	0	1	0	0	0	0	0%	10
dif20-cov30-LC50-es1-m1	0	1	2	0	0	1	1	0	0	1	1	0	3	0	6	60%	10
dif20-cov30-LC50-es10-m0	0	0	0	0	0	4	1	0	0	5	0	0	0	0	7.5	75%	10
dif20-cov30-LC50-es10-m1	0	0	7	0	0	0	3	0	0	0	0	0	0	0	10	100%	10
dif20-cov30-LC50-es50-m0	0	0	2	0	0	6	1	0	0	1	0	0	0	0	9.5	95%	10
dif20-cov30-LC50-es50-m1	0	0	4	0	0	0	6	0	0	0	0	0	0	0	10	100%	10
dif20-cov50-LC15-es1-m0	0	2	2	0	2	0	0	0	0	4	0	0	0	0	4	40%	10
dif20-cov50-LC15-es1-m1	0	2	1	0	1	1	0	0	0	5	0	0	0	0	4.5	45%	10
dif20-cov50-LC15-es10-m0	0	2	4	0	0	0	0	0	0	4	0	0	0	0	6	60%	10
dif20-cov50-LC15-es10-m1	0	0	3	0	0	4	1	0	0	2	0	0	0	0	9	90%	10
dif20-cov50-LC15-es50-m0	0	1	3	0	0	0	0	0	0	6	0	0	0	0	6	60%	10
dif20-cov50-LC15-es50-m1	0	0	3	0	0	2	3	0	0	0	0	2	0	0	9	90%	10
dif20-cov50-LC30-es1-m0	0	1	1	0	4	0	0	0	0	3	1	0	0	0	2.5	25%	10
dif20-cov50-LC30-es1-m1	0	1	1	0	1	1	0	1	0	5	0	0	0	0	4.5	45%	10
dif20-cov50-LC30-es10-m0	0	0	3	1	0	2	0	0	0	4	0	0	0	0	7	70%	10
dif20-cov50-LC30-es10-m1	1	0	5	0	0	1	3	0	0	0	0	0	0	0	9	90%	10

Table 4.1c: Distribution of replication for each result category and result type (continued)

simulation factor description	2 chains converged				2 chains NC same					2 chains NC different					R	R%	total
	WCLS	C	LSR	PM	C	LSR	BCLS	PM	WCLS	C+LSR	C+PM	C+BCLS	LSR+PM	LSR+WCLS			
dif20-cov50-LC30-es50-m0	0	0	6	0	0	1	3	0	0	0	0	0	0	0	10	100%	10
dif20-cov50-LC30-es50-m1	0	0	1	0	0	0	9	0	0	0	0	0	0	0	10	100%	10
dif20-cov50-LC50-es1-m0	0	4	0	0	3	1	0	0	0	2	0	0	0	0	2	20%	10
dif20-cov50-LC50-es1-m1	0	1	4	0	0	3	0	0	1	0	0	0	0	1	7.5	75%	10
dif20-cov50-LC50-es10-m0	0	1	2	0	0	4	0	0	1	1	1	0	0	0	6.5	65%	10
dif20-cov50-LC50-es10-m1	0	0	1	0	0	0	9	0	0	0	0	0	0	0	10	100%	10
dif20-cov50-LC50-es50-m0	0	0	5	0	0	1	4	0	0	0	0	0	0	0	10	100%	10
dif20-cov50-LC50-es50-m1	0	0	9	0	0	0	1	0	0	0	0	0	0	0	10	100%	10
dif40-cov30-LC15-es1-m0	0	1	0	0	7	0	0	0	0	2	0	0	0	0	1	10%	10
dif40-cov30-LC15-es1-m1	0	5	0	0	1	1	2	0	0	1	0	0	0	0	3.5	35%	10
dif40-cov30-LC15-es10-m0	0	0	4	0	1	0	3	0	0	1	0	1	0	0	8	80%	10
dif40-cov30-LC15-es10-m1	0	0	10	0	0	0	0	0	0	0	0	0	0	0	10	100%	10
dif40-cov30-LC15-es50-m0	0	0	7	0	0	0	3	0	0	0	0	0	0	0	10	100%	10
dif40-cov30-LC15-es50-m1	0	0	9	0	0	0	1	0	0	0	0	0	0	0	10	100%	10
dif40-cov30-LC30-es1-m0	1	1	6	0	0	0	0	0	0	0	0	0	2	0	7	70%	10
dif40-cov30-LC30-es1-m1	0	0	8	0	0	0	1	0	0	0	0	0	1	0	9.5	95%	10
dif40-cov30-LC30-es10-m0	0	0	2	0	0	1	7	0	0	0	0	0	0	0	10	100%	10
dif40-cov30-LC30-es10-m1	0	0	5	0	0	0	5	0	0	0	0	0	0	0	10	100%	10
dif40-cov30-LC30-es50-m0	0	0	5	0	0	0	5	0	0	0	0	0	0	0	10	100%	10
dif40-cov30-LC30-es50-m1	0	0	4	0	0	0	6	0	0	0	0	0	0	0	10	100%	10
dif40-cov30-LC50-es1-m0	2	0	7	0	0	1	0	0	0	0	0	0	0	0	8	80%	10
dif40-cov30-LC50-es1-m1	0	0	3	0	0	0	7	0	0	0	0	0	0	0	10	100%	10
dif40-cov30-LC50-es10-m0	0	0	4	0	0	0	6	0	0	0	0	0	0	0	10	100%	10
dif40-cov30-LC50-es10-m1	0	0	4	0	0	0	6	0	0	0	0	0	0	0	10	100%	10
dif40-cov30-LC50-es50-m0	0	0	4	0	0	0	6	0	0	0	0	0	0	0	10	100%	10
dif40-cov30-LC50-es50-m1	0	0	2	0	0	0	8	0	0	0	0	0	0	0	10	100%	10

Table 4.1c: Distribution of replication for each result category and result type (continued)

simulation factor description	2 chains converged				2 chains NC same					2 chains NC different					R	R%	total
	WCLS	C	LSR	PM	C	LSR	BCLS	PM	WC LS	C+ LSR	C+ PM	C+ BCLS	LSR+ PM	LSR+ WCLS			
dif40-cov50-LC15-es1-m0	0	2	1	0	4	0	0	0	0	3	0	0	0	0	2.5	25%	10
dif40-cov50-LC15-es1-m1	0	0	5	0	1	0	3	0	0	1	0	0	0	0	8.5	85%	10
dif40-cov50-LC15-es10-m0	0	3	2	0	1	0	0	0	0	1	0	3	0	0	4	40%	10
dif40-cov50-LC15-es10-m1	0	0	8	0	1	0	0	0	0	0	0	0	1	0	8.5	85%	10
dif40-cov50-LC15-es50-m0	0	0	8	0	0	1	1	0	0	0	0	0	0	0	10	100%	10
dif40-cov50-LC15-es50-m1	0	0	10	0	0	0	0	0	0	0	0	0	0	0	10	100%	10
dif40-cov50-LC30-es1-m0	0	0	1	1	0	3	4	0	0	1	0	0	0	0	8.5	85%	10
dif40-cov50-LC30-es1-m1	0	1	6	0	0	0	3	0	0	0	0	0	0	0	9	90%	10
dif40-cov50-LC30-es10-m0	0	0	8	0	0	0	1	0	0	1	0	0	0	0	9.5	95%	10
dif40-cov50-LC30-es10-m1	0	0	10	0	0	0	0	0	0	0	0	0	0	0	10	100%	10
dif40-cov50-LC30-es50-m0	0	0	4	0	0	0	6	0	0	0	0	0	0	0	10	100%	10
dif40-cov50-LC30-es50-m1	0	0	10	0	0	0	0	0	0	0	0	0	0	0	10	100%	10
dif40-cov50-LC50-es1-m0	0	0	3	0	0	0	4	0	0	0	1	0	2	0	8	80%	10
dif40-cov50-LC50-es1-m1	0	0	2	0	0	1	5	1	0	0	0	0	1	0	8.5	85%	10
dif40-cov50-LC50-es10-m0	0	0	9	0	0	0	1	0	0	0	0	0	0	0	10	100%	10
dif40-cov50-LC50-es10-m1	0	0	10	0	0	0	0	0	0	0	0	0	0	0	10	100%	10
dif40-cov50-LC50-es50-m0	0	0	3	0	0	1	6	0	0	0	0	0	0	0	10	100%	10
dif40-cov50-LC50-es50-m1	0	0	3	0	0	0	7	0	0	0	0	0	0	0	10	100%	10
	4	39	298	3	35	48	175	3	2	89	5	7	11	1	575	80%	720
	1%	5%	41%	0%	5%	7%	24%	0%	0%	12%	1%	1%	2%	0%			100%

Table 4.1d: List of poor mixing chains by simulation cell

result category	Cell ID	simulation factor description	number of MCMC chain
poor mixing	7	dif20-cov30-LC30-es1-m0	1
poor mixing	9	dif20-cov30-LC30-es10-m0	1
poor mixing	13	dif20-cov30-LC50-es1-m0	3
poor mixing	14	dif20-cov30-LC50-es1-m1	4
poor mixing	25	dif20-cov50-LC30-es1-m0	1
poor mixing	26	dif20-cov50-LC30-es1-m1	2
poor mixing	27	dif20-cov50-LC30-es10-m0	1
poor mixing	33	dif20-cov50-LC50-es10-m0	1
poor mixing	43	dif40-cov30-LC30-es1-m0	2
poor mixing	44	dif40-cov30-LC30-es1-m1	1
poor mixing	58	dif40-cov50-LC15-es10-m1	1
poor mixing	61	dif40-cov50-LC30-es1-m0	1
poor mixing	67	dif40-cov50-LC50-es1-m0	3
poor mixing	68	dif40-cov50-LC50-es1-m1	3
		total count	25

Table 4.1e: List of within-chain label-switching chains by simulation cell

result category	Cell ID	simulation factor description	number of MCMC chain
WCLS	28	dif20-cov50-LC30-es10-m1	1
WCLS	32	dif20-cov50-LC50-es1-m1	3
WCLS	33	dif20-cov50-LC50-es10-m0	2
WCLS	43	dif40-cov30-LC30-es1-m0	1
WCLS	49	dif40-cov30-LC50-es1-m0	2
		total count	9

Table 4.1f: List of simulation cells with recovery rate less than 50%

Cell ID	simulation factor description	Recovery rate <50%
2	dif20-cov30-LC15-es1-m1	45%
13	dif20-cov30-LC50-es1-m0	0%
19	dif20-cov50-LC15-es1-m0	40%
20	dif20-cov50-LC15-es1-m1	45%
25	dif20-cov50-LC30-es1-m0	25%
26	dif20-cov50-LC30-es1-m1	45%
31	dif20-cov50-LC50-es1-m0	20%
37	dif40-cov30-LC15-es1-m0	10%
38	dif40-cov30-LC15-es1-m1	35%
55	dif40-cov50-LC15-es1-m0	25%
57	dif40-cov50-LC15-es10-m0	40%

A regression analysis was used to model effects from the simulation conditions on the percentage of successful replications for each simulation cell. Table 4.2 below shows how simulation factors are recoded into dichotomous variables for the regression analysis. An ANOVA analysis without recoding the simulation factors will serve the same purpose of analyzing the relations between simulation cell recovery rate and simulation factors.

To ensure the homogeneity of variance assumption in multiple regression analysis, the arcsine transformation of the proportion of recovered runs is used as the dependent variable in the analysis. In this transformation, a minor correction of the proportion is used to avoid extreme values; specifically, a presumed value of 0.5 was added to the counts of both successful and unsuccessful recoveries in each cell. For cells with recovery rates equal to zero, this small value is added to it. And for cells with 100% recovery rates, the small value is extracted from it.

Table 4.2: Denotation of simulation conditions in the regression analysis

Simulation factors	Variable name	Recoded as '1' when value=	Recoded as '0' when value=
Percentage of items with large DIF effect	C_DIF	40%	20%
distribution of covariate	C_COV	50%:50%	30%:70%,
Distribution of two latent classes	C1_LC	30%:70%	15%:85%, 50%:50%
Distribution of two latent classes	C2_LC	50%:50%	15%:85%, 30%:70%
Magnitude of $\exp(\text{sl}[2])$	C1_ES	10	1,50
Magnitude of $\exp(\text{sl}[2])$	C2_ES	50	1,10
Distribution of subjects' latent ability	C_M	Normal(0,1) for one latent class and Normal (1,1) for the other latent class	Normal(0,1) for both latent classes

The output from the regression analysis of recovery rates is presented in Table 4.3. Except for distribution of covariate groups, named 'C\_COV', all other simulation factors have statistically significant effects on the proportion of recovered replications. Below is the summary of the findings:

1. For a simulation cell, more replications recovered the underlying structure when more items have large DIF effects. In this study, 40% of items have large DIF effects in half of the simulated data and 20% of items have large DIF effects in the other half of simulated data.

2. For the distribution of latent classes, when the latent class proportion is equal to 30%:70% or 50%:50%, the proportion of recovered replication is greater than under the 15%:85% condition.

3. When the log odds ratio between covariate groups and latent class proportion is equal to 50, which is the strongest connection between covariate group distribution and latent group distribution, the effect on percentage of successful replication doubled the effect from the condition when the odds ratio is equal to 10.

Table 4.3: Regression analysis of recovered replications by simulation factors

Parameter	Estimate	Standard Error	t Value	Pr >  t
Intercept	0.21	0.07	3.04	0
C_DIF	0.24 *	0.05	4.99	<.0001
C_COV	-0.04	0.05	-0.77	0.45
C1_LC	0.27*	0.06	4.47	<.0001
C2_LC	0.26*	0.06	4.38	<.0001
C1_ES	0.41*	0.06	6.97	<.0001
C2_ES	0.74*	0.08	9.47	<.0001
C_M	0.35*	0.06	5.93	<.0001
C2_ES*C_M	-0.26*	0.1	-2.49	0.02

Note: The \* denotes that the regression coefficient is statistical significant at the  $\alpha=0.05$  level.

For those cells with different average latent ability, the percentage of successful replication is higher than those cells with equal average latent ability for both latent classes. Interactions between different pairs of simulation factors are tested and the interaction between the odds ratio equal to 50 and different mean latent ability is found to be statistically significant. Based on the estimate coefficient in Table 4.3, when there is a

strong relation between the covariates and latent class distribution, the recovery rate decreases when the average latent abilities for two latent groups of subjects are different, compared with those for which the average latent abilities for both group are the same.

#### **4.2.2 Bias and RMSE: Regression Analysis**

After the results are categorized, those which can be used for final result summary of MRM parameter recovery are included in the calculation of the average bias and RMSE for each simulation cell. In the MRM with a covariate, there are five groups of independent variables involved in the modeling process. These parameters of interest include mixing proportions, item difficulty, DIF parameters, mean and variance of latent ability distributions. I use regression analysis to show the relations between simulation factors and the calculated average bias and RMSE. Following the result of non-significant interaction terms in the convergence rate analysis, the regression analysis in this section focuses on main effects. The graphs of results shown in the following section depict two-way interactions.

Table 4.4a and Table 4.4b below contain the output of the regression analysis related to the mixing proportions. As shown in Table 4.4a, the bias of mixing proportions is affected by proportions of large DIF effect items and the distribution of latent groups. Though the effect is statistically significant, the magnitude of effect from having more large DIF items is small. However, when the sizes of latent class are equal, the bias of the larger latent class decreased significantly compared to the case in which the sizes of latent classes are different. Regarding the RMSE shown in Table 4.4b, equal latent classes lead to smaller sizes of RMSE in the estimation of mixing proportions.

In this study about MRM, I assume two latent classes, which mean that there are two mixing proportions summed equal to one. The bias and RMSE for these two mixing proportions are essentially the same except that the signs of the regression coefficient and t-value are opposite of each other. Thus in Table 4.4a and Table 4.4b, I present the regression coefficients for one mixing proportion.

Table 4.4a: Regression analysis of bias by simulation factors for the mixing proportion

P.tot[1]			
Variable	Parameter Estimate	t Value	Pr >  t
Intercept	-0.2	-6.97	<.0001
C_DIF	0.07 *	3.3	0
C_COV	0.00	-0.01	0.99
C1_LC	-0.01	-0.33	0.74
C2_LC	0.10*	4.23	<.0001
C1_ES	0.01	0.48	0.63
C2_ES	0.02	0.62	0.54
C_M	0.01	0.69	0.49

Note: The \* denotes that the regression coefficient is statistical significant at the  $\alpha=0.05$  level.

Table 4.4b: Regression analysis of RMSE by simulation factors for the mixing proportion

P.tot[1]			
Variable	Parameter Estimate	t Value	Pr >  t
Intercept	0.31	7.79	<.0001
C_DIF	-0.05	-1.86	0.07
C_COV	-0.02	-0.56	0.57
C1_LC	-0.01	-0.19	0.85
C2_LC	-0.11 *	-3.28	0
C1_ES	-0.06	-1.76	0.08
C2_ES	-0.06	-1.69	0.1
C_M	0.00	0.07	0.95

Note: The \* denotes that the regression coefficient is statistical significant at the  $\alpha=0.05$  level.

As shown in Table 4.5a, the bias of mean of latent ability distribution is significantly influenced by the corresponding pre-specified simulation conditions. When the mean abilities are different across latent classes, the bias of the smaller latent class decreased

and the bias of precision for the larger latent class increased. In Table 4.5b, having different mean latent abilities for latent classes increased the RMSE for the estimated average latent abilities.

Table 4.5a: Regression analysis of bias by simulation factors for the mean of latent ability distributions

mut[1]				mut[2]			
Variable	Parameter Estimate	t Value	Pr >  t	Variable	Parameter Estimate	t Value	Pr >  t
Intercept	-0.03	-0.58	0.57	Intercept	0.00	0.01	0.99
C_DIF	0.07	1.98	0.05	C_DIF	-0.03	-0.99	0.33
C_COV	0.01	0.34	0.73	C_COV	-0.02	-0.83	0.41
C1_LC	-0.02	-0.48	0.63	C1_LC	0.02	0.48	0.63
C2_LC	-0.02	-0.45	0.66	C2_LC	0.08 *	2.43	0.02
C1_ES	0.01	0.25	0.80	C1_ES	-0.02	-0.48	0.64
C2_ES	0.00	-0.08	0.94	C2_ES	0.02	0.47	0.64
C_M	-0.30 *	-8.62	<.0001	C_M	0.21 *	7.54	<.0001

Note: The \* denotes that the regression coefficient is statistical significant at the  $\alpha=0.05$  level.

Table 4.5b: Regression analysis of RMSE by simulation factors for the mean of latent ability distributions

mut[1]				mut[2]			
Variable	Parameter Estimate	t Value	Pr >  t	Variable	Parameter Estimate	t Value	Pr >  t
Intercept	0.37	6.99	<.0001	Intercept	0.11	2.36	0.02
C_DIF	-0.15 *	-4.20	<.0001	C_DIF	-0.07 *	-2.00	0.05
C_COV	-0.02	-0.57	0.57	C_COV	-0.04	-1.30	0.20
C1_LC	-0.08	-1.79	0.08	C1_LC	0.03	0.70	0.48
C2_LC	-0.07	-1.45	0.15	C2_LC	0.09 *	2.31	0.02
C1_ES	-0.07	-1.64	0.11	C1_ES	-0.07	-1.69	0.10
C2_ES	-0.07	-1.49	0.14	C2_ES	-0.03	-0.73	0.47
C_M	0.35 *	9.42	<.0001	C_M	0.37 *	11.22	<.0001

Note: The \* denotes that the regression coefficient is statistical significant at the  $\alpha=0.05$  level.

Table 4.6a: Regression analysis of bias by simulation factors for the precision of latent ability distributions

tau[1]				tau[2]			
Variable	Parameter Estimate	t Value	Pr >  t	Variable	Parameter Estimate	t Value	Pr >  t
Intercept	0.16	3.86	0.00	Intercept	-0.03	-2.03	0.05
C_DIF	-0.04	-1.44	0.16	C_DIF	0.01	1.06	0.29
C_COV	0.03	0.91	0.37	C_COV	0.01	0.72	0.47
C1_LC	-0.09 *	-2.64	0.01	C1_LC	0.01	0.80	0.43
C2_LC	-0.09 *	-2.64	0.01	C2_LC	0.02	1.45	0.15
C1_ES	-0.05	-1.42	0.16	C1_ES	0.03 *	2.29	0.03
C2_ES	-0.05	-1.31	0.20	C2_ES	0.02	1.69	0.10
C_M	0.03	0.98	0.33	C_M	-0.01	-1.58	0.12

Note: The \* denotes that the regression coefficient is statistical significant at the  $\alpha=0.05$  level.

In Table 4.6a, distributions of both latent classes and covariate groups have statistically significant effect of the bias of precision but the effect is small and negligible in practical terms. In Table 4.6b, similar pattern are found from simulation conditions on the RMSE of precision in the prior normal distribution.

Table 4.6b: Regression analysis of RMSE by simulation factors for the precision of latent ability distributions

tau[1]				tau[2]			
Variable	Parameter Estimate	t Value	Pr >  t	Variable	Parameter Estimate	t Value	Pr >  t
Intercept	0.44	14.19	<.0001	Intercept	0.07	7.86	<.0001
C_DIF	-0.08 *	-3.70	0.00	C_DIF	0.00	0.32	0.75
C_COV	0.04	1.66	0.10	C_COV	0.00	-0.41	0.68
C1_LC	-0.13 *	-4.93	<.0001	C1_LC	0.02 *	2.33	0.02
C2_LC	-0.19 *	-7.18	<.0001	C2_LC	0.05 *	5.81	<.0001
C1_ES	-0.10 *	-3.95	0.00	C1_ES	-0.01	-1.00	0.32
C2_ES	-0.10 *	-3.67	0.00	C2_ES	-0.02 *	-2.24	0.03
C_M	-0.01	-0.62	0.54	C_M	0.02 *	2.60	0.01

Note: The \* denotes that the regression coefficient is statistical significant at the  $\alpha=0.05$  level.

Table 4.7a shows the regression analysis results on intercept and slope from the logit link function. The bias of intercept decreases when there are more large DIF effect items and the mean of latent ability is distributed differently across latent classes. For the intercept, its bias increased when the distribution of latent class is equal to 30%:70%. When the connection between the covariate groups and latent classes are moderate or strong, the bias of slopes decreased.

Table 4.7a: Regression analysis of bias by simulation factors for the intercept and slope in logit link

int[2]				sl[2]			
Variable	Parameter Estimate	t Value	Pr >  t	Variable	Parameter Estimate	t Value	Pr >  t
Intercept	1.43	10.31	<.0001	Intercept	-0.02	-0.16	0.87
C_DIF	-0.66 *	-6.86	<.0001	C_DIF	0.05	0.75	0.46
C_COV	-0.13	-1.34	0.19	C_COV	-0.05	-0.71	0.48
C1_LC	-0.17	-1.45	0.15	C1_LC	0.23 *	2.66	0.01
C2_LC	-0.20	-1.65	0.10	C2_LC	0.11	1.22	0.23
C1_ES	0.01	0.10	0.92	C1_ES	-0.50 *	-5.66	<.0001
C2_ES	0.09	0.78	0.44	C2_ES	-1.28 *	-14.56	<.0001
C_M	-0.47 *	-4.90	<.0001	C_M	0.12	1.73	0.09

Note: The \* denotes that the regression coefficient is statistical significant at the  $\alpha=0.05$  level.

As shown in Table 4.7b, the RMSE of intercept estimates are mainly influenced by the proportion of large DIF items, the relative size of covariate groups, the distribution of covariate groups among each latent class and the average latent abilities for each latent class. The RMSE decreased significantly when 40% of the items are large DIF items comparing to the 20% case. When the sizes of covariate groups are equal, the RMSE of intercept also decreased. And when there are connections between distribution of covariate groups and latent classes, the RMSE of intercept are smaller. Different mean latent abilities across latent classes lead to the smaller RMSE of the intercept, too.

For the slope in the logistic regression function, magnitude of connection between covariate groups and latent groups, the percentage of large DIF items, the distribution of latent classes and unequal average latent abilities affect the RMSE of slope significantly. As expected, the RMSE decreased when there are more large DIF effect items. Comparing to the latent class distribution as 15%:85%, when the size of latent classes equal to 30%:70% or 50%:50%, the RMSE of slope is smaller. In addition, the RMSE of slope increased when there was a strong connection between covariate groups and latent

classes. Different mean latent abilities across latent groups resulted in smaller RMSE for slope estimates.

Table 4.7b: Regression analysis of RMSE by simulation factors for the intercept and slope in logit link

int[2]				sl[2]			
Variable	Parameter Estimate	t Value	Pr >  t	Variable	Parameter Estimate	t Value	Pr >  t
Intercept	1.71	16.53	<.0001	Intercept	0.75	9.70	<.0001
C_DIF	-0.66 *	-9.22	<.0001	C_DIF	-0.19 *	-3.63	0.00
C_COV	-0.16 *	-2.16	0.03	C_COV	0.04	0.75	0.46
C1_LC	-0.13	-1.52	0.13	C1_LC	-0.33 *	-5.11	<.0001
C2_LC	0.08	0.85	0.40	C2_LC	-0.32 *	-4.92	<.0001
C1_ES	-0.23 *	-2.60	0.01	C1_ES	0.09	1.30	0.20
C2_ES	-0.19 *	-2.13	0.04	C2_ES	0.80 *	12.10	<.0001
C_M	-0.50 *	-6.98	<.0001	C_M	-0.20 *	-3.68	0.00

Note: The \* denotes that the regression coefficient is statistical significant at the  $\alpha=0.05$  level.

The last groups of variables of interest are the item difficulty parameters for latent class one and latent class two. As described in the previous chapter, the sum of item parameters is set equal to zero for parameter identification purpose. Thus only the means of RMSE from item parameters for each latent class are used as dependent variables in the regression analysis. The Table 4.8 shows that the average of RMSE of item parameters from latent class one are affected by latent class distribution, the connection between covariate and latent classes, and whether the mean latent abilities are the same across latent groups. All three of these factors contribute to the decrease of RMSE of item difficulty parameters from latent class one. For item difficulty parameter from latent class two, proportions of large DIF items and distribution of subjects within latent classes have significant effects. Note that the closer to equal sizes of latent groups, the higher the RMSE of item parameters from latent class two are. This effect is opposite of the effect from latent class distribution on RMSE of item parameter for latent class one.

Table 4.8: Regression analysis of RMSE by simulation factors for average item difficulty parameters in each latent class

b[LC1]				b[LC2]			
Variable	Parameter Estimate	t Value	Pr >  t	Variable	Parameter Estimate	t Value	Pr >  t
Intercept	0.58	18.48	<.0001	Intercept	0.17	6.00	<.0001
C_DIF	-0.04	-1.64	0.11	C_DIF	0.06*	3.12	0.00
C_COV	-0.02	-0.83	0.41	C_COV	-0.02	-1.02	0.31
C1_LC	-0.08*	-3.11	0.00	C1_LC	0.05*	2.17	0.03
C2_LC	-0.11*	-4.04	0.00	C2_LC	0.12*	4.98	<.0001
C1_ES	-0.10*	-3.65	0.00	C1_ES	-0.04	-1.63	0.11
C2_ES	-0.10*	-3.58	0.00	C2_ES	0.00	0.13	0.90
C_M	-0.05*	-2.45	0.02	C_M	0.03	1.29	0.20

Note: The \* denotes that the regression coefficient is statistical significant at the  $\alpha=0.05$  level.

### 4.2.3 Bias and RMSE: Graphic Presentation

In this research study, the main focus is on how including a covariate in the MRM model affects the estimated parameters of interest. In this section, I use charts (see Figure11 to Figure 18 below) to describe graphically how different levels of  $\exp(\beta_1)$ , which is one of the simulation conditions and equal to the odds ratio between latent proportions and covariate groups (see Appendix A), interacts with four other simulation conditions. These four simulation factors are:

- i. the distribution of the two latent classes in MRM;
- ii. the distribution of subjects in a single covariate;
- iii. the proportion of items with large DIF effect size;
- iv. the prior distribution of subjects' latent ability.

For Figure11 to Figure 18, these simulation factors are denoted in the Table 4.9 below.

Table 4.9: Simulation factors and their denotation

Factors varied in this simulation study	Denotation	Corresponding values
Odds ratio between latent proportions and covariate groups	$\exp(\text{sl}[2])$	1,10, 50
Prior distribution of subjects' latent ability	$N(0,1), N(1,1)$	Normal (0, 1), Normal (1,1)
Distribution of subjects between two latent classes	LC	15%:85%,30%:70%, 50%:50%
Percent of items with large DIF effect size	DIF%	20%, 40%,
Distribution of subjects in covariate groups	COV	50%:50%, 30%:70%

Mainly, I focus on two sets of parameters, one is the average large size DIF parameters from each simulation cell (see Figure 11- 14) and the other one is the mixing proportion of the mixture model (see Figure 15- 18). Please note that there are two latent classes in this MRM which means that there are two mixing proportion parameters. The bias and RMSE for either one of these two mixing proportions are essentially the same, as one is simply the complement of the other. So I select one of them to demonstrate the effect of different magnitudes of relation between covariate group distribution and latent group distribution upon the accuracy of estimating the mixing proportion in MRM.

Additionally, I adjust the scales of the bias and RMSE in Figure 11 to Figure 18 to create better visual effects. For bias, its values can vary from negative to positive along the scale while for RMSE, its value can only be positive. But the scales for all charts for bias are kept the same and all charts for RMSE are kept the same. The range of the charts for bias is -0.20 to +0.40 and the range of the charts for RMSE is 0.00 to +0.60.

For the average of DIF parameters with large effect size, the patterns shown in Figure 11 to Figure 14 are very clear. Both bias and RMSE decrease for all interactions between  $\exp(\text{sl}[2])$  and the four other simulation conditions. Significant decrease in bias and RMSE happen when  $\exp(\text{sl}[2])$  change from 1 to 10. And these decreases are shown in charts for both bias and RMSE. Please note that when the  $\exp(\text{sl}[2])$  is equal to 1, this

means that there is no relation in the distributions of subjects between covariate groups and latent groups. When the value of  $\exp(\beta_2)$  changes from 10 to 50, which means the connections between covariate groups and latent classes change from medium to stronger connection, Figure 11 to Figure 14 show difference between bias and RMSE. The difference is that when  $\exp(\beta_2)$  changes from 10 to 50, the values of bias for the estimated average large effect DIF are very similar. In contrast, when  $\exp(\beta_2)$  changes from 10 to 50, the values of RMSE for the estimated average large effect DIF decrease again.

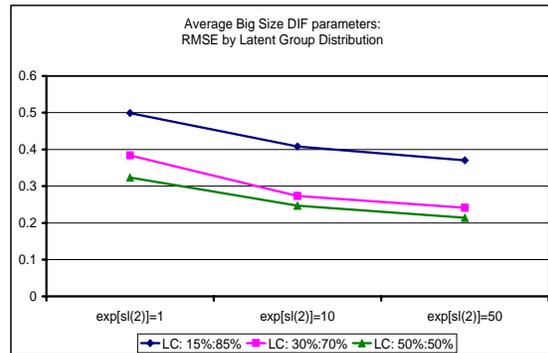
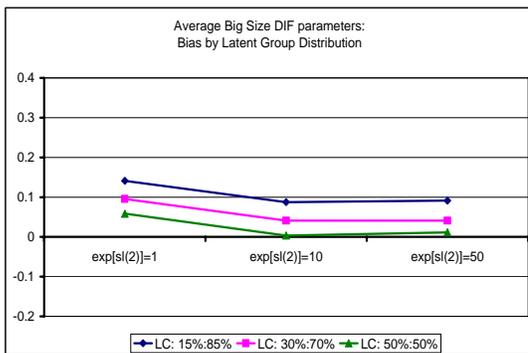


Figure 4.11a: Bias by Latent Group Distribution      Figure 4.11b: RMSE by Latent Group Distribution

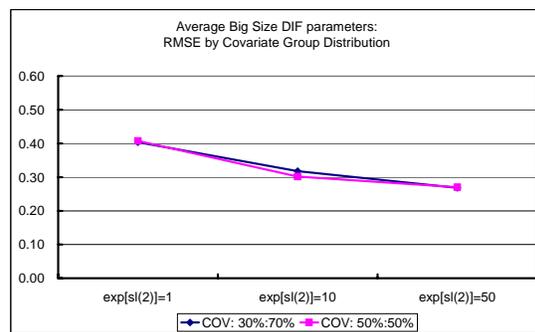
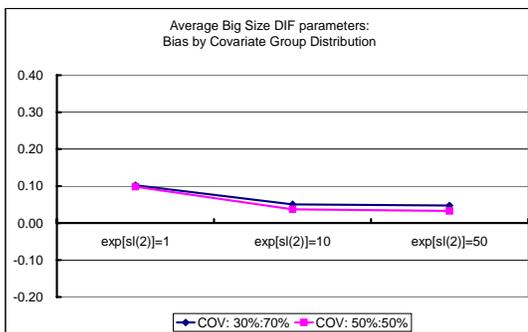


Figure 4.12a: Bias by Covariate Distribution      Figure 4.12b: RMSE by Covariate Distribution

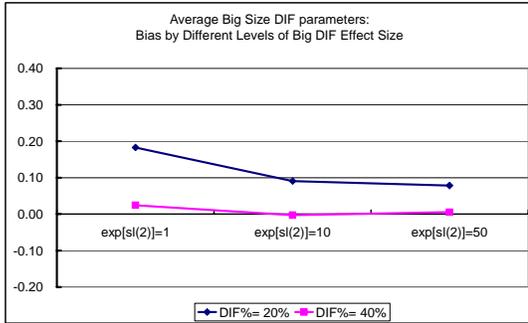


Figure 4.13a: Bias by Levels of Large DIF Effect

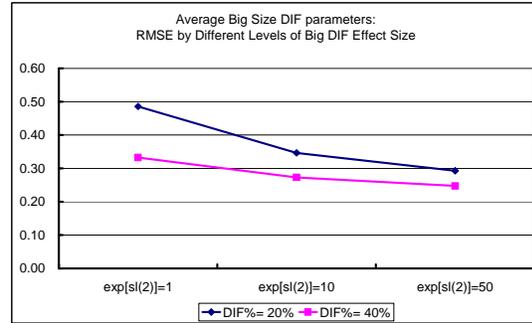


Figure 4.13b: RMSE by Levels of Large DIF Effect

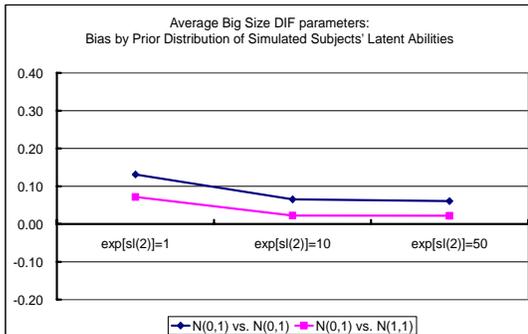


Figure 4.14a: Bias by Subjects' Latent Abilities

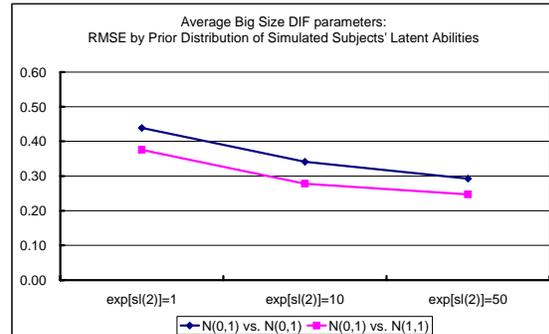


Figure 4.14b: RMSE by Subjects' Latent Abilities

For the mixing proportion, the patterns shown in Figure 15 to Figure 18 below are not as clear as that shown in Figure 11 to Figure 14 for the average DIF parameters with large effect sizes. Thus I use Table 4.10a-4.10d to summarize the patterns observed in the interaction between  $\exp[sl(2)]$  and the four other simulation factors. In my points of view, the different observed patterns between the average large effect size DIF parameters and the mixing proportion means the inclusion of a covariate in the MRM have a stronger effect on the estimation of DIF parameters than on the estimation of mixing proportions so that the estimations of DIF are more accurate when the covariate is included in MRM and when the distribution of subjects between two covariate groups has a strong connection with the distribution of subjects between two latent groups.

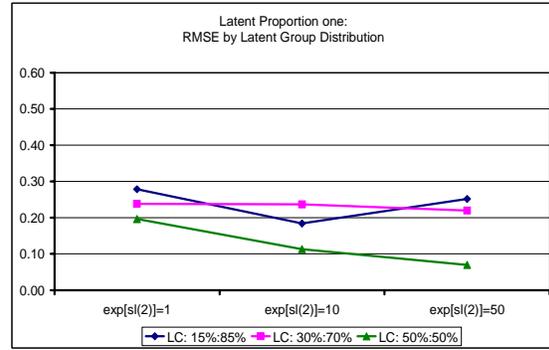
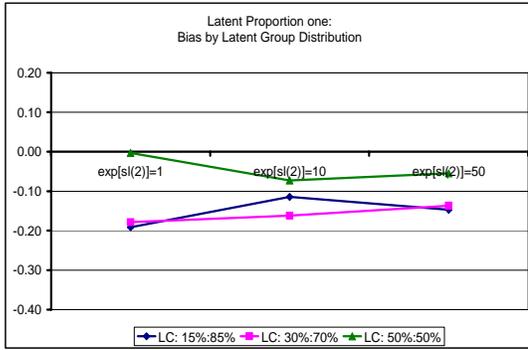


Figure 4.15a: Bias by Latent Group Distribution      Figure 4.15b: RMSE by Latent Group Distribution

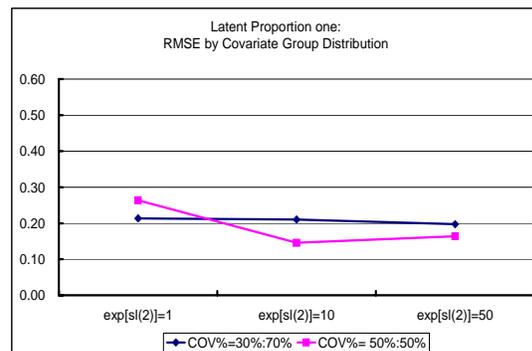
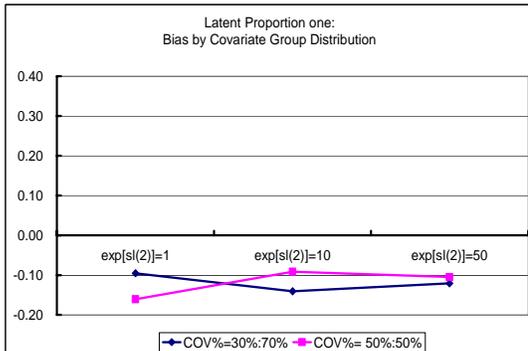


Figure 4.16a: Bias by Covariate Distribution      Figure 4.16b: RMSE by Covariate Distribution

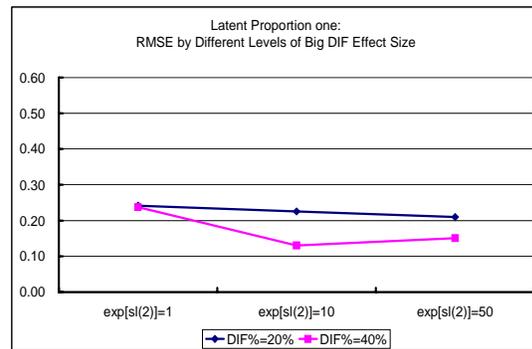
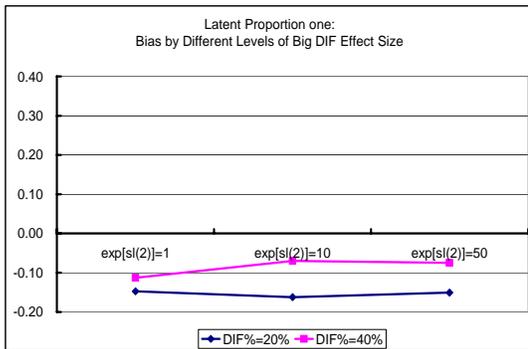


Figure 4.17a: Bias by Levels of Large DIF Effect      Figure 4.17b: RMSE by Levels of Large DIF Effect

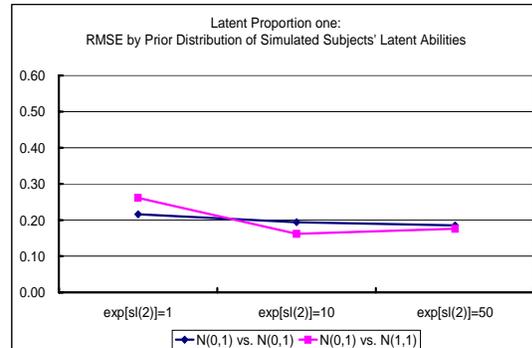
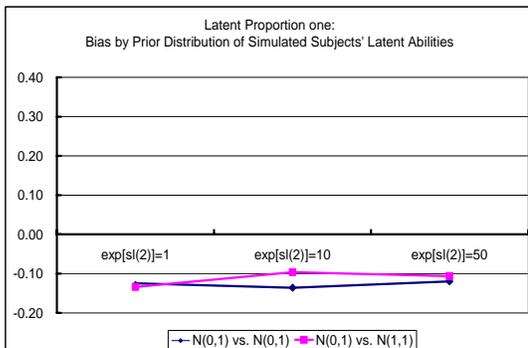


Figure 4.18a: Bias by Subjects' Latent Abilities      Figure 4.18b: RMSE by Subjects' Latent Abilities

Table 4.10a: Pattern observed in Figure 15-Figure 18 for the mixing proportion

LC% (15%:85%, 30%:70%, 50%:50%)		
for bias:	exp(sl[2])=1 ->10	exp(sl[2])=10 -> 50
15%:85%	decrease	increase
30%:70%	decrease	decrease
50%:50%	increase	decrease
for RMSE:	exp(sl[2])=1 ->10	exp(sl[2])=10 -> 50
15%:85%	decrease	increase
30%:70%	flat	flat
50%:50%	decrease	decrease

Table 4.10b: Pattern observed in Figure 15-Figure 18 for the mixing proportion

COV% (30%:70%, 50%:50%)		
for bias:	exp(sl[2])=1 ->10	exp(sl[2])=10 -> 50
30%:70%	increase	flat
50%:50%	decrease	flat
for RMSE:	exp(sl[2])=1 ->10	exp(sl[2])=10 -> 50
30%:70%	flat	flat
50%:50%	decrease	flat

Table 4.10c: Pattern observed in Figure 15-Figure 18 for the mixing proportion

DIF% (20%, 40%)		
for bias:	exp(sl[2])=1 ->10	exp(sl[2])=10 -> 50
20%	decrease	flat
40%	flat	flat
for RMSE:	exp(sl[2])=1 ->10	exp(sl[2])=10 -> 50
20%	decrease	decrease
40%	decrease	flat

Table 4.10d: Pattern observed in Figure 15-Figure 18 for the mixing proportion

N(0,1) , N(1,1)		
for bias:	exp(sl[2])=1 ->10	exp(sl[2])=10 -> 50
N(0,1) vs. N(0,1)	flat	flat
N(0,1) vs. N(1,1)	decrease	flat
for RMSE:	exp(sl[2])=1 ->10	exp(sl[2])=10 -> 50
N(0,1) vs. N(0,1)	flat	flat
N(0,1) vs. N(1,1)	decrease	flat

## Chapter 5 Discussion

In this chapter, I describe the research questions proposed in this study and how the findings in the simulation results answer these research questions. In addition, I discuss the effects of each simulation design characteristic on the overall model estimation and the relative importance of these factors regarding a good recovery of the underlying structure. Also, I provide insights on MRM estimation and advices for practitioners about dealing with the label-switching problem in this proposed model. Furthermore, I suggest research directions for analysts who are interested in studying MRM.

### 5.1 Discussion

This current study sought to explore the interaction between MRM with a covariate and various design characteristics, especially the effect of the covariate upon the model estimation. The two major research questions answered by this study are:

I. To what extent is a stable and accurate estimation of the MRM with a covariate obtained;

II. Under different combinations of design characteristics, what is to be expected from the estimation of MRM with a covariate.

To answer the first research question, I adopt a multi-faceted result screening process to identify those results in which I obtain a stable estimation and identify the true underlying structure. In this simulation study, I manipulated five simulation factors, which are:

- i. odds ratio between latent proportions and covariate groups,

- ii. prior distribution of subjects' latent ability,
- iii. distributions of subjects between two latent classes,
- iv. the number of items with large DIF effect size,
- v. distributions of subjects in covariate groups.

To answer the second research question, I use regression analysis to evaluate the relative significance of each simulation factor to the recovery of parameters. In this research study, I use bias and RMSE to evaluate parameter estimations.

In the MRM with a covariate, latent class modeling technique is combined with the Rasch model from item response theory. On one hand, by combining these two advanced statistical methods, subjects' response data can be analyzed to a great degree and the complex structure underlying the data can be recovered as well as is possible for the model. On the other hand, the more complicated the statistical model is the larger the sample size required to obtain a stable estimation. In addition, I let the model itself classify subjects into different latent classes. If the information contained in the response data is not sufficient for a stable estimation, this means this sophisticated statistical model is not able to classify subjects into latent classes effectively and the distinction among different latent classes are not clear enough from the information provided by the response data.

However, additional information from a covariate or covariates can supply and solve the difficulties when researchers use complicated statistical models to analyze real data. The inclusion of a covariate plays an important role by adding extra contextual information to the data. Based on the findings in this research, additional pieces of information improve the estimation of the MRM with a covariate and I obtain more

accurate and stable estimation. By combining quantitative information, which is the subjects' response data, and qualitative information which is the covariate, I use a more statistically sophisticated model to recover the complicated underlying data structure.

## **5.2 Summary of Results**

By conducting this study, I have four important research goals. First, I investigate the performance of MRM with a covariate under various simulation conditions. Second, I use a logistic regression function to relate the covariate directly to mixing proportions in MRM. Third, I compare 'MRM with a covariate' to 'MRM without a covariate' to demonstrate the advantages and possible drawbacks of using this complex statistical model. Fourth, I provide insights on the application and estimation of the MRM with a covariate.

To achieve the research goals, I extract two sets of dependent variables based on the simulation output. In the result summary process, I use these two dependent variables which are (1) the proportion of cases in which the generating mixture structure is recovered, and (2) the bias and RMSE of parameter estimates among those cases in which the structure is recovered. After these two sources of dependent variables are calculated, I conduct regression analyses to examine the effects of simulation factors.

In the modeling of the recovery rate for each simulation cell, I find out that four simulation factors have statistically significant effects on the recovery of underlying structure. The only non-significant effect is from the distributions of covariate group on the recovery rate.

To be more specific, I find out that the higher the percentage of items with large DIF effect size, the higher the recovery rate of underlying DIF and two-class structure. Comparing with DIF%=20%, simulation cells with DIF%=40% obtain higher correct recovery rates. For latent class distributions, when the two latent classes are distributed as 30%:70% or 50%:50%, the recovery rates from these two levels of this simulation factor are always higher than the recovery rate from the two-class structure with 15%:85% distribution. This indicates that when the sizes of two latent classes are quite different from each other, it is hard to obtain stable and accurate estimation after controlling for other simulation factors. As the focus of this study, the inclusion of collateral information in MRM has positive effects on correct identification of latent structure. Comparing with MRM without covariate, which is the simulation condition where the odds ratio between manifest groups and latent groups is equal to one (i.e.,  $\exp[\text{sl}(2)]=1$ ), MRM with covariate groups moderately (i.e.,  $\exp[\text{sl}(2)]=10$ ) or strongly (i.e.,  $\exp[\text{sl}(2)]=50$ ) related with the distributions of latent classes always achieve higher recovery rates. As expected, the stronger the connection between manifest group distribution and latent class distribution, the higher the chances of correctly identifying the latent structure.

Additionally, when the subjects' latent ability distribution is different across two latent classes, there are higher chances that a stable and reliable estimations of parameters in MRM with a covariate can be obtained. This is consistent with the expectation that the more distinct the latent classes are the easier the MRM model with a covariate can separate subjects into different latent groups.

As previously mentioned, I use bias and RMSE as the second category of dependent variables to evaluate the accuracy of parameter estimations. In the regression analyses of

bias and RMSE for mixing proportions, I find out that when the percentage of items with large DIF effect size is high and when the sizes of latent classes are equal across groups, the effects of these two levels of simulation factors on the bias of mixing proportion are statistically significant. For RMSE of mixing proportion, when the latent classes are distributed as 50%:50% in the sample, the RMSE of the mixing proportion is significantly lower than that from other simulation levels and simulation factors. From the analyses of the latent ability distribution, the corresponding RMSE for the average latent ability is higher when the distributions are different across latent classes. For bias of the average latent ability, similar effects are found from prior distributions of latent ability. Regarding the estimates of standard deviation in the prior distribution of latent abilities, noticeable effects can be found, stemming from the distributions of subjects across latent classes towards the bias and RMSE of this parameter.

The bias of the intercept in the logistic regression function decreases when the percentage of items with large DIF is high and when the latent ability distributions are different across latent groups. In contrast, for the slope parameter, the bias decreases when the odds ratio between latent groups is high and increases when the latent class distribution is equal to 30%:70%, comparing the other two levels of this simulation factor. Considering the RMSE for the intercept parameter, the corresponding values decrease when there are 40% of total items with large DIF effect size, the sizes of manifest groups are equal or the sizes of latent classes is not too different (i.e., 15%:85%). For the slope parameter, the RMSE decreases when there are more items with large DIF effect size, when the average latent abilities are different for the two latent classes, and when the sizes of latent classes are equal to 30%:70% or 50%:50%.

### 5.3 Insights on Model Estimation

Based on my experiences in conducting this simulation study, I have a few recommendations for practitioners who are interested in applying the mixture IRT model proposed in this study. First of all, aside from applying the mixture IRT model on real test data, it is very important to also use the non-mixture IRT model to estimate the item parameters. The reason for doing this is to make sure that when the ‘collapsed’ run occurs using the MRM, the researcher can compare the item parameters from the MRM and item parameters from the IRT model where the item parameters are constrained to be the same for all subjects. In this study, about 20% of the simulation outputs are identified as ‘collapsed’ runs using the known generating parameters and the item parameters estimated from the simple Rasch model based on the same simulation factors.

In addition, to identify the label-switching issues, I would recommend researchers to look at the detailed trace plot of the mixing proportion. In this study, the initial step in result reviewing process is to look at the history chart of mixing proportion from WinBUGS. The reason for doing this is that most of the time the mixing proportion serves as a good indicator of whether the label-switching occurs and, if so, what type of label-switching it is. In this study, to resolve the label-switching issue, I relabel the estimated parameters for the between-chain label-switching runs and exclude the within-chain label-switching runs in which only a small amount of simulation output belongs to this result category. As reviewed in Chapter 2 and discussed in Chapter 3, there are other available solutions to deal with the label-switching that occurs in modeling finite mixture models. I explored some of these solutions in my preliminary analysis and chose the

relabeling method for this study because this approach is effective in solving the between-chain label-switching in the proposed model.

In simulation studies, after the initial review, it is important to calculate correlation indexes between estimated DIF parameters and generated DIF parameters. This is because the comparison between estimated and true DIF structure is very informative in providing evidence about whether the underlying structure is correctly identified. Also, the calculation of the percentage of statistically significant DIF is meaningful in that this allows the comparison of the DIF directions from estimated DIF parameters and generated ones.

Furthermore, practitioners should be cautious when the estimated mixing proportions are close to zero or one. Because of the complexity of the MRM, it is quite possible that the information contained in the data is not sufficient to separate subjects into distinct latent classes or the sample size is not large enough for applying this model. Thus, when researchers have a very large latent class versus a small one, there is chance that the estimation will result in ‘collapsed’ runs.

Above all, it is also possible that when using the proposed MRM with a covariate, different MCMC chains turn out to be in different result categories after the MCMC chains achieve stationary distributions. This type of result is described and summarized in Chapter 4. Practitioners should be able to identify this type of result by looking at the trace plots of mixing proportions and by studying whether the solutions from different MCMC chains are equivalent after the estimated parameters are relabeled. If after switching labels for the mixing proportion parameters and recalculating the corresponding parameters for each latent class, the estimated values from different

MCMC chains are essentially the same, then the model identification issue encountered is the between-chain label-switching. Otherwise, the issue could be ‘poor mixing’, ‘collapsed’ solution, within-chain label-switching, or a different combination of these three types of results.

#### **5.4 Future Research**

To have this study completed within a manageable timeframe, there are four simulation factors that remain constant. These factors and their corresponding values are:

- i. the number of latent classes is equal to two;
- ii. the number of covariate is equal to one;
- iii. the test length is equal to 30;
- iv. the total number of subjects in the simulated response data is equal to 1000.

In future research studies, it is of interest to vary these simulation factors so that researchers can study effects from these constant simulation factors upon the estimation of parameters and recovery of underlying structure. For as complex a statistical model as the MRM with a covariate, increasing sample size and total number of items might not improve the accuracy of parameter estimations because this requires estimations of more parameters (i.e., subjects’ latent abilities and item parameters) which might exceed the computing capacity of the model itself.

Another direction for future research is the inclusion of more than one covariate in the logistic regression function. Since this simulation is exploratory in its nature, to study the effect of a single covariate on MRM, future research can build more complicated models based on the findings from this study, and having more than one covariate is one

research direction. Also in this research project, I study the situation where the number of latent classes is equal to two. Increasing to more than two latent classes would be another explanation of this simulation study.

In Chapter 4, I mention that the computing time of a single estimation of a replication takes about ten hours by using a highly efficient computer with dual-core CPU. I recommend those researchers who are interested in MRM with a covariate to explore other possibilities of statistical programming languages and see whether using other statistical software or programming languages will lead to the decrease in the computing time. In this way, researcher can include more simulation factors in their study.

Furthermore, note that this study is a simulation study in which parameters' true values are known. However, when fitting the model to real data, the item parameters and subjects' latent abilities are unknown. Thus I recommend researchers calculate model fit indexes, such as AIC and BIC, to evaluate how well the statistical model fits the data in a real world scenario.

Recovering the underlying latent structure is the main focus for this simulation study of MRM with a covariate. Another focus of studying mixture IRT model can be the calculation of correct classification of subjects into latent classes. In estimating MRM, it is possible that the underlying latent class structure and DIF structure are well recovered but the classification of subjects into generating latent class is incorrect. This can be another research direction for analysts who are interested in study MRM. A simulation study conducted by Ru & Jiao (2009) focuses on this line of research in MRM without covariate.

In this study, the covariate is directly incorporated in the MRM and a one-step process is thus used to link the covariate with subjects' latent class memberships. This is how the current study is different from the Cohen and Bolt (2005) study in which they use a two-step procedure to locate the cause of DIF. Thus, another research direction would be the comparison of the performance of MRM between the one-step process and the two-step process. From my experience in conducting this simulation research and as also suggested in Samuelsen, (2008), the one-step process is more difficult to execute. I would expect that, under certain simulation conditions, the two-step process (i.e., obtain item parameters and persons' latent class memberships before modeling collateral information with latent class classification) to including information from covariate might lead to more stable but less efficient parameter estimation.

## Appendix A: Inference on exp[sl(2)]

$$\ln(\text{PHI}[i,j]) <- \text{int}[j] + \text{sl}[j] \times \text{cov}[i] \quad (1)$$

$$P[i,j] <- \text{PHI}[i,j] / \text{sum}(\text{PHI}[i,]) \quad (2)$$

WINBUGS Constraint: int[1]=0; sl[1]=0;

- 'i' stands for the examinee, 'j' stands for the latent class the examinee belongs to.
- Int: intercept
- Sl: slope
- Cov: the dummy coded covariate (cov=0 or 1)
- So when cov[i]=0, the probability of belonging to latent class 1 is equal to

$$P[i, 1]_{\text{cov}=0} = \text{PHI}[i,1] / \text{sum}(\text{PHI}[i,])$$

- So when cov[i]=0, the probability of belonging to latent class 2 is equal to

$$P[i, 2]_{\text{cov}=0} = \text{PHI}[i,2] / \text{sum}(\text{PHI}[i,])$$

While  $\text{PHI}[i,1]_{\text{cov}=0} = \exp(\text{int}[1] + \text{sl}[1] \times \text{cov}[i])$

$$= \exp(0 + 0 \times 0)$$

$$= 1$$

$$\text{PHI}[i,2]_{\text{cov}=0} = \exp(\text{int}[2] + \text{sl}[2] \times \text{cov}[i])$$

$$= \exp(\text{int}[2] + \text{sl}[2] \times 0)$$

$$= \exp(\text{int}[2])$$

$$\text{sum}(\text{PHI}[i,])_{\text{cov}=0} = \text{PHI}[i, 1] + \text{PHI}[i, 2]$$

Thus,  $P[i,1]_{\text{cov}=0} = \text{PHI}[i,1] / \text{sum}(\text{PHI}[i,])$

$$= 1 / (1 + \exp(\text{int}[2]))$$

Thus,  $P[i,2]_{\text{cov}=0} = \text{PHI}[i,2] / \text{sum}(\text{PHI}[i,])$

$$= \exp(\text{int}[2]) / (1 + \exp(\text{int}[2]))$$

- So when cov[i]=1, the probability of belonging to latent class 1 is equal to

$$P[i, 1]_{\text{cov}=1} = \text{PHI}[i,1] / \text{sum}(\text{PHI}[i,])$$

- So when cov[i]=1, the probability of belonging to latent class 2 is equal to

$$P[i, 2]_{\text{cov}=1} = \text{PHI}[i,2] / \text{sum}(\text{PHI}[i,])$$

$$\begin{aligned} \text{While } \text{PHI}[i,1]_{\text{cov}=1} &= \exp(\text{int}[1] + \text{sl}[1] \times \text{cov}[i]) \\ &= \exp(0 + 0 \times 1) \\ &= 1 \end{aligned}$$

$$\begin{aligned} \text{PHI}[i,2]_{\text{cov}=1} &= \exp(\text{int}[2] + \text{sl}[2] \times \text{cov}[i]) \\ &= \exp(\text{int}[2] + \text{sl}[2] \times 1) \\ &= \exp(\text{int}[2] + \text{sl}[2]) \end{aligned}$$

$$\text{sum}(\text{PHI}[i,])_{\text{cov}=1} = \text{PHI}[i, 1] + \text{PHI}[i, 2]$$

$$\begin{aligned} \text{Thus, } P[i,1]_{\text{cov}=1} &= \text{PHI}[i,1] / \text{sum}(\text{PHI}[i,]) \\ &= 1 / (1 + \exp(\text{int}[2] + \text{sl}[2])) \end{aligned}$$

$$\begin{aligned} \text{Thus, } P[i,2]_{\text{cov}=1} &= \text{PHI}[i,2] / \text{sum}(\text{PHI}[i,]) \\ &= \exp(\text{int}[2] + \text{sl}[2]) / (1 + \exp(\text{int}[2] + \text{sl}[2])) \end{aligned}$$

	latent class 1	latent class 2
cov=0	$1 / (1 + \exp(\text{int}[2]))$	$\exp(\text{int}[2]) / (1 + \exp(\text{int}[2]))$
cov=1	$1 / (1 + \exp(\text{int}[2] + \text{sl}[2]))$	$\exp(\text{int}[2] + \text{sl}[2]) / (1 + \exp(\text{int}[2] + \text{sl}[2]))$

	latent class 1	latent class 2
cov=0	$P[i,1]_{\text{cov}=0}$	$P[i,2]_{\text{cov}=0}$
cov=1	$P[i,1]_{\text{cov}=1}$	$P[i,2]_{\text{cov}=1}$

$$\begin{aligned} \text{Odds ratio} &= \frac{(P[i,1]_{\text{cov}=0} / P[i,2]_{\text{cov}=0})}{(P[i,1]_{\text{cov}=1} / P[i,2]_{\text{cov}=1})} \\ &= \frac{\frac{1 / (1 + \exp(\text{int}[2]))}{\exp(\text{int}[2]) / (1 + \exp(\text{int}[2]))}}{\frac{1 / (1 + \exp(\text{int}[2] + \text{sl}[2]))}{\exp(\text{int}[2] + \text{sl}[2]) / (1 + \exp(\text{int}[2] + \text{sl}[2]))}} \\ &= \frac{\frac{1}{\exp(\text{int}[2])}}{\frac{1}{\exp(\text{int}[2] + \text{sl}[2])}} \\ &= \frac{\exp(\text{int}[2] + \text{sl}[2])}{\exp(\text{int}[2])} \\ &= \exp(\text{sl}[2]) \end{aligned}$$

→ Thus  $\log(\text{odds ratio}) = \ln(\exp(\text{sl}[2])) = \text{sl}[2]$

## Appendix B: Annotated SAS Code

```
/*dif40_cov30_LC50_es50_m0*/
data aa1;
do subject=1 to 287;
  c=0; x1=rannor(0);lc=1;output;end;
do subject=288 to 500;
  c=1; x1=rannor(0);lc=1;output;end;
run;
/* this is where to change N(0,1) to N (1,1)
x2=rannor(0)+1;*/
data aa2;
do subject=1 to 13;
  c=0; x2=rannor(0); lc=2;output;end;
do subject=14 to 500;
  c=1; x2=rannor(0); lc=2;output;end;
run;
/*DIF*/
data g1;
set aa1 ;array b (i) b1-b30;array prob (i) p1-p30;
b1=1.461;
b2=0.081;
b3=2.101;
b4=0.971;
b5=1.341;
b6=-0.089;
b7=0.271;
b8=-0.979;
b9=-1.309;
b10=-1.859;
b11=-1.479;
b12=0.381;
b13=0.541;
b14=0.901;
b15=1.901;
b16=-0.399;
b17=0.021;
b18=-0.619;
b19=0.001;
b20=0.331;
b21=-0.599;
b22=-1.509;
b23=-0.189;
b24=-0.159;
b25=-0.029;
b26=-0.769;
```

```

b27=-0.019;
b28=-0.499;
b29=0.851;
b30=-0.649;
do i=1 to 30;
prob=(exp(x1-b))/(1+ exp(x1-b));
end;
run;
data g2;
set aa2;array prob (ii) p1-p30;array bb (ii) bb1-bb30;
bb1=2.241;
bb2=0.861;
bb3=2.881;
bb4=1.751;
bb5=2.121;
bb6=0.691;
bb7=1.051;
bb8=-0.199;
bb9=-0.529;
bb10=-1.079;
bb11=-0.699;
bb12=1.161;
bb13=0.021;
bb14=0.381;
bb15=1.381;
bb16=-0.919;
bb17=-0.499;
bb18=-1.139;
bb19=-0.519;
bb20=-0.189;
bb21=-1.119;
bb22=-2.029;
bb23=-0.709;
bb24=-0.679;
bb25=-0.549;
bb26=-1.289;
bb27=-0.539;
bb28=-1.019;
bb29=0.331;
bb30=-1.169;

do ii=1 to 30;
prob=(exp(x2-bb))/(1+ exp(x2-bb));
end;
run;
/* mut */

```

```

data respg1;
set g1;
array prob (j) p1-p30;
array resp (j) r1-r30;
do j=1 to 30;
  x=ranuni(0);
  if x le prob then resp=1; else resp=0;
  /*where le means `less than or equal to'*/
end; run;
data respg2;
set g2;
array prob (j) p1-p30;
array resp (j) r1-r30;
do j=1 to 30;
  x=ranuni(0);
  if x le prob then resp=1; else resp=0;
end; run;
/*proc means data=aa1;run;proc means data=aa2;run;*/
data response;
SET respG1 respG2 ;
rand=ranuni(0)*1000000;
drop x1 x2 x subject b1-b30 bb1-bb30 i ii j p1-p30;
run;
proc sort data=response;
by rand;run;
data _Null_;
set response;
FILENAME OUTDATA 'x:\xxxx.DAT';
FILE OUTDATA NOTITLES NOPRINT;
put (R1-R30 c )(@1 31*2. );
run;
data _Null_;
set response;
FILENAME OUTDATA 'x:\xxx.DAT';
FILE OUTDATA NOTITLES NOPRINT;
put (R1-R30 c rand lc )(@1 31*2. @63 7. @70 2.);
run;

```

## Appendix C: Annotated WINBUGS Code

```

Model
{# Priors
# j: latent class
# m: item
# i: subject
# b: item difficulty parameter
# dif: DIF parameter
# mut: hyperparameter of mean in the latent ability distribution
# tau: hyperparameter of precision in the latent ability distribution
# int: intercept parameter in the logit function
# sl: slope parameter in the logit function
# theta: subject's latent ability parameter
# r[i,m]: subject 'i' response on item 'm'
# P.tot[j]: the mixing proportion for latent class 'j'

for (j in 1:J) { P.tot[j] <- sum(P[,j])/N ;
                mut[j]~ dnorm(0,1)
                tau[j]~ dgamma(1, 1);}
for (m in 1:M-1) { b[1,m] ~ dnorm(0,1);
                  b[2,m] <- b[1,m]+ dif[m]
                  dif[m]~ dnorm(0,0.5)}

b[1,M]<- -1*sum(b[1,1:(M-1)]);
b[2,M] <- -1*sum(b[2,1:(M-1)]);
dif[M]<-b[2,M]-b[1,M]

for (j in 2:J) { int[j] ~ dnorm(0,1); sl[j] ~ dnorm(0,1)}
int [1]<-0; sl[1] <- 0;
for (i in 1:N) { G[i]~dcat( P[i, ] );
                gmemcat1[i]<-equals(G[i],1)}

#G[1]<-1 #G[2]<-1 #G[3]<-2 #G[4]<-2

for (i in 1:N) {
for (j in 1:J) { P[i,j]<- PHI[i,j] / sum(PHI[i,]);
                log(PHI[i,j]) <- int[j] + sl[j]*sex[i];
theta[i,j]~dnorm(mut[G[i]], tau[G[i]]);}

for (m in 1:M) { logit(pi[i,m, 1]) <-theta[i, 1] - b[1,m];
                logit(pi[i,m, 2]) <-theta[i, 2] - b[2,m];
                p[i,m] <- pi[i,m,G[i]]
                r[i,m] ~ dbern(p[i,m])
}
}
}

```

## Bibliography

- Bock, R.D. (1972). Estimating item parameters and latent ability when response are scored in two or more nominal categories. *Psychometrika*, vol. 37, No.1, 29-51
- Camilli, G., & Shepard, L.A. (1994). *Methods for Identifying Biased Test Items*. Thousand Oaks, CA: SAGE Publications.
- Celeux, G. (1998). Bayesian inference for mixtures: the label-switching problem, in R. Payne & P. J. Green, (Eds.), `COMPSTAT 98', Physica, Heidelberg, 227-232.
- Celeux, G., Hurn, M. & Robert, C. P. (2000), Computational and inferential difficulties with mixture posterior distributions, *Journal of the American Statistical Association*. 95, 957-970.
- Cho, S, Cohen, A. S. and Kim, S (2007). Modeling Prior Information for Detection of Latent Classes In the Mixture Rasch Model. Paper Presented at the International Meeting of the Psychometric Society: The 71st annual meeting of the Psychometric Society, 2006, Montreal, CA.
- Chung, H., Loken, E., Schafer, J.L. (2004). Difficulties in drawing inferences with finite-mixture models: a simple example with a simple solution. *The American Statistician*, 2004, 58, 152(7)
- Cohen, A. S., & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement*, 42, 133-148.
- Congdon, P. (2003). *Applied Bayesian Modelling*, John Wiley
- Congdon, P. (2005). *Bayesian Methods for Categorical Data*, Wiley
- Congdon, P. (2007). Bayesian Statistics, in *Encyclopedia of Quantitative Risk Assessment* Wiley.
- Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika* 56, 463-474.
- Dayton, C. M. (1999). *Latent class scaling analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-126). Thousand Oaks, CA: Sage.
- DeAyala, R.J., Kim, S.H., Stapleton, L.M., & Dayton, C.M. (2002). Differential item functioning: A mixture distribution conceptualization. *International Journal of Testing*, 2(3&4), 243-276.
- Diebolt, J. & Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society*, 56, 363-375.

- Embretson, S. E. (2006). Mixed Rasch models for measurement in cognitive psychology. In von Davier, M. & Carstensen C. H. *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications*. New York, NY: Springer
- Fruhwirth-Schnatter, S. (2001), Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96, 194-209.
- Fruhwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York: Springer.
- Hancock, G., & Samuelsen, K. (2008). *Mixture models in latent variable research*. Greenwich, CT: Information Age Publishing, Inc.
- Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Technometrics* 8, 431-444.
- Hasselblad, V. (1969). Estimation of finite mixtures of distributions from the exponential family. *Journal of the American Statistical Association* 64, 1459-1471.
- Hurn, M., Justel, A. & Robert, C. P. (2003). Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics*, 12, 55-79.
- Jasra, A., Holmes, C. C. & Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, 20, 50-67.
- Kelderman, H., & Macready, G.B. (1990). The use of loglinear models for assessing differential item functioning across manifest and latent examinee groups. *Journal of Educational Measurement*, 27, 307-327.
- Kyllonen, P. C., Lohman, D. F & Snow, R. E. (1984). Effects of aptitudes, strategy training, and task facets on spatial task performance. *Journal of Educational Psychology*, 76, 130-145.
- Lazarsfeld, P.F., & Henry, N.W. (1968). *Latent structure analysis*. Mew York: Houghton-Mifflin.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. New Jersey: Lawrence Erlbaum.
- Lubke, G.H., Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10, 21-39.
- McCutcheon A. C. (1987). *Latent class analysis*. Beverly Hills, CA: Sage.

- Mislevy, R.J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, 55, 195-215.
- Mislevy, R.J., Levy, R., Kroopnick, M., & Rutstein, D. (2008). Evidentiary foundations of mixture item response theory models. In G. Hancock & K. Samuelsen (Eds.) *Mixture Models in Latent Variable Research*. Greenwich, CT: Information Age Publishing, Inc.
- Muthén, B. (2001a). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In Collins, L.M. & Sayer, A. (eds.), *New Methods for the Analysis of Change* (pp. 291-322). Washington, D.C.: APA. (#82)
- Muthén, B. (2001b). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (eds.), *New Developments and Techniques in Structural Equation Modeling* (pp. 1-33). Lawrence Erlbaum Associates. (#86)
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.
- Muthén, B. (2008). Latent variable hybrids: overview of old and new models. In G. Hancock & K. Samuelsen (Eds.) *Mixture Models in Latent Variable Research*. Greenwich, CT: Information Age Publishing, Inc.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Danish Institute for Educational Research, The University of Chicago Press.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, IV. Berkeley: University of Chicago Press, 1980.
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14, 271-282.
- Ru, L., & Jiao, H. (2009). Detecting DIF using the mixture Rasch model. Paper Presented at the 2009 annual meeting of National Council on Measurement in Education, San Diego, CA.
- Samuelsen, K. (2005). Examining differential item functioning from a latent class perspective. Doctoral dissertation, University of Maryland, College Park.
- Samuelsen, K. (2008). Examining differential item function from a latent class perspective. In G. Hancock & K. Samuelsen (Eds.) *Mixture Models in Latent Variable Research*. Greenwich, CT: Information Age Publishing, Inc.
- SAS Institute Inc. (2002). Computer Program SAS 9.0, SAS Institute Inc., Cary, NC, USA.

Smit, A., Kelderman, H., & Flier, H. (1999). Collateral information and mixture Rasch models. *Methods of Psychological Research Online*, 4.

Spiegelhalter, D., Thomas, A., & Best, N. (2000). WinBUGS version 1.4 [computer program].

Stephens, M. (1997). Bayesian methods for mixtures of normal distributions. Doctoral dissertation, Department of Statistics, University of Oxford.

von Davier, M. (2008). The mixture general diagnostic model. In G. Hancock & K. Samuelsen (Eds.) *Mixture Models in Latent Variable Research*. Greenwich, CT: Information Age Publishing, Inc.

Wolf, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, 5, 329-350.